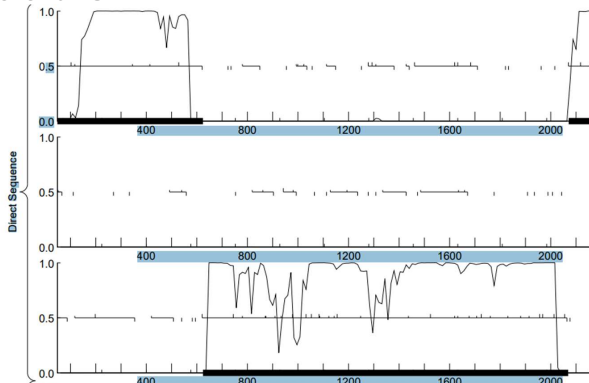
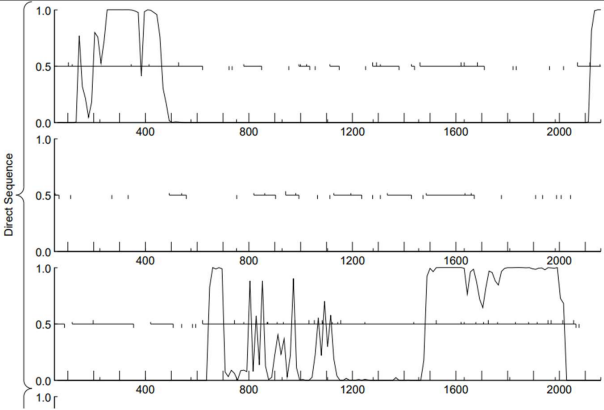
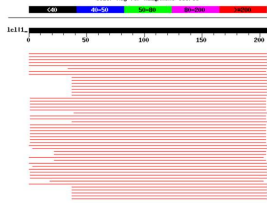


Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	1
Coordonnées du Stop	624
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	-1
Coordonnées du Start retenu	1
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenemarkS</p>  <p>Genemark folio</p>

	 <p>Pic fin dans les deux cas entre 40200 et 40300 Potentiel codant relativement faible sur la première moitié de l'ORF</p>																																																																																																																																																																																																																																				
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show details and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p>  <p>Sequences producing significant alignments:</p> <table><thead><tr><th>Score</th><th>E</th></tr><tr><th>(bits)</th><th>Value</th></tr></thead><tbody><tr><td>802</td><td>e-112</td></tr><tr><td>238</td><td>e-111</td></tr><tr><td>238</td><td>e-110</td></tr><tr><td>238</td><td>e-109</td></tr><tr><td>238</td><td>e-108</td></tr><tr><td>238</td><td>e-107</td></tr><tr><td>238</td><td>e-106</td></tr><tr><td>238</td><td>e-105</td></tr><tr><td>238</td><td>e-104</td></tr><tr><td>238</td><td>e-103</td></tr><tr><td>238</td><td>e-102</td></tr><tr><td>238</td><td>e-101</td></tr><tr><td>238</td><td>e-100</td></tr><tr><td>238</td><td>e-99</td></tr><tr><td>238</td><td>e-98</td></tr><tr><td>238</td><td>e-97</td></tr><tr><td>238</td><td>e-96</td></tr><tr><td>238</td><td>e-95</td></tr><tr><td>238</td><td>e-94</td></tr><tr><td>238</td><td>e-93</td></tr><tr><td>238</td><td>e-92</td></tr><tr><td>238</td><td>e-91</td></tr><tr><td>238</td><td>e-90</td></tr><tr><td>238</td><td>e-89</td></tr><tr><td>238</td><td>e-88</td></tr><tr><td>238</td><td>e-87</td></tr><tr><td>238</td><td>e-86</td></tr><tr><td>238</td><td>e-85</td></tr><tr><td>238</td><td>e-84</td></tr><tr><td>238</td><td>e-83</td></tr><tr><td>238</td><td>e-82</td></tr><tr><td>238</td><td>e-81</td></tr><tr><td>238</td><td>e-80</td></tr><tr><td>238</td><td>e-79</td></tr><tr><td>238</td><td>e-78</td></tr><tr><td>238</td><td>e-77</td></tr><tr><td>238</td><td>e-76</td></tr><tr><td>238</td><td>e-75</td></tr><tr><td>238</td><td>e-74</td></tr><tr><td>238</td><td>e-73</td></tr><tr><td>238</td><td>e-72</td></tr><tr><td>238</td><td>e-71</td></tr><tr><td>238</td><td>e-70</td></tr><tr><td>238</td><td>e-69</td></tr><tr><td>238</td><td>e-68</td></tr><tr><td>238</td><td>e-67</td></tr><tr><td>238</td><td>e-66</td></tr><tr><td>238</td><td>e-65</td></tr><tr><td>238</td><td>e-64</td></tr><tr><td>238</td><td>e-63</td></tr><tr><td>238</td><td>e-62</td></tr><tr><td>238</td><td>e-61</td></tr><tr><td>238</td><td>e-60</td></tr><tr><td>238</td><td>e-59</td></tr><tr><td>238</td><td>e-58</td></tr><tr><td>238</td><td>e-57</td></tr><tr><td>238</td><td>e-56</td></tr><tr><td>238</td><td>e-55</td></tr><tr><td>238</td><td>e-54</td></tr><tr><td>238</td><td>e-53</td></tr><tr><td>238</td><td>e-52</td></tr><tr><td>238</td><td>e-51</td></tr><tr><td>238</td><td>e-50</td></tr><tr><td>238</td><td>e-49</td></tr><tr><td>238</td><td>e-48</td></tr><tr><td>238</td><td>e-47</td></tr><tr><td>238</td><td>e-46</td></tr><tr><td>238</td><td>e-45</td></tr><tr><td>238</td><td>e-44</td></tr><tr><td>238</td><td>e-43</td></tr><tr><td>238</td><td>e-42</td></tr><tr><td>238</td><td>e-41</td></tr><tr><td>238</td><td>e-40</td></tr><tr><td>238</td><td>e-39</td></tr><tr><td>238</td><td>e-38</td></tr><tr><td>238</td><td>e-37</td></tr><tr><td>238</td><td>e-36</td></tr><tr><td>238</td><td>e-35</td></tr><tr><td>238</td><td>e-34</td></tr><tr><td>238</td><td>e-33</td></tr><tr><td>238</td><td>e-32</td></tr><tr><td>238</td><td>e-31</td></tr><tr><td>238</td><td>e-30</td></tr><tr><td>238</td><td>e-29</td></tr><tr><td>238</td><td>e-28</td></tr><tr><td>238</td><td>e-27</td></tr><tr><td>238</td><td>e-26</td></tr><tr><td>238</td><td>e-25</td></tr><tr><td>238</td><td>e-24</td></tr><tr><td>238</td><td>e-23</td></tr><tr><td>238</td><td>e-22</td></tr><tr><td>238</td><td>e-21</td></tr><tr><td>238</td><td>e-20</td></tr><tr><td>238</td><td>e-19</td></tr><tr><td>238</td><td>e-18</td></tr><tr><td>238</td><td>e-17</td></tr><tr><td>238</td><td>e-16</td></tr><tr><td>238</td><td>e-15</td></tr><tr><td>238</td><td>e-14</td></tr><tr><td>238</td><td>e-13</td></tr><tr><td>238</td><td>e-12</td></tr><tr><td>238</td><td>e-11</td></tr><tr><td>238</td><td>e-10</td></tr><tr><td>238</td><td>e-9</td></tr><tr><td>238</td><td>e-8</td></tr><tr><td>238</td><td>e-7</td></tr><tr><td>238</td><td>e-6</td></tr><tr><td>238</td><td>e-5</td></tr><tr><td>238</td><td>e-4</td></tr><tr><td>238</td><td>e-3</td></tr><tr><td>238</td><td>e-2</td></tr><tr><td>238</td><td>e-1</td></tr></tbody></table> <p>Non annotés mais nombreux homologues</p>	Score	E	(bits)	Value	802	e-112	238	e-111	238	e-110	238	e-109	238	e-108	238	e-107	238	e-106	238	e-105	238	e-104	238	e-103	238	e-102	238	e-101	238	e-100	238	e-99	238	e-98	238	e-97	238	e-96	238	e-95	238	e-94	238	e-93	238	e-92	238	e-91	238	e-90	238	e-89	238	e-88	238	e-87	238	e-86	238	e-85	238	e-84	238	e-83	238	e-82	238	e-81	238	e-80	238	e-79	238	e-78	238	e-77	238	e-76	238	e-75	238	e-74	238	e-73	238	e-72	238	e-71	238	e-70	238	e-69	238	e-68	238	e-67	238	e-66	238	e-65	238	e-64	238	e-63	238	e-62	238	e-61	238	e-60	238	e-59	238	e-58	238	e-57	238	e-56	238	e-55	238	e-54	238	e-53	238	e-52	238	e-51	238	e-50	238	e-49	238	e-48	238	e-47	238	e-46	238	e-45	238	e-44	238	e-43	238	e-42	238	e-41	238	e-40	238	e-39	238	e-38	238	e-37	238	e-36	238	e-35	238	e-34	238	e-33	238	e-32	238	e-31	238	e-30	238	e-29	238	e-28	238	e-27	238	e-26	238	e-25	238	e-24	238	e-23	238	e-22	238	e-21	238	e-20	238	e-19	238	e-18	238	e-17	238	e-16	238	e-15	238	e-14	238	e-13	238	e-12	238	e-11	238	e-10	238	e-9	238	e-8	238	e-7	238	e-6	238	e-5	238	e-4	238	e-3	238	e-2	238	e-1
Score	E																																																																																																																																																																																																																																				
(bits)	Value																																																																																																																																																																																																																																				
802	e-112																																																																																																																																																																																																																																				
238	e-111																																																																																																																																																																																																																																				
238	e-110																																																																																																																																																																																																																																				
238	e-109																																																																																																																																																																																																																																				
238	e-108																																																																																																																																																																																																																																				
238	e-107																																																																																																																																																																																																																																				
238	e-106																																																																																																																																																																																																																																				
238	e-105																																																																																																																																																																																																																																				
238	e-104																																																																																																																																																																																																																																				
238	e-103																																																																																																																																																																																																																																				
238	e-102																																																																																																																																																																																																																																				
238	e-101																																																																																																																																																																																																																																				
238	e-100																																																																																																																																																																																																																																				
238	e-99																																																																																																																																																																																																																																				
238	e-98																																																																																																																																																																																																																																				
238	e-97																																																																																																																																																																																																																																				
238	e-96																																																																																																																																																																																																																																				
238	e-95																																																																																																																																																																																																																																				
238	e-94																																																																																																																																																																																																																																				
238	e-93																																																																																																																																																																																																																																				
238	e-92																																																																																																																																																																																																																																				
238	e-91																																																																																																																																																																																																																																				
238	e-90																																																																																																																																																																																																																																				
238	e-89																																																																																																																																																																																																																																				
238	e-88																																																																																																																																																																																																																																				
238	e-87																																																																																																																																																																																																																																				
238	e-86																																																																																																																																																																																																																																				
238	e-85																																																																																																																																																																																																																																				
238	e-84																																																																																																																																																																																																																																				
238	e-83																																																																																																																																																																																																																																				
238	e-82																																																																																																																																																																																																																																				
238	e-81																																																																																																																																																																																																																																				
238	e-80																																																																																																																																																																																																																																				
238	e-79																																																																																																																																																																																																																																				
238	e-78																																																																																																																																																																																																																																				
238	e-77																																																																																																																																																																																																																																				
238	e-76																																																																																																																																																																																																																																				
238	e-75																																																																																																																																																																																																																																				
238	e-74																																																																																																																																																																																																																																				
238	e-73																																																																																																																																																																																																																																				
238	e-72																																																																																																																																																																																																																																				
238	e-71																																																																																																																																																																																																																																				
238	e-70																																																																																																																																																																																																																																				
238	e-69																																																																																																																																																																																																																																				
238	e-68																																																																																																																																																																																																																																				
238	e-67																																																																																																																																																																																																																																				
238	e-66																																																																																																																																																																																																																																				
238	e-65																																																																																																																																																																																																																																				
238	e-64																																																																																																																																																																																																																																				
238	e-63																																																																																																																																																																																																																																				
238	e-62																																																																																																																																																																																																																																				
238	e-61																																																																																																																																																																																																																																				
238	e-60																																																																																																																																																																																																																																				
238	e-59																																																																																																																																																																																																																																				
238	e-58																																																																																																																																																																																																																																				
238	e-57																																																																																																																																																																																																																																				
238	e-56																																																																																																																																																																																																																																				
238	e-55																																																																																																																																																																																																																																				
238	e-54																																																																																																																																																																																																																																				
238	e-53																																																																																																																																																																																																																																				
238	e-52																																																																																																																																																																																																																																				
238	e-51																																																																																																																																																																																																																																				
238	e-50																																																																																																																																																																																																																																				
238	e-49																																																																																																																																																																																																																																				
238	e-48																																																																																																																																																																																																																																				
238	e-47																																																																																																																																																																																																																																				
238	e-46																																																																																																																																																																																																																																				
238	e-45																																																																																																																																																																																																																																				
238	e-44																																																																																																																																																																																																																																				
238	e-43																																																																																																																																																																																																																																				
238	e-42																																																																																																																																																																																																																																				
238	e-41																																																																																																																																																																																																																																				
238	e-40																																																																																																																																																																																																																																				
238	e-39																																																																																																																																																																																																																																				
238	e-38																																																																																																																																																																																																																																				
238	e-37																																																																																																																																																																																																																																				
238	e-36																																																																																																																																																																																																																																				
238	e-35																																																																																																																																																																																																																																				
238	e-34																																																																																																																																																																																																																																				
238	e-33																																																																																																																																																																																																																																				
238	e-32																																																																																																																																																																																																																																				
238	e-31																																																																																																																																																																																																																																				
238	e-30																																																																																																																																																																																																																																				
238	e-29																																																																																																																																																																																																																																				
238	e-28																																																																																																																																																																																																																																				
238	e-27																																																																																																																																																																																																																																				
238	e-26																																																																																																																																																																																																																																				
238	e-25																																																																																																																																																																																																																																				
238	e-24																																																																																																																																																																																																																																				
238	e-23																																																																																																																																																																																																																																				
238	e-22																																																																																																																																																																																																																																				
238	e-21																																																																																																																																																																																																																																				
238	e-20																																																																																																																																																																																																																																				
238	e-19																																																																																																																																																																																																																																				
238	e-18																																																																																																																																																																																																																																				
238	e-17																																																																																																																																																																																																																																				
238	e-16																																																																																																																																																																																																																																				
238	e-15																																																																																																																																																																																																																																				
238	e-14																																																																																																																																																																																																																																				
238	e-13																																																																																																																																																																																																																																				
238	e-12																																																																																																																																																																																																																																				
238	e-11																																																																																																																																																																																																																																				
238	e-10																																																																																																																																																																																																																																				
238	e-9																																																																																																																																																																																																																																				
238	e-8																																																																																																																																																																																																																																				
238	e-7																																																																																																																																																																																																																																				
238	e-6																																																																																																																																																																																																																																				
238	e-5																																																																																																																																																																																																																																				
238	e-4																																																																																																																																																																																																																																				
238	e-3																																																																																																																																																																																																																																				
238	e-2																																																																																																																																																																																																																																				
238	e-1																																																																																																																																																																																																																																				
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION :</p> <ul style="list-style-type: none">-plus de 50pb-pas de chevauchement de plus de 30pb-même direction que le gène aval																																																																																																																																																																																																																																				
DECISION:	OUI																																																																																																																																																																																																																																				

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 1 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :1																																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Start #</th><th>Raw Score</th><th>SD</th><th>Genomic Z Value</th><th>Spacer Distance</th><th>Final Score</th><th>Sequence of the Region Upstream of the Start</th><th>Start Codon</th><th>Start Position</th><th>ORF Length</th></tr><tr><td>1</td><td>-3,026</td><td>2,393</td><td>40394</td><td>-3,801</td><td></td><td></td><td>ATG</td><td>1</td><td>624</td></tr><tr><td>2</td><td>-3,745</td><td>2,047</td><td>8</td><td>-4,967</td><td></td><td>GTCTGGCTGTAAAGCGGACGGCC</td><td>ATG</td><td>103</td><td>522</td></tr><tr><td>3</td><td>-4,705</td><td>1,597</td><td>9</td><td>-5,480</td><td></td><td>GACGGCCATGCGCTGGTCTGACC</td><td>GTG</td><td>118</td><td>507</td></tr><tr><td>4</td><td>-6,201</td><td>0,897</td><td>13</td><td>-7,247</td><td></td><td>CATCGGCACGGCTGACCAGTTGG</td><td>GTG</td><td>346</td><td>279</td></tr><tr><td>5</td><td>-4,141</td><td>1,862</td><td>13</td><td>-5,187</td><td></td><td>GAGACGGAAGAAGTGGCTGGC</td><td>GTG</td><td>415</td><td>210</td></tr><tr><td>6</td><td>-4,013</td><td>1,921</td><td>14</td><td>-5,360</td><td></td><td>GGTTGACACGAAGCGTCTGGCC</td><td>ATG</td><td>529</td><td>96</td></tr></table> <p>start : 1 longueur: 624pb nature: ATG score: -3.801 z=2.383 Longueur la plus importante, bon score, z-value parmi les plus élevées</p>	Start #	Raw Score	SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length	1	-3,026	2,393	40394	-3,801			ATG	1	624	2	-3,745	2,047	8	-4,967		GTCTGGCTGTAAAGCGGACGGCC	ATG	103	522	3	-4,705	1,597	9	-5,480		GACGGCCATGCGCTGGTCTGACC	GTG	118	507	4	-6,201	0,897	13	-7,247		CATCGGCACGGCTGACCAGTTGG	GTG	346	279	5	-4,141	1,862	13	-5,187		GAGACGGAAGAAGTGGCTGGC	GTG	415	210	6	-4,013	1,921	14	-5,360		GGTTGACACGAAGCGTCTGGCC	ATG	529	96
Start #	Raw Score	SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length																																																														
1	-3,026	2,393	40394	-3,801			ATG	1	624																																																														
2	-3,745	2,047	8	-4,967		GTCTGGCTGTAAAGCGGACGGCC	ATG	103	522																																																														
3	-4,705	1,597	9	-5,480		GACGGCCATGCGCTGGTCTGACC	GTG	118	507																																																														
4	-6,201	0,897	13	-7,247		CATCGGCACGGCTGACCAGTTGG	GTG	346	279																																																														
5	-4,141	1,862	13	-5,187		GAGACGGAAGAAGTGGCTGGC	GTG	415	210																																																														
6	-4,013	1,921	14	-5,360		GGTTGACACGAAGCGTCTGGCC	ATG	529	96																																																														
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start 1 permet une ORF de 624pb																																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Oui</p>																																																																						

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Score = 402 bits (1032), Expect = e-112 Identities = 207/207 (100%), Positives = 207/207 (100%)</p> <p>Query: 1 MPSQQTSDQDGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV 60 MPSQQTSDQDGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV Sbjct: 1 MPSQQTSDQDGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV 60</p> <p>Query: 61 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE 120 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE Sbjct: 61 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE 120</p> <p>Query: 121 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA 180 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA Sbjct: 121 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA 180</p> <p>Query: 181 RETARILKAFPGISVDEVAHEVAKRAS 207 RETARILKAFPGISVDEVAHEVAKRAS Sbjct: 181 RETARILKAFPGISVDEVAHEVAKRAS 207</p> <p>>Nucci_1, function unknown, 207 Length = 207</p> <p>Score = 398 bits (1023), Expect = e-111 Identities = 205/207 (99%), Positives = 206/207 (99%)</p> <p>Query: 1 MPSQQTSDQDGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV 60 MPSQQT+ DQGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV Sbjct: 1 MPSQQTTPDQGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV 60</p> <p>Query: 61 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE 120 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE Sbjct: 61 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE 120</p> <p>Query: 121 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA 180 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA Sbjct: 121 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA 180</p> <p>Query: 181 RETARILKAFPGISVDEVAHEVAKRAS 207 RETARILKAFPGISVDEVAHEVAKRAS Sbjct: 181 RETARILKAFPGISVDEVAHEVAKRAS 207</p> <p>>Quartz_1, function unknown, 207 Length = 207</p> <p>Score = 394 bits (1013), Expect = e-110 Identities = 203/207 (98%), Positives = 205/207 (99%)</p> <p>Query: 1 MPSQQTSDQDGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV 60 MPSQQ + DQGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV Sbjct: 1 MPSQQTTPDQGSKPSRRRVCKALTTSGSRCKRTAMPGLTVCKSHGGGTAASVRKSTKARV 60</p> <p>Query: 61 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE 120 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE Sbjct: 61 SQQAQLWGISDAGSISVEEELQKLARNKLTIIALRTYISEGTRSHIGTLTSSVEDTE 120</p> <p>Query: 121 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA 180 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA Sbjct: 121 QEVGEETYQTRKTKKSPGVSPHWQELHKAQELVAILRLLLQEVTTGGTDEVOTKRLRMQTA 180</p> <p>Query: 181 RETARILKAFPGISVDEVAHEVAKRAS 207</p> <p>Start 1 conservé</p>
DECISION:	<p>1:</p> <ul style="list-style-type: none">-meilleure longueur-chevauchement de moins de 30pb (et donc proximité des ORF)-bon score et Z value-conservé chez des homologues

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Phage Db :

Sequences producing significant alignments:

score

E

(bits) Value

Carostasia_Draft_1, function unknown, 207

402

e-112

Nucci_1, function unknown, 207

398

e-111

Quartz_1, function unknown, 207

394

e-110

Mandalorian_1, function unknown, 207

394

e-110

Yuuy_1, function unknown, 206

358

e-102

Juanyo_1, function unknown, 173

325

3e-89

Schubert_1, function unknown, 205

286

2e-77

Ixel_1, function unknown, 209

250

1e-66

Zenitsu_1, function unknown, 169

249

3e-66

Shamu_1, function unknown, 169

249

4e-66

Saratos_1, function unknown, 169

249

4e-66

Eleri_1, function unknown, 169

249

4e-66

ColaCorta_1, function unknown, 169

249

4e-66

ChikPic_1, function unknown, 169

249

4e-66

Andromedas_1, function unknown, 169

249

4e-66

Lucky3_1, function unknown, 204

248

5e-66

Golden_1, function unknown, 204

248

5e-66

Pikmin_1, function unknown, 208

248

9e-66

Pajaza_1, function unknown, 208

248

9e-66

Casey_1, function unknown, 208

248

9e-66

Finny_1, function unknown, 167

246

2e-65

SirVictor_1, function unknown, 204

246

3e-65

Guetzie_1, function unknown, 204

246

3e-65

Sansa_62, function unknown, 169

244

8e-65

BouleyBill_Draft_1, function unknown, 204

240

1e-63

Sinatra_1, function unknown, 204

240

2e-63

PrincePhergus_1, function unknown, 204

240

2e-63

Pherbot_1, function unknown, 204

240

2e-63

Koji_1, function unknown, 204

240

2e-63

Query= Carostasia_1

(207 letters)

Distribution of 100 Blast Hits on the Query Sequence

395887 Nucci_1, function unknown, 207, 5-398 E=4e-111

Color Key for Alignment Scores

40-50

50-60

60-70

70-80

80-90

90-100

100

50

200

150

200

100

50

200

150

200

nr :

Go to alignment for hypothetical protein SEA_XUUY_1 (Microbacterium phage Yuuy)

Description

Scientific Name

Max Score

Total Score

Query Cover %

E value

Per Ident %

Ass. Len

Accession

☐ hypothetical protein SEA_NUCCI_1 (Microbacterium phage Nucci)

Microbacterium phage Nucci

414

414

100%

2e-145

99.03%

207

OC802394.1

☐ hypothetical protein SEA_MANDALORIAN_1 (Microbacterium phage Mandalorian)

Microbacterium phage Mandalorian

410

410

100%

5e-144

98.07%

207

QX013595.1

☐ hypothetical protein SEA_YUUY_1 (Microbacterium phage Yuuy)

Microbacterium phage Yuuy

381

381

100%

2e-132

92.27%

206

QFP96330.1

☐ hypothetical protein SEA_JUANYO_1 (Microbacterium phage Juanyo)

Microbacterium phage Juanyo

339

339

83%

2e-116

97.11%

173

LSH44752.1

☐ hypothetical protein HCV00_0e01 (Microbacterium phage Schubert)

Microbacterium phage Schubert

293

293

99%

1e-97

71.36%

205

YP_00818833.1

☐ hypothetical protein SEA_ZENITSU_1 (Microbacterium phage Zenitsu)

Microbacterium phage Zenitsu

258

258

82%

2e-84

78.24%

169

WNN93885.1

☐ hypothetical protein FDIJ4_0e01 (Microbacterium phage Eleri)

Microbacterium phage Eleri

257

257

82%

3e-84

78.95%

169

YP_009623039.1

☐ hypothetical protein SEA_FINNY_1 (Microbacterium phage Finny)

Microbacterium phage Finny

255

255

81%

3e-83

78.57%

167

QGF18730.1

☐ hypothetical protein FDI42_0e01 (Microbacterium phage Golden)

Microbacterium phage Golden

256

256

99%

3e-83

83.77%

204

YP_009624162.1

☐ hypothetical protein SEA_SIRVICTOR_1 (Microbacterium phage SirVictor)

Microbacterium phage SirVictor

254

254

99%

2e-82

83.29%

204

WNN93791.1

☐ hypothetical protein SEA_SANSA_62 (Microbacterium phage Sansa)

Microbacterium phage Sansa

253

253

82%

3e-82

75.88%

169

OC536957.1

☐ hypothetical protein FDI44_0e01 (Microbacterium phage Pherbot)

Microbacterium phage Pherbot

250

250

90%

1e-80

70.05%

208

YP_009624258.1

☐ hypothetical protein FDI43_0e01 (Microbacterium phage Koji)

Microbacterium phage Koji

247

247

99%

1e-79

80.87%

204

YP_009624200.1

☐ hypothetical protein SEA_KAUJALA_1 (Microbacterium phage Kavala)

Microbacterium phage Kavala

247

247

99%

2e-79

80.39%

204

QNL31013.1

☐ hypothetical protein ISEL_1 (Microbacterium phage Ixel)

Microbacterium phage Ixel

239

239

100%

3e-76

68.20%

209

QHN87702.1

☐ hypothetical protein SEA_CHICKENKING_1 (Microbacterium phage ChickenKing)

Microbacterium phage ChickenKing

234

234

88%

9e-75

63.04%

181

QF506871.1

☐ hypothetical protein SEA_CHEETOT_1 (Microbacterium phage Cheetot)

Microbacterium phage Cheetot

233

233

88%

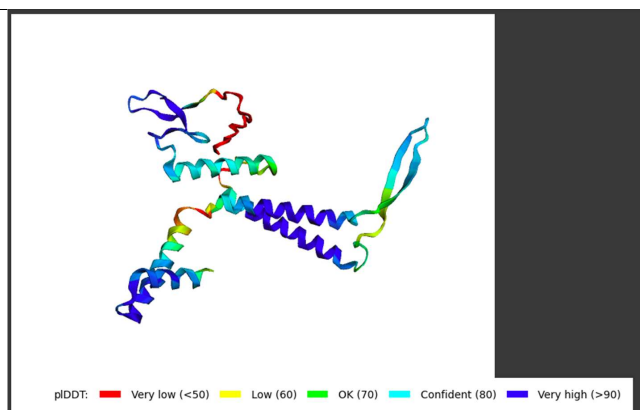
3e-74

63.59%

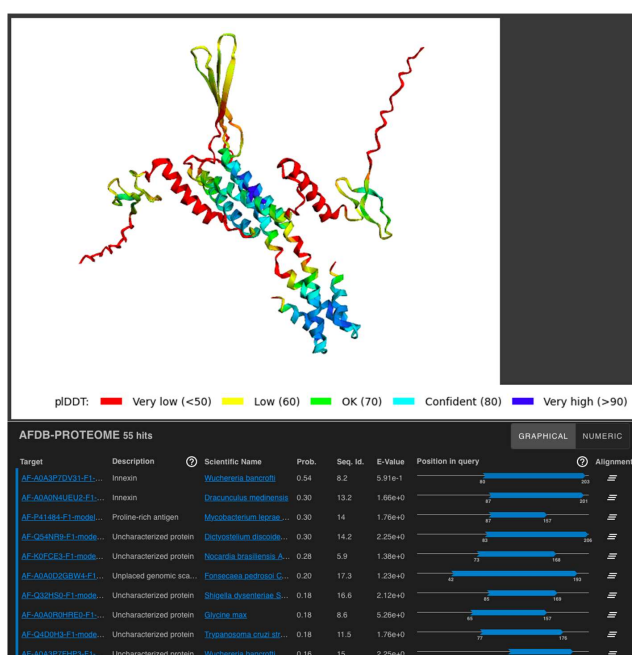
181

LVG34587.1

	<div><div>Carostasia_Draft (EA10)</div><div>Mandalorian (EA10)</div><div>Nucci (EA10)</div></div>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div><div>DeepTMHMM - Most Likely Topology Type: Globular</div><div>DeepTMHMM - Posterior Probabilities</div></div> <div>Non, la protéine prédite serait globulaire.</div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Les protéines hypothétiques sont approuvées par SEA-PHAGES.
DECISION:	NKF
Alphafolds	J'ai utilisé le logiciel AlphaFold afin de voir si une structure secondaire fiable pouvait être prédite, le logiciel est confiant sur une structure, néanmoins celle ci est "étendue" et ne pourrait réellement exister que si elle était transmembranaire, mais deep TMHMM assure qu'elle ne l'est pas. Ce serait une protéine globulaire, on peut émettre l'hypothèse qu'elle serait en fait fonctionnelle sous forme de polymère. Structure monomérique :



Structure dimérique :



J'ai ensuite cherché des homologues de la structure dimérique grâce à alphaseek mais aucun résultat n'est concluant.

Rien de concluant avec Dalisearch non plus.

Summary

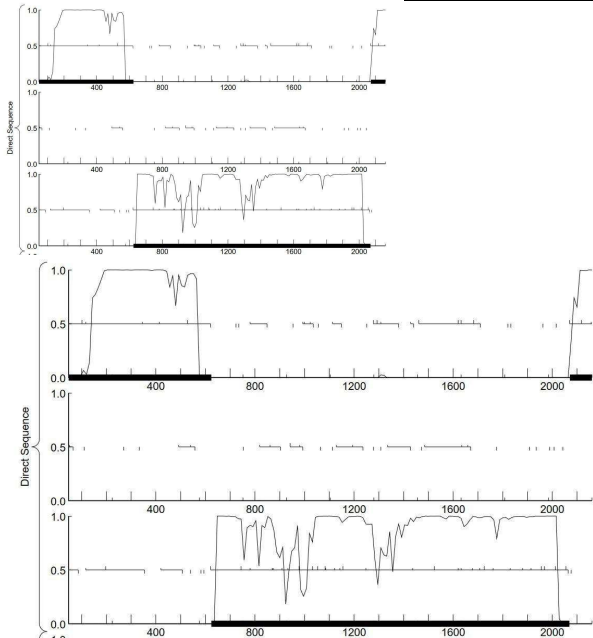
No:	Chain	Z	rmsd	lali	nres	lid	PDB	Description
1:	8Emv-D	8.9	8.3	129	665	7	PDB	MOLECULE: PROBABLE MULTIDRUG RESISTANCE ABC TRANSPORTER ATP
2:	7a59a-C	8.7	2.2	61	68	8	PDB	MOLECULE: EMC8;
3:	8pm4-A	8.5	23.0	114	604	11	PDB	MOLECULE: TRANSPOSASE;
4:	8ap9-G	8.4	3.7	82	279	5	PDB	MOLECULE: ATP SYNTHASE GAMMA SUBUNIT;
5:	7xp4-R	8.4	9.5	95	299	8	PDB	MOLECULE: GUANINE NUCLEOTIDE-BINDING PROTEIN G(T) SUBUNIT A
6:	6ek4-A	8.4	5.7	89	342	10	PDB	MOLECULE: PAB3;
7:	8j80-A	8.3	8.0	80	290	13	PDB	MOLECULE: ZINC TRANSPORTER 7;
8:	5yke-B	8.3	15.4	89	845	9	PDB	MOLECULE: ATP-SENSITIVE INWARD RECTIFIER POTASSIUM CHANNEL
9:	6n9h-A	8.2	2.1	63	75	14	PDB	MOLECULE: AMANTADINE-BINDING PROTEIN;
10:	5h5e-A	8.2	2.6	70	98	9	PDB	MOLECULE: PFC_05175;
11:	3mav-A	8.2	10.6	93	352	8	PDB	MOLECULE: NUCLEAR IMPORT ADAPTOR, NRO1;
12:	6s2e-A	8.1	20.7	97	1006	5	PDB	MOLECULE: DNA POLYMERASE EPSILON CATALYTIC SUBUNIT A;
13:	5h69-A	8.0	3.3	77	252	5	PDB	MOLECULE: CHROMOSOME PARTITION PROTEIN SMC;
14:	7m7-A	8.0	4.1	86	546	3	PDB	MOLECULE: FLAGELLIN;
15:	6s80-C	7.9	3.8	70	574	11	PDB	MOLECULE: FLAGELLIN A;
16:	8gji-A	7.7	5.5	70	167	9	PDB	MOLECULE: GCG BINDER;
17:	8ij9-C	7.6	2.9	62	71	15	PDB	MOLECULE: RAS-RELATED PROTEIN RAB-6B;
18:	7vvg-D	7.6	2.9	59	60	14	PDB	MOLECULE: FCCA;
19:	1cii-A	7.6	5.9	97	602	12	PDB	MOLECULE: COLICIN 1A;
20:	6rw8-B	7.6	10.2	110	2337	14	PDB	MOLECULE: A COMPONENT OF INSECTICIDAL TOXIN COMPLEX (TC);
21:	7mi6-A	7.5	16.7	96	2419	9	PDB	MOLECULE: FUSION PROTEIN OF DYNEIN AND ENDOLYSIN;
22:	6d8e-A	7.5	2.2	64	77	14	PDB	MOLECULE: DSD131 A;
23:	6clb-A	7.5	13.2	107	969	4	PDB	MOLECULE: O6-ALKYLGUANINE-DNA ALKYLTRANSFERASE MUTANT,DYMC2
24:	6wjy-A	7.4	24.4	108	1951	7	PDB	MOLECULE: DNA POLYMERASE EPSILON CATALYTIC SUBUNIT A;
25:	6v9e-A	7.4	13.0	88	715	3	PDB	MOLECULE: ABC-TYPE BACTERIOCIN TRANSPORTER;
26:	6lcs-A	7.3	8.9	83	377	10	PDB	MOLECULE: SOLOUT CHARACTER FAMILY 39 (IRON-REGULATED TRANSPOR
27:	3ets-B	7.3	3.8	75	115	8	PDB	MOLECULE: ADHESIN A;
28:	3p8e-F	7.3	4.4	89	156	6	PDB	MOLECULE: CYTOPLASMIC FMRI-INTERACTING PROTEIN 1;
29:	6tpi-A	7.2	6.7	74	380	14	PDB	MOLECULE: MUREIN HYDROLASE ACTIVATOR ENVG;
30:	3ero-A	7.1	7.3	71	336	6	PDB	MOLECULE: GODEF DOMAIN PROTEIN;
31:	2g8D-D	7.1	3.9	82	162	4	PDB	MOLECULE: CHARGED MULTIVESICULAR BODY PROTEIN 3;

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	2
Coordonnées du Stop	2069
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 1
Coordonnées du Start retenu	624
Fonction prédite	Terminase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Potentiel codant détecté par Genemark sur toute la longueur de l'ORF.</i>

																																					
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>OUI</div> <div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_2, function unknown, 481</td><td>965</td><td>0.0</td></tr><tr><td>Mandalorian_2, terminase, 481</td><td>962</td><td>0.0</td></tr><tr><td>Quartz_2, terminase, 481</td><td>960</td><td>0.0</td></tr><tr><td>Nucci_2, terminase, 481</td><td>957</td><td>0.0</td></tr><tr><td>Juanyo_2, terminase, 480</td><td>931</td><td>0.0</td></tr><tr><td>YuuY_2, terminase, 481</td><td>930</td><td>0.0</td></tr><tr><td>Finny_2, terminase, 482</td><td>842</td><td>0.0</td></tr><tr><td>Shamu_Draft_2, function unknown, 482</td><td>840</td><td>0.0</td></tr><tr><td>Sansa_1, terminase, 482</td><td>839</td><td>0.0</td></tr><tr><td>Saratos_2, terminase, 482</td><td>839</td><td>0.0</td></tr><tr><td>Glamour_1, terminase, 482</td><td>838</td><td>0.0</td></tr></table>		Score (bits)	E Value	Carostasia_Draft_2, function unknown, 481	965	0.0	Mandalorian_2, terminase, 481	962	0.0	Quartz_2, terminase, 481	960	0.0	Nucci_2, terminase, 481	957	0.0	Juanyo_2, terminase, 480	931	0.0	YuuY_2, terminase, 481	930	0.0	Finny_2, terminase, 482	842	0.0	Shamu_Draft_2, function unknown, 482	840	0.0	Sansa_1, terminase, 482	839	0.0	Saratos_2, terminase, 482	839	0.0	Glamour_1, terminase, 482	838	0.0
	Score (bits)	E Value																																			
Carostasia_Draft_2, function unknown, 481	965	0.0																																			
Mandalorian_2, terminase, 481	962	0.0																																			
Quartz_2, terminase, 481	960	0.0																																			
Nucci_2, terminase, 481	957	0.0																																			
Juanyo_2, terminase, 480	931	0.0																																			
YuuY_2, terminase, 481	930	0.0																																			
Finny_2, terminase, 482	842	0.0																																			
Shamu_Draft_2, function unknown, 482	840	0.0																																			
Sansa_1, terminase, 482	839	0.0																																			
Saratos_2, terminase, 482	839	0.0																																			
Glamour_1, terminase, 482	838	0.0																																			

Distribution of 100 Blast Hits on the Query Sequence

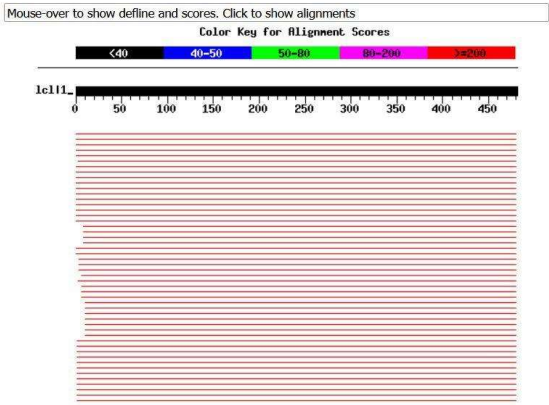
Mouse-over to show define and scores. Click to show alignments

Color Key for Alignment Scores

<40	40-50	50-60	60-700	>7000
-----	-------	-------	--------	-------

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																																																																																																				
Quel start Glimmer et GeneMark suggèrent-ils ?	624 pour glimmer et genemark																																																																																																																																																																																																																																																																																																																																				
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<p>Le start retenu est le même que celui proposé par Glimmer et Genemark car celui-ci correspond à l'ORF la plus grande et correspond aux critères de chevauchement/espaces entre les ORFs. Le final score n'est pas le plus proche de 0 et la Z-value n'est pas la plus élevée, mais ce start correspond également au début du potentiel codant prédit par GeneMark.</p> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,608</td><td>1,643</td><td>10</td><td>-5,303</td><td>AAGTTGGGAGCGGGCCAGCTG</td><td>ATG</td><td>624</td><td>1446</td></tr><tr><td>2</td><td>-3,985</td><td>1,935</td><td>16</td><td>-5,781</td><td>GACCGAGGTCTGACGTCCACA</td><td>ATG</td><td>747</td><td>1323</td></tr><tr><td>3</td><td>-7,368</td><td>0,351</td><td>10</td><td>-8,063</td><td>TCTCGACACCCCGCTCGGCTGT</td><td>GTG</td><td>783</td><td>1287</td></tr><tr><td>4</td><td>-5,112</td><td>1,407</td><td>5</td><td>-7,112</td><td>CGTTCACATCGGCTGGCAG</td><td>TTG</td><td>849</td><td>1221</td></tr><tr><td>5</td><td>-5,805</td><td>1,083</td><td>10</td><td>-6,500</td><td>GATTGCAATCCGTGGCTCAAG</td><td>GTG</td><td>873</td><td>1197</td></tr><tr><td>6</td><td>-5,702</td><td>1,131</td><td>11</td><td>-6,459</td><td>TGCAATCCGTGGCTCAAGGTG</td><td>GTG</td><td>876</td><td>1194</td></tr><tr><td>7</td><td>-3,778</td><td>2,031</td><td>11</td><td>-4,535</td><td>CTCGAGGCCAGGGCTAACAGC</td><td>GTG</td><td>912</td><td>1158</td></tr><tr><td>8</td><td>-5,724</td><td>1,120</td><td>13</td><td>-6,770</td><td>GTCTCGACAGATCCGCGGCTG</td><td>GTG</td><td>936</td><td>1134</td></tr><tr><td>9</td><td>-4,502</td><td>1,693</td><td>11</td><td>-5,259</td><td>GTCGGGACAGCAGCAACGTCCAG</td><td>ATG</td><td>981</td><td>1089</td></tr><tr><td>10</td><td>-4,489</td><td>1,699</td><td>9</td><td>-5,264</td><td>GGCTGCTGGTGGTGGTGGCT</td><td>ATG</td><td>1035</td><td>1035</td></tr><tr><td>11</td><td>-5,319</td><td>1,310</td><td>15</td><td>-6,921</td><td>TATGACCGGCTTCGGTTCGGAC</td><td>ATG</td><td>1056</td><td>1014</td></tr><tr><td>12</td><td>-4,960</td><td>1,478</td><td>16</td><td>-6,756</td><td>CATTGACGCCCATCAAGGAC</td><td>ATG</td><td>1086</td><td>984</td></tr><tr><td>13</td><td>-2,757</td><td>2,509</td><td>7</td><td>-4,280</td><td>TGACGACCCCATCAAGGACATG</td><td>GTG</td><td>1089</td><td>981</td></tr><tr><td>14</td><td>-5,550</td><td>1,202</td><td>9</td><td>-6,325</td><td>CGAGCTCATCCGTGAAGACACA</td><td>GTG</td><td>1125</td><td>945</td></tr><tr><td>15</td><td>-4,000</td><td>1,928</td><td>14</td><td>-5,347</td><td>AGTGGAGTGGTTGAGTCCGTT</td><td>GTG</td><td>1146</td><td>924</td></tr><tr><td>16</td><td>-5,741</td><td>1,112</td><td>10</td><td>-6,436</td><td>CAGCTCGGTTGTGCTGACCCGT</td><td>ATG</td><td>1158</td><td>912</td></tr><tr><td>17</td><td>-6,193</td><td>0,901</td><td>13</td><td>-7,238</td><td>GAGGCGGCTTGACGCCAGCTTC</td><td>GTG</td><td>1251</td><td>819</td></tr><tr><td>18</td><td>-3,398</td><td>2,209</td><td>10</td><td>-4,093</td><td>TGCCAACGGGACGGGTGACATA</td><td>TTG</td><td>1317</td><td>753</td></tr><tr><td>19</td><td>-2,654</td><td>2,558</td><td>12</td><td>-3,489</td><td>ACTTTACCAAGGTGACCCACAG</td><td>GTG</td><td>1440</td><td>630</td></tr><tr><td>20</td><td>-4,775</td><td>1,565</td><td>7</td><td>-6,297</td><td>CCTGGACCGGCGCAAGGGCAGC</td><td>ATG</td><td>1527</td><td>543</td></tr><tr><td>21</td><td>-3,800</td><td>2,021</td><td>10</td><td>-4,495</td><td>CATGCTCAGGCTGGACCGGCGCC</td><td>TTG</td><td>1548</td><td>522</td></tr><tr><td>22</td><td>-4,855</td><td>1,527</td><td>9</td><td>-5,629</td><td>GAGGGCCAGCAGCGGCGACTAT</td><td>GTG</td><td>1623</td><td>447</td></tr><tr><td>23</td><td>-4,070</td><td>1,895</td><td>6</td><td>-5,815</td><td>CGGCGACTATGTGGCCGACAT</td><td>GTG</td><td>1635</td><td>435</td></tr><tr><td>24</td><td>-3,958</td><td>1,947</td><td>17</td><td>-5,958</td><td>CRAGTGGGTTCTCGTTGACCGC</td><td>GTG</td><td>1680</td><td>390</td></tr><tr><td>25</td><td>-5,296</td><td>1,321</td><td>7</td><td>-6,818</td><td>CCGGTTCACCTTCACCGAGACC</td><td>GTG</td><td>1710</td><td>360</td></tr><tr><td>26</td><td>-3,014</td><td>2,389</td><td>10</td><td>-3,709</td><td>GACCGTGACCCAGGTACAGAG</td><td>ATG</td><td>1728</td><td>342</td></tr><tr><td>27</td><td>-4,553</td><td>1,669</td><td>8</td><td>-5,775</td><td>GCCCCAGACGACCGGGCTCTAC</td><td>GTG</td><td>1764</td><td>306</td></tr><tr><td>28</td><td>-5,524</td><td>1,027</td><td>11</td><td>-6,681</td><td>CGGGGCTGCACTGATCAAGCGG</td><td>GTG</td><td>1833</td><td>237</td></tr><tr><td>29</td><td>-5,918</td><td>1,030</td><td>16</td><td>-7,714</td><td>GGGCTGGGCATCCAGCGGCTC</td><td>GTG</td><td>1884</td><td>186</td></tr><tr><td>30</td><td>-6,193</td><td>0,901</td><td>7</td><td>-7,716</td><td>CACGTTGGCATCTTGCACACC</td><td>GTG</td><td>1917</td><td>153</td></tr><tr><td>31</td><td>-2,611</td><td>2,577</td><td>18</td><td>-4,912</td><td>TGAAGGGGACACCGGCTCGCC</td><td>ATG</td><td>1959</td><td>111</td></tr><tr><td>32</td><td>-5,309</td><td>1,315</td><td>8</td><td>-6,530</td><td>CGGCTCGGCATGCTGAACGAG</td><td>ATG</td><td>1971</td><td>99</td></tr><tr><td>33</td><td>-5,472</td><td>1,239</td><td>13</td><td>-6,518</td><td>GCAAGACGACGAGTTGACGCC</td><td>ATG</td><td>2016</td><td>54</td></tr><tr><td>34</td><td>-5,545</td><td>1,204</td><td>13</td><td>-6,591</td><td>GGGCGCAATTGACTACTTCAAG</td><td>ATG</td><td>2058</td><td>12</td></tr></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,608	1,643	10	-5,303	AAGTTGGGAGCGGGCCAGCTG	ATG	624	1446	2	-3,985	1,935	16	-5,781	GACCGAGGTCTGACGTCCACA	ATG	747	1323	3	-7,368	0,351	10	-8,063	TCTCGACACCCCGCTCGGCTGT	GTG	783	1287	4	-5,112	1,407	5	-7,112	CGTTCACATCGGCTGGCAG	TTG	849	1221	5	-5,805	1,083	10	-6,500	GATTGCAATCCGTGGCTCAAG	GTG	873	1197	6	-5,702	1,131	11	-6,459	TGCAATCCGTGGCTCAAGGTG	GTG	876	1194	7	-3,778	2,031	11	-4,535	CTCGAGGCCAGGGCTAACAGC	GTG	912	1158	8	-5,724	1,120	13	-6,770	GTCTCGACAGATCCGCGGCTG	GTG	936	1134	9	-4,502	1,693	11	-5,259	GTCGGGACAGCAGCAACGTCCAG	ATG	981	1089	10	-4,489	1,699	9	-5,264	GGCTGCTGGTGGTGGTGGCT	ATG	1035	1035	11	-5,319	1,310	15	-6,921	TATGACCGGCTTCGGTTCGGAC	ATG	1056	1014	12	-4,960	1,478	16	-6,756	CATTGACGCCCATCAAGGAC	ATG	1086	984	13	-2,757	2,509	7	-4,280	TGACGACCCCATCAAGGACATG	GTG	1089	981	14	-5,550	1,202	9	-6,325	CGAGCTCATCCGTGAAGACACA	GTG	1125	945	15	-4,000	1,928	14	-5,347	AGTGGAGTGGTTGAGTCCGTT	GTG	1146	924	16	-5,741	1,112	10	-6,436	CAGCTCGGTTGTGCTGACCCGT	ATG	1158	912	17	-6,193	0,901	13	-7,238	GAGGCGGCTTGACGCCAGCTTC	GTG	1251	819	18	-3,398	2,209	10	-4,093	TGCCAACGGGACGGGTGACATA	TTG	1317	753	19	-2,654	2,558	12	-3,489	ACTTTACCAAGGTGACCCACAG	GTG	1440	630	20	-4,775	1,565	7	-6,297	CCTGGACCGGCGCAAGGGCAGC	ATG	1527	543	21	-3,800	2,021	10	-4,495	CATGCTCAGGCTGGACCGGCGCC	TTG	1548	522	22	-4,855	1,527	9	-5,629	GAGGGCCAGCAGCGGCGACTAT	GTG	1623	447	23	-4,070	1,895	6	-5,815	CGGCGACTATGTGGCCGACAT	GTG	1635	435	24	-3,958	1,947	17	-5,958	CRAGTGGGTTCTCGTTGACCGC	GTG	1680	390	25	-5,296	1,321	7	-6,818	CCGGTTCACCTTCACCGAGACC	GTG	1710	360	26	-3,014	2,389	10	-3,709	GACCGTGACCCAGGTACAGAG	ATG	1728	342	27	-4,553	1,669	8	-5,775	GCCCCAGACGACCGGGCTCTAC	GTG	1764	306	28	-5,524	1,027	11	-6,681	CGGGGCTGCACTGATCAAGCGG	GTG	1833	237	29	-5,918	1,030	16	-7,714	GGGCTGGGCATCCAGCGGCTC	GTG	1884	186	30	-6,193	0,901	7	-7,716	CACGTTGGCATCTTGCACACC	GTG	1917	153	31	-2,611	2,577	18	-4,912	TGAAGGGGACACCGGCTCGCC	ATG	1959	111	32	-5,309	1,315	8	-6,530	CGGCTCGGCATGCTGAACGAG	ATG	1971	99	33	-5,472	1,239	13	-6,518	GCAAGACGACGAGTTGACGCC	ATG	2016	54	34	-5,545	1,204	13	-6,591	GGGCGCAATTGACTACTTCAAG	ATG	2058	12
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																																																																																																																																													
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																																																																																																																																													
1	-4,608	1,643	10	-5,303	AAGTTGGGAGCGGGCCAGCTG	ATG	624	1446																																																																																																																																																																																																																																																																																																																													
2	-3,985	1,935	16	-5,781	GACCGAGGTCTGACGTCCACA	ATG	747	1323																																																																																																																																																																																																																																																																																																																													
3	-7,368	0,351	10	-8,063	TCTCGACACCCCGCTCGGCTGT	GTG	783	1287																																																																																																																																																																																																																																																																																																																													
4	-5,112	1,407	5	-7,112	CGTTCACATCGGCTGGCAG	TTG	849	1221																																																																																																																																																																																																																																																																																																																													
5	-5,805	1,083	10	-6,500	GATTGCAATCCGTGGCTCAAG	GTG	873	1197																																																																																																																																																																																																																																																																																																																													
6	-5,702	1,131	11	-6,459	TGCAATCCGTGGCTCAAGGTG	GTG	876	1194																																																																																																																																																																																																																																																																																																																													
7	-3,778	2,031	11	-4,535	CTCGAGGCCAGGGCTAACAGC	GTG	912	1158																																																																																																																																																																																																																																																																																																																													
8	-5,724	1,120	13	-6,770	GTCTCGACAGATCCGCGGCTG	GTG	936	1134																																																																																																																																																																																																																																																																																																																													
9	-4,502	1,693	11	-5,259	GTCGGGACAGCAGCAACGTCCAG	ATG	981	1089																																																																																																																																																																																																																																																																																																																													
10	-4,489	1,699	9	-5,264	GGCTGCTGGTGGTGGTGGCT	ATG	1035	1035																																																																																																																																																																																																																																																																																																																													
11	-5,319	1,310	15	-6,921	TATGACCGGCTTCGGTTCGGAC	ATG	1056	1014																																																																																																																																																																																																																																																																																																																													
12	-4,960	1,478	16	-6,756	CATTGACGCCCATCAAGGAC	ATG	1086	984																																																																																																																																																																																																																																																																																																																													
13	-2,757	2,509	7	-4,280	TGACGACCCCATCAAGGACATG	GTG	1089	981																																																																																																																																																																																																																																																																																																																													
14	-5,550	1,202	9	-6,325	CGAGCTCATCCGTGAAGACACA	GTG	1125	945																																																																																																																																																																																																																																																																																																																													
15	-4,000	1,928	14	-5,347	AGTGGAGTGGTTGAGTCCGTT	GTG	1146	924																																																																																																																																																																																																																																																																																																																													
16	-5,741	1,112	10	-6,436	CAGCTCGGTTGTGCTGACCCGT	ATG	1158	912																																																																																																																																																																																																																																																																																																																													
17	-6,193	0,901	13	-7,238	GAGGCGGCTTGACGCCAGCTTC	GTG	1251	819																																																																																																																																																																																																																																																																																																																													
18	-3,398	2,209	10	-4,093	TGCCAACGGGACGGGTGACATA	TTG	1317	753																																																																																																																																																																																																																																																																																																																													
19	-2,654	2,558	12	-3,489	ACTTTACCAAGGTGACCCACAG	GTG	1440	630																																																																																																																																																																																																																																																																																																																													
20	-4,775	1,565	7	-6,297	CCTGGACCGGCGCAAGGGCAGC	ATG	1527	543																																																																																																																																																																																																																																																																																																																													
21	-3,800	2,021	10	-4,495	CATGCTCAGGCTGGACCGGCGCC	TTG	1548	522																																																																																																																																																																																																																																																																																																																													
22	-4,855	1,527	9	-5,629	GAGGGCCAGCAGCGGCGACTAT	GTG	1623	447																																																																																																																																																																																																																																																																																																																													
23	-4,070	1,895	6	-5,815	CGGCGACTATGTGGCCGACAT	GTG	1635	435																																																																																																																																																																																																																																																																																																																													
24	-3,958	1,947	17	-5,958	CRAGTGGGTTCTCGTTGACCGC	GTG	1680	390																																																																																																																																																																																																																																																																																																																													
25	-5,296	1,321	7	-6,818	CCGGTTCACCTTCACCGAGACC	GTG	1710	360																																																																																																																																																																																																																																																																																																																													
26	-3,014	2,389	10	-3,709	GACCGTGACCCAGGTACAGAG	ATG	1728	342																																																																																																																																																																																																																																																																																																																													
27	-4,553	1,669	8	-5,775	GCCCCAGACGACCGGGCTCTAC	GTG	1764	306																																																																																																																																																																																																																																																																																																																													
28	-5,524	1,027	11	-6,681	CGGGGCTGCACTGATCAAGCGG	GTG	1833	237																																																																																																																																																																																																																																																																																																																													
29	-5,918	1,030	16	-7,714	GGGCTGGGCATCCAGCGGCTC	GTG	1884	186																																																																																																																																																																																																																																																																																																																													
30	-6,193	0,901	7	-7,716	CACGTTGGCATCTTGCACACC	GTG	1917	153																																																																																																																																																																																																																																																																																																																													
31	-2,611	2,577	18	-4,912	TGAAGGGGACACCGGCTCGCC	ATG	1959	111																																																																																																																																																																																																																																																																																																																													
32	-5,309	1,315	8	-6,530	CGGCTCGGCATGCTGAACGAG	ATG	1971	99																																																																																																																																																																																																																																																																																																																													
33	-5,472	1,239	13	-6,518	GCAAGACGACGAGTTGACGCC	ATG	2016	54																																																																																																																																																																																																																																																																																																																													
34	-5,545	1,204	13	-6,591	GGGCGCAATTGACTACTTCAAG	ATG	2058	12																																																																																																																																																																																																																																																																																																																													
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	<p>Oui, le start prédit correspond à l'ORF la plus longue et correspond au début du potentiel codant prédit par GeneMark. Start à la position 624.</p> <p>Longueur de l'ORF : 1446pb</p> <p>Chevauchement : -1</p>																																																																																																																																																																																																																																																																																																																																				

Est-ce que le start est conservé chez les homologues voir Starterator ?	Ce gène chez Carostasia est le seul annoté dans son groupe, starterator ne peut donc pas nous aider à trancher par définition.
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Le start est conservé chez une grande majorité des homologues trouvés par BLASTp</p> <p>L'alignement commence au premier AA pour la query et la majorité des subjects.</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p>  <p>>Mandalorian_2, terminase, 481 Length = 481</p> <p>Score = 962 bits (2486), Expect = 0.0 Identities = 479/481 (99%), Positives = 480/481 (99%)</p> <p>Query: 1 MSILIEDLVEYSYEDVSEDFTLFCKANGIRESALTEVLSTMDPDALDTPLGCVAYSTPP 60 MSILIEDLVEYSYEDVSEDFTLFCKANGIRESALTEVLSTMDPDALDTPLGCVAYSTPP Sbjct: 1 MSILIEDLVEYSYEDVSEDFTLFCKANGIRESALTEVLSTMDPDALDTPLGCVAYSTPP 60</p>
DECISION:	Le start retenu est le 624. Il correspond à l'ORF la plus longue et est conservé chez la majorité des homologues.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
------------------------------------	-------------

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

PhagesDB :

>Mandalorian_2, terminase, 481
Length = 481

Score = 962 bits (2486), Expect = 0.0
Identities = 479/481 (99%), Positives = 480/481 (99%)

Query: 1 MSILIEDLVEYSYEDVSEDFTLFCKANGIRESALTEVLTSTMDPDALDTPLGCVAYSTPP 60
MSILIEDLVEYSYEDVSEDFTLFCKANGIRESALTEVLTSTMDPDALDTPLGCVAYSTPP
Sbjct: 1 MSILIEDLVEYSYEDVSEDFTLFCKANGIRESALTEVLTSTMDPDALDTPLGCVAYSTPP 60

Query: 61 QEGKTTWIVHYIAWQLIRNPWLKVVYATYSQARANSVSRQIRALVQAWTPLKSGSSINVQM 120
QEGKTTWIVHYIAWQLIRNPWLKVVYATYSQARANSVSRQIRALVQAWTPLKSGSSINVQM
Sbjct: 61 QEGKTTWIVHYIAWQLIRNPWLKVVYATYSQARANSVSRQIRALVQAWTPLKSGSSINVQM 120

Query: 121 WQTRGGGLLAAGRGSAmtGFRSDMTVIDDPIDMVEAQSELIRETTVEWFSVVLTMA 180
WQTRGGGLLAAGRGSAmtGFRSDMTVIDDPIDMVEAQSELIRETTVEWFSVVLTMA
Sbjct: 121 WQTRGGGLLAAGRGSAmtGFRSDMTVIDDPIDMVEAQSELIRETTVEWFSVVLTMA 180

Query: 181 SLSQIIIVIAIRWHKDDLIAHVQKALDATFVNIQAQTHDEHPDANGQDILGRAVGEWL 240
SLSQIIIVIAIRWHKDDLIAHVQKALDATFVNIQAQTHDEHPDANGQDILGRAVGEWL
Sbjct: 181 SLSQIIIVIAIRWHKDDLIAHVQKALDATFVNIQAQTHDEHPDANGQDILGRAVGEWL 240

Query: 241 PSVQNRDTSWALIKSAVGPPYVWQALYQGDQVTTGGSYINVDKIDVLPWEQVVYLDQRTG 300
PSVQNRDTSWALIKSAVGPPYVWQALYQGDQVTTGGSYINVDKIDVLPWEQVVYLDQRTG
Sbjct: 241 PSVQNRDTSWALIKSAVGPPYVWQALYQGDQVTTGGSYINVDKIDVLPWEQVVYLDQRTG 300

Query: 301 SMLTLDRALVIQSWDLTFGNIQNGKKKASSGDYVAGHWIVLGGTKWLVDRVHGRFTFT 360
SMLTLDRALVIQSWDLTFGNIQNGKKKASSGDYVAGHWIVLGGTKWLVDRVHGRFTFT
Sbjct: 301 SMLTLDRALVIQSWDLTFGNIQNGKKKASSGDYVAGHWIVLGGTKWLVDRVHGRFTFT 360

Query: 361 ETVTVQVKMAARWPQTSRVVVEKAANGTALLNLRKRAALIKPVTAGGSKEVRALAIQPV 420
ETVTVQVKMAARWPQTSRVVVEKAANGTALLNLRKRAALIKPVTAGGSKEVRALAIQPV
Sbjct: 361 ETVTVQVKMAARWPQTSRVVVEKAANGTALLNLRKRAALIKPVTAGGSKEVRALAIQPV 420

Query: 421 VDEGNVAILDVTLEAQDAEGATGLAMLNEMRDFPFKGHDDVDAMTQAIGQGRIDYFKMG 480
VDEGNVAILDVTLEAQDAEGATGLAMLNEMRDFPFKGHDDVDAMTQAIGQGRIDYFKMG
Sbjct: 421 VDEGNVAILDVTLEAQDAEGATGLAMLNEMRDFPFKGHDDVDAMTQAIGQGRIDYFKMG 480

Query: 481 S 481
+
Sbjct: 481 N 481

BlastP :

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession
<input checked="" type="checkbox"/>	terminase (Mandobacterium shawii (Mandobacterium))	Mandobacterium...	984	984	99%	0.0	99.98%	481	Q2013306.1
<input checked="" type="checkbox"/>	terminase (Mandobacterium shawii (Mandobacterium))	Mandobacterium...	982	982	99%	0.0	99.98%	481	U955922.1
<input checked="" type="checkbox"/>	terminase (Mandobacterium shawii (Mandobacterium))	Mandobacterium...	979	979	99%	0.0	99.17%	481	Q2013306.1

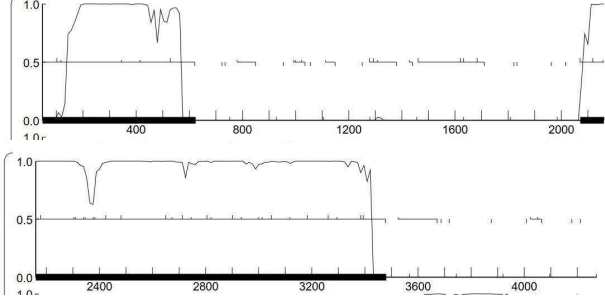
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

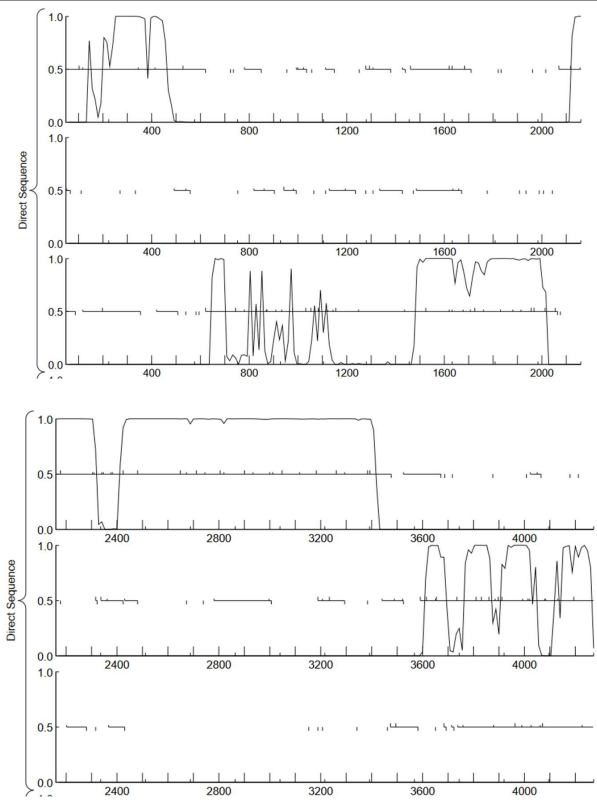
	<p>PDB :</p> <p>Template alignment Template 3D Structure PDBe</p> <div><input type="checkbox"/> 1. 8DGC_G Terminase, large subunit; phage defense, pattern-recognition receptor; nlr; stand. atpase, ANTIVIRAL PROTEIN; HET: ATP;(Salmonella enterica) Probability: 100%, E-value: 1.3e-38, Score: 317.35, Aligned cols: 455, Identities: 18%, Similarity: 0.159, Template Neff: 11.7</div> <p>UniProt :</p> <p>Template alignment</p> <div><input type="checkbox"/> 2. P10310 TERL_BPT3 Terminase, large subunit OS=Enterobacteria phage T3 OX=10759 GN=19 PE=1 SV=1 Probability: 100%, E-value: 1.9e-38, Score: 316.12, Aligned cols: 449, Identities: 16%, Similarity: 0.142, Template Neff:</div> <p>Pfam :</p> <p>Template alignment CDD</p> <div><input type="checkbox"/> 40. PF00441.2; TerL_nuclease ; Terminase large subunit, endonuclease domain Probability: 99.69%, E-value: 1.3e-14, Score: 132.27, Aligned cols: 208, Identities: 12%, Similarity: -0.025, Template Neff: 10.6</div> <p>CD :</p> <p>Template alignment</p> <div><input type="checkbox"/> 57. cd18032 DEXHc_RE_I_III_res; DEXH-box helicase domain of type III restriction enzyme res subunit. Members of this cd includes both type I and type III restriction enzymes. Probability: 98.66%, E-value: 0.0000011, Score: 72.37, Aligned cols: 151, Identities: 8%, Similarity: -0.081, Template Neff: 13.1</div>
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Oui, protéines de capsid
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	Non
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Oui, terminase
DECISION:	Terminase

QuaQStudent Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	3
Coordonnées du Stop	3480
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	gap 1
Coordonnées du Start retenu	2071
Fonction prédite	Portal

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GeneMark et GeneMarks detectent un pouvoir codant sur toute la longueur de l'ORF.</i></p> 

	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	

	<p>OUI</p> <p><u>Distribution of 100 Blast Hits on the Query Sequence</u></p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <p><40 40-50 50-80 80-200 >200</p> <p>1c111</p>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	NON
DECISION:	OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																																																																																																				
Quel start Glimmer et GeneMark suggèrent-ils ?	2071 pour les deux.																																																																																																																																																																																																																																																																																																																																				
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>OUI, Z value et Final score proches de 0.</div> <table><tr><td>Stats: 31</td><td>ORF Start : 2020</td><td>Cdn1</td><td>Cdn2</td><td>Cdn3</td><td>Length</td><td>SD Scoring Matrix</td><td>Kibler6</td><td>Explore</td></tr><tr><td>Selected: 1</td><td>ORF Stop : 3480</td><td>5' End</td><td>22.2</td><td>66.7</td><td>33.3</td><td>27</td><td>Spacing Weight Matrix</td><td>Karin Medium</td></tr><tr><td></td><td>ORF Length: 1461</td><td>3' End</td><td>62.8</td><td>41.1</td><td>88.1</td><td>1410</td><td></td><td>Document</td></tr></table> <table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,267</td><td>2,371</td><td>8</td><td>-4,488</td><td>CAGGCGATCGGTCAAGGCGGAA</td><td>TTG</td><td>2044</td><td>1497</td></tr><tr><td>2</td><td>-3,698</td><td>2,069</td><td>11</td><td>-4,465</td><td>TACTTCAGATGGAGACTAAC</td><td>ATG</td><td>2071</td><td>1410</td></tr><tr><td>3</td><td>-5,021</td><td>1,450</td><td>13</td><td>-6,067</td><td>GGATCCAGCGGCTGTCCAGC</td><td>ATG</td><td>2119</td><td>1362</td></tr><tr><td>4</td><td>-4,013</td><td>1,921</td><td>11</td><td>-4,770</td><td>CAGCGGAGCAGCAAGACTAC</td><td>GTG</td><td>2158</td><td>1323</td></tr><tr><td>5</td><td>-4,735</td><td>1,594</td><td>6</td><td>-6,479</td><td>CGTGGCTACTACAAGGCAAG</td><td>ATG</td><td>2179</td><td>1302</td></tr><tr><td>6</td><td>-3,325</td><td>2,290</td><td>10</td><td>-3,920</td><td>TGACCTGTACGGGAGAACTGT</td><td>GTG</td><td>2305</td><td>1176</td></tr><tr><td>7</td><td>-3,325</td><td>2,290</td><td>16</td><td>-5,021</td><td>GTACGGGAGAACTGTGCGCC</td><td>GTG</td><td>2311</td><td>1170</td></tr><tr><td>8</td><td>-2,611</td><td>2,577</td><td>13</td><td>-3,657</td><td>TGAGCTGAAGGGGTTCAAGAA</td><td>GTG</td><td>2341</td><td>1140</td></tr><tr><td>9</td><td>-3,590</td><td>1,932</td><td>11</td><td>-4,747</td><td>GAAGGGGTTCAAGAACTGTCT</td><td>GTG</td><td>2347</td><td>1134</td></tr><tr><td>10</td><td>-3,699</td><td>2,069</td><td>10</td><td>-4,393</td><td>TCCTTCGGCTGGGAGTGGCGG</td><td>GTG</td><td>2377</td><td>1104</td></tr><tr><td>11</td><td>-3,699</td><td>2,069</td><td>13</td><td>-4,744</td><td>TTGGCTGGGAGTGGCGGCTG</td><td>TTG</td><td>2380</td><td>1101</td></tr><tr><td>12</td><td>-3,967</td><td>1,943</td><td>7</td><td>-5,490</td><td>GGTGGGAGTGGCGGCTGTG</td><td>GTG</td><td>2383</td><td>1098</td></tr><tr><td>13</td><td>-5,454</td><td>1,247</td><td>13</td><td>-6,500</td><td>CGACTTCCCGAGGCTACAGAG</td><td>ATG</td><td>2425</td><td>1056</td></tr><tr><td>14</td><td>-4,975</td><td>1,471</td><td>14</td><td>-6,322</td><td>GTTCACCGGCTGCTCAAGGAG</td><td>ATG</td><td>2482</td><td>999</td></tr><tr><td>15</td><td>-3,620</td><td>2,105</td><td>13</td><td>-4,666</td><td>GGACGACAAAGGCTTGGGGCC</td><td>ATG</td><td>2650</td><td>831</td></tr><tr><td>16</td><td>-2,915</td><td>2,435</td><td>7</td><td>-4,438</td><td>GCTGGAGGCGCTCAGGACGGG</td><td>GTG</td><td>2674</td><td>807</td></tr><tr><td>17</td><td>-4,817</td><td>1,545</td><td>13</td><td>-5,963</td><td>CGACCGAGCGCTCATGGCGAG</td><td>ATG</td><td>2713</td><td>768</td></tr><tr><td>18</td><td>-3,387</td><td>2,214</td><td>14</td><td>-4,734</td><td>GGCTTCTGGTATCTGTGAAC</td><td>GTG</td><td>2746</td><td>735</td></tr><tr><td>19</td><td>-5,112</td><td>1,407</td><td>12</td><td>-5,948</td><td>TGCCAGCGTGGCAGCGAGGCC</td><td>ATG</td><td>2806</td><td>675</td></tr><tr><td>20</td><td>-3,642</td><td>2,095</td><td>13</td><td>-4,687</td><td>CGAGGCGATGGACGAGCAGAG</td><td>GTG</td><td>2821</td><td>660</td></tr><tr><td>21</td><td>-4,853</td><td>1,528</td><td>5</td><td>-6,853</td><td>GGACCGCGCCACCATCAGCGAC</td><td>ATG</td><td>2902</td><td>579</td></tr><tr><td>22</td><td>-4,439</td><td>1,722</td><td>14</td><td>-5,786</td><td>CCACGACAGCATCATGACAGAG</td><td>GTG</td><td>2935</td><td>546</td></tr><tr><td>23</td><td>-5,878</td><td>1,048</td><td>18</td><td>-8,179</td><td>CACGGGCGAGCGCTCCGACTGGT</td><td>GTG</td><td>3001</td><td>480</td></tr><tr><td>24</td><td>-3,208</td><td>2,298</td><td>15</td><td>-4,810</td><td>TCCGACTGGTGTGGGCTGAAG</td><td>GTG</td><td>3013</td><td>468</td></tr><tr><td>25</td><td>-4,439</td><td>1,722</td><td>11</td><td>-5,196</td><td>CTTCACAGACAGATGCGCGCC</td><td>ATG</td><td>3049</td><td>432</td></tr><tr><td>26</td><td>-4,666</td><td>1,616</td><td>11</td><td>-5,423</td><td>GGGCTGCCAGAGCAGAGGAA</td><td>GTG</td><td>3118</td><td>363</td></tr><tr><td>27</td><td>-2,071</td><td>2,831</td><td>10</td><td>-2,765</td><td>CGTCAGGACAGGAGCCCGAC</td><td>ATG</td><td>3184</td><td>297</td></tr><tr><td>28</td><td>-6,534</td><td>0,741</td><td>11</td><td>-7,291</td><td>CGACGGCATCAGGCTCAGCGAC</td><td>ATG</td><td>3242</td><td>219</td></tr><tr><td>29</td><td>-5,472</td><td>1,339</td><td>16</td><td>-7,268</td><td>CGTGGAGTACATGGCTCGGTC</td><td>GTG</td><td>3295</td><td>186</td></tr><tr><td>30</td><td>-6,856</td><td>0,591</td><td>13</td><td>-7,902</td><td>TGCTGGGACAGCGCGCGGAC</td><td>ATG</td><td>3385</td><td>96</td></tr><tr><td>31</td><td>-4,819</td><td>1,544</td><td>8</td><td>-6,041</td><td>CGAGCGCGCGCATGAGGCA</td><td>ATG</td><td>3394</td><td>87</td></tr></table>	Stats: 31	ORF Start : 2020	Cdn1	Cdn2	Cdn3	Length	SD Scoring Matrix	Kibler6	Explore	Selected: 1	ORF Stop : 3480	5' End	22.2	66.7	33.3	27	Spacing Weight Matrix	Karin Medium		ORF Length: 1461	3' End	62.8	41.1	88.1	1410		Document	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,267	2,371	8	-4,488	CAGGCGATCGGTCAAGGCGGAA	TTG	2044	1497	2	-3,698	2,069	11	-4,465	TACTTCAGATGGAGACTAAC	ATG	2071	1410	3	-5,021	1,450	13	-6,067	GGATCCAGCGGCTGTCCAGC	ATG	2119	1362	4	-4,013	1,921	11	-4,770	CAGCGGAGCAGCAAGACTAC	GTG	2158	1323	5	-4,735	1,594	6	-6,479	CGTGGCTACTACAAGGCAAG	ATG	2179	1302	6	-3,325	2,290	10	-3,920	TGACCTGTACGGGAGAACTGT	GTG	2305	1176	7	-3,325	2,290	16	-5,021	GTACGGGAGAACTGTGCGCC	GTG	2311	1170	8	-2,611	2,577	13	-3,657	TGAGCTGAAGGGGTTCAAGAA	GTG	2341	1140	9	-3,590	1,932	11	-4,747	GAAGGGGTTCAAGAACTGTCT	GTG	2347	1134	10	-3,699	2,069	10	-4,393	TCCTTCGGCTGGGAGTGGCGG	GTG	2377	1104	11	-3,699	2,069	13	-4,744	TTGGCTGGGAGTGGCGGCTG	TTG	2380	1101	12	-3,967	1,943	7	-5,490	GGTGGGAGTGGCGGCTGTG	GTG	2383	1098	13	-5,454	1,247	13	-6,500	CGACTTCCCGAGGCTACAGAG	ATG	2425	1056	14	-4,975	1,471	14	-6,322	GTTCACCGGCTGCTCAAGGAG	ATG	2482	999	15	-3,620	2,105	13	-4,666	GGACGACAAAGGCTTGGGGCC	ATG	2650	831	16	-2,915	2,435	7	-4,438	GCTGGAGGCGCTCAGGACGGG	GTG	2674	807	17	-4,817	1,545	13	-5,963	CGACCGAGCGCTCATGGCGAG	ATG	2713	768	18	-3,387	2,214	14	-4,734	GGCTTCTGGTATCTGTGAAC	GTG	2746	735	19	-5,112	1,407	12	-5,948	TGCCAGCGTGGCAGCGAGGCC	ATG	2806	675	20	-3,642	2,095	13	-4,687	CGAGGCGATGGACGAGCAGAG	GTG	2821	660	21	-4,853	1,528	5	-6,853	GGACCGCGCCACCATCAGCGAC	ATG	2902	579	22	-4,439	1,722	14	-5,786	CCACGACAGCATCATGACAGAG	GTG	2935	546	23	-5,878	1,048	18	-8,179	CACGGGCGAGCGCTCCGACTGGT	GTG	3001	480	24	-3,208	2,298	15	-4,810	TCCGACTGGTGTGGGCTGAAG	GTG	3013	468	25	-4,439	1,722	11	-5,196	CTTCACAGACAGATGCGCGCC	ATG	3049	432	26	-4,666	1,616	11	-5,423	GGGCTGCCAGAGCAGAGGAA	GTG	3118	363	27	-2,071	2,831	10	-2,765	CGTCAGGACAGGAGCCCGAC	ATG	3184	297	28	-6,534	0,741	11	-7,291	CGACGGCATCAGGCTCAGCGAC	ATG	3242	219	29	-5,472	1,339	16	-7,268	CGTGGAGTACATGGCTCGGTC	GTG	3295	186	30	-6,856	0,591	13	-7,902	TGCTGGGACAGCGCGCGGAC	ATG	3385	96	31	-4,819	1,544	8	-6,041	CGAGCGCGCGCATGAGGCA	ATG	3394	87
Stats: 31	ORF Start : 2020	Cdn1	Cdn2	Cdn3	Length	SD Scoring Matrix	Kibler6	Explore																																																																																																																																																																																																																																																																																																																													
Selected: 1	ORF Stop : 3480	5' End	22.2	66.7	33.3	27	Spacing Weight Matrix	Karin Medium																																																																																																																																																																																																																																																																																																																													
	ORF Length: 1461	3' End	62.8	41.1	88.1	1410		Document																																																																																																																																																																																																																																																																																																																													
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																																																																																																																																													
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																																																																																																																																													
1	-3,267	2,371	8	-4,488	CAGGCGATCGGTCAAGGCGGAA	TTG	2044	1497																																																																																																																																																																																																																																																																																																																													
2	-3,698	2,069	11	-4,465	TACTTCAGATGGAGACTAAC	ATG	2071	1410																																																																																																																																																																																																																																																																																																																													
3	-5,021	1,450	13	-6,067	GGATCCAGCGGCTGTCCAGC	ATG	2119	1362																																																																																																																																																																																																																																																																																																																													
4	-4,013	1,921	11	-4,770	CAGCGGAGCAGCAAGACTAC	GTG	2158	1323																																																																																																																																																																																																																																																																																																																													
5	-4,735	1,594	6	-6,479	CGTGGCTACTACAAGGCAAG	ATG	2179	1302																																																																																																																																																																																																																																																																																																																													
6	-3,325	2,290	10	-3,920	TGACCTGTACGGGAGAACTGT	GTG	2305	1176																																																																																																																																																																																																																																																																																																																													
7	-3,325	2,290	16	-5,021	GTACGGGAGAACTGTGCGCC	GTG	2311	1170																																																																																																																																																																																																																																																																																																																													
8	-2,611	2,577	13	-3,657	TGAGCTGAAGGGGTTCAAGAA	GTG	2341	1140																																																																																																																																																																																																																																																																																																																													
9	-3,590	1,932	11	-4,747	GAAGGGGTTCAAGAACTGTCT	GTG	2347	1134																																																																																																																																																																																																																																																																																																																													
10	-3,699	2,069	10	-4,393	TCCTTCGGCTGGGAGTGGCGG	GTG	2377	1104																																																																																																																																																																																																																																																																																																																													
11	-3,699	2,069	13	-4,744	TTGGCTGGGAGTGGCGGCTG	TTG	2380	1101																																																																																																																																																																																																																																																																																																																													
12	-3,967	1,943	7	-5,490	GGTGGGAGTGGCGGCTGTG	GTG	2383	1098																																																																																																																																																																																																																																																																																																																													
13	-5,454	1,247	13	-6,500	CGACTTCCCGAGGCTACAGAG	ATG	2425	1056																																																																																																																																																																																																																																																																																																																													
14	-4,975	1,471	14	-6,322	GTTCACCGGCTGCTCAAGGAG	ATG	2482	999																																																																																																																																																																																																																																																																																																																													
15	-3,620	2,105	13	-4,666	GGACGACAAAGGCTTGGGGCC	ATG	2650	831																																																																																																																																																																																																																																																																																																																													
16	-2,915	2,435	7	-4,438	GCTGGAGGCGCTCAGGACGGG	GTG	2674	807																																																																																																																																																																																																																																																																																																																													
17	-4,817	1,545	13	-5,963	CGACCGAGCGCTCATGGCGAG	ATG	2713	768																																																																																																																																																																																																																																																																																																																													
18	-3,387	2,214	14	-4,734	GGCTTCTGGTATCTGTGAAC	GTG	2746	735																																																																																																																																																																																																																																																																																																																													
19	-5,112	1,407	12	-5,948	TGCCAGCGTGGCAGCGAGGCC	ATG	2806	675																																																																																																																																																																																																																																																																																																																													
20	-3,642	2,095	13	-4,687	CGAGGCGATGGACGAGCAGAG	GTG	2821	660																																																																																																																																																																																																																																																																																																																													
21	-4,853	1,528	5	-6,853	GGACCGCGCCACCATCAGCGAC	ATG	2902	579																																																																																																																																																																																																																																																																																																																													
22	-4,439	1,722	14	-5,786	CCACGACAGCATCATGACAGAG	GTG	2935	546																																																																																																																																																																																																																																																																																																																													
23	-5,878	1,048	18	-8,179	CACGGGCGAGCGCTCCGACTGGT	GTG	3001	480																																																																																																																																																																																																																																																																																																																													
24	-3,208	2,298	15	-4,810	TCCGACTGGTGTGGGCTGAAG	GTG	3013	468																																																																																																																																																																																																																																																																																																																													
25	-4,439	1,722	11	-5,196	CTTCACAGACAGATGCGCGCC	ATG	3049	432																																																																																																																																																																																																																																																																																																																													
26	-4,666	1,616	11	-5,423	GGGCTGCCAGAGCAGAGGAA	GTG	3118	363																																																																																																																																																																																																																																																																																																																													
27	-2,071	2,831	10	-2,765	CGTCAGGACAGGAGCCCGAC	ATG	3184	297																																																																																																																																																																																																																																																																																																																													
28	-6,534	0,741	11	-7,291	CGACGGCATCAGGCTCAGCGAC	ATG	3242	219																																																																																																																																																																																																																																																																																																																													
29	-5,472	1,339	16	-7,268	CGTGGAGTACATGGCTCGGTC	GTG	3295	186																																																																																																																																																																																																																																																																																																																													
30	-6,856	0,591	13	-7,902	TGCTGGGACAGCGCGCGGAC	ATG	3385	96																																																																																																																																																																																																																																																																																																																													
31	-4,819	1,544	8	-6,041	CGAGCGCGCGCATGAGGCA	ATG	3394	87																																																																																																																																																																																																																																																																																																																													
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	Non, un autre START est proposé proposant une plus grande ORF, mais c'est un TTG et cela donnerait une ORF chevauchante avec celle en amont (>30bp)																																																																																																																																																																																																																																																																																																																																				
Est-ce que le start est conservé chez les homologues voir Starterator ?	Ce gène chez Carostasia est le seul annoté dans son groupe, Starterator ne peut donc pas nous aider à trancher par définition.																																																																																																																																																																																																																																																																																																																																				
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Le start est conservé chez une partie des homologues. L'alignement commence au premier AA pour la query et une partie des subjects.																																																																																																																																																																																																																																																																																																																																				

	<p>>Quartz_3, portal protein, 469 Length = 469</p> <p>Score = 938 bits (2424), Expect = 0.0 Identities = 464/469 (98%), Positives = 466/469 (99%)</p> <p>Query: 1 MVLEIEAPAVSTALSSMEAYLLSRDSKTYVPYNGKMSYALHGKSWDEYVKEAFPDLAGQ 60 MVLEIEAPAVSTALSSMEAYLLSRDSKTYVPYNGKMSYALHGKSWDEYVKEAFPDLAGQ 60 Sbjct: 1 MVLEIEAPAVSTALSSMEAYLLSRDSKTYVPYNGKMSYALHGKSWDEYVKEAFPDLAGQ 60</p> <p>Query: 61 TTSENIFKSVIDLVAENLVPVQDELKGFSLVPLLSRGCEPVLVDSTGTPHFPERYEMI 120 TTSENIFKSVIDLVAENLVPVQDELKGFSLVPLLSRGCEPVLVDSTGTPHFPERYEMI 120 Sbjct: 61 TTSENIFKSVIDLVAENLVPVQDELKGFSLVPLLSRGCEPVLVDSTGTPHFPERYEMI 120</p> <p>Query: 121 SDGTYSVAALFTRSLKEHNDIVIFCYTDGTTTLFQKDIPQDFAPADKEGYTFVEELHGST 180 SDGTYSVAALFTRSLKEHNDIVIFCYTDGTTTLFQKDIPQDFAPADKEGYTFVEELHGST 180 Sbjct: 121 SDGTYSVAALFTRSLKEHNDIVIFCYTDGTTTLFQKDIPQDFAPADKEGYTFVEELHGST 180</p> <p>Query: 181 LLRFALDDKGFAMLAALQDRVNHHSIIDQTVIAEHYARPFWYLLNVEIPTKNPFLPAGAQ 240 LLRFALDDKGFAMLAALQDRVNHHSIIDQTVIAEHYARPFWYLLNVEIPTKNPFLPAGAQ 240 Sbjct: 181 LLRFALDDKGFAMLAALQDRVNHHSIIDQTVIAEHYARPFWYLLNVEIPTKNPFLPAGAQ 240</p> <p>Query: 241 PGTEAMDEQKVDGATGRIFTSSGEGPFGQLDPPTISDMVGYHDSIIDKVSQSSGIPQHYF 300 PGTEAMDEQKVDGATGRIFTSSGEGPFGQLDPPTISDMVGYHDSIIDKVSQSSGIPQHYF 300 Sbjct: 241 PGTEAMDEQKVDGATGRIFTSSGEGPFGQLDPPTISDMVGYHDSIIDKVSQSSGIPQHYF 300</p> <p>Query: 301 KPGTGTPTGVALKVLKSRFNKNIARMRDDIQDTLEELASLLGVPKNKEVDRPKGTATDE 360 KPGTGTPTGVALKVLKSRFNKNIARMRDDIQDTLEELASLLGVPKNKEVDRPKGTATDE 360 Sbjct: 301 KPGTGTPTGVALKVLKSRFNKNIARMRDDIQDTLEELASLLGVPKNKEVDRPKGTATDE 360</p> <p>Query: 361 DAEDVKDEEPMEDYFWDPSDDLQESLDAHGISLSQMGYPLEYIASVVTGPNLDDYAE 420 DAEDV DEEPMEDYFWDPSDDLQESLDAHGISLSQMGYPLEYIASVVTGPNLDDYAE 420 Sbjct: 361 DAEDVTDEEPMEDYFWDPSDDLQESLDAHGISLSQMGYPLEYIASVVTGPNLDDYAE 420</p> <p>Query: 421 EEAGPELGTDPADQPADMTAMGQPLVATPGQVAAYGANPGQRAKATR 469 EEAGPELGTDPADQPADMTAMGQPLVATPGQVAAYGANPGQRAKATR 469 Sbjct: 421 EEAGPELGTDPADQPADMTAMGQPLVATPGQVAAYGANPGQRAKATR 469</p>
DECISION:	<p><i>Start à 2071pb retenu, bon score RBS, ne confère par l'ORF la plus longue mais le TTG en amont confère un chevauchement trop important avec le gène précédent.</i></p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	Résultats nr :



PhagesDB :

Query= Carostasia_3
(469 letters)

Distribution of 100 Blast Hits on the Query Sequence

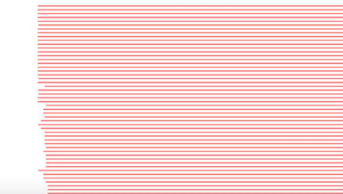
Mouse-over to show define and scores. Click to show alignments

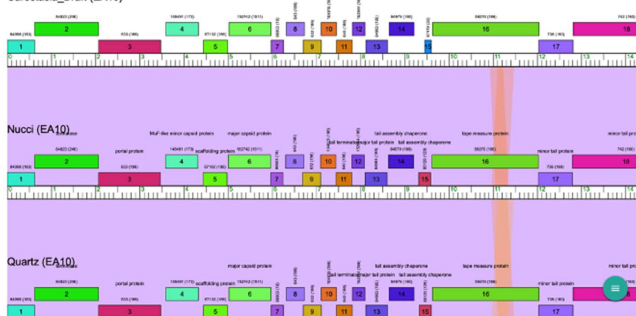

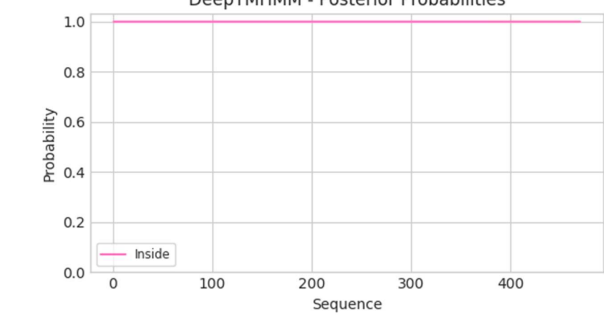
Color Key For Alignment Scores

■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >200

1c111

0 50 100 150 200 250 300 350 400 450



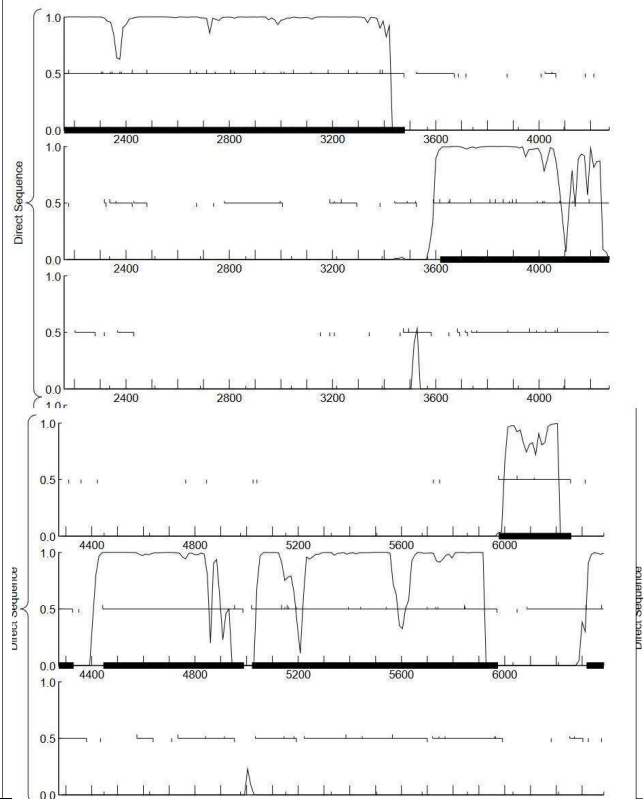
	<p><i>gènes à l'air conservé.</i></p> <p>Carostasia_Draft (EA10)</p>  <p>Nuccl (EA10)</p> <p>Quartz (EA10)</p>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p><i>Non, elle est prédite comme protéine globulaire interne.</i></p> <p>DeepTMHMM - Most Likely Topology Type: Globular</p>  <p>DeepTMHMM - Posterior Probabilities</p> 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<p><i>OUI.</i></p>
DECISION:	<p><i>Ce gène code pour une portal.</i></p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	4
Coordonnées du Stop	4330
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 113
Coordonnées du Start retenu	3593
Fonction prédite	NFK

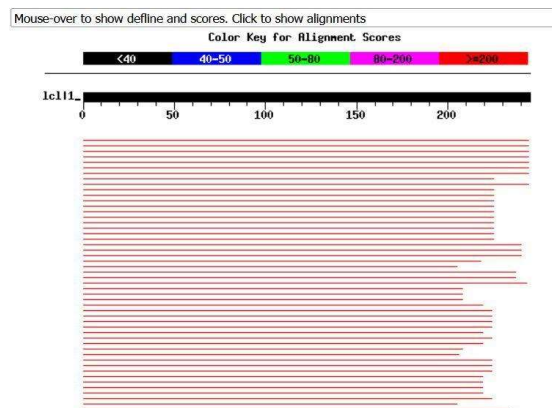
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Genemark detecte un potentiel codant tout le long de l'ORF, le pouvoir codant commence un peu avant le start proposé par Genemark</i>



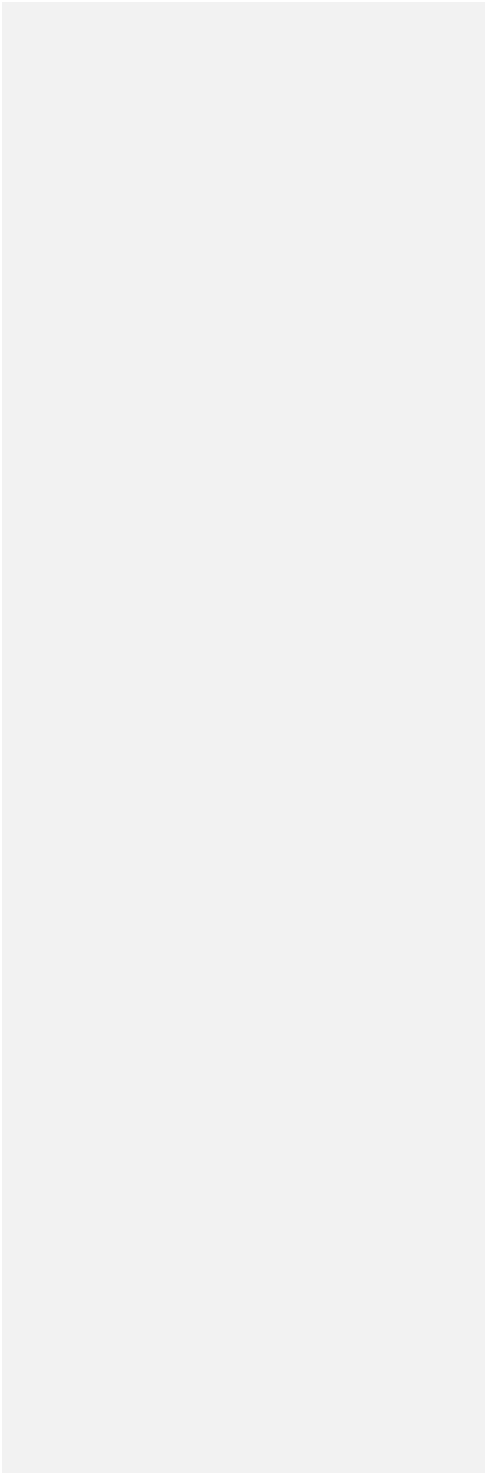
Oui

Distribution of 100 Blast Hits on the Query Sequence



Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?

Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non
DECISION:	<i>Oui</i>



Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	Glimmer : 3593, GeneMark : 3617																																																																																																																																																																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Starts: 20 Selected: 1</div><div>ORF Start : 3812 ORF Stop : 4330 ORF Length: 519</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 62.5 75.0 62.5 24 3' End 65.5 52.1 83.2 714</div><div>SD Scoring Matrix Kibler6</div><div>Explore</div><div>Spacing Weight Matrix Karlin Medium</div><div>Document</div></div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,377</td><td>2,487</td><td>11</td><td>-3,134</td><td>CGACACCGAAGGATAGTCATC</td><td>ATG</td><td>3593</td><td>738</td></tr><tr><td>2</td><td>-2,572</td><td>2,596</td><td>6</td><td>-4,317</td><td>GGGAACCGTACCCACCGGAGCC</td><td>ATG</td><td>3617</td><td>714</td></tr><tr><td>3</td><td>-6,055</td><td>0,966</td><td>16</td><td>-7,851</td><td>GGCAAGCTCTACCTCTCTGGG</td><td>GTG</td><td>3653</td><td>678</td></tr><tr><td>4</td><td>-4,283</td><td>1,795</td><td>5</td><td>-6,283</td><td>CAGGCTCTACCTCTCTGGGTG</td><td>ATG</td><td>3656</td><td>675</td></tr><tr><td>5</td><td>-3,952</td><td>1,950</td><td>16</td><td>-5,747</td><td>GAGTGAGAAGCTGATCGCCAGC</td><td>ATG</td><td>3737</td><td>594</td></tr><tr><td>6</td><td>-5,691</td><td>1,136</td><td>6</td><td>-7,436</td><td>ACGGGCTGCTGGCTCGGTGCC</td><td>TTC</td><td>3767</td><td>564</td></tr><tr><td>7</td><td>-7,369</td><td>0,951</td><td>9</td><td>-9,143</td><td>GGGGGACTTCCCGCTCCCAAG</td><td>TTC</td><td>3791</td><td>540</td></tr><tr><td>8</td><td>-3,800</td><td>2,021</td><td>16</td><td>-5,586</td><td>GTTCCTGGAGCTCAGCCGCTAC</td><td>ATG</td><td>3812</td><td>519</td></tr><tr><td>9</td><td>-4,928</td><td>1,493</td><td>14</td><td>-6,275</td><td>CATGGCGAAGATCTACGATCAG</td><td>ATG</td><td>3833</td><td>498</td></tr><tr><td>10</td><td>-4,654</td><td>1,422</td><td>10</td><td>-5,348</td><td>GGCCGTCATCCAGCTGGTATC</td><td>ATG</td><td>3863</td><td>468</td></tr><tr><td>11</td><td>-6,180</td><td>0,907</td><td>11</td><td>-6,937</td><td>GGCGGGGCTCAATAGCGCTGCC</td><td>GTG</td><td>3887</td><td>444</td></tr><tr><td>12</td><td>-6,082</td><td>0,953</td><td>10</td><td>-6,777</td><td>TAGCGGTGCGGTCGCCAGGCC</td><td>ATG</td><td>3895</td><td>432</td></tr><tr><td>13</td><td>-4,421</td><td>1,730</td><td>8</td><td>-5,643</td><td>CGAGGCACTGTTCGGGGCTGGG</td><td>ATG</td><td>3914</td><td>417</td></tr><tr><td>14</td><td>-4,705</td><td>1,597</td><td>8</td><td>-5,527</td><td>CTGGGCGACAGCTGGTGGAGT</td><td>GTG</td><td>3955</td><td>336</td></tr><tr><td>15</td><td>-5,369</td><td>1,287</td><td>13</td><td>-6,414</td><td>TGTGGCGAGCTGGCTGACATC</td><td>GTG</td><td>4016</td><td>315</td></tr><tr><td>16</td><td>-5,931</td><td>1,024</td><td>7</td><td>-7,454</td><td>CGAGCTGCTGACATGCTGCTG</td><td>GTG</td><td>4022</td><td>309</td></tr><tr><td>17</td><td>-3,638</td><td>2,097</td><td>11</td><td>-4,385</td><td>CTGTGCTGACCGGGATGGGCTG</td><td>GTG</td><td>4082</td><td>249</td></tr><tr><td>18</td><td>-3,638</td><td>2,097</td><td>14</td><td>-4,385</td><td>TGTGACCCGGGATGGGCTGGT</td><td>GTG</td><td>4085</td><td>246</td></tr><tr><td>19</td><td>-4,286</td><td>1,794</td><td>15</td><td>-5,888</td><td>CCCGAAGCGGGGCTGACCCCTC</td><td>GTG</td><td>4133</td><td>198</td></tr><tr><td>20</td><td>-5,224</td><td>1,355</td><td>12</td><td>-6,059</td><td>CGGGGCTGACGGCAGGCTCTAC</td><td>ATG</td><td>4196</td><td>135</td></tr></table> <p>Le Start proposé par Glimmer possède un RBS de meilleur score, mais ce n'est pas une différence très importante.</p>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,377	2,487	11	-3,134	CGACACCGAAGGATAGTCATC	ATG	3593	738	2	-2,572	2,596	6	-4,317	GGGAACCGTACCCACCGGAGCC	ATG	3617	714	3	-6,055	0,966	16	-7,851	GGCAAGCTCTACCTCTCTGGG	GTG	3653	678	4	-4,283	1,795	5	-6,283	CAGGCTCTACCTCTCTGGGTG	ATG	3656	675	5	-3,952	1,950	16	-5,747	GAGTGAGAAGCTGATCGCCAGC	ATG	3737	594	6	-5,691	1,136	6	-7,436	ACGGGCTGCTGGCTCGGTGCC	TTC	3767	564	7	-7,369	0,951	9	-9,143	GGGGGACTTCCCGCTCCCAAG	TTC	3791	540	8	-3,800	2,021	16	-5,586	GTTCCTGGAGCTCAGCCGCTAC	ATG	3812	519	9	-4,928	1,493	14	-6,275	CATGGCGAAGATCTACGATCAG	ATG	3833	498	10	-4,654	1,422	10	-5,348	GGCCGTCATCCAGCTGGTATC	ATG	3863	468	11	-6,180	0,907	11	-6,937	GGCGGGGCTCAATAGCGCTGCC	GTG	3887	444	12	-6,082	0,953	10	-6,777	TAGCGGTGCGGTCGCCAGGCC	ATG	3895	432	13	-4,421	1,730	8	-5,643	CGAGGCACTGTTCGGGGCTGGG	ATG	3914	417	14	-4,705	1,597	8	-5,527	CTGGGCGACAGCTGGTGGAGT	GTG	3955	336	15	-5,369	1,287	13	-6,414	TGTGGCGAGCTGGCTGACATC	GTG	4016	315	16	-5,931	1,024	7	-7,454	CGAGCTGCTGACATGCTGCTG	GTG	4022	309	17	-3,638	2,097	11	-4,385	CTGTGCTGACCGGGATGGGCTG	GTG	4082	249	18	-3,638	2,097	14	-4,385	TGTGACCCGGGATGGGCTGGT	GTG	4085	246	19	-4,286	1,794	15	-5,888	CCCGAAGCGGGGCTGACCCCTC	GTG	4133	198	20	-5,224	1,355	12	-6,059	CGGGGCTGACGGCAGGCTCTAC	ATG	4196	135
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																															
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																															
1	-2,377	2,487	11	-3,134	CGACACCGAAGGATAGTCATC	ATG	3593	738																																																																																																																																																																																															
2	-2,572	2,596	6	-4,317	GGGAACCGTACCCACCGGAGCC	ATG	3617	714																																																																																																																																																																																															
3	-6,055	0,966	16	-7,851	GGCAAGCTCTACCTCTCTGGG	GTG	3653	678																																																																																																																																																																																															
4	-4,283	1,795	5	-6,283	CAGGCTCTACCTCTCTGGGTG	ATG	3656	675																																																																																																																																																																																															
5	-3,952	1,950	16	-5,747	GAGTGAGAAGCTGATCGCCAGC	ATG	3737	594																																																																																																																																																																																															
6	-5,691	1,136	6	-7,436	ACGGGCTGCTGGCTCGGTGCC	TTC	3767	564																																																																																																																																																																																															
7	-7,369	0,951	9	-9,143	GGGGGACTTCCCGCTCCCAAG	TTC	3791	540																																																																																																																																																																																															
8	-3,800	2,021	16	-5,586	GTTCCTGGAGCTCAGCCGCTAC	ATG	3812	519																																																																																																																																																																																															
9	-4,928	1,493	14	-6,275	CATGGCGAAGATCTACGATCAG	ATG	3833	498																																																																																																																																																																																															
10	-4,654	1,422	10	-5,348	GGCCGTCATCCAGCTGGTATC	ATG	3863	468																																																																																																																																																																																															
11	-6,180	0,907	11	-6,937	GGCGGGGCTCAATAGCGCTGCC	GTG	3887	444																																																																																																																																																																																															
12	-6,082	0,953	10	-6,777	TAGCGGTGCGGTCGCCAGGCC	ATG	3895	432																																																																																																																																																																																															
13	-4,421	1,730	8	-5,643	CGAGGCACTGTTCGGGGCTGGG	ATG	3914	417																																																																																																																																																																																															
14	-4,705	1,597	8	-5,527	CTGGGCGACAGCTGGTGGAGT	GTG	3955	336																																																																																																																																																																																															
15	-5,369	1,287	13	-6,414	TGTGGCGAGCTGGCTGACATC	GTG	4016	315																																																																																																																																																																																															
16	-5,931	1,024	7	-7,454	CGAGCTGCTGACATGCTGCTG	GTG	4022	309																																																																																																																																																																																															
17	-3,638	2,097	11	-4,385	CTGTGCTGACCGGGATGGGCTG	GTG	4082	249																																																																																																																																																																																															
18	-3,638	2,097	14	-4,385	TGTGACCCGGGATGGGCTGGT	GTG	4085	246																																																																																																																																																																																															
19	-4,286	1,794	15	-5,888	CCCGAAGCGGGGCTGACCCCTC	GTG	4133	198																																																																																																																																																																																															
20	-5,224	1,355	12	-6,059	CGGGGCTGACGGCAGGCTCTAC	ATG	4196	135																																																																																																																																																																																															
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	Oui le start choisit prédit l'ORF la plus longue.																																																																																																																																																																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?	Oui le start est conservé chez les homologues																																																																																																																																																																																																						

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p><i>Oui le start est conservé chez les homologues</i></p> <pre> >Quartz_4, function unknown, 245 Length = 245 Score = 498 bits (1281), Expect = e-141 Identities = 244/245 (99%), Positives = 244/245 (99%) Query: 1 MGTVPTGAMERELRKLYLSWVMGLSWDINGSIHDQLTAFVIQSEKLIASMGGRAALGALG 60 MGTVPTGAMERELRKLYLSWVMGLSWDINGSIHDQLTAFVIQSEKLIASMGGR AALGALG Sbjct: 1 MGTVPTGAMERELRKLYLSWVMGLSWDINGSIHDQLTAFVIQSEKLIASMGGRTAALGALG 60 Query: 61 DFPAPKLLDLSPYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGMDKSYRRLERLAR 120 DFPAPKLLDLSPYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGMDKSYRRLERLAR Sbjct: 61 DFPAPKLLDLSPYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGMDKSYRRLERLAR 120 Query: 121 TETTNAVWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180 TETTNAVWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL Sbjct: 121 TETTNAVWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180 Query: 181 VPTHR5QLEYKGTLRADGSVYMDPRWADQRPVGARAKASAGPTTPEQRDPLSGRRNPAP 240 VPTHR5QLEYKGTLRADGSVYMDPRWADQRPVGARAKASAGPTTPEQRDPLSGRRNPAP Sbjct: 181 VPTHR5QLEYKGTLRADGSVYMDPRWADQRPVGARAKASAGPTTPEQRDPLSGRRNPAP 240 Query: 241 SRAAP 245 SRAAP Sbjct: 241 SRAAP 245 >Mandalorian_4, Muf-like minor capsid protein, 245 Length = 245 Score = 498 bits (1281), Expect = e-141 Identities = 244/245 (99%), Positives = 244/245 (99%) Query: 1 MGTVPTGAMERELRKLYLSWVMGLSWDINGSIHDQLTAFVIQSEKLIASMGGRAALGALG 60 MGTVPTGAMERELRKLYLSWVMGLSWDINGSIHDQLTAFVIQSEKLIASMGGR AALGALG Sbjct: 1 MGTVPTGAMERELRKLYLSWVMGLSWDINGSIHDQLTAFVIQSEKLIASMGGRTAALGALG 60 Query: 61 DFPAPKLLDLSPYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGMDKSYRRLERLAR 120 DFPAPKLLDLSPYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGMDKSYRRLERLAR Sbjct: 61 DFPAPKLLDLSPYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGMDKSYRRLERLAR 120 Query: 121 TETTNAVWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180 TETTNAVWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL Sbjct: 121 TETTNAVWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180 Query: 181 VPTHR5QLEYKGTLRADGSVYMDPRWADQRPVGARAKASAGPTTPEQRDPLSGRRNPAP 240 VPTHR5QLEYKGTLRADGSVYMDPRWADQRPVGARAKASAGPTTPEQRDPLSGRRNPAP Sbjct: 181 VPTHR5QLEYKGTLRADGSVYMDPRWADQRPVGARAKASAGPTTPEQRDPLSGRRNPAP 240 Query: 241 SRAAP 245 SRAAP Sbjct: 241 SRAAP 245 </pre>
DECISION:	<i>Start 3593, bon score RBS et prédit l'ORF la plus longue sans chevauchement excessif et est conservé chez les homologues.</i>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s’aligne avec	Listez le meilleur hit Blastp pour chaque source :

une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

*** *PhagesDB* :

>Quartz_4, function unknown, 245
Length = 245

Score = 498 bits (1281), Expect = e-141
Identities = 244/245 (99%), Positives = 244/245 (99%)

Query: 1 MGTVPTGAMERELRKLYLSWVWGLSWDNGSIHQDLTAFVIQSEKLIASMGGRAALGALG 60
MGTVPTGAMERELRKLYLSWVWGLSWDNGSIHQDLTAFVIQSEKLIASMGGR AALGALG
Sbjct: 1 MGTVPTGAMERELRKLYLSWVWGLSWDNGSIHQDLTAFVIQSEKLIASMGRTAALGALG 60

Query: 61 DFPAPKLLDLSPLYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGIDKSYRRLERLAR 120
DFPAPKLLDLSPLYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGIDKSYRRLERLAR
Sbjct: 61 DFPAPKLLDLSPLYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGIDKSYRRLERLAR 120

Query: 121 TETTNAVYWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180
TETTNAVYWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL
Sbjct: 121 TETTNAVYWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180

Query: 181 VPTHRSQLLEYKGTLRADGSVYVDPRIADQRPVGARAKASAGPTTPEQRDPLSGRRNPAAP 240
VPTHRSQLLEYKGTLRADGSVYVDPRIADQRPVGARAKASAGPTTPEQRDPLSGRRNPAAP
Sbjct: 181 VPTHRSQLLEYKGTLRADGSVYVDPRIADQRPVGARAKASAGPTTPEQRDPLSGRRNPAAP 240

Query: 241 SRAAP 245
SRAAP
Sbjct: 241 SRAAP 245

>Mandalorian_4, MuF-like minor capsid protein, 245
Length = 245

Score = 498 bits (1281), Expect = e-141
Identities = 244/245 (99%), Positives = 244/245 (99%)

Query: 1 MGTVPTGAMERELRKLYLSWVWGLSWDNGSIHQDLTAFVIQSEKLIASMGGRAALGALG 60
MGTVPTGAMERELRKLYLSWVWGLSWDNGSIHQDLTAFVIQSEKLIASMGGR AALGALG
Sbjct: 1 MGTVPTGAMERELRKLYLSWVWGLSWDNGSIHQDLTAFVIQSEKLIASMGRTAALGALG 60

Query: 61 DFPAPKLLDLSPLYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGIDKSYRRLERLAR 120
DFPAPKLLDLSPLYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGIDKSYRRLERLAR
Sbjct: 61 DFPAPKLLDLSPLYMGKIYDQMQQAVIQAGIMAGLNSRAVAQAMFRAGIDKSYRRLERLAR 120

Query: 121 TETTNAVYWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180
TETTNAVYWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL
Sbjct: 121 TETTNAVYWRNTWSSVAELPDIVLVWGAEESARTCAYCRDRDGLVDDPSIRDHPNGRCTL 180

Query: 181 VPTHRSQLLEYKGTLRADGSVYVDPRIADQRPVGARAKASAGPTTPEQRDPLSGRRNPAAP 240
VPTHRSQLLEYKGTLRADGSVYVDPRIADQRPVGARAKASAGPTTPEQRDPLSGRRNPAAP
Sbjct: 181 VPTHRSQLLEYKGTLRADGSVYVDPRIADQRPVGARAKASAGPTTPEQRDPLSGRRNPAAP 240

Query: 241 SRAAP 245
SRAAP
Sbjct: 241 SRAAP 245

*** *nr* :

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> MF-the minor capsid protein [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	501	501	99%	9e-179	98.59%	245	QX013558.1
<input checked="" type="checkbox"/> MF-the minor capsid protein [Microbacterium phage Rucci]	Microbacterium phage Rucci	486	486	99%	9e-175	98.33%	245	QX002387.1

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

UniProt :

[Template alignment](#)

☐ 1. [Q01259 GPF](#), BPMU Putative capsid assembly protein F O5-Escherichia phage Mu OX-10677 GN-F PE-2 SV-2
Probability: 99.81%, E-value: 7.5e-18, Score: 147.56, Aligned cols: 114, Identities: 13%, Similarity: 0.113, Template Neff: 8.4

Pfam :

[Template alignment](#) | [CDD](#)

☐ 3. [PF04233.18](#): Phage Mu_F; Phage Mu protein F like protein
Probability: 99.59%, E-value: 2.8e-14, Score: 99.45, Aligned cols: 100, Identities: 20%, Similarity: 0.195, Template Neff: 12

CD :

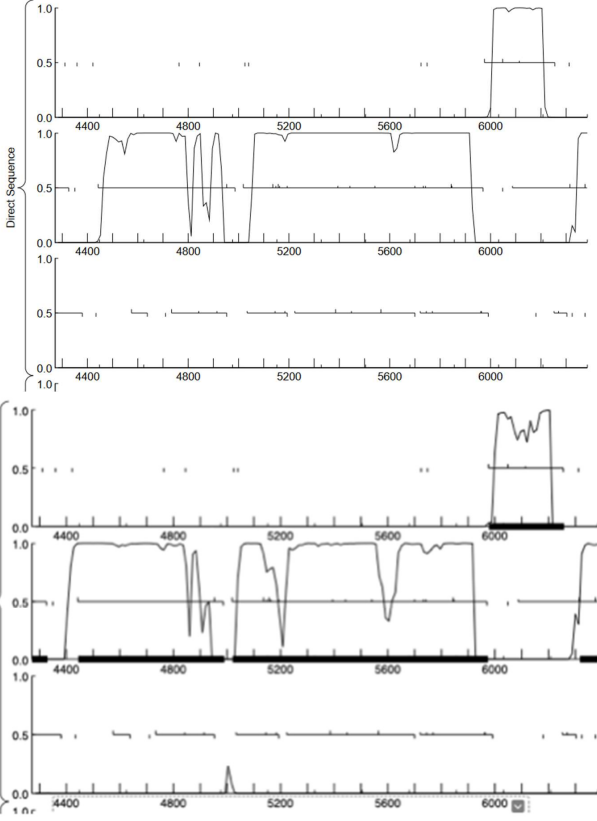
	<p>Template alignment</p> <p><input type="checkbox"/> 13.cd16378 CcmH_N: N-terminal domain of cytochrome c-type biogenesis protein CcmH and related proteins. Probability: 35.09%, E-value: 250, Score: 18.37, Aligned cols: 30, Identities: 10%, Similarity: 0.038, Template Neff: 8.5</p> <p>PDB :</p> <p>Template alignment Template 3D Structure PDBe</p> <p><input type="checkbox"/> 12.7JRQ_A MPP1; metalloprotein, porphyrin, oxidation, sulfoxidation, high-talent, helical bundle, DE NOVO PROTEIN; HET: SMU; 1.75A (Escherichia coli) Probability: 38.66%, E-value: 320, Score: 20.7, Aligned cols: 101, Identities: 10%, Similarity: 0.048, Template Neff: 1.8</p>		
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Oui, terminase et protéines de capsid		
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	NON		
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	minor capsid protein		
	Hypothetical Protein	MuF-like minor capsid protein	If an HHPred alignment to gp15 of phage D29, see capsid maturation protease
DECISION:	NFK		

Student Gene Annotation Worksheet

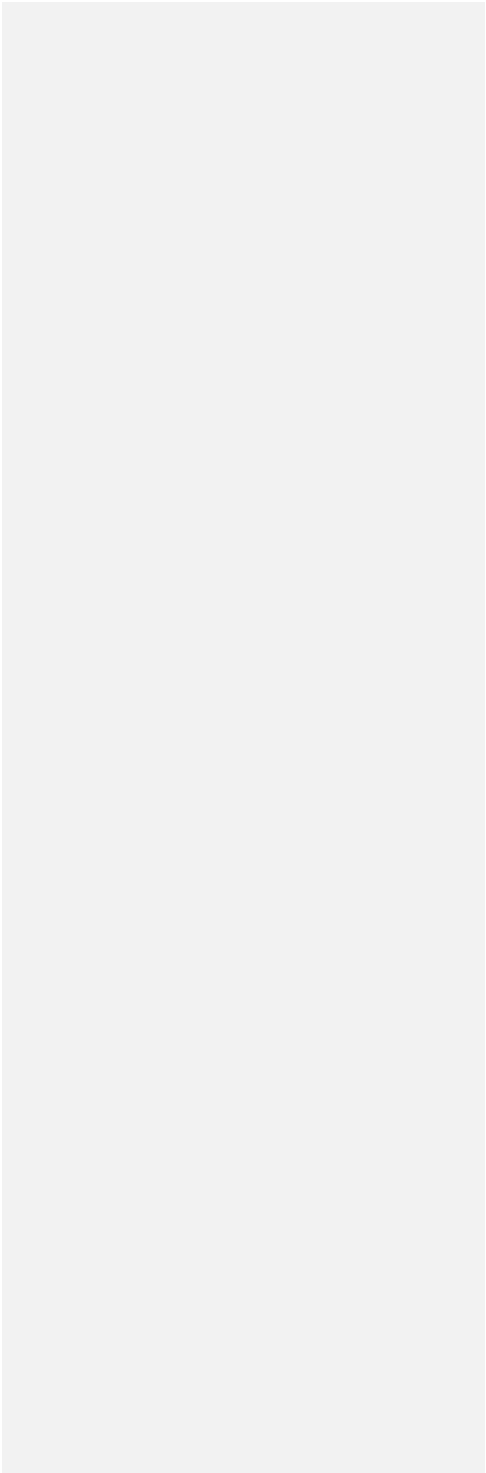
Basic Phage Information	
Nom du Phage	Carostasia
Gène #	5
Coordonnées du Stop	4990
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 114
Coordonnées du Start retenu	4445
Fonction prédite	Scaffold

Décision #1 : Est-ce un gène ?

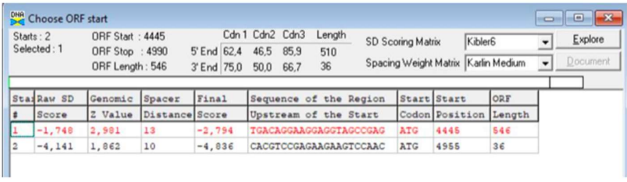

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Oui, GeneMark et GeneMarks détectent un potentiel codant sur une partie de l'ORF proposée par DNA master. Le pouvoir codant commence néanmoins plus tôt que le start proposé, et sur la partie 3' le potentiel détecté est variable/faible.</i>

	
<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<p>Quelques homologues sont retrouvés, néanmoins pour la majorité les scores sont entre 80 et 200.</p>

	<div><p><u>Distribution of 100 Blast Hits on the Query Sequence</u></p><p>Mouse-over to show define and scores. Click to show alignments</p><p>Color Key for Alignment Scores</p><table><tr><td><40</td><td>40-50</td><td>50-80</td><td>80-200</td><td>>=200</td></tr></table><p>lc111</p></div>	<40	40-50	50-80	80-200	>=200
<40	40-50	50-80	80-200	>=200		
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Non</i>					
DECISION:	<i>Oui.</i>					



Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	4445.
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<p>Oui, c'est le meilleur score proposé il est proche de 0. La Z-value aussi.</p> 
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Oui le start prédit la plus longue ORF (546pb), choisir ce start conduit à un gap de 15pb, ce qui n'est pas excessif.</p>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Oui, le start est conservé chez les homologues.</p> 
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Oui le start est conservé chez d'autres homologues.</p>

	<p>>Quartz_5, scaffolding protein, 181 Length = 181</p> <p>Score = 352 bits (902), Expect = 4e-97 Identities = 180/181 (99%), Positives = 180/181 (99%)</p> <p>Query: 1 MATNDTFETGSESQDTSQDQQTSGDSTDETSQGEQGSUTQQDASQEDPTKGLKSAIQ 60 MATNDTFETGSESQDTSQDQQTSGDSTDETSQGEQGSUTQQDASQEDPTKGLKSAIQ Sbjct: 1 MATNDTFETGSESQDTSQDQQTSGDSTDETSQGEQGSUTQQDASQEDPTKGLKSAIQ 60</p> <p>Query: 61 GARADLNKKKKHKATTARVAEAEQVADLTFRKETTDAIQERYDRLEAFIQALPGPLGK 120 GARADLNKKKKHKATTARVAEAEQV DLTFRKETTDAIQERYDRLEAFIQALPGPLGK Sbjct: 61 GARADLNKKKKHKATTARVAEAEQVADLTFRKETTDAIQERYDRLEAFIQALPGPLGK 120</p> <p>Query: 121 ALDSKSFTEKLFESEDRVSDI:IKDWNKANPTATASALGGGTGATSEKSNHMDLLRAAAR 180 ALDSKSFTEKLFESEDRVSDI:IKDWNKANPTATASALGGGTGATSEKSNHMDLLRAAAR Sbjct: 121 ALDSKSFTEKLFESEDRVSDI:IKDWNKANPTATASALGGGTGATSEKSNHMDLLRAAAR 180</p> <p>Query: 181 G 181 G Sbjct: 181 G 181</p> <p>>Mandalorian_5, scaffolding protein, 181 Length = 181</p> <p>Score = 345 bits (884), Expect = 4e-95 Identities = 176/181 (97%), Positives = 178/181 (98%)</p> <p>Query: 1 MATNDTFETGSESQDTSQDQQTSGDSTDETSQGEQGSUTQQDASQEDPTKGLKSAIQ 60 MATNDTFETGSESQDTSQDQQTSGDSTDETSQGEQGSUTQQDASQEDPTKGLKSAIQ Sbjct: 1 MATNDTFETGSESQDTSQDQQTSGDSTDETSQGEQGSUTQQDASQEDPTKGLKSAIQ 60</p> <p>Query: 61 GARADLNKKKKHKATTARVAEAEQVADLTFRKETTDAIQERYDRLEAFIQALPGPLGK 120 GARADLNKKKKHKRA+ RV ELEAQVADLTFRKETTDAIQERYDRLEAFIQALPGPLGK Sbjct: 61 GARADLNKKKKHKAKKRVTELEAQVADLTFRKETTDAIQERYDRLEAFIQALPGPLGK 120</p> <p>Query: 121 ALDSKSFTEKLFESEDRVSDI:IKDWNKANPTATASALGGGTGATSEKSNHMDLLRAAAR 180 ALDSKSFTEKLFESEDRVSDI:IKDWNKANPTAT+ALGGGTGATSEKSNHMDLLRAAAR Sbjct: 121 ALDSKSFTEKLFESEDRVSDI:IKDWNKANPTATAALGGGTGATSEKSNHMDLLRAAAR 180</p> <p>Query: 181 G 181 G Sbjct: 181 G 181</p>
DECISION:	<p>Le start retenu est celui à 4445 pb. Il est conservé chez d'autres homologues, est associé à un RBS de bon score et prédit la plus longue ORF.</p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																																	
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Résultat Phage DB :</p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_5, function unknown, 181</td><td>353</td><td>9e-98</td></tr><tr><td>Quartz_5, scaffolding protein, 181</td><td>352</td><td>4e-97</td></tr><tr><td>Mandalorian_5, scaffolding protein, 181</td><td>345</td><td>4e-95</td></tr><tr><td>Nucci_5, scaffolding protein, 180</td><td>337</td><td>7e-93</td></tr><tr><td>Yuuf_5, scaffolding protein, 182</td><td>328</td><td>4e-90</td></tr><tr><td>Juanyo_5, scaffolding protein, 182</td><td>325</td><td>4e-89</td></tr><tr><td>Zenitsu_5, scaffolding protein, 192</td><td>197</td><td>1e-50</td></tr><tr><td>Sansa_4, scaffolding protein, 208</td><td>197</td><td>1e-50</td></tr><tr><td>MCubed_5, scaffolding protein, 209</td><td>197</td><td>1e-50</td></tr><tr><td>Shamu_5, scaffolding protein, 191</td><td>197</td><td>1e-50</td></tr><tr><td>Saratos_5, scaffolding protein, 208</td><td>197</td><td>1e-50</td></tr><tr><td>Glamour_4, scaffolding protein, 208</td><td>197</td><td>1e-50</td></tr><tr><td>Eleri_5, scaffolding protein, 208</td><td>197</td><td>1e-50</td></tr><tr><td>ColaCorta_5, scaffolding protein, 208</td><td>197</td><td>1e-50</td></tr><tr><td>Andromedas_5, scaffolding protein, 208</td><td>197</td><td>1e-50</td></tr><tr><td>ChikPic_5, scaffolding protein, 192</td><td>196</td><td>3e-50</td></tr><tr><td>Finny_5, scaffolding protein, 191</td><td>187</td><td>9e-48</td></tr><tr><td>Schubert_6, scaffolding protein, 185</td><td>174</td><td>1e-43</td></tr><tr><td>Schimmels22_8, scaffolding protein, 183</td><td>173</td><td>2e-43</td></tr><tr><td>HerculesXL_8, scaffolding protein, 184</td><td>171</td><td>7e-43</td></tr><tr><td>Tinyman4_8, scaffolding protein, 184</td><td>171</td><td>1e-42</td></tr><tr><td>Theresita_6, scaffolding protein, 188</td><td>170</td><td>1e-42</td></tr><tr><td>WestPM_Draft_7, function unknown, 183</td><td>167</td><td>9e-42</td></tr><tr><td>Kauala_5, scaffolding protein, 182</td><td>160</td><td>2e-39</td></tr><tr><td>PrincePhergus_5, scaffolding protein, 182</td><td>160</td><td>2e-39</td></tr><tr><td>Pherbot_5, scaffolding protein, 182</td><td>160</td><td>2e-39</td></tr><tr><td>Bustleton_5, scaffolding protein, 182</td><td>160</td><td>2e-39</td></tr><tr><td>Sinatra_5, scaffolding protein, 182</td><td>159</td><td>3e-39</td></tr><tr><td>Koji_5, scaffolding protein, 182</td><td>157</td><td>1e-38</td></tr><tr><td>BouleyBill_Draft_5, function unknown, 174</td><td>155</td><td>6e-38</td></tr><tr><td>Ixel_7, scaffolding protein, 180</td><td>152</td><td>4e-37</td></tr><tr><td>Pajaza_5, scaffolding protein, 179</td><td>147</td><td>1e-35</td></tr><tr><td>Pikmin_5, scaffolding protein, 179</td><td>145</td><td>4e-35</td></tr><tr><td>Casey_5, scaffolding protein, 179</td><td>145</td><td>4e-35</td></tr><tr><td>Alakazam_8, scaffolding protein, 189</td><td>140</td><td>1e-33</td></tr></table> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <table><tr><td><40</td><td>40-50</td><td>50-80</td><td>80-200</td><td>>=200</td></tr></table> <p>Résultats nr :</p>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_5, function unknown, 181	353	9e-98	Quartz_5, scaffolding protein, 181	352	4e-97	Mandalorian_5, scaffolding protein, 181	345	4e-95	Nucci_5, scaffolding protein, 180	337	7e-93	Yuuf_5, scaffolding protein, 182	328	4e-90	Juanyo_5, scaffolding protein, 182	325	4e-89	Zenitsu_5, scaffolding protein, 192	197	1e-50	Sansa_4, scaffolding protein, 208	197	1e-50	MCubed_5, scaffolding protein, 209	197	1e-50	Shamu_5, scaffolding protein, 191	197	1e-50	Saratos_5, scaffolding protein, 208	197	1e-50	Glamour_4, scaffolding protein, 208	197	1e-50	Eleri_5, scaffolding protein, 208	197	1e-50	ColaCorta_5, scaffolding protein, 208	197	1e-50	Andromedas_5, scaffolding protein, 208	197	1e-50	ChikPic_5, scaffolding protein, 192	196	3e-50	Finny_5, scaffolding protein, 191	187	9e-48	Schubert_6, scaffolding protein, 185	174	1e-43	Schimmels22_8, scaffolding protein, 183	173	2e-43	HerculesXL_8, scaffolding protein, 184	171	7e-43	Tinyman4_8, scaffolding protein, 184	171	1e-42	Theresita_6, scaffolding protein, 188	170	1e-42	WestPM_Draft_7, function unknown, 183	167	9e-42	Kauala_5, scaffolding protein, 182	160	2e-39	PrincePhergus_5, scaffolding protein, 182	160	2e-39	Pherbot_5, scaffolding protein, 182	160	2e-39	Bustleton_5, scaffolding protein, 182	160	2e-39	Sinatra_5, scaffolding protein, 182	159	3e-39	Koji_5, scaffolding protein, 182	157	1e-38	BouleyBill_Draft_5, function unknown, 174	155	6e-38	Ixel_7, scaffolding protein, 180	152	4e-37	Pajaza_5, scaffolding protein, 179	147	1e-35	Pikmin_5, scaffolding protein, 179	145	4e-35	Casey_5, scaffolding protein, 179	145	4e-35	Alakazam_8, scaffolding protein, 189	140	1e-33	<40	40-50	50-80	80-200	>=200
Sequences producing significant alignments:	Score (bits)	E Value																																																																																																																
Carostasia_Draft_5, function unknown, 181	353	9e-98																																																																																																																
Quartz_5, scaffolding protein, 181	352	4e-97																																																																																																																
Mandalorian_5, scaffolding protein, 181	345	4e-95																																																																																																																
Nucci_5, scaffolding protein, 180	337	7e-93																																																																																																																
Yuuf_5, scaffolding protein, 182	328	4e-90																																																																																																																
Juanyo_5, scaffolding protein, 182	325	4e-89																																																																																																																
Zenitsu_5, scaffolding protein, 192	197	1e-50																																																																																																																
Sansa_4, scaffolding protein, 208	197	1e-50																																																																																																																
MCubed_5, scaffolding protein, 209	197	1e-50																																																																																																																
Shamu_5, scaffolding protein, 191	197	1e-50																																																																																																																
Saratos_5, scaffolding protein, 208	197	1e-50																																																																																																																
Glamour_4, scaffolding protein, 208	197	1e-50																																																																																																																
Eleri_5, scaffolding protein, 208	197	1e-50																																																																																																																
ColaCorta_5, scaffolding protein, 208	197	1e-50																																																																																																																
Andromedas_5, scaffolding protein, 208	197	1e-50																																																																																																																
ChikPic_5, scaffolding protein, 192	196	3e-50																																																																																																																
Finny_5, scaffolding protein, 191	187	9e-48																																																																																																																
Schubert_6, scaffolding protein, 185	174	1e-43																																																																																																																
Schimmels22_8, scaffolding protein, 183	173	2e-43																																																																																																																
HerculesXL_8, scaffolding protein, 184	171	7e-43																																																																																																																
Tinyman4_8, scaffolding protein, 184	171	1e-42																																																																																																																
Theresita_6, scaffolding protein, 188	170	1e-42																																																																																																																
WestPM_Draft_7, function unknown, 183	167	9e-42																																																																																																																
Kauala_5, scaffolding protein, 182	160	2e-39																																																																																																																
PrincePhergus_5, scaffolding protein, 182	160	2e-39																																																																																																																
Pherbot_5, scaffolding protein, 182	160	2e-39																																																																																																																
Bustleton_5, scaffolding protein, 182	160	2e-39																																																																																																																
Sinatra_5, scaffolding protein, 182	159	3e-39																																																																																																																
Koji_5, scaffolding protein, 182	157	1e-38																																																																																																																
BouleyBill_Draft_5, function unknown, 174	155	6e-38																																																																																																																
Ixel_7, scaffolding protein, 180	152	4e-37																																																																																																																
Pajaza_5, scaffolding protein, 179	147	1e-35																																																																																																																
Pikmin_5, scaffolding protein, 179	145	4e-35																																																																																																																
Casey_5, scaffolding protein, 179	145	4e-35																																																																																																																
Alakazam_8, scaffolding protein, 189	140	1e-33																																																																																																																
<40	40-50	50-80	80-200	>=200																																																																																																														

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	scaffolding protein [Mycobacterium phage Quartz]	Mycobacterium phage Quartz	358	358	100%	3e-124	99.45%	181	U0038824.1
✓	scaffolding protein [Mycobacterium phage Mandelorian]	Mycobacterium phage Mandelorian	350	350	100%	3e-121	97.24%	181	QX013599.1
✓	scaffolding protein [Mycobacterium phage Nucsi]	Mycobacterium phage Nucsi	343	343	100%	2e-119	96.13%	180	QX022398.1
✓	scaffolding protein [Mycobacterium phage YauY]	Mycobacterium phage YauY	333	333	100%	2e-114	92.86%	182	GF95534.1
✓	scaffolding protein [Mycobacterium phage Juavro]	Mycobacterium phage Juavro	303	303	100%	1e-102	90.86%	182	U0164736.1
✓	scaffolding protein [Mycobacterium phage Zecitau]	Mycobacterium phage Zecitau	181	181	100%	9e-54	95.44%	182	W09893808.1
✓	scaffolding protein [Mycobacterium phage MCalabi]	Mycobacterium phage MCalabi	181	181	100%	7e-54	95.44%	208	Q26937725.1
✓	scaffolding protein [Mycobacterium phage Banaa]	Mycobacterium phage Banaa	181	181	100%	8e-54	95.21%	208	K0568972.2
✓	scaffolding protein [Mycobacterium phage Elari]	Mycobacterium phage Elari	180	180	100%	1e-53	95.21%	208	YP_009623043.1
✓	scaffolding protein [Mycobacterium phage Saratos]	Mycobacterium phage Saratos	180	180	100%	1e-53	95.21%	208	U0327896.1
✓	scaffolding protein [Mycobacterium phage ChbPis]	Mycobacterium phage ChbPis	179	179	100%	2e-53	94.92%	182	W09893769.1
✓	scaffolding protein [Mycobacterium phage Finxy]	Mycobacterium phage Finxy	172	172	93%	1e-50	97.89%	191	GF15734.1
✓	scaffolding protein [Mycobacterium phage HerculesX]	Mycobacterium phage HerculesX	164	164	98%	8e-48	93.51%	184	U0333793.1
✓	scaffolding protein [Mycobacterium phage Truymand]	Mycobacterium phage Truymand	162	162	98%	9e-47	92.97%	184	W09895029.1
✓	scaffolding protein [Mycobacterium phage Schubert]	Mycobacterium phage Schubert	162	162	92%	1e-46	91.79%	180	YP_009819636.1
✓	scaffolding protein [Mycobacterium phage Schirmers22]	Mycobacterium phage Schirmers22	156	156	99%	2e-44	92.79%	183	W09895032.1
✓	scaffolding protein [Mycobacterium phage Kowale]	Mycobacterium phage Kowale	148	148	96%	2e-41	47.22%	182	Q9L31017.1

Des protéines scaffold de forte homologie sont retrouvées avec des deux DB. Les e-values sont très faibles ($<<<10^{-4}$) et le query coverage est supérieur à 95% pour de nombreux hits.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un criblé HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

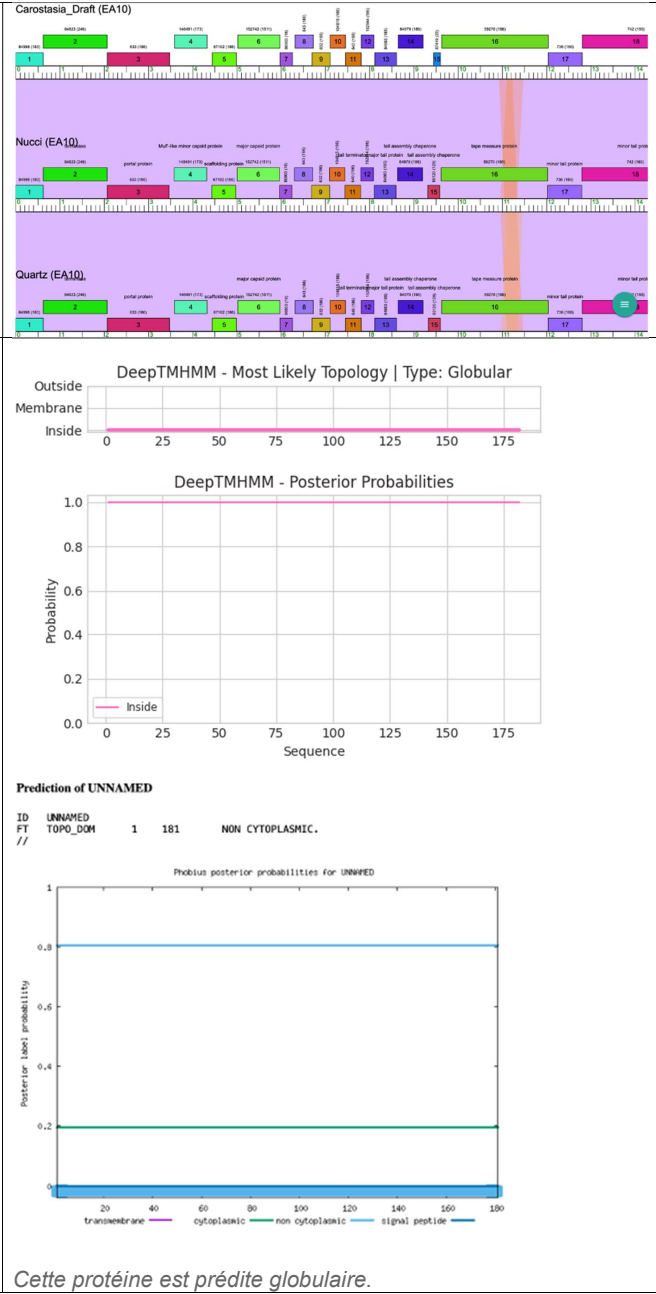
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	ZAKF_B	Coronin-1A; coiled coil, coronin 1, PROTEIN BINDING; 1.2A (N/A)	95.62	0.026	38.15	2.4	31	32
2	PF08954.15	; Trimer_CC ; Trimerisation motif	94.82	0.11	37.98	3.8	36	48
3	PF06810.15	; Phage_GP20 ; Phage minor structural protein GP20	93.9	3.2	33.24	10.2	109	147
4	PF10805.12	; DUF2730 ; Protein of unknown function (DUF2730)	93.3	0.75	34.56	5.7	49	95
5	Q9T1W9	GP25_BPMU Uncharacterized protein gp25 OS=Escherichia phage Mu OX=10677 GN=Mup25 PE=2 SV=1	92.86	0.91	34.78	5.7	49	99
6	Q05222	SCAF_BPMLS Probable capsid assembly scaffolding protein OS=Mycobacterium phage L5 OX=31757 GN=16 PE=1 SV=2	90.07	19	30.33	13.7	154	173
7	O64209	SCAF_BPMD2 Probable capsid assembly scaffolding protein OS=Mycobacterium phage D29 OX=28369 GN=16 PE=3 SV=1	89.76	21	30.39	11.8	157	185
8	Q04763	SCAF_BPLH Capsid assembly scaffolding protein OS=Lactococcus phage LL-H OX=12348 GN=g20 PE=3 SV=1	85.97	12	31.56	7.5	95	178

Le résultat avec HHPred n'est pas aussi concluant à première vue, il trouve des hits proteins binding. Mais aussi des protéines scaffold avec une haute probabilité ($>90\%$ pour deux).

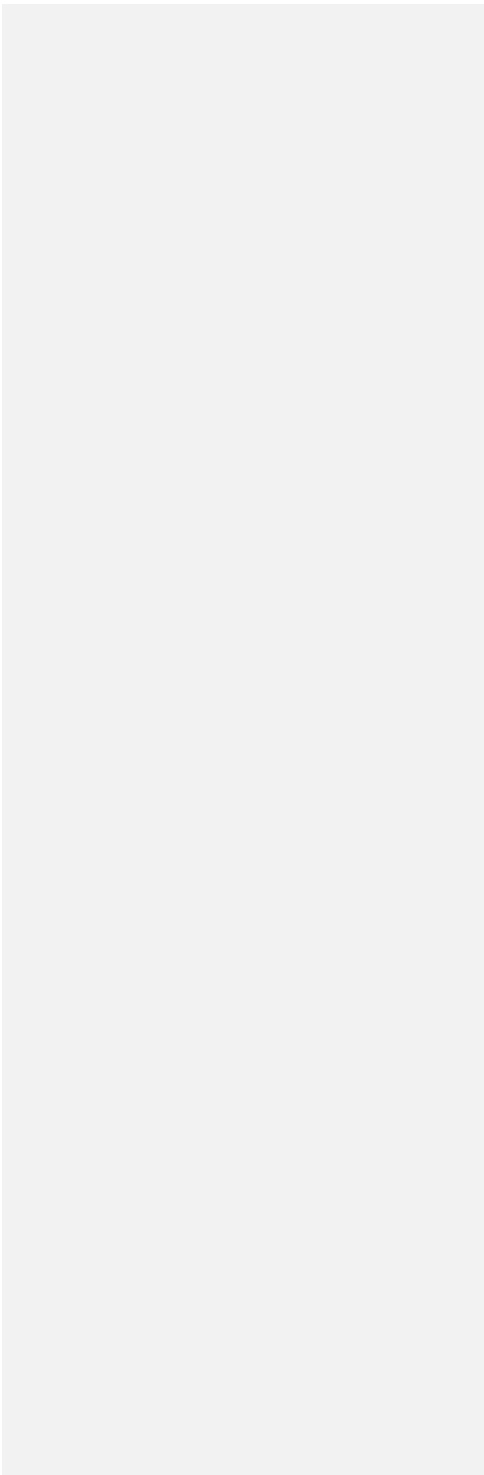
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Chez les phages du même groupe, les gènes 4 et 6 codent respectivement pour une Muf-like minor capsid protein et une major capsid protein. C'est en lien avec les protéines scaffold, il y aurait une forte conservation de l'ordre des gènes.

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



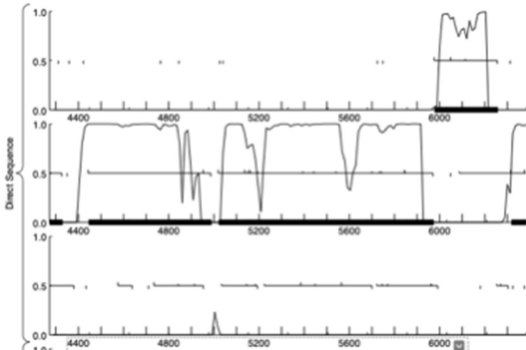
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Oui.</i>
DECISION:	<i>Ce gène code pour une protéine Scaffold.</i>

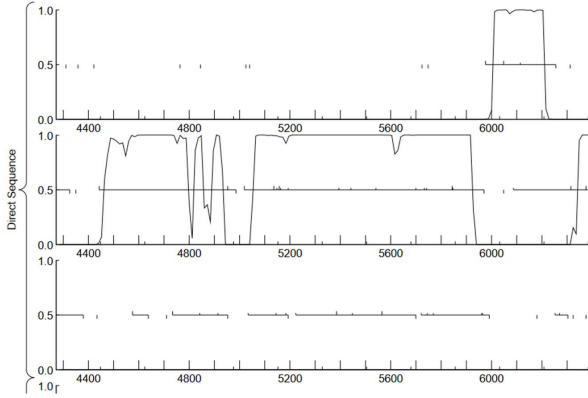
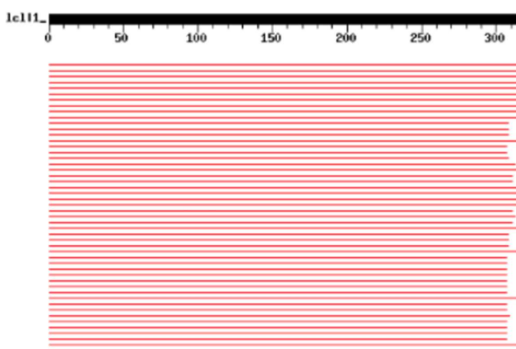


Student Gene Annotation Worksheet

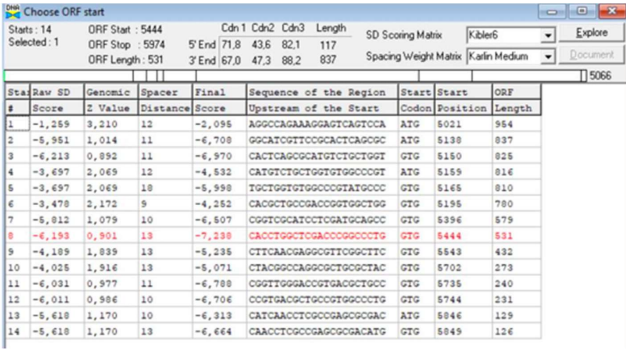

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	6
Coordonnées du Stop	5974
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 30
Coordonnées du Start retenu	5021
Fonction prédite	Major capsid protein

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>Oui, GeneMark détecte un potentiel codant tout le long de l'ORF. Il est néanmoins bas en certains points.</i></p> 

						
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p><i>Oui le candidat est retrouvé chez d'autres génomes annotés. Les hits sont tous de bons scores.</i></p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show details and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <table><tr><td><40</td><td>40-50</td><td>50-100</td><td>100-200</td><td>>200</td></tr></table> 	<40	40-50	50-100	100-200	>200
<40	40-50	50-100	100-200	>200		
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Non</i>					
DECISION:	<i>Oui.</i>					

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	5021 par Glimmer et GeneMark
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<p>Le start est un ATG, le RBS associé est de bon score.</p> 
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Oui le start prédit confère la plus grande ORF = 954pb. Il mène à un gap de 31 pb ce qui n'est pas excessif.</p>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Oui, le start est conservé chez une partie des homologues.</p> 

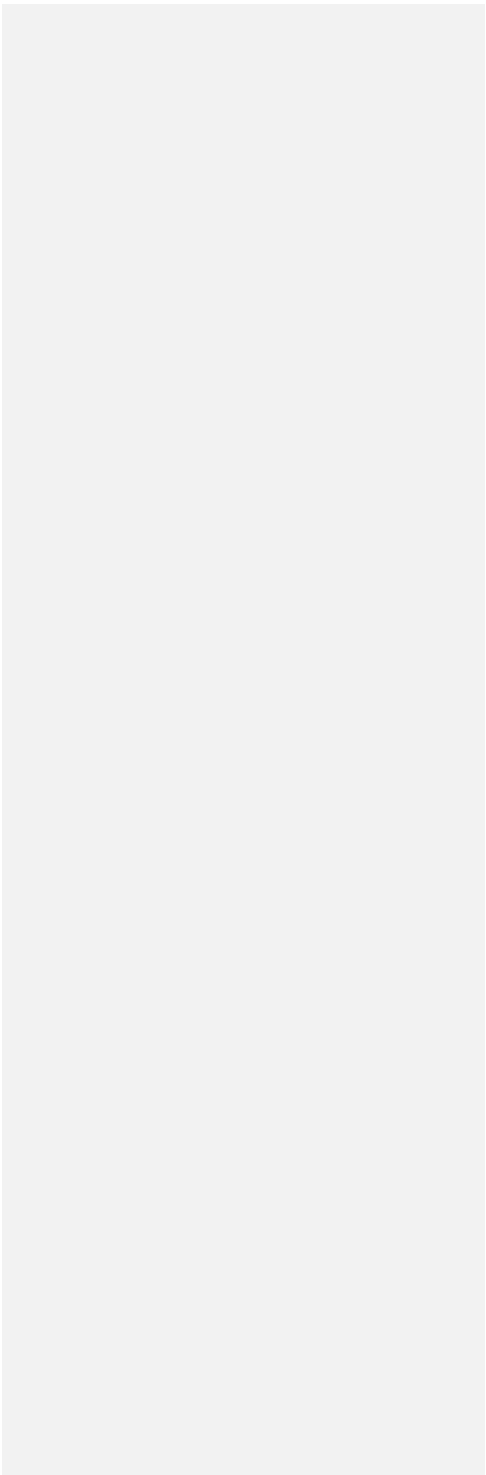
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Oui le start est conservé chez d'autres homologues.</p> <pre> Score = 623 bits (1607), Expect = e-178 Identities = 317/317 (100%), Positives = 317/317 (100%) Query: 1 MADITRADALALLAQDIDEIKFDVAGSVLASFRTQRMAGVARNPVLATLPTGQWYT 60 MADITRADALALLAQDIDEIKFDVAGSVLASFRTQRMAGVARNPVLATLPTGQWYT Sbjct: 1 MADITRADALALLAQDIDEIKFDVAGSVLASFRTQRMAGVARNPVLATLPTGQWYT 60 Query: 61 DNNSTGASGVFTSKVVTOKELIAEIIAVIVFIRENTLADSNFDIMGEVRFLVAQEFGR 120 DNNSTGASGVFTSKVVTOKELIAEIIAVIVFIRENTLADSNFDIMGEVRFLVAQEFGR Sbjct: 61 DNNSTGASGVFTSKVVTOKELIAEIIAVIVFIRENTLADSNFDIMGEVRFLVAQEFGR 120 Query: 121 ILDAAVFFGTNKPFTWLPALVPGAIAGNEQVEGTGGTETGLAADFHEAFGVDEGGF 180 ILDAAVFFGTNKPFTWLPALVPGAIAGNEQVEGTGGTETGLAADFHEAFGVDEGGF Sbjct: 121 ILDAAVFFGTNKPFTWLPALVPGAIAGNEQVEGTGGTETGLAADFHEAFGVDEGGF 180 Query: 181 DVNAAFTGRFLRKRLGLRDOTQPIYLDVAVSDGSTASITQALRYVNLGWRDAAVA 240 DVNAAFTGRFLRKRLGLRDOTQPIYLDVAVSDGSTASITQALRYVNLGWRDAAVA Sbjct: 181 DVNAAFTGRFLRKRLGLRDOTQPIYLDVAVSDGSTASITQALRYVNLGWRDAAVA 240 Query: 241 LVGDASKVVIQVREDTQVLLTEATVGGINLAERDMVALRFKRVAFATATSTAAGQATD 300 LVGDASKVVIQVREDTQVLLTEATVGGINLAERDMVALRFKRVAFATATSTAAGQATD Sbjct: 241 LVGDASKVVIQVREDTQVLLTEATVGGINLAERDMVALRFKRVAFATATSTAAGQATD 300 Query: 301 YFFAVITPAAAEIEGGA 317 YFFAVITPAAAEIEGGA Sbjct: 301 YFFAVITPAAAEIEGGA 317 >Carostasia_Draft_6, function unknown, 317 Length = 317 Score = 623 bits (1607), Expect = e-178 Identities = 317/317 (100%), Positives = 317/317 (100%) Query: 1 MADITRADALALLAQDIDEIKFDVAGSVLASFRTQRMAGVARNPVLATLPTGQWYT 60 MADITRADALALLAQDIDEIKFDVAGSVLASFRTQRMAGVARNPVLATLPTGQWYT Sbjct: 1 MADITRADALALLAQDIDEIKFDVAGSVLASFRTQRMAGVARNPVLATLPTGQWYT 60 Query: 61 DNNSTGASGVFTSKVVTOKELIAEIIAVIVFIRENTLADSNFDIMGEVRFLVAQEFGR 120 DNNSTGASGVFTSKVVTOKELIAEIIAVIVFIRENTLADSNFDIMGEVRFLVAQEFGR Sbjct: 61 DNNSTGASGVFTSKVVTOKELIAEIIAVIVFIRENTLADSNFDIMGEVRFLVAQEFGR 120 Query: 121 ILDAAVFFGTNKPFTWLPALVPGAIAGNEQVEGTGGTETGLAADFHEAFGVDEGGF 180 ILDAAVFFGTNKPFTWLPALVPGAIAGNEQVEGTGGTETGLAADFHEAFGVDEGGF Sbjct: 121 ILDAAVFFGTNKPFTWLPALVPGAIAGNEQVEGTGGTETGLAADFHEAFGVDEGGF 180 Query: 181 DVNAAFTGRFLRKRLGLRDOTQPIYLDVAVSDGSTASITQALRYVNLGWRDAAVA 240 DVNAAFTGRFLRKRLGLRDOTQPIYLDVAVSDGSTASITQALRYVNLGWRDAAVA Sbjct: 181 DVNAAFTGRFLRKRLGLRDOTQPIYLDVAVSDGSTASITQALRYVNLGWRDAAVA 240 Query: 241 LVGDASKVVIQVREDTQVLLTEATVGGINLAERDMVALRFKRVAFATATSTAAGQATD 300 LVGDASKVVIQVREDTQVLLTEATVGGINLAERDMVALRFKRVAFATATSTAAGQATD Sbjct: 241 LVGDASKVVIQVREDTQVLLTEATVGGINLAERDMVALRFKRVAFATATSTAAGQATD 300 Query: 301 YFFAVITPAAAEIEGGA 317 YFFAVITPAAAEIEGGA Sbjct: 301 YFFAVITPAAAEIEGGA 317 </pre>
DECISION:	<p>Le start retenu est celui à 5021 pb.</p> <p>Il est conservé chez d'autres homologues, est associé à un RBS de bon score et prédit la plus longue ORF.</p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	Listez le meilleur hit Blastp pour chaque source :

	<div>*** PhagesDB :</div> <div>>Quartz_6, major capsid protein, 317 Length = 317</div> <div>Score = 623 bits (1607), Expect = e-178 Identities = 317/317 (100%), Positives = 317/317 (100%)</div> <div>Query: 1 MADITRADALALLARQDIDEIIKPDVAQSVALASFRTQRMISAGVARNPVLATLPTGGWVT 60 MADITRADALALLARQDIDEIIKPDVAQSVALASFRTQRMISAGVARNPVLATLPTGGWVT Sbjct: 1 MADITRADALALLARQDIDEIIKPDVAQSVALASFRTQRMISAGVARNPVLATLPTGGWVT 60</div> <div>Query: 61 DNNSTDASGVKPTSKVTWTDKELIAEEIAVIVPIHENTLADSNFDIWEVRPLVAQEFGR 120 DNNSTDASGVKPTSKVTWTDKELIAEEIAVIVPIHENTLADSNFDIWEVRPLVAQEFGR Sbjct: 61 DNNSTDASGVKPTSKVTWTDKELIAEEIAVIVPIHENTLADSNFDIWEVRPLVAQEFGR 120</div> <div>Query: 121 ILDAAVFFGTNKPPTTWLDPALVPGAIAAGNEQVEGTGDTETDLAADFNEAFGFVEDDGF 180 ILDAAVFFGTNKPPTTWLDPALVPGAIAAGNEQVEGTGDTETDLAADFNEAFGFVEDDGF Sbjct: 121 ILDAAVFFGTNKPPTTWLDPALVPGAIAAGNEQVEGTGDTETDLAADFNEAFGFVEDDGF 180</div> <div>Query: 181 DVNAAFTGRFLRRQLRGLRDDTGQPIYLDVAVSDGSTASIVGQALRYVNNLWDRDAAVA 240 DVNAAFTGRFLRRQLRGLRDDTGQPIYLDVAVSDGSTASIVGQALRYVNNLWDRDAAVA Sbjct: 181 DVNAAFTGRFLRRQLRGLRDDTGQPIYLDVAVSDGSTASIVGQALRYVNNLWDRDAAVA 240</div> <div>Query: 241 LVGDASKVVIGVREDYQVKLLTEATVGGINLAERDMVALRFKFRVAFATAYSTAAGQATD 300 LVGDASKVVIGVREDYQVKLLTEATVGGINLAERDMVALRFKFRVAFATAYSTAAGQATD Sbjct: 241 LVGDASKVVIGVREDYQVKLLTEATVGGINLAERDMVALRFKFRVAFATAYSTAAGQATD 300</div> <div>Query: 301 YPFAVITPAAAAEEGGA 317 YPFAVITPAAAAEEGGA Sbjct: 301 YPFAVITPAAAAEEGGA 317</div> <div>*** nr :</div> <table><thead><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident.</th><th>Acc. Len</th><th>Accession</th></tr></thead><tbody><tr><td><input checked="" type="checkbox"/></td><td>major capsid protein [Microbacterium phage Quartz]</td><td>Microbacterium phage Quartz</td><td>632</td><td>632</td><td>89%</td><td>0.0</td><td>100.00%</td><td>317</td><td>U959522.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>major capsid protein [Microbacterium phage Nucsi]</td><td>Microbacterium phage Nucsi</td><td>629</td><td>629</td><td>89%</td><td>0.0</td><td>99.88%</td><td>318</td><td>Q3602399.1</td></tr></tbody></table>		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident.	Acc. Len	Accession	<input checked="" type="checkbox"/>	major capsid protein [Microbacterium phage Quartz]	Microbacterium phage Quartz	632	632	89%	0.0	100.00%	317	U959522.1	<input checked="" type="checkbox"/>	major capsid protein [Microbacterium phage Nucsi]	Microbacterium phage Nucsi	629	629	89%	0.0	99.88%	318	Q3602399.1	
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident.	Acc. Len	Accession																							
<input checked="" type="checkbox"/>	major capsid protein [Microbacterium phage Quartz]	Microbacterium phage Quartz	632	632	89%	0.0	100.00%	317	U959522.1																							
<input checked="" type="checkbox"/>	major capsid protein [Microbacterium phage Nucsi]	Microbacterium phage Nucsi	629	629	89%	0.0	99.88%	318	Q3602399.1																							
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<div>PDB :</div> <div>Template alignment Template 3D Structure PDBe</div> <div><input type="checkbox"/> 1. BECO_B Major capsid protein; HK97-fold, T=7, tailed bacteriophage, VIRUS; 2.2A (Microbacterium phage October96) Probability: 100%, E-value: 2.8e-31, Score: 226.24, Aligned cols: 301, Identities: 75%, Similarity: 1.178, Template Neff: 11</div> <div>UniProt :</div> <div>Template alignment</div> <div><input type="checkbox"/> 11. 064210 CAPSD, BPMD2 Probable major capsid protein gp17 OS=Mycobacterium phage D29 OX=28369 GN=17 PE=3 SV=1 Probability: 99.96%, E-value: 1.5e-26, Score: 198.18, Aligned cols: 291, Identities: 18%, Similarity: 0.261, Template Neff: 11.2</div> <div>Pfam :</div> <div>Template alignment CDD</div> <div><input type="checkbox"/> 21. PF05065.17; Phage_capsid; Phage capsid family Probability: 99.94%, E-value: 4.4e-25, Score: 185.07, Aligned cols: 264, Identities: 14%, Similarity: 0.1, Template Neff: 11.6</div> <div>CD :</div> <div>Template alignment</div> <div><input type="checkbox"/> 1. cd10145 TFIIA_gamma_N; Gamma subunit of transcription initiation factor IIA, N-terminal helical domain. Probability: 2.03%, E-value: 650, Score: 12.46, Aligned cols: 17, Identities: 18%, Similarity: 0.21, Template Neff: 6.4</div>																															
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Oui, proche de la minor capsid protein																															
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	Non																															
Est-ce que la fonction proposée fait partie de liste	Oui, major capsid protein																															

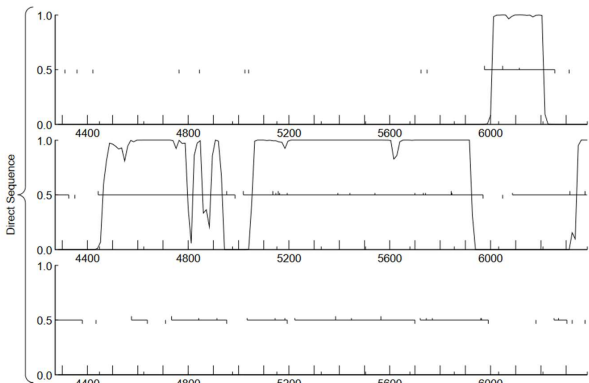
de fonctions approuvées par SEA-PHAGES ?	
DECISION:	Major capsid protein

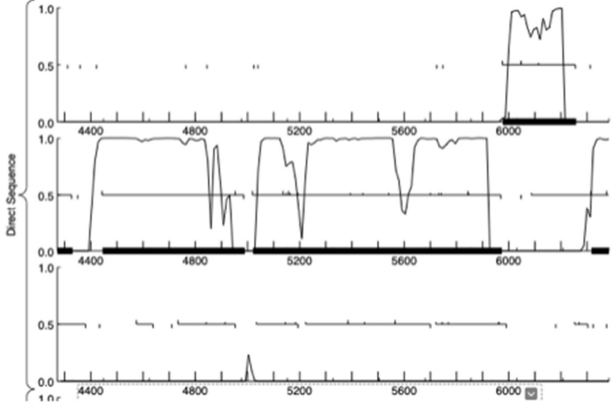
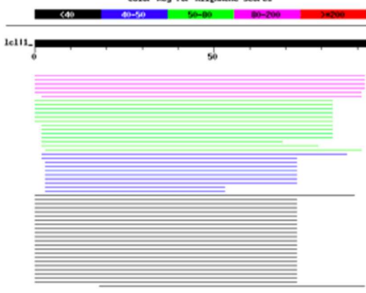


Student Gene Annotation Worksheet

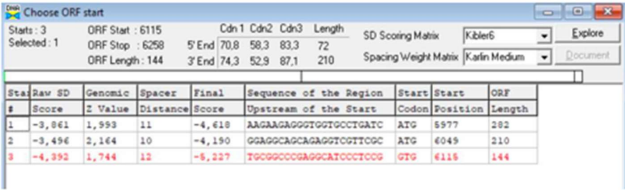
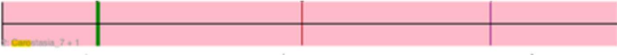
Basic Phage Information	
Nom du Phage	Carostasia
Gène #	7
Coordonnées du Stop	6258
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 2
Coordonnées du Start retenu	5977
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>Oui, GeneMark et GeneMarks détectent un potentiel codant tout le long de l'ORF.</i></p> 

	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Seulement quelques hits avec un score supérieur à 80 sont trouvés.</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <p>0-40 40-50 50-60 60-70 70-80 80-90 90-100</p> 
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non
DECISION:	Oui.

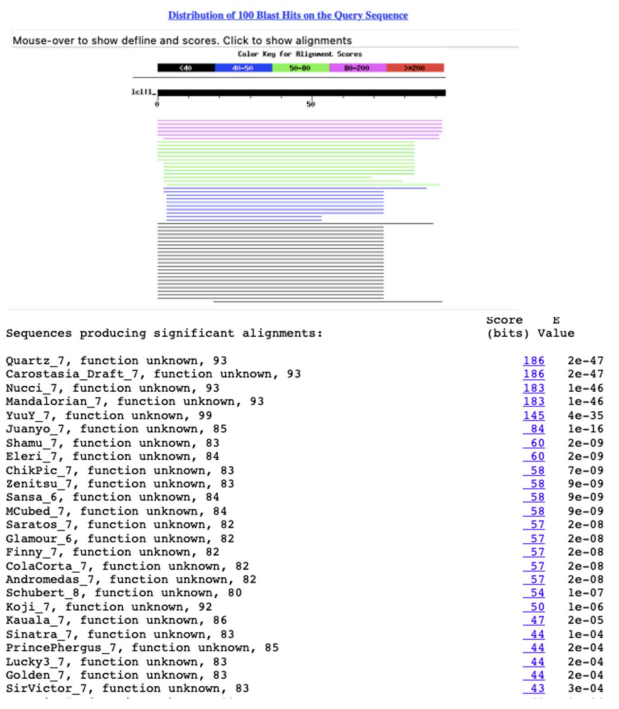
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	5977 par Glimmer et GeneMark
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<p>Le start est un ATG, le RBS associé est de bon score.</p> 
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Oui le start prédit confère la plus grande ORF = 282pb. Gap de 3pb.
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Oui, le start est conservé chez tous les homologues.</p> 
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Oui le start est conservé chez d'autres homologues.

	<pre>Score = 183 bits (445), Expect = 1e-46 Identities = 92/93 (98%), Positives = 92/93 (98%) Query: 1 MASTSAEFRAQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 MASTSAEFRAQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 Sbjct: 1 MASTSAEFRAQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 Query: 61 VENDGTSGFDGQEDGHTQTEDQCTPPPTQESSE 93 VENDGTSGFD GEDGHTQTEDQCTPPPTQESSE 93 Sbjct: 61 VENDGTSGFDGQEDGHTQTEDQCTPPPTQESSE 93 >Yuu_7, function unknown, 99 Length = 99 Score = 145 bits (345), Expect = 4e-35 Identities = 71/92 (77%), Positives = 77/92 (83%) Query: 1 MASTSAEFRAQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 MAST++EFAQKAKKAAEAAYVVMHAAEAANKL ATRA ANGIPFVE P E EPVGTGF 60 Sbjct: 1 MASTSAEFRAQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 Query: 61 VENDGTSGFDGQEDGHTQTEDQCTPPPTQESSE 92 ENDGT+GFD Q DCHT+ E QCTPP Q = 92 Sbjct: 61 VENDGTSGFDGQEDGHTQTEDQCTPPPTQESSE 92 >Juanyo_7, function unknown, 85 Length = 85 Score = 84.0 bits (206), Expect = 1e-16 Identities = 51/92 (55%), Positives = 60/92 (65%), Gaps = 12/92 (13%) Query: 3 STSAEFRAQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 STS E+R QKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 60 Sbjct: 2 STSTENRQKAKKAAEAAYVVMHAAEAANKLRAATRAAANGIPFVEEPAEVEPVGTGF 61 Query: 61 VENDGTSGFDGQEDGHTQTEDQCTPPPTQESSE 92 V +DCHT+ + QCTPP + SE 92 Sbjct: 62 V-----TQCHTEDQCTPPPTQESSE 83</pre>
DECISION:	<i>Le start retenu est à 5977pb, il est associé à un RBS de score satisfaisant, est retrouvé chez les homologues, et confère l'ORF la plus grande.</i>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<i>Phage DB :</i>



nr :

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_7 [Microbacterium phage Quartz]	Microbacterium phage Quartz	184	184	100%	2e-58	100.00%	93	U0558228.1
<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_7 [Microbacterium phage Nucci]	Microbacterium phage Nucci	182	182	100%	1e-57	98.92%	93	QD052450.1
<input checked="" type="checkbox"/> hypothetical protein SEA_YUUY_7 [Microbacterium phage YuuY]	Microbacterium phage YuuY	141	141	94%	3e-41	80.68%	99	GF295338.1
<input checked="" type="checkbox"/> hypothetical protein SEA_JUANYO_7 [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	67.0	67.0	96%	6e-12	53.26%	85	U5144736.1
<input checked="" type="checkbox"/> hypothetical protein ESJ046_007 [Microbacterium phage Esj046_007]	Microbacterium phage Esj046_007	53.0	53.0	99%	1e-06	97.14%	84	YP_006250665.1
<input checked="" type="checkbox"/> hypothetical protein SEA_CHAMPIC_7 [Microbacterium phage Champic]	Microbacterium phage Champic	51.2	51.2	99%	9e-06	95.36%	83	V0069390.1
<input checked="" type="checkbox"/> hypothetical protein SEA_SANSA_6 [Microbacterium phage Sansa]	Microbacterium phage Sansa	50.8	50.8	99%	1e-05	95.36%	84	AX148572.1
<input checked="" type="checkbox"/> hypothetical protein SEA_ZENITSU_7 [Microbacterium phage Zenitsu]	Microbacterium phage Zenitsu	50.8	50.8	99%	1e-05	95.36%	83	V0069390.1
<input checked="" type="checkbox"/> hypothetical protein SEA_MCUBED_7 [Microbacterium phage MCubed]	Microbacterium phage MCubed	50.8	50.8	99%	1e-05	95.36%	84	QD052450.1
<input checked="" type="checkbox"/> hypothetical protein SEA_FINNY_7 [Microbacterium phage Finny]	Microbacterium phage Finny	50.1	50.1	73%	2e-05	47.37%	82	QD051576.1
<input checked="" type="checkbox"/> hypothetical protein SEA_COLACORTA_7 [Microbacterium phage ColaCorta]	Microbacterium phage ColaCorta	49.7	49.7	96%	4e-05	95.56%	82	AX173566.1
<input checked="" type="checkbox"/> hypothetical protein HOV09_008 [Microbacterium phage Schubert]	Microbacterium phage Schubert	48.1	48.1	67%	2e-04	47.62%	80	YP_00918840.1

Beaucoup de hits sont trouvés avec la phage DB, néanmoins la majorité ont de très faibles scores d'identité.

Les E-values sont $<10^{-4}$.

La query cover est acceptable pour quelques hits contre la nr, un plus avec la

Est-ce que la séquence protéique s'aligne avec une protéine de fonction

annotée venant de la PDB
 ou autre base de données
 lors d'un crible HHPred
 avec une proba $\geq 90\%$ et
 une couverture acceptable
 ?

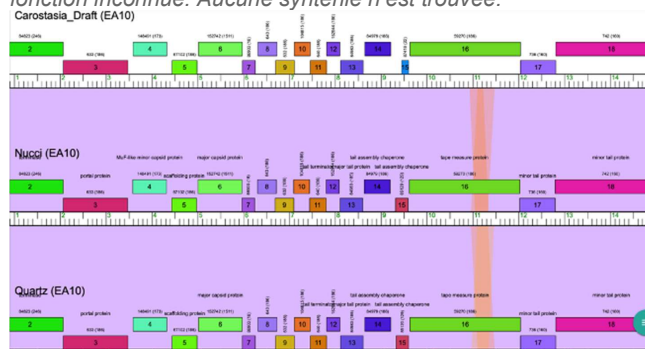
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	2KK7_A	V-type ATP synthase subunit E; subunit E, Methanocaldococcus jannaschii, A1A0 ATP synthase, ATP synthesis, Hydrogen ion	52.96	91	18.45	5.9	42	52
2	2DM9_B	V-type ATP synthase subunit E; A-ATPase, Structural genomics, NPPSFA, National Project on Protein Structural and Functio	49	160	20.17	6.1	42	198
3	5Y6P_C1	LRC4; phycobilisome, PHOTOSYNTHESIS; HET: PUB, PEB, CYC, PVN; 3.5A (Griffithsia pacifica)	46.04	110	26.3	4.5	55	192
4	4EFA_G	V-type proton ATPase subunit G; heterotrimer, peripheral stalk, vacuolar ATPase, HYDROLASE; HET: SO4; 2.8163A (Saccharom	42.99	210	19.7	5.7	41	119

[illegible][illegible]

Aucune target n'est au-dessus de 70% de probabilité, elles ne sont pas intéressantes à évaluer.

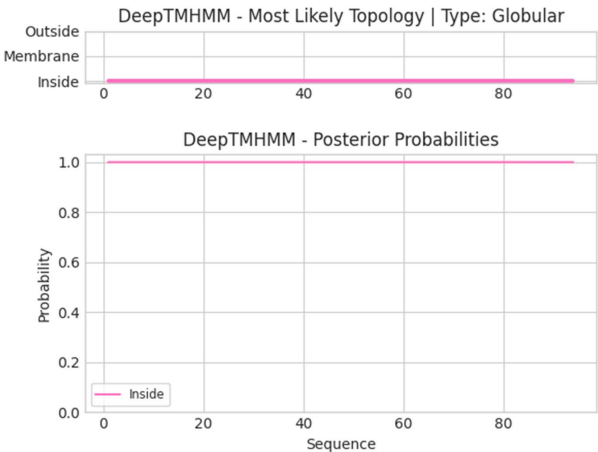
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Chez les phages du même groupe EA10, le gène 6 code pour une major capsid protein et le gène 8 pour une protéine de fonction inconnue. Aucune synténie n'est trouvée.



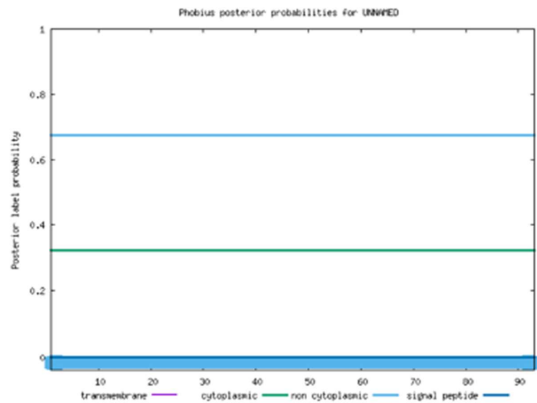
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

Non, ce serait une protéine globulaire.



Prediction of UNNAMED

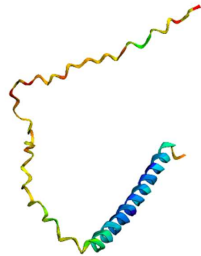
ID UNNAMED 1 93 NON CYTOPLASMIC.
FT TOPO_DOM
//



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Oui.

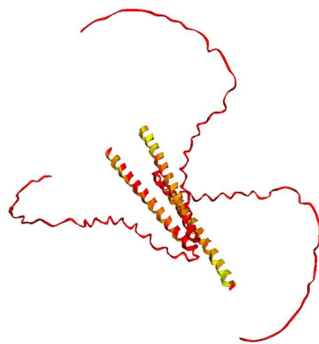
Alphafolds



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

Protéine complètement déstructurée.

Structure trimérique :

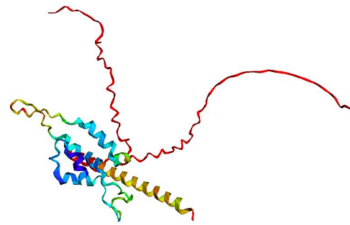


iDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

Cette structure est encore moins fiable, néanmoins c'est peu probable qu'il y est un gap de 240 pbs. Sans aucune expertise, on a l'impression qu'elle pourrait agir avec une protéine codée par un autre gène.

il faudrait utiliser la synténie pour déduire sa fonction... Peut-être qu'elle fonctionne avec le produit du gène 8, qui est aussi inconnue...

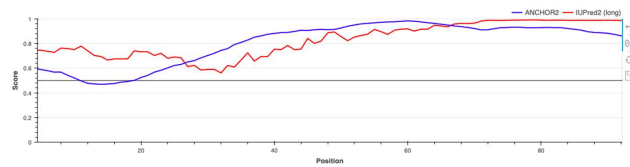
En dimère avec le gène 8 :



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

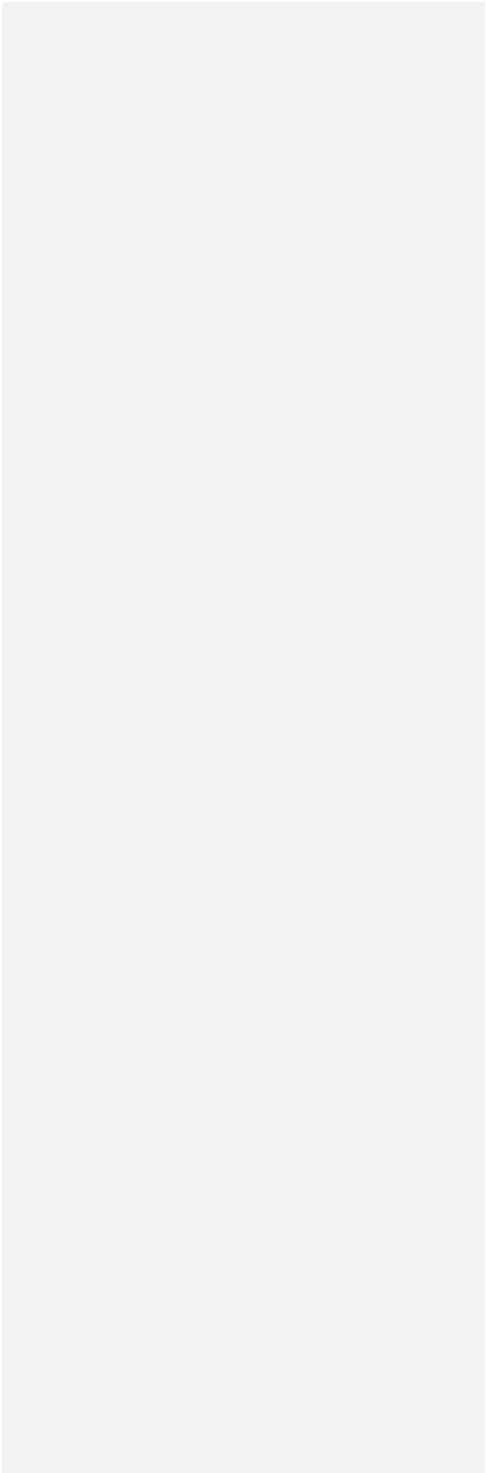
La structure reste très déstructurée.

Résultat IUPred2A :



IUPred précise que la protéine est bien déstructurée en tant que monomère, ce n'est pas dû à un manque de données chez alphafold.

DECISION:	<i>NFK.</i>
------------------	-------------

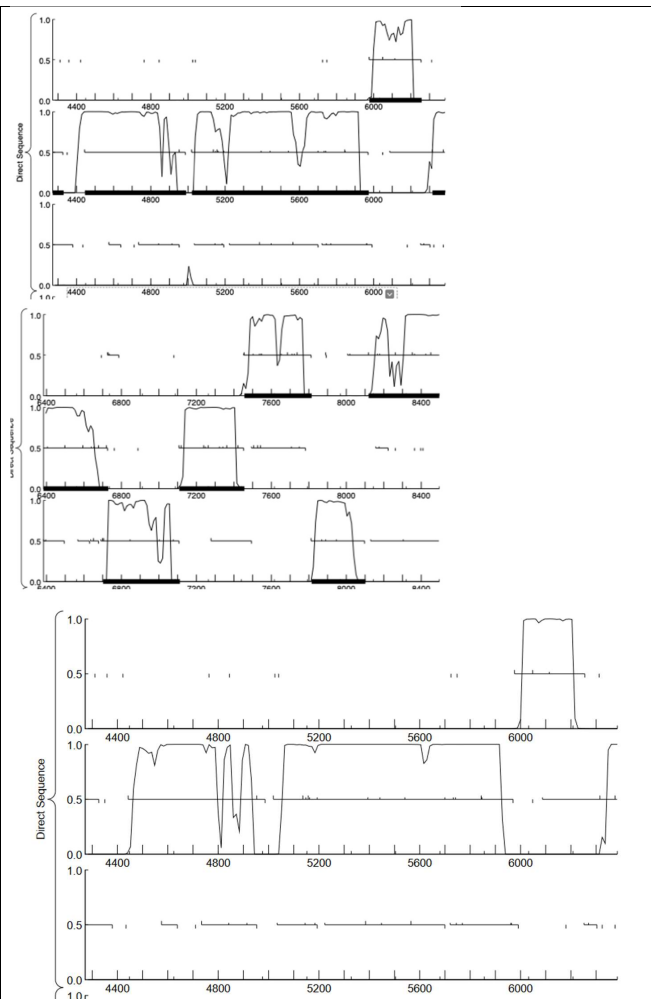


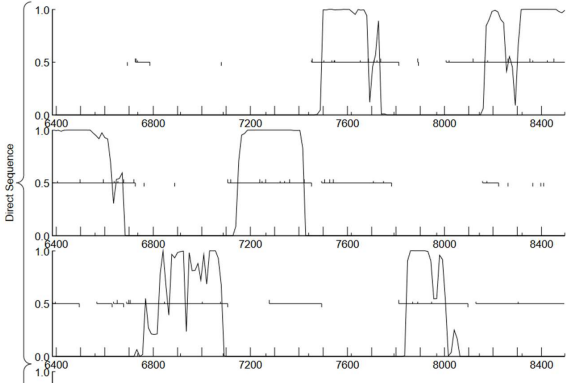

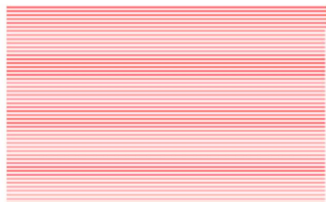
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	8
Coordonnées du Stop	6730
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 59
Coordonnées du Start retenu	6317
Fonction prédite	NFK

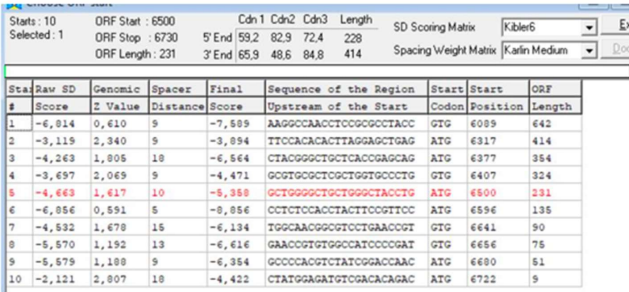
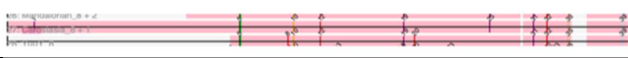
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Oui, GeneMark et GeneMarks détectent un potentiel codant tout le long de l'ORF.</i>



	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p><i>Oui des homologues sont trouvés par Blastp, tous de très bons scores.</i></p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Aligned Scores</p>  
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Non</i>
DECISION:	<i>Oui.</i>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	6317 par Glimmer et GeneMark
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<p>Le start est un ATG, le RBS associé est de bon score.</p> 
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Non un start GTG se trouve en amont. Mais le score RBS est bien plus faible et le potentiel codant n'est pas détecté à partir de cette position.</p>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Oui, le start est conservé chez tous les homologues.</p> 
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Oui le start est conservé chez d'autres homologues.</p>

	<pre>>Quartz_8, function unknown, 137 Length = 137 Score = 285 bits (729), Expect = 3e-77 Identities = 136/137 (99%), Positives = 137/137 (100%) Query: 1 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 Sbjct: 1 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 Query: 61 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 Sbjct: 61 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 Query: 121 NMNHGGERYGDVDTDMG 137 NMNHGGERYGD+DTDMG Sbjct: 121 NMNHGGERYGDIDTDMG 137 >Mandalorian_8, function unknown, 137 Length = 137 Score = 283 bits (725), Expect = 8e-77 Identities = 136/137 (99%), Positives = 137/137 (100%) Query: 1 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 Sbjct: 1 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 Query: 61 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 Sbjct: 61 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 Query: 121 NMNHGGERYGDVDTDMG 137 +NMNHGGERYGDVDTDMG Sbjct: 121 DMNHGGERYGDVDTDMG 137</pre>
DECISION:	<i>Start retenu : 6317. Associé à un bon RBS, correspond au potentiel codant, est conservé chez les homologues.</i>

Décision #3 : Quelle est la fonction de la protéine putative ?

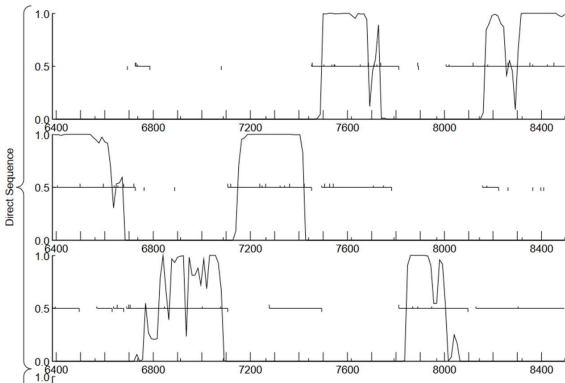
Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <pre>>Quartz_8, function unknown, 137 Length = 137 Score = 285 bits (729), Expect = 3e-77 Identities = 136/137 (99%), Positives = 137/137 (100%) Query: 1 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 Sbjct: 1 MDLNLALLSPLPGYGLLTEQMKQALAGALVPDSFGINPGQPGYETTYDPYFAALGLLGY 60 Query: 61 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 Sbjct: 61 LMAQPVLRLQTSSEGTSAVDAPDWGALSTYFRSMSPICAATGNGVLNRVAIPDVPHVYRT 120 Query: 121 NMNHGGERYGDVDTDMG 137 NMNHGGERYGD+DTDMG Sbjct: 121 NMNHGGERYGDIDTDMG 137 *** nr :</pre>

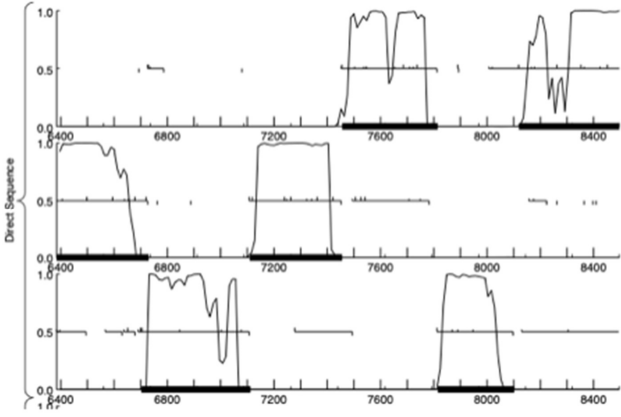
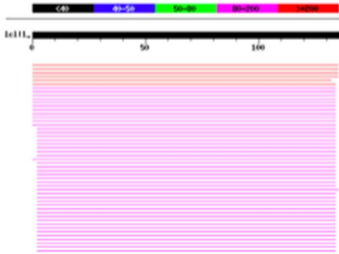
	<table><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E-value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_8 [Microbacterium rhysos Quartz]</td><td>Microbacterium rhysos Quartz</td><td>281</td><td>281</td><td>99%</td><td>4e-95</td><td>99.27%</td><td>137</td><td>U955922.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_8 [Microbacterium rhysos Mandalorian]</td><td>Microbacterium rhysos Mandalorian</td><td>279</td><td>279</td><td>99%</td><td>1e-94</td><td>99.27%</td><td>137</td><td>OR013602.1</td></tr></table>	Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_8 [Microbacterium rhysos Quartz]	Microbacterium rhysos Quartz	281	281	99%	4e-95	99.27%	137	U955922.1	<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_8 [Microbacterium rhysos Mandalorian]	Microbacterium rhysos Mandalorian	279	279	99%	1e-94	99.27%	137	OR013602.1																																				
Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Per. Ident	Acc. Len	Accession																																																								
<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_8 [Microbacterium rhysos Quartz]	Microbacterium rhysos Quartz	281	281	99%	4e-95	99.27%	137	U955922.1																																																								
<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_8 [Microbacterium rhysos Mandalorian]	Microbacterium rhysos Mandalorian	279	279	99%	1e-94	99.27%	137	OR013602.1																																																								
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/> 1</td><td>1P9L_A</td><td>cortaxillin I/GCN4 hybrid peptide; coiled-coil; UNKNOWN FUNCTION; 1.17A (N/A) SCDP: h.1.10.1</td><td>71.9</td><td>1.8</td><td>27.85</td><td>-1.1</td><td>28</td><td>31</td></tr><tr><td><input type="checkbox"/> 2</td><td>PF14081.10</td><td>; DUF4262 ; Domain of unknown function (DUF4262)</td><td>43.21</td><td>42</td><td>25.29</td><td>1.7</td><td>15</td><td>123</td></tr><tr><td><input type="checkbox"/> 3</td><td>PF20449.2</td><td>; DUF6706 ; Family of unknown function (DUF6706)</td><td>42.61</td><td>290</td><td>21.28</td><td>6.2</td><td>86</td><td>110</td></tr><tr><td><input type="checkbox"/> 4</td><td>PF04604.17</td><td>; L_biotic_typeA ; Type-A lantibiotic</td><td>34.89</td><td>77</td><td>22.46</td><td>1.8</td><td>22</td><td>50</td></tr><tr><td><input type="checkbox"/> 5</td><td>PF09860.13</td><td>; DUF2087 ; Uncharacterized protein conserved in bacteria (DUF2087)</td><td>22.96</td><td>280</td><td>19.32</td><td>2.9</td><td>25</td><td>65</td></tr><tr><td><input type="checkbox"/> 6</td><td>PF18881.4</td><td>; DUF5646 ; Family of unknown function (DUF5646)</td><td>22.33</td><td>180</td><td>22.58</td><td>2</td><td>29</td><td>69</td></tr></table> <p>Pas de résultats acceptables</p>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/> 1	1P9L_A	cortaxillin I/GCN4 hybrid peptide; coiled-coil; UNKNOWN FUNCTION; 1.17A (N/A) SCDP: h.1.10.1	71.9	1.8	27.85	-1.1	28	31	<input type="checkbox"/> 2	PF14081.10	; DUF4262 ; Domain of unknown function (DUF4262)	43.21	42	25.29	1.7	15	123	<input type="checkbox"/> 3	PF20449.2	; DUF6706 ; Family of unknown function (DUF6706)	42.61	290	21.28	6.2	86	110	<input type="checkbox"/> 4	PF04604.17	; L_biotic_typeA ; Type-A lantibiotic	34.89	77	22.46	1.8	22	50	<input type="checkbox"/> 5	PF09860.13	; DUF2087 ; Uncharacterized protein conserved in bacteria (DUF2087)	22.96	280	19.32	2.9	25	65	<input type="checkbox"/> 6	PF18881.4	; DUF5646 ; Family of unknown function (DUF5646)	22.33	180	22.58	2	29	69
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																								
<input type="checkbox"/> 1	1P9L_A	cortaxillin I/GCN4 hybrid peptide; coiled-coil; UNKNOWN FUNCTION; 1.17A (N/A) SCDP: h.1.10.1	71.9	1.8	27.85	-1.1	28	31																																																								
<input type="checkbox"/> 2	PF14081.10	; DUF4262 ; Domain of unknown function (DUF4262)	43.21	42	25.29	1.7	15	123																																																								
<input type="checkbox"/> 3	PF20449.2	; DUF6706 ; Family of unknown function (DUF6706)	42.61	290	21.28	6.2	86	110																																																								
<input type="checkbox"/> 4	PF04604.17	; L_biotic_typeA ; Type-A lantibiotic	34.89	77	22.46	1.8	22	50																																																								
<input type="checkbox"/> 5	PF09860.13	; DUF2087 ; Uncharacterized protein conserved in bacteria (DUF2087)	22.96	280	19.32	2.9	25	65																																																								
<input type="checkbox"/> 6	PF18881.4	; DUF5646 ; Family of unknown function (DUF5646)	22.33	180	22.58	2	29	69																																																								
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Aucune synténie observée.																																																															
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	Non																																																															
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Non																																																															
DECISION:	NFK																																																															

Student Gene Annotation Worksheet


Basic Phage Information	
Nom du Phage	Carostasia
Gène #	9
Coordonnées du Stop	7112
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 29
Coordonnées du Start retenu	6702
Fonction prédite	Head to tail adaptor

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>Oui, GeneMark et GeneMarks détectent un potentiel codant tout le long de l'ORF.</i></p> 

	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Oui le candidat est retrouvé chez d'autres génomes annotés, avec un bon score d'alignement.</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> 
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non
DECISION:	Oui.

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	6702 par Glimmer et GeneMark
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	Le start est un ATG, le RBS associé est de bon score (final score = -4,812, il n'y a pas de meilleur score dans les autres starts proposés).
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Non un start est proposé plus en amont, mais c'est un GTG et chevauchait l'ORF en amont, pas excessivement. (28pb)
Est-ce que le start est conservé chez les homologues voir Starterator ?	Le start n'est pas retrouvé chez tous les homologues après analyses sur starterator. 
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Oui le start est conservé chez d'autres homologues avec Blastp. <pre>>Sham_Draft_9, function unknown, 136 Length = 136 Score = 183 bits (465), Expect = 1e-46 Identities = 97/135 (71%), Positives = 111/135 (82%), Gaps = 1/135 (0%) Query: 1 MKNSTQWAGAGLQAVLVDTLQELDNGEPVTSCTKVVRLTPVGEFVAGLVQTTSLA 60 ME+STQW +A L VL D+*QI SW EPVT G V R+LTPVGE+ GLVQTT+LA Sbjct: 1 MEISTQWAAAMTLA-TVLFDFVQITDVMEPTTGVAVTRALTTPVGEFVAGLVQTTSLA 59 Query: 61 NAAELSTVNVVTVVYVARTVIFAGQAKELACQEPOLACKVLLDRVDSQGLAVIRKAV 120 NAAELST+V YVYVYAR T I AGQA+++L C+ EP L GRVLLDRVDSQGL+A+IRKAV Sbjct: 60 NAAELSTINTTVVYVARTDISAGQAVVLEQLQEPVLVORVLLDRVDSQGLAVIRKAV 119 Query: 121 ASDTTVYVDSKEVL 135 ASD TVYVDSKEVL Sbjct: 120 ASDATVYVDSKEVL 134 >Saratos_9, function unknown, 136 Length = 136 Score = 183 bits (465), Expect = 1e-46 Identities = 97/135 (71%), Positives = 111/135 (82%), Gaps = 1/135 (0%) Query: 1 MKNSTQWAGAGLQAVLVDTLQELDNGEPVTSCTKVVRLTPVGEFVAGLVQTTSLA 60 ME+STQW +A L VL D+*QI SW EPVT G V R+LTPVGE+ GLVQTT+LA Sbjct: 1 MEISTQWAAAMTLA-TVLFDFVQITDVMEPTTGVAVTRALTTPVGEFVAGLVQTTSLA 59 Query: 61 NAAELSTVNVVTVVYVARTVIFAGQAKELACQEPOLACKVLLDRVDSQGLAVIRKAV 120 NAAELST+V YVYVYAR T I AGQA+++L C+ EP L GRVLLDRVDSQGL+A+IRKAV Sbjct: 60 NAAELSTINTTVVYVARTDISAGQAVVLEQLQEPVLVORVLLDRVDSQGLAVIRKAV 119 Query: 121 ASDTTVYVDSKEVL 135 ASD TVYVDSKEVL Sbjct: 120 ASDATVYVDSKEVL 134</pre>
DECISION:	Le start conservé est celui positionné à 6702 pb, c'est un ATG associé à un bon RBS et est conservé chez les homologues.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p><i>Phage DB :</i></p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_9, function unknown, 136</td><td>257</td><td>8e-69</td></tr><tr><td>Quartz_9, function unknown, 136</td><td>255</td><td>2e-68</td></tr><tr><td>Mandalorian_9, function unknown, 136</td><td>255</td><td>3e-68</td></tr><tr><td>Nucci_9, function unknown, 136</td><td>250</td><td>1e-66</td></tr><tr><td>Yuyu_9, function unknown, 136</td><td>231</td><td>5e-61</td></tr><tr><td>Juanyo_9, function unknown, 136</td><td>229</td><td>2e-60</td></tr><tr><td>Glamour_8, function unknown, 136</td><td>184</td><td>5e-47</td></tr><tr><td>ColaCorta_9, function unknown, 136</td><td>184</td><td>5e-47</td></tr><tr><td>Andromeda_9, function unknown, 136</td><td>184</td><td>5e-47</td></tr><tr><td>Finny_9, function unknown, 136</td><td>184</td><td>7e-47</td></tr><tr><td>Zenitsu_9, function unknown, 136</td><td>183</td><td>1e-46</td></tr><tr><td>Shamu_9, function unknown, 136</td><td>183</td><td>1e-46</td></tr><tr><td>Saratos_9, function unknown, 136</td><td>183</td><td>1e-46</td></tr><tr><td>Sansa_8, function unknown, 136</td><td>183</td><td>1e-46</td></tr><tr><td>MCubed_9, function unknown, 136</td><td>183</td><td>1e-46</td></tr><tr><td>ChikPic_9, function unknown, 136</td><td>183</td><td>1e-46</td></tr><tr><td>Eleri_9, function unknown, 136</td><td>182</td><td>3e-46</td></tr><tr><td>SonOfLevi_10, function unknown, 137</td><td>177</td><td>1e-44</td></tr><tr><td>Natties_10, function unknown, 137</td><td>177</td><td>1e-44</td></tr><tr><td>Klimt_10, function unknown, 133</td><td>177</td><td>1e-44</td></tr><tr><td>Gelo_10, function unknown, 137</td><td>177</td><td>1e-44</td></tr><tr><td>Etta_10, function unknown, 137</td><td>177</td><td>1e-44</td></tr><tr><td>Erta_10, function unknown, 137</td><td>177</td><td>1e-44</td></tr><tr><td>Chamuel_10, function unknown, 133</td><td>177</td><td>1e-44</td></tr><tr><td>Baines_10, function unknown, 137</td><td>177</td><td>1e-44</td></tr><tr><td>SoilGremlin Draft 10, function unknown, 137</td><td>175</td><td>4e-44</td></tr><tr><td>MillyPhilly_10, function unknown, 137</td><td>175</td><td>4e-44</td></tr><tr><td>Garey24_10, function unknown, 137</td><td>175</td><td>4e-44</td></tr></table> <p>Query: Carostasia_9 (136 letters)</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>nr:</p> <table><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_9 [Microbacterium phage Quartz]</td><td>Microbacterium phage Quartz</td><td>265</td><td>265</td><td>100%</td><td>9e-69</td><td>89.20%</td><td>136</td><td>U9595028.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_9 [Microbacterium phage Mandalorian]</td><td>Microbacterium phage Mandalorian</td><td>264</td><td>264</td><td>100%</td><td>1e-68</td><td>88.93%</td><td>136</td><td>QX013903.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_9 [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>258</td><td>258</td><td>100%</td><td>4e-66</td><td>97.06%</td><td>136</td><td>QX020402.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_JUANYO_9 [Microbacterium phage Juanyo]</td><td>Microbacterium phage Juanyo</td><td>237</td><td>237</td><td>99%</td><td>5e-78</td><td>88.89%</td><td>136</td><td>U9544769.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_YUYU_9 [Microbacterium phage Yuyu]</td><td>Microbacterium phage Yuyu</td><td>237</td><td>237</td><td>97%</td><td>6e-78</td><td>90.23%</td><td>136</td><td>QEP95338.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_COLACORTA_9 [Microbacterium phage ColaCorta]</td><td>Microbacterium phage ColaCorta</td><td>191</td><td>191</td><td>99%</td><td>2e-59</td><td>72.99%</td><td>136</td><td>AX010562.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_FINNY_9 [Microbacterium phage Finny]</td><td>Microbacterium phage Finny</td><td>190</td><td>190</td><td>99%</td><td>2e-59</td><td>71.85%</td><td>136</td><td>QGP15738.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MCUDED_9 [Microbacterium phage MCubed]</td><td>Microbacterium phage MCubed</td><td>189</td><td>189</td><td>99%</td><td>4e-59</td><td>71.85%</td><td>136</td><td>QX020870.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_SANSA_8 [Microbacterium phage Sansa]</td><td>Microbacterium phage Sansa</td><td>189</td><td>189</td><td>99%</td><td>5e-59</td><td>71.85%</td><td>136</td><td>AX048581.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein FOJH_069 [Microbacterium phage Eleri]</td><td>Microbacterium phage Eleri</td><td>188</td><td>188</td><td>99%</td><td>2e-58</td><td>71.11%</td><td>136</td><td>YP_009623047.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein PBI_BAINES_10 [Microbacterium phage Baines]</td><td>Microbacterium phage Baines</td><td>182</td><td>182</td><td>97%</td><td>3e-56</td><td>69.92%</td><td>137</td><td>AY082722.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_KLINT_10 [Microbacterium phage Klimt]</td><td>Microbacterium phage Klimt</td><td>182</td><td>182</td><td>97%</td><td>3e-56</td><td>69.92%</td><td>133</td><td>QGP77658.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MILLYPHILLY_10 [Microbacterium phage MillyPhilly]</td><td>Microbacterium phage MillyPhilly</td><td>181</td><td>181</td><td>97%</td><td>2e-56</td><td>69.92%</td><td>137</td><td>QX022998.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein PBI_ADLER_10 [Microbacterium phage Adler]</td><td>Microbacterium phage Adler</td><td>179</td><td>179</td><td>97%</td><td>4e-55</td><td>69.17%</td><td>137</td><td>AY040401.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_GAREY24_10 [Microbacterium phage Garey24]</td><td>Microbacterium phage Garey24</td><td>179</td><td>179</td><td>97%</td><td>5e-55</td><td>69.17%</td><td>137</td><td>WNC08252.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein PBI_BONINO_10 [Microbacterium phage Bonino]</td><td>Microbacterium phage Bonino</td><td>179</td><td>179</td><td>97%</td><td>9e-55</td><td>68.42%</td><td>137</td><td>AY08212.1</td></tr></table>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_9, function unknown, 136	257	8e-69	Quartz_9, function unknown, 136	255	2e-68	Mandalorian_9, function unknown, 136	255	3e-68	Nucci_9, function unknown, 136	250	1e-66	Yuyu_9, function unknown, 136	231	5e-61	Juanyo_9, function unknown, 136	229	2e-60	Glamour_8, function unknown, 136	184	5e-47	ColaCorta_9, function unknown, 136	184	5e-47	Andromeda_9, function unknown, 136	184	5e-47	Finny_9, function unknown, 136	184	7e-47	Zenitsu_9, function unknown, 136	183	1e-46	Shamu_9, function unknown, 136	183	1e-46	Saratos_9, function unknown, 136	183	1e-46	Sansa_8, function unknown, 136	183	1e-46	MCubed_9, function unknown, 136	183	1e-46	ChikPic_9, function unknown, 136	183	1e-46	Eleri_9, function unknown, 136	182	3e-46	SonOfLevi_10, function unknown, 137	177	1e-44	Natties_10, function unknown, 137	177	1e-44	Klimt_10, function unknown, 133	177	1e-44	Gelo_10, function unknown, 137	177	1e-44	Etta_10, function unknown, 137	177	1e-44	Erta_10, function unknown, 137	177	1e-44	Chamuel_10, function unknown, 133	177	1e-44	Baines_10, function unknown, 137	177	1e-44	SoilGremlin Draft 10, function unknown, 137	175	4e-44	MillyPhilly_10, function unknown, 137	175	4e-44	Garey24_10, function unknown, 137	175	4e-44	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_9 [Microbacterium phage Quartz]	Microbacterium phage Quartz	265	265	100%	9e-69	89.20%	136	U9595028.1	<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_9 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	264	264	100%	1e-68	88.93%	136	QX013903.1	<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_9 [Microbacterium phage Nucci]	Microbacterium phage Nucci	258	258	100%	4e-66	97.06%	136	QX020402.1	<input checked="" type="checkbox"/> hypothetical protein SEA_JUANYO_9 [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	237	237	99%	5e-78	88.89%	136	U9544769.1	<input checked="" type="checkbox"/> hypothetical protein SEA_YUYU_9 [Microbacterium phage Yuyu]	Microbacterium phage Yuyu	237	237	97%	6e-78	90.23%	136	QEP95338.1	<input checked="" type="checkbox"/> hypothetical protein SEA_COLACORTA_9 [Microbacterium phage ColaCorta]	Microbacterium phage ColaCorta	191	191	99%	2e-59	72.99%	136	AX010562.1	<input checked="" type="checkbox"/> hypothetical protein SEA_FINNY_9 [Microbacterium phage Finny]	Microbacterium phage Finny	190	190	99%	2e-59	71.85%	136	QGP15738.1	<input checked="" type="checkbox"/> hypothetical protein SEA_MCUDED_9 [Microbacterium phage MCubed]	Microbacterium phage MCubed	189	189	99%	4e-59	71.85%	136	QX020870.1	<input checked="" type="checkbox"/> hypothetical protein SEA_SANSA_8 [Microbacterium phage Sansa]	Microbacterium phage Sansa	189	189	99%	5e-59	71.85%	136	AX048581.1	<input checked="" type="checkbox"/> hypothetical protein FOJH_069 [Microbacterium phage Eleri]	Microbacterium phage Eleri	188	188	99%	2e-58	71.11%	136	YP_009623047.1	<input checked="" type="checkbox"/> hypothetical protein PBI_BAINES_10 [Microbacterium phage Baines]	Microbacterium phage Baines	182	182	97%	3e-56	69.92%	137	AY082722.1	<input checked="" type="checkbox"/> hypothetical protein SEA_KLINT_10 [Microbacterium phage Klimt]	Microbacterium phage Klimt	182	182	97%	3e-56	69.92%	133	QGP77658.1	<input checked="" type="checkbox"/> hypothetical protein SEA_MILLYPHILLY_10 [Microbacterium phage MillyPhilly]	Microbacterium phage MillyPhilly	181	181	97%	2e-56	69.92%	137	QX022998.1	<input checked="" type="checkbox"/> hypothetical protein PBI_ADLER_10 [Microbacterium phage Adler]	Microbacterium phage Adler	179	179	97%	4e-55	69.17%	137	AY040401.1	<input checked="" type="checkbox"/> hypothetical protein SEA_GAREY24_10 [Microbacterium phage Garey24]	Microbacterium phage Garey24	179	179	97%	5e-55	69.17%	137	WNC08252.1	<input checked="" type="checkbox"/> hypothetical protein PBI_BONINO_10 [Microbacterium phage Bonino]	Microbacterium phage Bonino	179	179	97%	9e-55	68.42%	137	AY08212.1
Sequences producing significant alignments:	Score (bits)	E Value																																																																																																																																																																																																																																															
Carostasia_Draft_9, function unknown, 136	257	8e-69																																																																																																																																																																																																																																															
Quartz_9, function unknown, 136	255	2e-68																																																																																																																																																																																																																																															
Mandalorian_9, function unknown, 136	255	3e-68																																																																																																																																																																																																																																															
Nucci_9, function unknown, 136	250	1e-66																																																																																																																																																																																																																																															
Yuyu_9, function unknown, 136	231	5e-61																																																																																																																																																																																																																																															
Juanyo_9, function unknown, 136	229	2e-60																																																																																																																																																																																																																																															
Glamour_8, function unknown, 136	184	5e-47																																																																																																																																																																																																																																															
ColaCorta_9, function unknown, 136	184	5e-47																																																																																																																																																																																																																																															
Andromeda_9, function unknown, 136	184	5e-47																																																																																																																																																																																																																																															
Finny_9, function unknown, 136	184	7e-47																																																																																																																																																																																																																																															
Zenitsu_9, function unknown, 136	183	1e-46																																																																																																																																																																																																																																															
Shamu_9, function unknown, 136	183	1e-46																																																																																																																																																																																																																																															
Saratos_9, function unknown, 136	183	1e-46																																																																																																																																																																																																																																															
Sansa_8, function unknown, 136	183	1e-46																																																																																																																																																																																																																																															
MCubed_9, function unknown, 136	183	1e-46																																																																																																																																																																																																																																															
ChikPic_9, function unknown, 136	183	1e-46																																																																																																																																																																																																																																															
Eleri_9, function unknown, 136	182	3e-46																																																																																																																																																																																																																																															
SonOfLevi_10, function unknown, 137	177	1e-44																																																																																																																																																																																																																																															
Natties_10, function unknown, 137	177	1e-44																																																																																																																																																																																																																																															
Klimt_10, function unknown, 133	177	1e-44																																																																																																																																																																																																																																															
Gelo_10, function unknown, 137	177	1e-44																																																																																																																																																																																																																																															
Etta_10, function unknown, 137	177	1e-44																																																																																																																																																																																																																																															
Erta_10, function unknown, 137	177	1e-44																																																																																																																																																																																																																																															
Chamuel_10, function unknown, 133	177	1e-44																																																																																																																																																																																																																																															
Baines_10, function unknown, 137	177	1e-44																																																																																																																																																																																																																																															
SoilGremlin Draft 10, function unknown, 137	175	4e-44																																																																																																																																																																																																																																															
MillyPhilly_10, function unknown, 137	175	4e-44																																																																																																																																																																																																																																															
Garey24_10, function unknown, 137	175	4e-44																																																																																																																																																																																																																																															
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_9 [Microbacterium phage Quartz]	Microbacterium phage Quartz	265	265	100%	9e-69	89.20%	136	U9595028.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_9 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	264	264	100%	1e-68	88.93%	136	QX013903.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_9 [Microbacterium phage Nucci]	Microbacterium phage Nucci	258	258	100%	4e-66	97.06%	136	QX020402.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_JUANYO_9 [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	237	237	99%	5e-78	88.89%	136	U9544769.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_YUYU_9 [Microbacterium phage Yuyu]	Microbacterium phage Yuyu	237	237	97%	6e-78	90.23%	136	QEP95338.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_COLACORTA_9 [Microbacterium phage ColaCorta]	Microbacterium phage ColaCorta	191	191	99%	2e-59	72.99%	136	AX010562.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_FINNY_9 [Microbacterium phage Finny]	Microbacterium phage Finny	190	190	99%	2e-59	71.85%	136	QGP15738.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_MCUDED_9 [Microbacterium phage MCubed]	Microbacterium phage MCubed	189	189	99%	4e-59	71.85%	136	QX020870.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_SANSA_8 [Microbacterium phage Sansa]	Microbacterium phage Sansa	189	189	99%	5e-59	71.85%	136	AX048581.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein FOJH_069 [Microbacterium phage Eleri]	Microbacterium phage Eleri	188	188	99%	2e-58	71.11%	136	YP_009623047.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein PBI_BAINES_10 [Microbacterium phage Baines]	Microbacterium phage Baines	182	182	97%	3e-56	69.92%	137	AY082722.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_KLINT_10 [Microbacterium phage Klimt]	Microbacterium phage Klimt	182	182	97%	3e-56	69.92%	133	QGP77658.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_MILLYPHILLY_10 [Microbacterium phage MillyPhilly]	Microbacterium phage MillyPhilly	181	181	97%	2e-56	69.92%	137	QX022998.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein PBI_ADLER_10 [Microbacterium phage Adler]	Microbacterium phage Adler	179	179	97%	4e-55	69.17%	137	AY040401.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein SEA_GAREY24_10 [Microbacterium phage Garey24]	Microbacterium phage Garey24	179	179	97%	5e-55	69.17%	137	WNC08252.1																																																																																																																																																																																																																																									
<input checked="" type="checkbox"/> hypothetical protein PBI_BONINO_10 [Microbacterium phage Bonino]	Microbacterium phage Bonino	179	179	97%	9e-55	68.42%	137	AY08212.1																																																																																																																																																																																																																																									

Descriptions

Graphic Summary

Alignments

Taxonomy

hover to see the title

click to show alignments

Show Conserved Domains

Alignment Scores

< 40

40 - 50

50 - 80

80 - 200

>= 200

71 sequences selected

No putative conserved domains have been detected

Distribution of the top 71 Blast Hits on 71 subject sequences

Les homologues trouvés avec les deux DB sont des protéines putatives/ de fonction inconnue. Les e-values sont $<10^{-4}$ avec un query coverage $\geq 97\%$ pour une grande partie des hits.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Un seul subject à une probabilité $>90\%$, c'est une protéine de fonction inconnue. e.value eleve

Visualization

Hitlist

Show

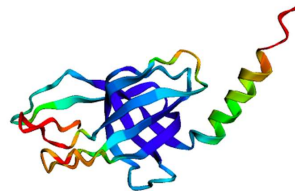
25

Entries

Search:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	PF19586.3	; DUF6093 ; Family of unknown function (DUF6093)	96.67	0.007	43.42	4	101	118
2	PF13856.10	; Gfsy-2 ; ATP-binding sugar transporter from pro-phage	75.5	26	23.98	4.9	79	99
3	PF19637.3	; DUF6140 ; Family of unknown function (DUF6140)	67.94	17	26.25	2.9	22	70
4	3K77_G	DNA repair protein XRCC1; xrcc1, base excision repair; scaffolding protein, DNA damage, DNA repair; Nucleus, Phosphoprot	56.38	33	28.2	3	51	161

	<div>Template alignment CDD</div> <div>1. PF19586.3 : DUF6093 : Family of unknown function (DUF6093)</div> <div>Probability: 96.67%, E-value: 0.007, Score: 43.37, Aligned cols: 101, Identities: 15%, Similarity: 0.147, Template Neff: 8.6</div> <div><div>Q ss_pred</div><div>Q Q_1942888</div><div>Q Consensus</div><div>T Consensus</div><div>T PF19586.3</div><div>T ss_pred</div></div> <div><div>Q ss_pred</div><div>Q Q_1942888</div><div>Q Consensus</div><div>T Consensus</div><div>T PF19586.3</div><div>T ss_pred</div></div>
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<div>Aucune synténie trouvée.</div> <div>Caroslasia_Drait (EA10)</div> <div>Nupcl(EA10)</div> <div>Quatz(EA10)</div>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div>Non, ce serait une protéine globulaire.</div> <div>DeepTMHMM - Most Likely Topology Type: Globular</div> <div>DeepTMHMM - Posterior Probabilities</div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Oui.
AlphaFold	Structure monomérique :



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

La structure trouvée par alphafold est fiable et globulaire comme prédite par DeepTMHMM.

Recherche d'homologie de structure avec Foldseek :

AFDB-PROTEOME 33 hits

GRAPHICAL NUMERIC

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
AF-WASP3GURAF1-model	Phage head-tail adaptor pr...	Enterococcus faecium	1.00	13.8	2.72e-4	1-123	
AF-WASP3GURAF1-model	Uncharacterized protein	Klebsiella pneumoniae sub...	0.99	6.6	1.50e-2	1-121	
AF-WASP3GURAF1-model	Uncharacterized protein	Enterococcus faecium	0.41	6.9	2.56e-1	1-122	

Un hit avec une probabilité de 100%, e-value 2.10^{-4} , "Phage head-tail adaptor".

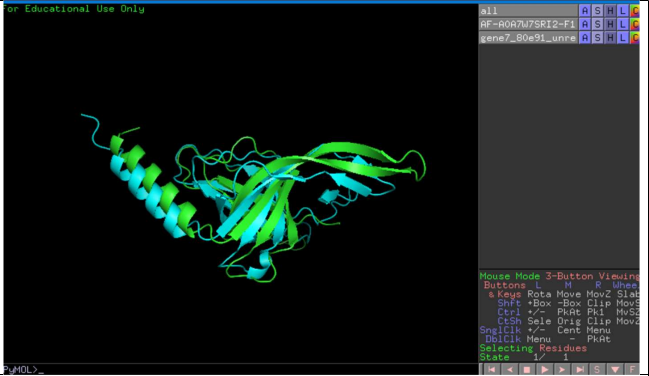
De plus, la structure du premier hit HHPred ressemble fortement à celle de la protéine du gène 9. Les hits trouvés par Seek Fold sont aussi des head to tail adaptor.

AFDB-PROTEOME 20 hits

GRAPHICAL NUMERIC

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
AF-WASP3GURAF1-model	Phage head-tail adaptor pr...	Enterococcus faecium	1.00	10.6	1.81e-3	1-123	
AF-Q2FQ9P-F1-model_v4	Conserved hypothetical ph...	Staphylococcus aureus sub...	1.00	14.4	9.92e-3	1-122	
AF-Q2G8S4-F1-model_v4	SLT or 110-like protein	Staphylococcus aureus sub...	1.00	12.5	6.75e-3	1-121	
AF-Q2FQ9P-F1-model_v4	Phage head-tail adaptor, p...	Staphylococcus aureus sub...	1.00	9	1.31e-2	1-122	

Alignement des deux structures grâce à Pymol :

	<div><div>for Educational Use Only</div><div></div></div>
DECISION:	<p><i>D'après l'homologie structurale ce serait une head to tail adaptor.</i></p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	10
Coordonnées du Stop	7456
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove 4
Coordonnées du Start retenu	7109
Fonction prédite	Minor capsid protein

Décision #1 : Est-ce un gène ?

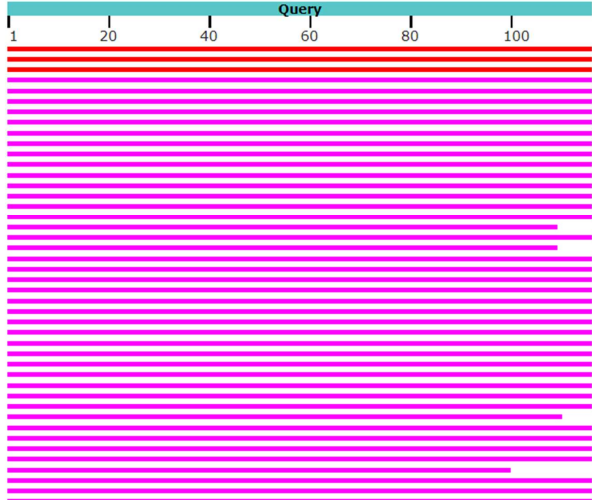
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both Regardez les sorties des pg d'auto-annotation et indiquez (YES BOTH, YES Glimmer only, YES GeneMark only, Neither)</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Les 2 proposent un potentiel codant</i>

Guiding Principles of

	<i>(pham similaire) dans un phage annoté du même groupe dans Phamerator? Indiquez le nom du phage, le numéro de gène.</i>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Discutez s'il y a des conflits avec le Guiding Principles of Genome Annotation. Y a-t-il un chevauchement significatif avec d'autres gènes? Est-il suffisamment long ? Les gènes avant et après ce gène sont-ils dans la même direction ? non</i>
DECISION:	<i>Répondre clairement OUI ou NON à la question « Est-ce un gène ? » oui</i>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																												
Quel start Glimmer et GeneMark suggèrent-ils ?	<p>7109</p> <p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :</p> <p>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :</p>																																																																																																												
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-5,296</td><td>1,321</td><td>10</td><td>-5,991</td><td>GCCAGGCCATCCGGCTGCTGGC</td><td>TTG</td><td>6974</td><td>483</td></tr><tr><td>2</td><td>-1,748</td><td>2,981</td><td>12</td><td>-2,584</td><td>AGGAAGGCCAAGGAGTCTCTGT</td><td>ATG</td><td>7109</td><td>348</td></tr><tr><td>3</td><td>-4,580</td><td>1,656</td><td>8</td><td>-5,802</td><td>AGGTCTCTGTATGAGCATCACG</td><td>ATG</td><td>7121</td><td>336</td></tr><tr><td>4</td><td>-6,377</td><td>0,815</td><td>10</td><td>-7,072</td><td>GGTCAAGCCGAGATCCAGGGC</td><td>ATG</td><td>7241</td><td>216</td></tr><tr><td>5</td><td>-3,778</td><td>2,031</td><td>13</td><td>-4,824</td><td>CGAGATCCAGGGCATGCACGCG</td><td>GTG</td><td>7250</td><td>207</td></tr><tr><td>6</td><td>-3,964</td><td>1,944</td><td>13</td><td>-5,010</td><td>GCACGCGGTGGACACCGGCACC</td><td>ATG</td><td>7265</td><td>192</td></tr><tr><td>7</td><td>-7,220</td><td>0,421</td><td>10</td><td>-7,914</td><td>CACGTACTCTGTGCGCCGCTCG</td><td>GTG</td><td>7325</td><td>132</td></tr><tr><td>8</td><td>-3,964</td><td>1,944</td><td>16</td><td>-5,760</td><td>GTGGTGGACTACGGCGCCTAC</td><td>GTG</td><td>7343</td><td>114</td></tr><tr><td>9</td><td>-6,936</td><td>0,553</td><td>11</td><td>-7,693</td><td>CGTGGCCCTCGGCACACGCGG</td><td>ATG</td><td>7364</td><td>93</td></tr><tr><td>10</td><td>-4,164</td><td>1,851</td><td>10</td><td>-4,859</td><td>GCTGGCTGCACAGGCTGGGAC</td><td>ATG</td><td>7424</td><td>33</td></tr></table> <p>Meilleur score RBS : -2.584, correspond au start retenu (ATG) qui correspond au start retenu.</p> <p>Listez les scores finaux des RBS et les Z-scores start prédits en utilisant la table Kibler6/Karlin Medium. Indiquez dans votre réponse si vous avez retenu le meilleur score ou non et pourquoi. Remarque : si vous proposez plusieurs starts, indiquez pourquoi et fournissez l'information pour chaque start.</p>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-5,296	1,321	10	-5,991	GCCAGGCCATCCGGCTGCTGGC	TTG	6974	483	2	-1,748	2,981	12	-2,584	AGGAAGGCCAAGGAGTCTCTGT	ATG	7109	348	3	-4,580	1,656	8	-5,802	AGGTCTCTGTATGAGCATCACG	ATG	7121	336	4	-6,377	0,815	10	-7,072	GGTCAAGCCGAGATCCAGGGC	ATG	7241	216	5	-3,778	2,031	13	-4,824	CGAGATCCAGGGCATGCACGCG	GTG	7250	207	6	-3,964	1,944	13	-5,010	GCACGCGGTGGACACCGGCACC	ATG	7265	192	7	-7,220	0,421	10	-7,914	CACGTACTCTGTGCGCCGCTCG	GTG	7325	132	8	-3,964	1,944	16	-5,760	GTGGTGGACTACGGCGCCTAC	GTG	7343	114	9	-6,936	0,553	11	-7,693	CGTGGCCCTCGGCACACGCGG	ATG	7364	93	10	-4,164	1,851	10	-4,859	GCTGGCTGCACAGGCTGGGAC	ATG	7424	33
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																					
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																					
1	-5,296	1,321	10	-5,991	GCCAGGCCATCCGGCTGCTGGC	TTG	6974	483																																																																																																					
2	-1,748	2,981	12	-2,584	AGGAAGGCCAAGGAGTCTCTGT	ATG	7109	348																																																																																																					
3	-4,580	1,656	8	-5,802	AGGTCTCTGTATGAGCATCACG	ATG	7121	336																																																																																																					
4	-6,377	0,815	10	-7,072	GGTCAAGCCGAGATCCAGGGC	ATG	7241	216																																																																																																					
5	-3,778	2,031	13	-4,824	CGAGATCCAGGGCATGCACGCG	GTG	7250	207																																																																																																					
6	-3,964	1,944	13	-5,010	GCACGCGGTGGACACCGGCACC	ATG	7265	192																																																																																																					
7	-7,220	0,421	10	-7,914	CACGTACTCTGTGCGCCGCTCG	GTG	7325	132																																																																																																					
8	-3,964	1,944	16	-5,760	GTGGTGGACTACGGCGCCTAC	GTG	7343	114																																																																																																					
9	-6,936	0,553	11	-7,693	CGTGGCCCTCGGCACACGCGG	ATG	7364	93																																																																																																					
10	-4,164	1,851	10	-4,859	GCTGGCTGCACAGGCTGGGAC	ATG	7424	33																																																																																																					
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	<p>Conduit pas à l'orf la plus longue mais l'ORF plus longue fait un chev de 135pb</p> <p>Indiquez la longueur de l'ORF avec le start prédit ainsi que l'écart/chevauchement avec le codon stop le plus proche de l'ORF en amont. Le start proposé présente-t-il un écart/chevauchement avec le gène situé juste en amont qui n'entre pas en conflit avec le Guiding Principles of Genome Annotation ?</p> <p>Remarque : si vous proposez plusieurs starts, indiquez pourquoi et fournissez l'information pour chaque start.</p>																																																																																																												
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>19: Carostasia_10</p> <p>Pas comparaison possible avec starteratorVous devrez également fournir les informations suivantes à partir de Starterator: le start correspond-il au start consensus prédit par Starterator ? Si non, le start consensus est-il trouvé dans cet ORF ? Si non, y a-t-il une meilleure option pour le start que celui prédit par Starterator et pourquoi ? Si Starterator n'a pas fourni de start consensus, vous pouvez préciser que Starterator n'a pas fourni d'informations.</p>																																																																																																												

	<p>Remarque : si vous proposez plusieurs starts, indiquez pourquoi et fournissez l'information pour chaque start.</p>
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<p>Distribution of the top 100 Blast Hits on 100 subject sequenc</p>  <p>Start conservé chez les homologues</p> <p>Si nécessaire, fournissez le meilleur hit BlastP de NCBI, PhagesDB avec l'alignement correspondant dans le format (Q#: S#), où Q (requête) est la séquence que vous analysez et S (sujet) est la correspondance dans la base de données. Donnez la e-value de l'alignement et la localisation du start retenu. Justifiez votre démarche.</p> <p>Remarque : si vous proposez plusieurs starts, indiquez pourquoi et fournissez l'information pour chaque start.</p>
<p>DECISION:</p>	<p>Indiquez le start retenu et justifiez brièvement.</p> <p>On a retenu le start 7109 car correspond au potentiel codant des 2 genMark, bon score rbs et pas de divergence avec les règles et homologues avec le même start</p>

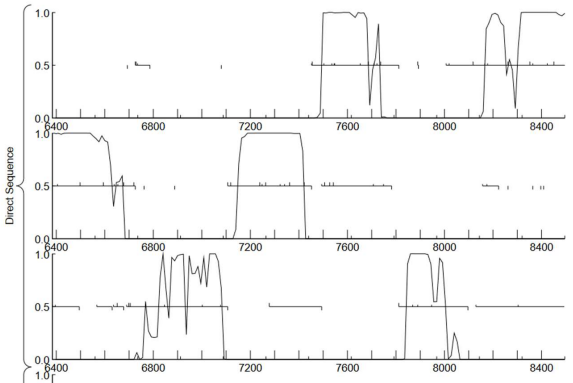
Décision #3 : Quelle est la fonction de la protéine putative ?

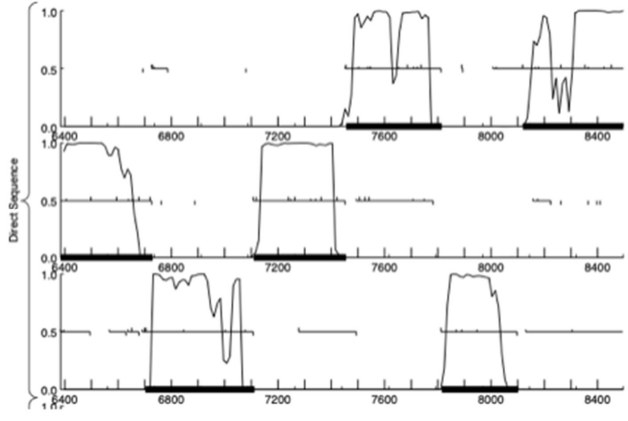
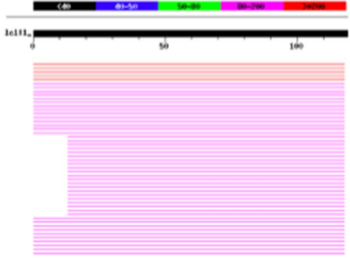
Collection des éléments de réponse	Rationnelle																																							
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?</p>	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>>Quartz_10, function unknown, 115 Length = 115</p> <p>Score = 224 bits (571), Expect = 6e-59 Identities = 115/115 (100%), Positives = 115/115 (100%)</p> <p>Query: 1 MSITMGQLASKFADAASKVPSVAEETVRTLAQVGVLVKREIQGMHAVDTGTMNSTDAE 60 MSITMGQLASKFADAASKVPSVAEETVRTLAQVGVLVKREIQGMHAVDTGTMNSTDAE 60 Sbjct: 1 MSITMGQLASKFADAASKVPSVAEETVRTLAQVGVLVKREIQGMHAVDTGTMNSTDAE 60</p> <p>Query: 61 KVGRATYLVGPSVDYAPYVALGTSRMPARPFHVAAAKKLAQAQDMGFDTDKLGL 115 KVGRATYLVGPSVDYAPYVALGTSRMPARPFHVAAAKKLAQAQDMGFDTDKLGL 115 Sbjct: 61 KVGRATYLVGPSVDYAPYVALGTSRMPARPFHVAAAKKLAQAQDMGFDTDKLGL 115</p> <p>*** nr :</p> <table><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per Ident</th><th>Acc Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_10 [Microbacterium phage Mandelorian]</td><td>Microbacterium phage Mandelorian</td><td>232</td><td>232</td><td>99%</td><td>2e-76</td><td>100.00%</td><td>115</td><td>QX013604.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_10 [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>230</td><td>230</td><td>99%</td><td>1e-75</td><td>99.13%</td><td>115</td><td>QX002403.1</td></tr></table> <p>Les premiers hits matches avec hypothetical proteins</p> <p>Mais il est possible de trouver :</p> <p>minor capsid protein [Microbacterium phage HerculesXL] Sequence ID: UVG33797.1 Length: 114 Number of Matches: 1 See 1 more title(s) See all Identical Proteins(IPG)</p> <p>Range 1: 1 to 114 GenPept Graphics Next Match</p> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>79.0 bits(193)</td><td>6e-16</td><td>Composition-based stats.</td><td>81/114(71%)</td><td>96/114(84%)</td><td>6/114(5%)</td></tr></table> <p>Query 1 MSITMGQLASKFADAASKVPSVAEETVRTLAQVGVLVKREIQGMHAVDTGTMNSTDAE 60 MSIT+GQLAS+FA AA KV V+++ +RTLAQVGVLVK EIQGMHAVDTGTMNST AE Sbjct 1 MSITLQGLASRFALAAQKVPVSDQKLRTLAQVGVLVKSEIQGMHAVDTGTMNSTTAE 60</p> <p>Query 61 KVGRATYLVGPSVDYAPYVALGTSRMPARPFHVAAAKKLAQAQDMGF 108 + G++T+LVGP+V YAPYVALGTSRMPARPFH+ AA+KL+ Q AGD+G Sbjct 61 QAGKSTFLVGPTVKYAPYVALGTSRMPARPFHIEAARKLSKQADEILNAGDLGL 114</p> <p>Evalue : 10⁻¹⁶</p>	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession	<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_10 [Microbacterium phage Mandelorian]	Microbacterium phage Mandelorian	232	232	99%	2e-76	100.00%	115	QX013604.1	<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_10 [Microbacterium phage Nucci]	Microbacterium phage Nucci	230	230	99%	1e-75	99.13%	115	QX002403.1	Score	Expect	Method	Identities	Positives	Gaps	79.0 bits(193)	6e-16	Composition-based stats.	81/114(71%)	96/114(84%)	6/114(5%)
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession																																
<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_10 [Microbacterium phage Mandelorian]	Microbacterium phage Mandelorian	232	232	99%	2e-76	100.00%	115	QX013604.1																																
<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_10 [Microbacterium phage Nucci]	Microbacterium phage Nucci	230	230	99%	1e-75	99.13%	115	QX002403.1																																
Score	Expect	Method	Identities	Positives	Gaps																																			
79.0 bits(193)	6e-16	Composition-based stats.	81/114(71%)	96/114(84%)	6/114(5%)																																			
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?</p>	<p>Pfam :</p> <p>Template alignment CDD</p> <p><input type="checkbox"/> 1. PF11114.12 : Minor capsid 2 ; Minor capsid protein Probability: 99.57%, E-value: 6.7e-14, Score: 79.81, Aligned cols: 87, Identities: 11%, Similarity: -0.061, Template Neff: 9.6</p> <p>UniProt :</p> <p>Template alignment</p> <p><input type="checkbox"/> 3. Q01261 GPG_BPMU Putative capsid assembly protein G OS=Escherichia phage Mu OX=10677 GN=G PE=2 SV=1 Probability: 99.28%, E-value: 5.8e-11, Score: 72.14, Aligned cols: 66, Identities: 20%, Similarity: 0.159, Template Neff: 10.1</p> <p>PDB :</p> <p>Template alignment Template 3D Structure PDBe</p> <p><input type="checkbox"/> 15.1EZE_A CHOLESTERYL ESTER TRANSFERASE INHIBITOR PROTEIN; amphipathic helix, TRANSFERASE INHIBITOR; NMR (N/A) SCOP: J.39.1.1 Probability: 35.85%, E-value: 120, Score: 14.77, Aligned cols: 27, Identities: 11%, Similarity: 0.185, Template Neff: 5.4</p>																																							

Student Gene Annotation Worksheet


Basic Phage Information	
Nom du Phage	Carostasia
Gène #	11
Coordonnées du Stop	7815
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	ove 1
Coordonnées du Start retenu	7456
Fonction prédite	Tail terminator

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both.
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarks detects a coding potential along the entire length of the ORF, but with a downward peak in the middle of the ORF.</p> 

	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Yes, the candidate is found in other annotated genomes, with a good alignment score. Many of the hits do not cover the entire 5' query.</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> 
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	No
DECISION:	Yes

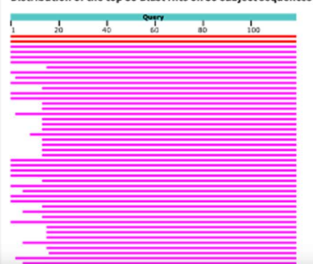
Décision #2 : Quel est le start du gène ?

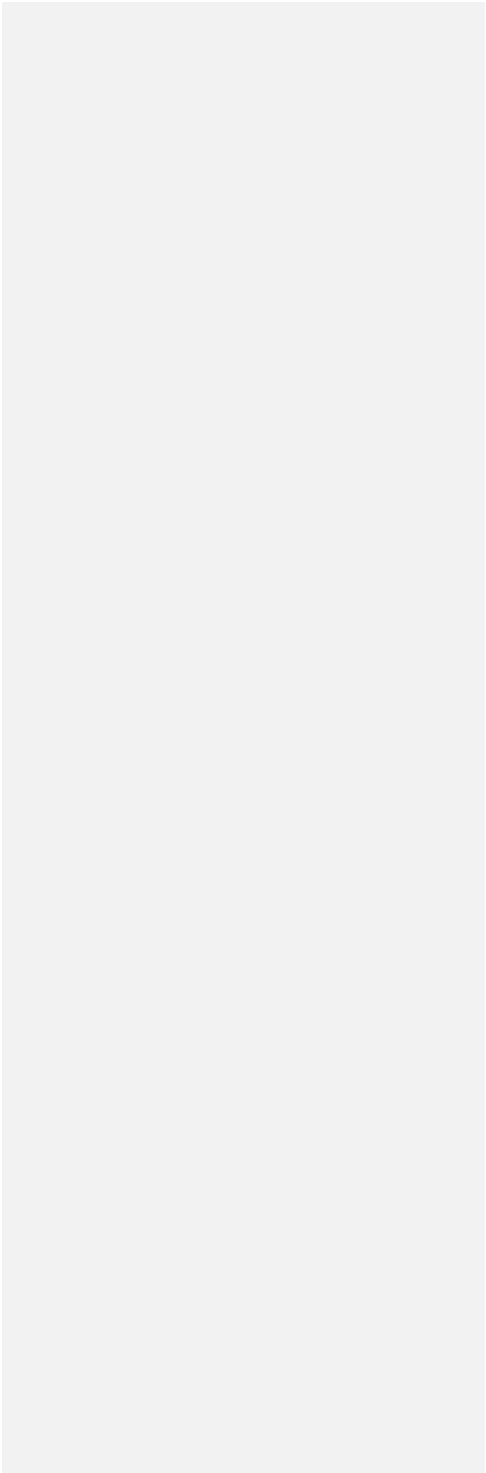
Collection des éléments de réponse	Rationnelle																																																																																																												
Quel start Glimmer et GeneMark suggèrent-ils ?	7456 by Glimmer and GeneMark																																																																																																												
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Start #</th><th>Raw SD Score</th><th>Genomic Z Value</th><th>Spacer Distance</th><th>Final Score</th><th>Sequence of the Region Upstream of the Start</th><th>Start Codon</th><th>Start Position</th><th>ORF Length</th></tr><tr><td>1</td><td>-3,654</td><td>2,089</td><td>6</td><td>-5,399</td><td>TOGACACGGACAGCTGGGGCT</td><td>GTG</td><td>7453</td><td>363</td></tr><tr><td>2</td><td>-3,654</td><td>2,089</td><td>9</td><td>-4,429</td><td>ACACGGACAGCTGGGGCTGTG</td><td>ATG</td><td>7456</td><td>360</td></tr><tr><td>3</td><td>-5,202</td><td>1,365</td><td>9</td><td>-5,977</td><td>GCTGGTGGGCATCGGTGACCCG</td><td>GTG</td><td>7504</td><td>312</td></tr><tr><td>4</td><td>-4,502</td><td>1,693</td><td>11</td><td>-5,259</td><td>CTTGGCCCCACGAATGGGAAG</td><td>GTG</td><td>7537</td><td>279</td></tr><tr><td>5</td><td>-2,654</td><td>2,558</td><td>10</td><td>-3,348</td><td>CACGAATGGGAAGTGGCCCTAT</td><td>GTG</td><td>7546</td><td>270</td></tr><tr><td>6</td><td>-2,654</td><td>2,558</td><td>13</td><td>-3,699</td><td>GAATGGGAAGTGGCCCTATGTG</td><td>GTG</td><td>7549</td><td>267</td></tr><tr><td>7</td><td>-3,924</td><td>1,963</td><td>9</td><td>-4,699</td><td>CGCGTACTGGCGGGGGCGAGC</td><td>GTG</td><td>7654</td><td>162</td></tr><tr><td>8</td><td>-4,515</td><td>1,686</td><td>6</td><td>-6,260</td><td>CTTCAACCTGGCGAAAGCGGTC</td><td>ATG</td><td>7687</td><td>129</td></tr><tr><td>9</td><td>-2,972</td><td>2,409</td><td>10</td><td>-3,667</td><td>GGGCAACCTCCAGGACAGGGT</td><td>GTG</td><td>7711</td><td>105</td></tr><tr><td>10</td><td>-5,570</td><td>1,192</td><td>10</td><td>-6,265</td><td>GGGACAGCGTGTGGCCGGTTAC</td><td>GTG</td><td>7723</td><td>93</td></tr><tr><td>11</td><td>-6,882</td><td>0,579</td><td>13</td><td>-7,927</td><td>CGGTTACGTGCTTTCAACGTCC</td><td>ATG</td><td>7739</td><td>78</td></tr></table> <p>Yes, this start is associated with an RBS with a final score of -4.429. This is an ATG.</p>	Start #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length	1	-3,654	2,089	6	-5,399	TOGACACGGACAGCTGGGGCT	GTG	7453	363	2	-3,654	2,089	9	-4,429	ACACGGACAGCTGGGGCTGTG	ATG	7456	360	3	-5,202	1,365	9	-5,977	GCTGGTGGGCATCGGTGACCCG	GTG	7504	312	4	-4,502	1,693	11	-5,259	CTTGGCCCCACGAATGGGAAG	GTG	7537	279	5	-2,654	2,558	10	-3,348	CACGAATGGGAAGTGGCCCTAT	GTG	7546	270	6	-2,654	2,558	13	-3,699	GAATGGGAAGTGGCCCTATGTG	GTG	7549	267	7	-3,924	1,963	9	-4,699	CGCGTACTGGCGGGGGCGAGC	GTG	7654	162	8	-4,515	1,686	6	-6,260	CTTCAACCTGGCGAAAGCGGTC	ATG	7687	129	9	-2,972	2,409	10	-3,667	GGGCAACCTCCAGGACAGGGT	GTG	7711	105	10	-5,570	1,192	10	-6,265	GGGACAGCGTGTGGCCGGTTAC	GTG	7723	93	11	-6,882	0,579	13	-7,927	CGGTTACGTGCTTTCAACGTCC	ATG	7739	78
Start #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length																																																																																																					
1	-3,654	2,089	6	-5,399	TOGACACGGACAGCTGGGGCT	GTG	7453	363																																																																																																					
2	-3,654	2,089	9	-4,429	ACACGGACAGCTGGGGCTGTG	ATG	7456	360																																																																																																					
3	-5,202	1,365	9	-5,977	GCTGGTGGGCATCGGTGACCCG	GTG	7504	312																																																																																																					
4	-4,502	1,693	11	-5,259	CTTGGCCCCACGAATGGGAAG	GTG	7537	279																																																																																																					
5	-2,654	2,558	10	-3,348	CACGAATGGGAAGTGGCCCTAT	GTG	7546	270																																																																																																					
6	-2,654	2,558	13	-3,699	GAATGGGAAGTGGCCCTATGTG	GTG	7549	267																																																																																																					
7	-3,924	1,963	9	-4,699	CGCGTACTGGCGGGGGCGAGC	GTG	7654	162																																																																																																					
8	-4,515	1,686	6	-6,260	CTTCAACCTGGCGAAAGCGGTC	ATG	7687	129																																																																																																					
9	-2,972	2,409	10	-3,667	GGGCAACCTCCAGGACAGGGT	GTG	7711	105																																																																																																					
10	-5,570	1,192	10	-6,265	GGGACAGCGTGTGGCCGGTTAC	GTG	7723	93																																																																																																					
11	-6,882	0,579	13	-7,927	CGGTTACGTGCTTTCAACGTCC	ATG	7739	78																																																																																																					
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	No, another start is proposed upstream, but it's a GTG and has a lower RBS score. Nevertheless, retaining this start would not lead to excessive overlap with the upstream ORF (3bp).																																																																																																												
Est-ce que le start est conservé chez les homologues voir Starterator ?	Start is not found in all homologs after starterator analysis. 																																																																																																												
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Yes, the start is preserved in other homologs with Blastp.																																																																																																												

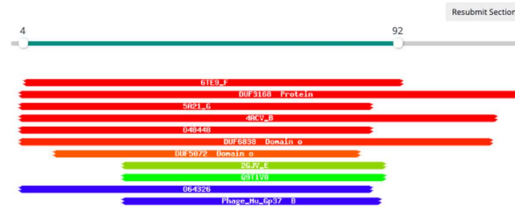
	<pre>>Juanyo_11, tail terminator, 119 Length = 119 Score = 215 bits (548), Expect = 3e-56 Identities = 106/119 (89%), Positives = 113/119 (94%) Query: 1 MSLQAIFKALVIGIGDPVVGAFPTNARYFTTVLSPFILOLDEGPAAGNAISWDSQVAA 60 MSLQAIFK AL GIGDPV+VGFAP+NAKYFTTVLSP FILOLDEGPA+AGNAI+WD+QV+A Sbjct: 1 MSLQAIFKALGGIGDPVVGAFPTNARYFTTVLSPFILOLDEGPAAGNAITWDSQVSA 60 Query: 61 YCAGGSVEASFNLAKAVNGLQGGVAGTVLSTENGTGAPVEGHEAQVTIQSNQGGI 119 YCAGGSVEASFNLAKAVNG LQG RYA TVLSTENGTGAPVEGHEAQVTIQ+MQGGI Sbjct: 61 YCAGGSVEASFNLAKAVNGLQGGVAGTVLSTENGTGAPVEGHEAQVTIQSNQGGI 119 >YuuY_11, tail terminator, 117 Length = 117 Score = 187 bits (474), Expect = 1e-47 Identities = 92/119 (77%), Positives = 102/119 (85%), Gaps = 2/119 (1%) Query: 1 MSLQAIFKALVIGIGDPVVGAFPTNARYFTTVLSPFILOLDEGPAAGNAISWDSQVAA 60 MSLQAIFKAA+ GIGDPVVG+AF+NAKYFTTVLSP +LD +G LAGNAI WD+Q A Sbjct: 1 MSLQAIFKAAAGIGDPVVGAFPTNARYFTTVLSPFILOLDEGPAAGNAITWDSQVSA 58 Query: 61 YCAGGSVEASFNLAKAVNGLQGGVAGTVLSTENGTGAPVEGHEAQVTIQSNQGGI 119 Y GGSVEASFNLA+AVN QC+RVAGTVLSTENGT GAPVEGHE QVTIQ+MQGGI Sbjct: 59 YACGASVEASFNLAKAVNGLQGGVAGTVLSTENGTGAPVEGHEQVTIQSNQGGI 117 >Schubert_12, tail terminator, 117 Length = 117 Score = 159 bits (402), Expect = 2e-39 Identities = 79/119 (66%), Positives = 94/119 (78%), Gaps = 2/119 (1%) Query: 1 MSLQAIFKALVIGIGDPVVGAFPTNARYFTTVLSPFILOLDEGPAAGNAISWDSQVAA 60 MSLQ +K L +G PV+VG+AF++AR+FTTVLSP +D D A+ G+AI WD+Q A Sbjct: 1 MSLQTLKGLASLOVFTVGTAFENARLPFTTVLSPFILOLDEGPAAGNAITWDSQVSA 58 Query: 61 YCAGGSVEASFNLAKAVNGLQGGVAGTVLSTENGTGAPVEGHEAQVTIQSNQGGI 119 Y AG SVEASFNLA+AVNG LQG R+ GYVL TSNGT GA VEGHE+QVTIQ+MQGGI Sbjct: 59 YAAGASVEASFNLAKAVNGLQGGVAGTVLSTENGTGAPVEGHEQVTIQSNQGGI 117</pre>
DECISION:	<p>The start conserved is at position 7456. It is an ATG associated with a good-scoring RBS and is preserved in the homologs.</p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	Phage DB :

	<div><div><div>Descriptions</div><div>Graphic Summary</div><div>Alignments</div><div>Taxonomy</div></div><div><div>Click to see the title</div><div>Click to show alignments</div><div>Show Conserved Domains</div></div><div>Alignment Scores</div><div><div>< 40</div><div>40 - 50</div><div>50 - 80</div><div>80 - 200</div><div>> 200</div></div><div>59 sequences selected</div><div>No putative conserved domains have been detected</div><div><div>Distribution of the top 59 Blast Hits on 59 subject sequences</div></div></div> <div><p>Avec les deux DB on trouve des homologues Tail terminator avec de e-values $<10^{-4}$ et un query coverage $>80\%$ pour la majorité. On trouve également quelques protéines de fonction inconnue.</p></div>
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHpred avec une proba $\geq 90\%$ et une couverture acceptable ?</p>	<p>Le meilleur hit avec HHpred est une tail terminator avec une probabilité de 97.6%. mais e value élevée</p>








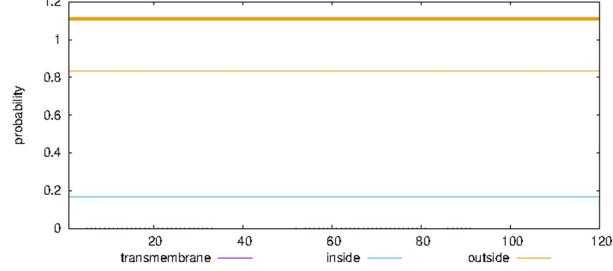
Show 25 Entries Search:

Nr	HIT	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	6TES_F	Tail terminator protein Rcc01690; "neck", "portal", "capsid", "tail tube", VIRUS; 3.58A (Rhodobacter capsulatus)	97.6	0.0014	42.63	7.4	85	135
2	PF11367.12	: DUF3168 ; Protein of unknown function (DUF3168)	96.83	0.042	36.22	8.2	100	116
3	5A21_G	TAIL-TO-HEAD JOINING PROTEIN GP17; VIRAL PROTEIN, VIRAL INFECTION, TAILED BACTERIOPHAGE, SIPHOVIRIDAE, SPP1, VIRAL ASSEM	96.46	0.065	35.49	7.1	76	134
4	4ACV_B	PROPHAGE LAMBDA M01, ANTIGEN B; IMMUNE SYSTEM, SURFACE, PHAGE-RELATED; 2.401A (LISTERIA MONOCYTOGENES)	96.21	0.34	33.45	9.7	102	129

[illegible]

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

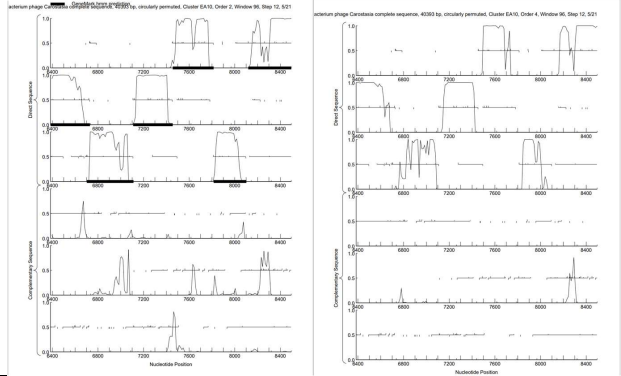
Aucune synthèse trouvée.

	<div><div>Carostasia_Draft (EA10)</div><div></div><div><div>Nucqi (EA10)</div><div></div></div><div><div>Quatz (EA10)</div><div></div></div></div>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div><p>Le serveur DeepTMHMM était saturé donc j'ai utilisé le regular TMHMM...</p><p>La protéine est prédite comme non transmembranaire.</p><div><div>TMHMM posterior probabilities for WEBSEQUENCE</div></div></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<p>Oui.</p>
DECISION:	<p>Ce gène code pour une tail Terminator.</p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	12
Coordonnées du Stop	8102
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove1
Coordonnées du Start retenu	7815
Fonction prédite	NFK

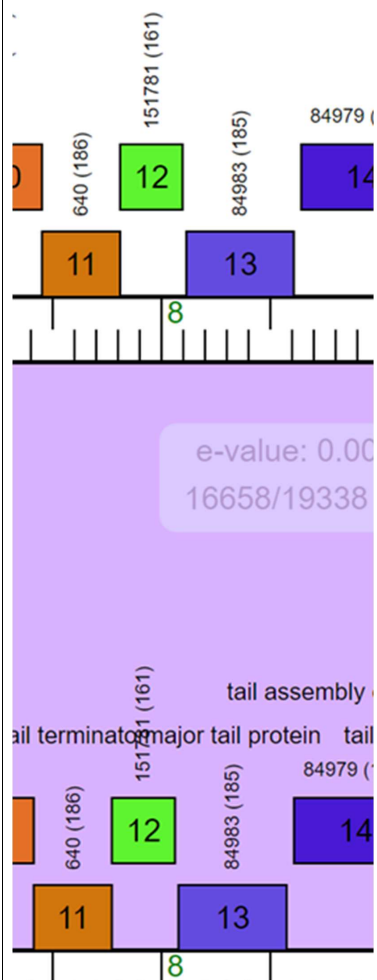
Décision #1 : Est-ce un gène ?

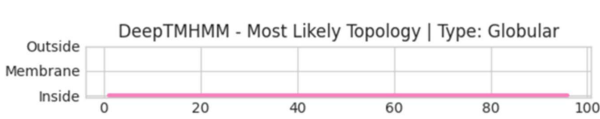
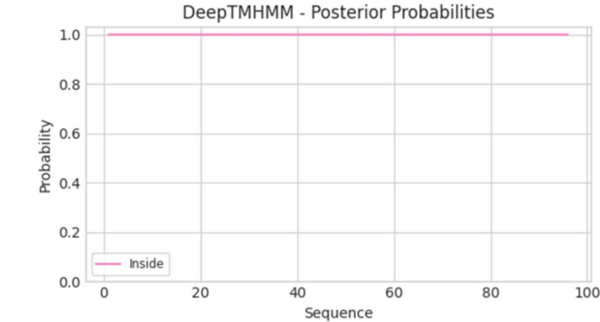
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	<p>We have a coding potential proposed by the 2</p> 

	<p>Other related phages do have this gene:</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <div><div>3059 Acosta_Draft_14, function unknown, 76. S=89.0 E=4e-18</div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>200</div></div><div><div>1c111</div><div>050</div></div></div>																																																
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div>Sequences producing significant alignments:</div><div><table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_12, function unknown, 95</td><td>191</td><td>7e-49</td></tr><tr><td>Nucci_12, function unknown, 95</td><td>191</td><td>7e-49</td></tr><tr><td>Mandalorian_12, function unknown, 95</td><td>191</td><td>7e-49</td></tr><tr><td>Carostasia_Draft_12, function unknown, 95</td><td>191</td><td>7e-49</td></tr><tr><td>YuuY_12, function unknown, 96</td><td>147</td><td>1e-35</td></tr><tr><td>Juanyo_12, function unknown, 96</td><td>118</td><td>6e-27</td></tr><tr><td>Luna18_14, function unknown, 104</td><td>105</td><td>4e-23</td></tr><tr><td>KatChan_14, function unknown, 104</td><td>105</td><td>4e-23</td></tr><tr><td>Alakazam_15, function unknown, 101</td><td>104</td><td>9e-23</td></tr><tr><td>Chepli_14, function unknown, 104</td><td>103</td><td>1e-22</td></tr><tr><td>WestPM_Draft_13, function unknown, 100</td><td>99</td><td>3e-21</td></tr><tr><td>Tinyman4_14, function unknown, 100</td><td>99</td><td>3e-21</td></tr><tr><td>HerculesXL_14, function unknown, 100</td><td>99</td><td>3e-21</td></tr><tr><td>Schimmels22_14, function unknown, 100</td><td>99</td><td>4e-21</td></tr><tr><td>WilliamStrong_12, function unknown, 101</td><td>98</td><td>8e-21</td></tr></table></div></div>		Score (bits)	E Value	Quartz_12, function unknown, 95	191	7e-49	Nucci_12, function unknown, 95	191	7e-49	Mandalorian_12, function unknown, 95	191	7e-49	Carostasia_Draft_12, function unknown, 95	191	7e-49	YuuY_12, function unknown, 96	147	1e-35	Juanyo_12, function unknown, 96	118	6e-27	Luna18_14, function unknown, 104	105	4e-23	KatChan_14, function unknown, 104	105	4e-23	Alakazam_15, function unknown, 101	104	9e-23	Chepli_14, function unknown, 104	103	1e-22	WestPM_Draft_13, function unknown, 100	99	3e-21	Tinyman4_14, function unknown, 100	99	3e-21	HerculesXL_14, function unknown, 100	99	3e-21	Schimmels22_14, function unknown, 100	99	4e-21	WilliamStrong_12, function unknown, 101	98	8e-21
	Score (bits)	E Value																																															
Quartz_12, function unknown, 95	191	7e-49																																															
Nucci_12, function unknown, 95	191	7e-49																																															
Mandalorian_12, function unknown, 95	191	7e-49																																															
Carostasia_Draft_12, function unknown, 95	191	7e-49																																															
YuuY_12, function unknown, 96	147	1e-35																																															
Juanyo_12, function unknown, 96	118	6e-27																																															
Luna18_14, function unknown, 104	105	4e-23																																															
KatChan_14, function unknown, 104	105	4e-23																																															
Alakazam_15, function unknown, 101	104	9e-23																																															
Chepli_14, function unknown, 104	103	1e-22																																															
WestPM_Draft_13, function unknown, 100	99	3e-21																																															
Tinyman4_14, function unknown, 100	99	3e-21																																															
HerculesXL_14, function unknown, 100	99	3e-21																																															
Schimmels22_14, function unknown, 100	99	4e-21																																															
WilliamStrong_12, function unknown, 101	98	8e-21																																															
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	no																																																
DECISION:	yes																																																

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	Start : 7815																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Stats: 4 Selected: 1</div><div>ORF Start : 7950 ORF Stop : 8102 ORF Length: 153</div><div>Cdn1 Cdn2 Cdn3 5' End 57.9 36.8 94.7 3' End 66.2 50.6 81.8</div><div>Length 57 231</div><div>SD Scoring Matrix Kibler6</div><div>Spacing/Weight Matrix Karlin Medium</div></div> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,748</td><td>2,981</td><td>11</td><td>-2,505</td><td>AGTCCACCCAAGGAGGCATCTG</td><td>ATG</td><td>7815</td><td>288</td></tr><tr><td>2</td><td>-1,748</td><td>2,981</td><td>10</td><td>-2,443</td><td>CGTCGACCCACCAAGGAGACCGGC</td><td>GTG</td><td>7872</td><td>231</td></tr><tr><td>3</td><td>-7,105</td><td>0,474</td><td>12</td><td>-7,940</td><td>CGTCGGCTACGCCATCTCCGAT</td><td>GTG</td><td>7893</td><td>210</td></tr><tr><td>4</td><td>-4,910</td><td>1,502</td><td>9</td><td>-5,684</td><td>GGACCTCAAGCCTGGCGAGTCG</td><td>GTG</td><td>7950</td><td>153</td></tr></table> <p>The predicted start has the 2nd best RBS score: -2.505 (ATG) but very close to the best: -2.443 (GTG)</p>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,748	2,981	11	-2,505	AGTCCACCCAAGGAGGCATCTG	ATG	7815	288	2	-1,748	2,981	10	-2,443	CGTCGACCCACCAAGGAGACCGGC	GTG	7872	231	3	-7,105	0,474	12	-7,940	CGTCGGCTACGCCATCTCCGAT	GTG	7893	210	4	-4,910	1,502	9	-5,684	GGACCTCAAGCCTGGCGAGTCG	GTG	7950	153
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																															
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																															
1	-1,748	2,981	11	-2,505	AGTCCACCCAAGGAGGCATCTG	ATG	7815	288																																															
2	-1,748	2,981	10	-2,443	CGTCGACCCACCAAGGAGACCGGC	GTG	7872	231																																															
3	-7,105	0,474	12	-7,940	CGTCGGCTACGCCATCTCCGAT	GTG	7893	210																																															
4	-4,910	1,502	9	-5,684	GGACCTCAAGCCTGGCGAGTCG	GTG	7950	153																																															
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	<i>The predicted start gives the longest ORF (288) and just -1 overlap with the preceding gene.</i>																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?	Gene: <u>Carostasia_12</u> Start: 7815, Stop: 8102, Start Num: 11 Candidate Starts for <u>Carostasia_12</u> : (Start: 11 @7815 has 92 MA's), (14, 7872), (16, 7893), (20, 7950), <i>Carostasia not present in tarcks</i>																																																						

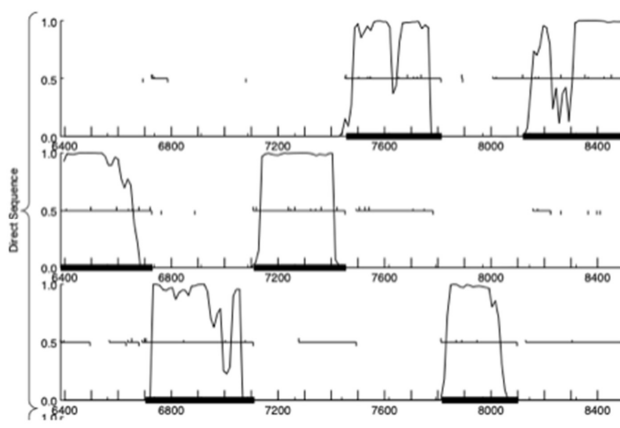


Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div>Non</div> <div></div> <div></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Non
DECISION:	NKF

Student Gene Annotation Worksheet


Basic Phage Information	
Nom du Phage	Carostasia
Gène #	13
Coordonnées du Stop	8613
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 16
Coordonnées du Start retenu	8119
Fonction prédite	Major tail protein

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GeneMarks detects a coding potential along the entire length of the ORF, but it is weak at 5'..</i></p> 

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Yes, the candidate is found in other annotated genomes, with a very good alignment score.</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	No
DECISION:	Yes

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	8119 by Glimmer and GeneMark
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<i>This start is an ATG associated with a good-scoring RBS. Final score = -3.025. This is the best score proposed.</i>
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? <i>Si le plus long pas retenu, quel est l'écart intergène résultant ?</i>	<i>No, this start does not lead to the longest ORF. A GTG start is proposed at position 8008, but this would lead to too much overlap with the upstream ORF. The start at 8119 confers an ORF of 495bp.</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<i>Start is found in all homologs after starterator analysis.</i> 
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<i>Yes, the start is preserved in other homologs with Blastp.</i>

	<p>>Quartz_13, major tail protein, 164 Length = 164</p> <p>Score = 330 bits (846), Expect = 9e-91 Identities = 163/164 (99%), Positives = 164/164 (100%)</p> <p>Query: 1 MFLTQWNPATQISRGNVAVGVAFVVELDAPALSELTTCIGLDCSITTNKOTSSTDSSEI 60 Sbjct: 1 MFLTQWNPATQISRGNVAVGVAFVVELDAPALSELTTCIGLDCSITTNKOTSSTDSSEI 60</p> <p>Query: 61 DNLCDPASEQLPGSTTHAMGOLVINGTQGGADLIAGLNGDVVTVWRDGMHDTAPAA 120 Sbjct: 61 DNLCDPASEQLPGSTTHAMGOLVINGTQGGADLIAGLNGDVVTVWRDGMHDTAPAA 120</p> <p>Query: 121 GQFVWVKVIITSIDPLEANNITVGITARITVLAREKTAIAA 164 Sbjct: 121 GQFVWVKVIITSIDPLEANNITVGITARITVLAREKTAIAA 164</p> <p>>Mandalorian_13, major tail protein, 164 Length = 164</p> <p>Score = 330 bits (846), Expect = 9e-91 Identities = 163/163 (100%), Positives = 163/163 (100%)</p> <p>Query: 1 MFLTQWNPATQISRGNVAVGVAFVVELDAPALSELTTCIGLDCSITTNKOTSSTDSSEI 60 Sbjct: 1 MFLTQWNPATQISRGNVAVGVAFVVELDAPALSELTTCIGLDCSITTNKOTSSTDSSEI 60</p> <p>Query: 61 DNLCDPASEQLPGSTTHAMGOLVINGTQGGADLIAGLNGDVVTVWRDGMHDTAPAA 120 Sbjct: 61 DNLCDPASEQLPGSTTHAMGOLVINGTQGGADLIAGLNGDVVTVWRDGMHDTAPAA 120</p> <p>Query: 121 GQFVWVKVIITSIDPLEANNITVGITARITVLAREKTAIAA 163 Sbjct: 121 GQFVWVKVIITSIDPLEANNITVGITARITVLAREKTAIAA 163</p> <p>>Juanyo_13, major tail protein, 165 Length = 165</p> <p>Score = 305 bits (781), Expect = 3e-83 Identities = 147/164 (89%), Positives = 156/164 (95%)</p> <p>Query: 1 MFLTQWNPATQISRGNVAVGVAFVVELDAPALSELTTCIGLDCSITTNKOTSSTDSSEI 60 Sbjct: 1 MFLTQWNPATQISRGNVAVGVAFVVELDAPALSELTTCIGLDCSITTNKOTSSTDSSEI 60</p> <p>Query: 61 DNLCDPASEQLPGSTTHAMGOLVINGTQGGADLIAGLNGDVVTVWRDGMHDTAPAA 120 Sbjct: 61 DNLCDPASEQLPGSTTHAMGOLVINGTQGGADLIAGLNGDVVTVWRDGMHDTAPAA 120</p> <p>Query: 121 GQFVWVKVIITSIDPLEANNITVGITARITVLAREKTAIAA 164 Sbjct: 121 GQFVWVKVIITSIDPLEANNITVGITARITVLAREKTAIAA 164</p>
DECISION:	<p><i>The start selected is that at position 8119. It confers an ORF of satisfactory size and is associated with a good RBS. It is also conserved in other homologs.</i></p>

Décision #3 : Quelle est la fonction de la protéine putative ?

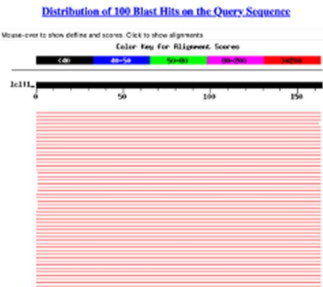
Collection des éléments de réponse	Rationnelle
------------------------------------	-------------

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Phage DB :

Sequences producing significant alignments:		Score	E
		(bits)	Value
Nucci_13, major tail protein, 164		332	3e-91
Carostasia_Draft_13, function unknown, 164		332	3e-91
Quartz_13, major tail protein, 164		330	9e-91
Mandalorian_13, major tail protein, 164		330	9e-91
Juanyo_13, major tail protein, 165		305	3e-83
Yuuy_13, major tail protein, 164		305	4e-83
Finny_13, major tail protein, 164		303	9e-83
Shamu_13, major tail protein, 164		302	2e-82
Saratos_13, major tail protein, 164		302	2e-82
Glamour_12, major tail protein, 164		302	2e-82
Eleri_13, major tail protein, 164		302	2e-82
ColaCorta_13, major tail protein, 164		302	2e-82
ChikPic_13, major tail protein, 164		302	2e-82
Andromedas_13, major tail protein, 164		302	2e-82
Zenitsu_13, major tail protein, 165		300	8e-82
Sansa_12, major tail protein, 165		300	8e-82
MCubed_13, major tail protein, 165		300	8e-82
Tinyman4_15, major tail protein, 166		300	1e-81
HerculesXL_15, major tail protein, 166		298	3e-81
WestPM_Draft_14, function unknown, 166		298	4e-81
Ixel_14, major tail protein, 165		298	5e-81
Schimmels22_15, major tail protein, 166		297	7e-81
BouleyBill_Draft_14, function unknown, 166		297	7e-81
Clock_Draft_15, function unknown, 164		296	1e-80
Neferthens_15, major tail protein, 164		294	6e-80
PrincePhergus_14, major tail protein, 166		293	9e-80
Pherbot_14, major tail protein, 166		293	2e-79
Bustleton_14, major tail protein, 166		293	2e-79
Luna18_15, major tail protein, 164		292	2e-79
KatChan_15, major tail protein, 164		292	2e-79

Dorey= Carostasia_13
(144 letters)



nr :

select all 65 sequences selected		GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer			
	Description	Scientific Name	Max Score	Total Score	Coverage	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Nucci)	Microbacterium phage Nucci	328	328	100%	4e-113	100.00%	164	QXK22456.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Mandalorian)	Microbacterium phage Mandalorian	327	327	99%	1e-112	100.00%	164	QXK13607.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Quartz)	Microbacterium phage Quartz	326	326	100%	3e-112	99.99%	164	UWS92362.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Tinyman4)	Microbacterium phage Tinyman4	298	298	99%	5e-101	87.12%	166	WNN95216.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage HerculesXL)	Microbacterium phage HerculesXL	296	296	99%	2e-100	87.12%	166	UWQ33800.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Ixel)	Microbacterium phage Ixel	294	294	99%	1e-99	87.73%	165	QXK87715.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Neferthens)	Microbacterium phage Neferthens	291	291	100%	2e-98	84.76%	164	YP_00908191.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage PrincePhergus)	Microbacterium phage PrincePhergus	289	289	99%	1e-97	85.28%	166	QSZ72883.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Pherbot)	Microbacterium phage Pherbot	288	288	99%	3e-97	85.28%	166	QSZ73065.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Nebulous)	Microbacterium phage Nebulous	288	288	100%	4e-97	84.19%	164	QXK87778.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Juicer)	Microbacterium phage Juicer	287	287	100%	1e-96	82.93%	164	QZQ99663.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Beethoven)	Microbacterium phage Beethoven	286	286	100%	2e-96	82.93%	165	AY555153.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Alakazam)	Microbacterium phage Alakazam	285	285	100%	4e-96	82.32%	164	QXK93838.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage AnPud)	Microbacterium phage AnPud	284	284	100%	1e-95	81.71%	165	ALX82663.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Hermeti)	Microbacterium phage Hermeti	284	284	100%	1e-95	82.32%	165	YP_009623115.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Gerdent)	Microbacterium phage Gerdent	284	284	100%	1e-95	81.71%	165	WGS89572.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage YuuY)	Microbacterium phage YuuY	284	284	100%	2e-95	80.24%	164	QFP95362.1
<input checked="" type="checkbox"/>	major tail protein (Microbacterium phage Juanyo)	Microbacterium phage Juanyo	284	284	100%	2e-95	89.63%	165	U5H44764.1

Cette protéine présente une forte homologie avec des major tail protein d'après des Blastp contre la nr et la phagesDB. Les e-values sont $<10^{-4}$ et le query coverage est $>98\%$ pour une majorité des hits contre la nr.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Le meilleur hit est une major tail protein avec une probabilité de 78.8%, ce qui est assez faible mais en accord avec ce qui est observé précédemment avec Blastp.

Visualization

Hitlist

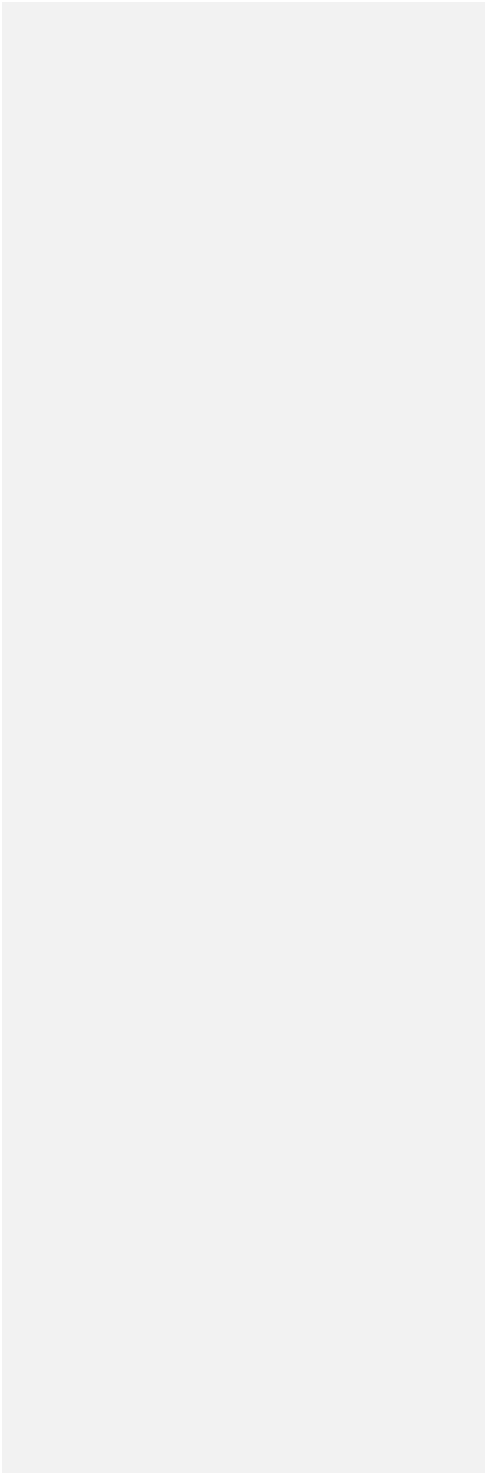
Show 25 Entries

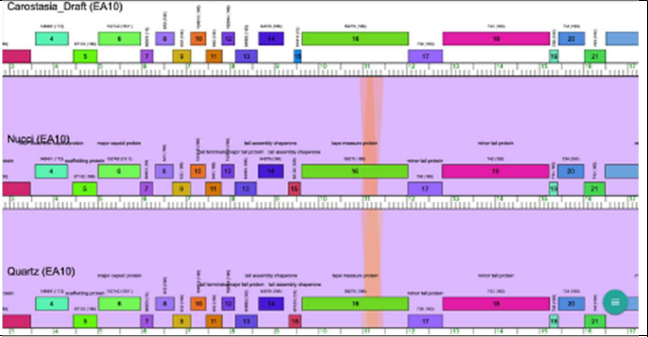
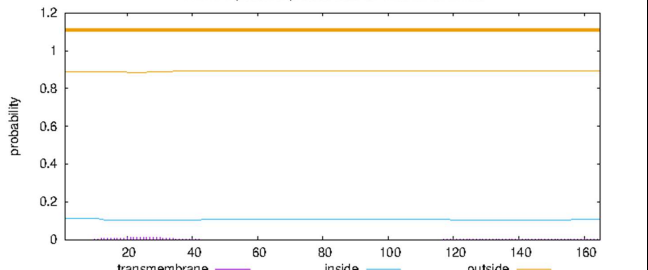
Search:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	8GTC_F	Major tail protein; Marine bacteriophage, Cryo-EM, Siphophage, Baseplate, Megatron protein, Tail fibre protein, Distal t	78.88	43	24.56	8.6	113	130
2	P85503	STRU3_BPPA Structural protein 3 OS=Pseudomonas phage PAJ02 QX=504346 PE=1 SV=2	71.58	89	24.67	11.8	120	164
3	PF08813.15	; Phage_tail_3; Phage tail tube protein, TTP	67.13	110	23.88	10.7	108	161

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une

Les gènes en amont et en aval chez les phages du même groupe codent respectivement pour une tail terminator et une tail assembly chaperone. Il y a conservation de l'ordre des gènes.



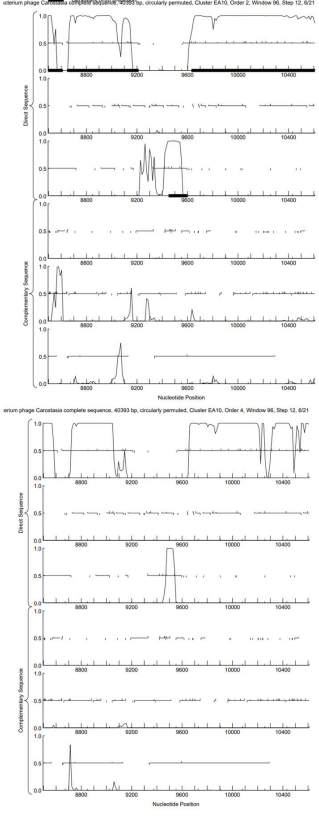

forte conservation de l'ordre des gènes ?	
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p>TMHMM posterior probabilities for WEBSEQUENCE</p>  <p><i>Ce n'est pas une protéine transmembranaire.</i></p>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<p><i>oui.</i></p>
DECISION:	<p><i>Ce gène code pour une major tail protein.</i></p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	14
Coordonnées du Stop	9213
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 33
Coordonnées du Start retenu	8647
Fonction prédite	Tail assembly chaperone

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both</i>

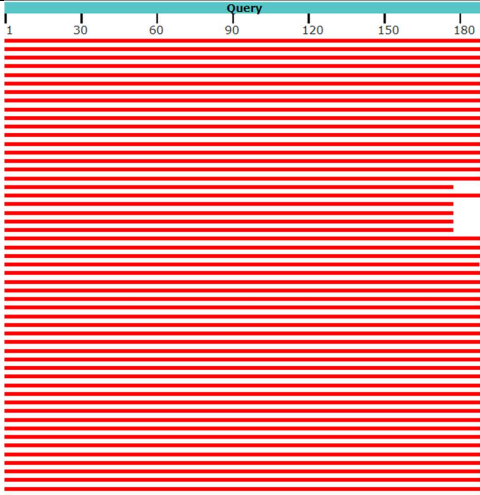
<p>Y-a-t-il des éléments supportant un potentiel codant ?</p>													
<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <table border="1"><thead><tr><th>Score Range</th><th>Color</th></tr></thead><tbody><tr><td><40</td><td>Black</td></tr><tr><td>40-50</td><td>Blue</td></tr><tr><td>50-60</td><td>Green</td></tr><tr><td>60-200</td><td>Red</td></tr><tr><td>>200</td><td>Yellow</td></tr></tbody></table> <p>1c111</p> 	Score Range	Color	<40	Black	40-50	Blue	50-60	Green	60-200	Red	>200	Yellow
Score Range	Color												
<40	Black												
40-50	Blue												
50-60	Green												
60-200	Red												
>200	Yellow												

	<p>Sequences producing significant alignments:</p> <p>Nucci_14, tail assembly chaperone, 188 372 e-103</p> <p>Carostasia_Draft_14, function unknown, 188 372 e-103</p> <p>Mandalorian_14, tail assembly chaperone, 188 371 e-103</p> <p>Quartz_14, tail assembly chaperone, 188 371 e-103</p> <p>YuuY_14, tail assembly chaperone, 188 366 e-101</p> <p>Juanyo_14, tail assembly chaperone, 189 344 8e-95</p> <p>Zenitsu_14, tail assembly chaperone, 186 281 5e-76</p> <p>Sansa_13, tail assembly chaperone, 186 281 5e-76</p> <p>MCubed_14, tail assembly chaperone, 186 281 5e-76</p>	Score (bits)	E Value
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>no</i>		
DECISION:	<i>yes</i>		

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																			
Quel start Glimmer et GeneMark suggèrent-ils ?	8647																																																																																																			
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Choose ORF start</div><div><div>Stats: 9</div><div>ORF Start : 8647</div><div>Cdn1 Cdn2 Cdn3 Length</div><div>Selected: 1</div><div>ORF Stop : 9213</div><div>5' End 68.6 33.3 76.8 207</div><div>ORF Length: 567</div><div>3' End 60.8 45.8 85.8 360</div><div>SD Scoring Matrix Kibler6</div><div>Spacing Weight Matrix Karlin Medium</div></div></div> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,748</td><td>2,981</td><td>9</td><td>-2,523</td><td>ACCAAGCCCCCAAGGAGGCATC</td><td>ATG</td><td>8647</td><td>567</td></tr><tr><td>2</td><td>-4,495</td><td>1,696</td><td>7</td><td>-6,018</td><td>CCTGGAGCAGCTCAAGGCTCGT</td><td>GTG</td><td>8854</td><td>360</td></tr><tr><td>3</td><td>-4,263</td><td>1,805</td><td>12</td><td>-5,098</td><td>GCTGGCTACGGGCGCTGTGCAG</td><td>TTG</td><td>8911</td><td>303</td></tr><tr><td>4</td><td>-5,618</td><td>1,170</td><td>10</td><td>-6,313</td><td>TCTGCAGTTGGGCGAGTGGTCC</td><td>ATG</td><td>8926</td><td>288</td></tr><tr><td>5</td><td>-3,620</td><td>2,105</td><td>13</td><td>-4,666</td><td>GACGTTCAGGGCGCTCTACGGT</td><td>GTG</td><td>9007</td><td>207</td></tr><tr><td>6</td><td>-3,654</td><td>2,089</td><td>6</td><td>-5,359</td><td>GGAAACGATCCTGCTGGGGCC</td><td>ATG</td><td>9124</td><td>90</td></tr><tr><td>7</td><td>-3,654</td><td>2,089</td><td>18</td><td>-5,955</td><td>GCCTGGGGCCATGCTCCACGTC</td><td>GTG</td><td>9136</td><td>78</td></tr><tr><td>8</td><td>-6,213</td><td>0,892</td><td>7</td><td>-7,736</td><td>TGGGGCCATGCTCCACGTGTG</td><td>GTG</td><td>9139</td><td>75</td></tr><tr><td>9</td><td>-3,861</td><td>1,993</td><td>13</td><td>-4,906</td><td>CCACGTGCTGGTGAAGCGCTTC</td><td>ATG</td><td>9151</td><td>63</td></tr></table> <p>Best RBS score: -2.523 corresponds to the selected start</p>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,748	2,981	9	-2,523	ACCAAGCCCCCAAGGAGGCATC	ATG	8647	567	2	-4,495	1,696	7	-6,018	CCTGGAGCAGCTCAAGGCTCGT	GTG	8854	360	3	-4,263	1,805	12	-5,098	GCTGGCTACGGGCGCTGTGCAG	TTG	8911	303	4	-5,618	1,170	10	-6,313	TCTGCAGTTGGGCGAGTGGTCC	ATG	8926	288	5	-3,620	2,105	13	-4,666	GACGTTCAGGGCGCTCTACGGT	GTG	9007	207	6	-3,654	2,089	6	-5,359	GGAAACGATCCTGCTGGGGCC	ATG	9124	90	7	-3,654	2,089	18	-5,955	GCCTGGGGCCATGCTCCACGTC	GTG	9136	78	8	-6,213	0,892	7	-7,736	TGGGGCCATGCTCCACGTGTG	GTG	9139	75	9	-3,861	1,993	13	-4,906	CCACGTGCTGGTGAAGCGCTTC	ATG	9151	63
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																												
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																												
1	-1,748	2,981	9	-2,523	ACCAAGCCCCCAAGGAGGCATC	ATG	8647	567																																																																																												
2	-4,495	1,696	7	-6,018	CCTGGAGCAGCTCAAGGCTCGT	GTG	8854	360																																																																																												
3	-4,263	1,805	12	-5,098	GCTGGCTACGGGCGCTGTGCAG	TTG	8911	303																																																																																												
4	-5,618	1,170	10	-6,313	TCTGCAGTTGGGCGAGTGGTCC	ATG	8926	288																																																																																												
5	-3,620	2,105	13	-4,666	GACGTTCAGGGCGCTCTACGGT	GTG	9007	207																																																																																												
6	-3,654	2,089	6	-5,359	GGAAACGATCCTGCTGGGGCC	ATG	9124	90																																																																																												
7	-3,654	2,089	18	-5,955	GCCTGGGGCCATGCTCCACGTC	GTG	9136	78																																																																																												
8	-6,213	0,892	7	-7,736	TGGGGCCATGCTCCACGTGTG	GTG	9139	75																																																																																												
9	-3,861	1,993	13	-4,906	CCACGTGCTGGTGAAGCGCTTC	ATG	9151	63																																																																																												
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	Yes, Length of 567 with start at 8647 Gap of +34																																																																																																			
Est-ce que le start est conservé chez les homologues voir Starterator ?	Gene: Carostasia_14 Start: 8647, Stop: 9213, Start Num: 9 Candidate Starts for Carostasia_14: (Start: 9 @8647 has 140 MA's), (23, 8854), (27, 8911), (29, 8926), (40, 9007), (58, 9124), (60, 9136), (61, 9139), (62, 9151),																																																																																																			

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?



[Download](#) [GenPept](#) [Graphics](#)

tail assembly chaperone [Microbacterium phage Nucci]

Sequence ID: [Q0K02407.1](#) Length: 188 Number of Matches: 1

Range 1: 1 to 188 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
387 bits(993)	2e-135	Compositional matrix adjust.	188/188(100%)	188/188(100%)	0/188(0%)
Query 1	HSFSSVEELKAANDERRKDI	LTTEVDLGARYSQEHEDAKKELAQAEATQKL	AGGGGQGLN	60	
Sbjct 1	HSFSSVEELKAANDERRKDI	LTTEVDLGARYSQEHEDAKKELAQAEATQKL	AGGGGQGLN	60	
Query 61	DNLQKARVAETKPEANSI	MLRYGRQLAENSHLTKATGLTPIDQYEKVL	PQTFKGVVG	120	
Sbjct 61	DNLQKARVAETKPEANSI	MLRYGRQLAENSHLTKATGLTPIDQYEKVL	PQTFKGVVG	120	
Query 121	VDPTAEDEEGNLLHPDAE	PLTTDARAVSSRSEETILPGARLHVWNAFTHQNSG	GETSI	180	
Sbjct 121	VDPTAEDEEGNLLHPDAE	PLTTDARAVSSRSEETILPGARLHVWNAFTHQNSG	GETSI	180	
Query 181	RPTKSGRV			188	
Sbjct 181	RPTKSGRV			188	

[Download](#) [GenPept](#) [Graphics](#)

tail assembly chaperone [Microbacterium phage Mandalorian]

Sequence ID: [Q013608.1](#) Length: 188 Number of Matches: 1

Range 1: 1 to 188 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
385 bits(990)	7e-135	Compositional matrix adjust.	187/188(99%)	188/188(100%)	0/188(0%)
Query 1	HSFSSVEELKAANDERRKDI	LTTEVDLGARYSQEHEDAKKELAQAEATQKL	AGGGGQGLN	60	
Sbjct 1	HSFSSVEELKAANDERRKDI	LTTEVDLGARYSQEHEDAKKELAQAEATQKL	AGGGGQGLN	60	
Query 61	DNLQKARVAETKPEANSI	MLRYGRQLAENSHLTKATGLTPIDQYEKVL	PQTFKGVVG	120	
Sbjct 61	DNLQKARVAETKPEANSI	MLRYGRQLAENSHLTKATGLTPIDQYEKVL	PQTFKGVVG	120	
Query 121	VDPTAEDEEGNLLHPDAE	PLTTDARAVSSRSEETILPGARLHVWNAFTHQNSG	GETSI	180	
Sbjct 121	VDPTAEDEEGNLLHPDAE	PLTTDARAVSSRSEETILPGARLHVWNAFTHQNSG	GETSI	180	
Query 181	RPTKSGRV			188	
Sbjct 181	RPTKSGRV			188	

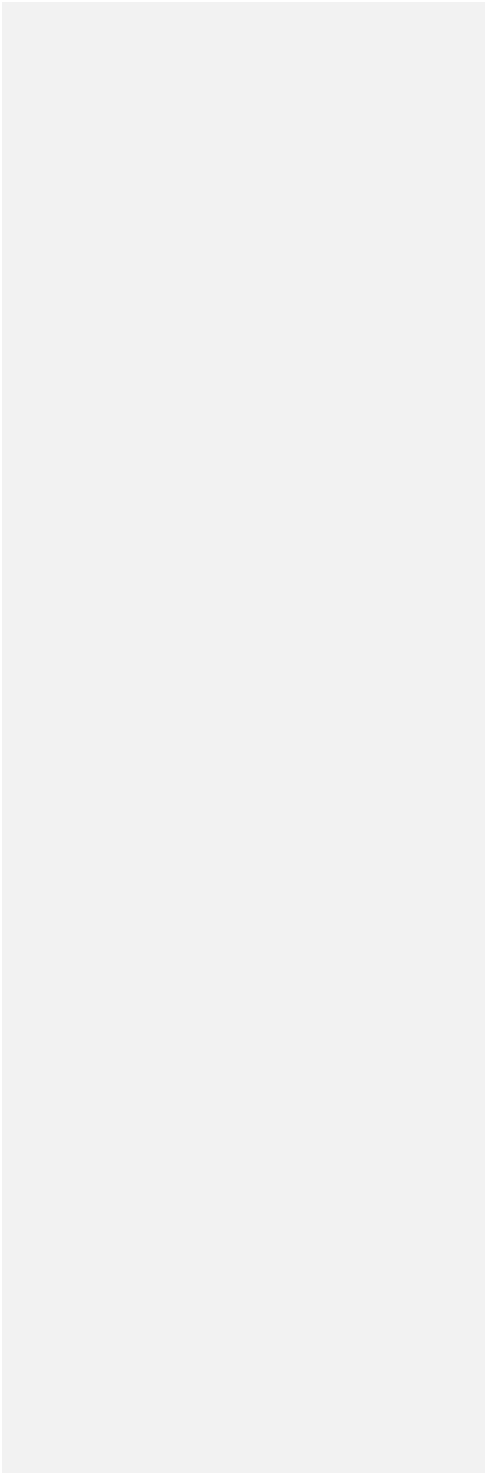
DECISION:

Start of 8647 retained despite 33 nucleotides gap

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																													
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** <i>PhagesDB</i> :</p> <p>>Nucci_14, tail assembly chaperone, 188 Length = 188</p> <p>Score = 372 bits (956), Expect = e-103 Identities = 188/188 (100%), Positives = 188/188 (100%)</p> <p>Query: 1 MSFSSYEELKAAVDERRKDILTIEVDLGARYSQEHEDAKKELAQAEAIQKLAGGGQGFLN 60 MSFSSYEELKAAVDERRKDILTIEVDLGARYSQEHEDAKKELAQAEAIQKLAGGGQGFLN Sbjct: 1 MSFSSYEELKAAVDERRKDILTIEVDLGARYSQEHEDAKKELAQAEAIQKLAGGGQGFLN 60</p> <p>Query: 61 DNLEQLKARVAETKPEANSIWLRYGRQLAEWSMLTKATGLTPIDQYEKVLPTQTFKGVYG 120 DNLEQLKARVAETKPEANSIWLRYGRQLAEWSMLTKATGLTPIDQYEKVLPTQTFKGVYG Sbjct: 61 DNLEQLKARVAETKPEANSIWLRYGRQLAEWSMLTKATGLTPIDQYEKVLPTQTFKGVYG 120</p> <p>Query: 121 VDPTAEDEEGNLLHPDAEPLTTDARAVSSRSEETILPGAMLHVVVNAFMTWQNSSGEISI 180 VDPTAEDEEGNLLHPDAEPLTTDARAVSSRSEETILPGAMLHVVVNAFMTWQNSSGEISI Sbjct: 121 VDPTAEDEEGNLLHPDAEPLTTDARAVSSRSEETILPGAMLHVVVNAFMTWQNSSGEISI 180</p> <p>Query: 181 RPTKSGRV 188 RPTKSGRV Sbjct: 181 RPTKSGRV 188</p> <p>*** <i>nr</i> :</p> <table><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/></td><td>tail assembly chaperone [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>387</td><td>387</td><td>99%</td><td>2e-135</td><td>100.00%</td><td>188</td><td>GQ002407.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>tail assembly chaperone [Microbacterium phage Mendocino]</td><td>Microbacterium phage Mendocino</td><td>385</td><td>385</td><td>99%</td><td>7e-135</td><td>99.47%</td><td>188</td><td>GQ013608.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>tail assembly chaperone [Microbacterium phage Quantz]</td><td>Microbacterium phage Quantz</td><td>385</td><td>385</td><td>99%</td><td>1e-134</td><td>98.94%</td><td>188</td><td>LN959233.1</td></tr></table>		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/>	tail assembly chaperone [Microbacterium phage Nucci]	Microbacterium phage Nucci	387	387	99%	2e-135	100.00%	188	GQ002407.1	<input checked="" type="checkbox"/>	tail assembly chaperone [Microbacterium phage Mendocino]	Microbacterium phage Mendocino	385	385	99%	7e-135	99.47%	188	GQ013608.1	<input checked="" type="checkbox"/>	tail assembly chaperone [Microbacterium phage Quantz]	Microbacterium phage Quantz	385	385	99%	1e-134	98.94%	188	LN959233.1					
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																					
<input checked="" type="checkbox"/>	tail assembly chaperone [Microbacterium phage Nucci]	Microbacterium phage Nucci	387	387	99%	2e-135	100.00%	188	GQ002407.1																																					
<input checked="" type="checkbox"/>	tail assembly chaperone [Microbacterium phage Mendocino]	Microbacterium phage Mendocino	385	385	99%	7e-135	99.47%	188	GQ013608.1																																					
<input checked="" type="checkbox"/>	tail assembly chaperone [Microbacterium phage Quantz]	Microbacterium phage Quantz	385	385	99%	1e-134	98.94%	188	LN959233.1																																					
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/></td><td>1</td><td>cd14725 ZIP_Gal4-like_2; Leucine zipper Dimerization coil of Gal4-like transcription factors. The Gal4p family of transcription</td><td>59.52</td><td>21</td><td>21.98</td><td>1.9</td><td>16</td><td>24</td></tr><tr><td><input type="checkbox"/></td><td>2</td><td>O21877 HCP2_BPLSK Probable head completion protein 2 OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1</td><td>54.67</td><td>19</td><td>28.09</td><td>1.4</td><td>15</td><td>56</td></tr><tr><td><input type="checkbox"/></td><td>3</td><td>PF08748.15 ; Phage_TAC_4 ; Phage tail assembly chaperone</td><td>50.21</td><td>100</td><td>23.2</td><td>4.5</td><td>36</td><td>103</td></tr><tr><td><input type="checkbox"/></td><td>4</td><td>8K5L_B tRNA (cmo5U34)-methyltransferase; tRNA (cmo5U34)-methyltransferase, TRANSFERASE; 1.67A (Fusobacterium nucleatum)</td><td>49.48</td><td>170</td><td>21.73</td><td>5.4</td><td>55</td><td>211</td></tr></table> <p>Pas de résultats acceptables</p>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/>	1	cd14725 ZIP_Gal4-like_2; Leucine zipper Dimerization coil of Gal4-like transcription factors. The Gal4p family of transcription	59.52	21	21.98	1.9	16	24	<input type="checkbox"/>	2	O21877 HCP2_BPLSK Probable head completion protein 2 OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1	54.67	19	28.09	1.4	15	56	<input type="checkbox"/>	3	PF08748.15 ; Phage_TAC_4 ; Phage tail assembly chaperone	50.21	100	23.2	4.5	36	103	<input type="checkbox"/>	4	8K5L_B tRNA (cmo5U34)-methyltransferase; tRNA (cmo5U34)-methyltransferase, TRANSFERASE; 1.67A (Fusobacterium nucleatum)	49.48	170	21.73	5.4	55	211
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																						
<input type="checkbox"/>	1	cd14725 ZIP_Gal4-like_2; Leucine zipper Dimerization coil of Gal4-like transcription factors. The Gal4p family of transcription	59.52	21	21.98	1.9	16	24																																						
<input type="checkbox"/>	2	O21877 HCP2_BPLSK Probable head completion protein 2 OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1	54.67	19	28.09	1.4	15	56																																						
<input type="checkbox"/>	3	PF08748.15 ; Phage_TAC_4 ; Phage tail assembly chaperone	50.21	100	23.2	4.5	36	103																																						
<input type="checkbox"/>	4	8K5L_B tRNA (cmo5U34)-methyltransferase; tRNA (cmo5U34)-methyltransferase, TRANSFERASE; 1.67A (Fusobacterium nucleatum)	49.48	170	21.73	5.4	55	211																																						
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Oui, proche de la tape measure protein et de la tail assembly chaperone																																													
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	Non																																													
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Oui, tail assembly chaperone																																													
	<table><tr><td>tail assembly chaperone</td><td>Tail scaffolding protein</td><td>Evidence needed to call TAC: Please see Bioinformatics Guide for what evidence is needed</td><td></td></tr><tr><td></td><td></td><td>do not call TAC when there is</td><td></td></tr></table>	tail assembly chaperone	Tail scaffolding protein	Evidence needed to call TAC: Please see Bioinformatics Guide for what evidence is needed				do not call TAC when there is																																						
tail assembly chaperone	Tail scaffolding protein	Evidence needed to call TAC: Please see Bioinformatics Guide for what evidence is needed																																												
		do not call TAC when there is																																												

DECISION:	Tail assembly chaperone
------------------	-------------------------



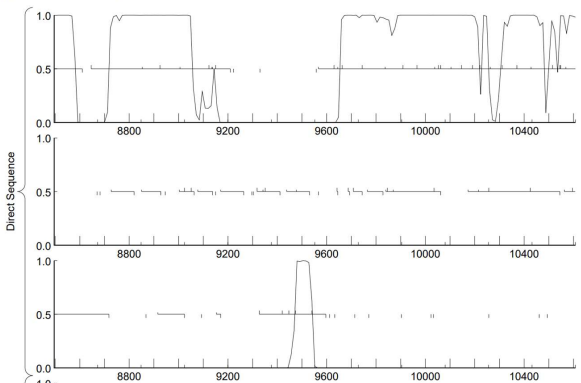
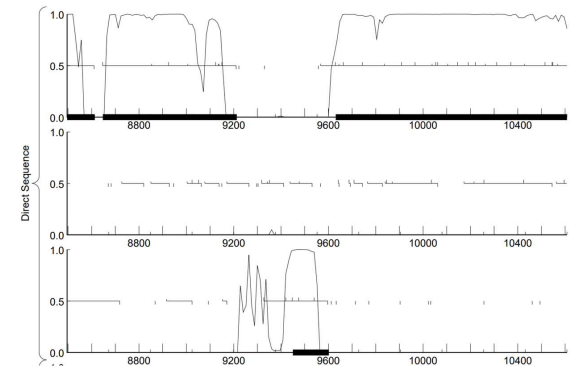
Student Gene Annotation Worksheet

Carostasia-gene-15

Basic Phage Information	
Nom du Phage	carostasia
Gène #	15
Coordonnées du Stop	9602
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 116
Coordonnées du Start retenu	9330
Fonction prédite	Tail assembly chaperone

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	

<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<p>arium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 6/21</p>  <p>acternum phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 6/21</p>  <p><i>Small coding power observed on frame +3, but start position is not the same on both sites</i></p>
<p>Est-ce que le candidat est en contradiction avec les principes d'annotation ?</p>	<p><i>Yes, with the start at 9450 we have a gap of 237bp With the new start we have a gap of 116bp</i></p>
<p>DECISION:</p>	<p><i>yes</i></p>

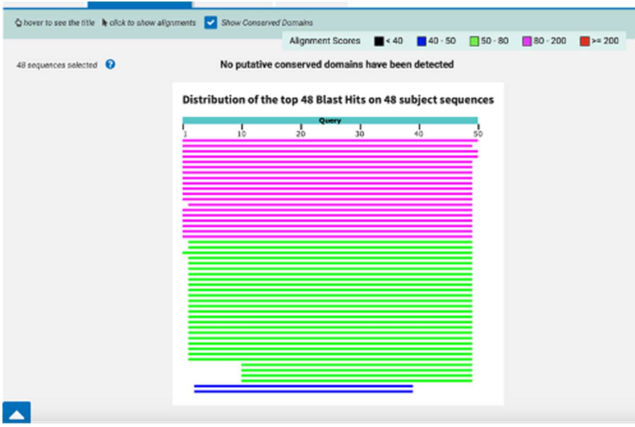
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																	
Quel start Glimmer et GeneMark suggèrent-ils ?	9450																																																																																	
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	Yes , score = -3.837 but 4 other RBS below it																																																																																	
	<div>Stats: 7 Selected: 1</div> <div>ORF Start : 9543 ORF Stop : 9602 ORF Length: 60</div> <div>Cdn 1 Cdn2 Cdn3 Length 5' End 62.5 50.0 81.3 48 3' End 57.3 46.7 70.7 225</div> <div>SD Scoring Matrix Kibler6 Spacing Weight Matrix Karlin Medium</div> <div>Explore Document</div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,262</td><td>1,806</td><td>8</td><td>-5,494</td><td>GGTCTCTCTGCAATGGCTTCC</td><td>ATG</td><td>9330</td><td>273</td></tr><tr><td>2</td><td>-5,918</td><td>1,030</td><td>10</td><td>-6,613</td><td>CGGGCGACCGCTGGCCAGCAC</td><td>TTG</td><td>9378</td><td>225</td></tr><tr><td>3</td><td>-7,016</td><td>0,516</td><td>14</td><td>-8,363</td><td>CCAGCACTTGCACACTCAAGA</td><td>TTG</td><td>9393</td><td>210</td></tr><tr><td>4</td><td>-5,145</td><td>1,352</td><td>10</td><td>-5,839</td><td>GGAAGAGCTCTCTGAAGACTAC</td><td>ATG</td><td>9423</td><td>180</td></tr><tr><td>5</td><td>-2,791</td><td>2,493</td><td>13</td><td>-3,837</td><td>TTGGTCGCTGGAGTGTCTCTCA</td><td>ATG</td><td>9450</td><td>153</td></tr><tr><td>6</td><td>-6,211</td><td>0,893</td><td>9</td><td>-6,985</td><td>GGCCATCGCCATGGTCAGGAC</td><td>ATG</td><td>9477</td><td>126</td></tr><tr><td>7</td><td>-6,082</td><td>0,953</td><td>10</td><td>-6,777</td><td>GGGCAACGGCCCTGATCTGGG</td><td>ATG</td><td>9543</td><td>60</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,262	1,806	8	-5,494	GGTCTCTCTGCAATGGCTTCC	ATG	9330	273	2	-5,918	1,030	10	-6,613	CGGGCGACCGCTGGCCAGCAC	TTG	9378	225	3	-7,016	0,516	14	-8,363	CCAGCACTTGCACACTCAAGA	TTG	9393	210	4	-5,145	1,352	10	-5,839	GGAAGAGCTCTCTGAAGACTAC	ATG	9423	180	5	-2,791	2,493	13	-3,837	TTGGTCGCTGGAGTGTCTCTCA	ATG	9450	153	6	-6,211	0,893	9	-6,985	GGCCATCGCCATGGTCAGGAC	ATG	9477	126	7	-6,082	0,953	10	-6,777	GGGCAACGGCCCTGATCTGGG	ATG	9543	60
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																										
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																										
1	-4,262	1,806	8	-5,494	GGTCTCTCTGCAATGGCTTCC	ATG	9330	273																																																																										
2	-5,918	1,030	10	-6,613	CGGGCGACCGCTGGCCAGCAC	TTG	9378	225																																																																										
3	-7,016	0,516	14	-8,363	CCAGCACTTGCACACTCAAGA	TTG	9393	210																																																																										
4	-5,145	1,352	10	-5,839	GGAAGAGCTCTCTGAAGACTAC	ATG	9423	180																																																																										
5	-2,791	2,493	13	-3,837	TTGGTCGCTGGAGTGTCTCTCA	ATG	9450	153																																																																										
6	-6,211	0,893	9	-6,985	GGCCATCGCCATGGTCAGGAC	ATG	9477	126																																																																										
7	-6,082	0,953	10	-6,777	GGGCAACGGCCCTGATCTGGG	ATG	9543	60																																																																										
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	No, other RBS give a longer ORF Longer ORF does not lead to overlap																																																																																	
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																		
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div>Télécharger GenPep Graphique</div> <div>chaperon de l'assemblage de queue [Microbacterium phage Nucci]</div> <div>ID de séquence : QDK02408.1 Longueur : 90 Nombre de correspondances : 1</div> <div>Voir 2 titres supplémentaires Voir toutes les protéines identiques (IPG)</div> <div>Plage 1 : 1 à 90 graphiques GenPep Prochain match Match précédent</div> <table><tr><th>Score</th><th>Attendre</th><th>Méthode</th><th>Identités</th><th>Points positifs</th><th>Lacunes</th></tr><tr><td>189 bits(481)</td><td>2e-60</td><td>Ajustement de la matrice de composition.</td><td>90/90(100%)</td><td>90/90(100%)</td><td>0/90(0%)</td></tr></table> <div>Requête 1 MKEVKCSGGRPLAQLHNSRLGREETPEDYMAWSLECPAMQAIATGQDMWKTANKTAIE 60 MKEVKCSGGRPLAQLHNSRLGREETPEDYMAWSLECPAMQAIATGQDMWKTANKTAIE Sbjct 1 MKEVKCSGGRPLAQLHNSRLGREETPEDYMAWSLECPAMQAIATGQDMWKTANKTAIE 60</div> <div>Requête 61 SHQKNGPDGPGKGVYVLSQGEVLPPTPET 90 SHQKNGPDGPGKGVYVLSQGEVLPPTPET Sbjct 61 SHQKNGPDGPGKGVYVLSQGEVLPPTPET 90</div>	Score	Attendre	Méthode	Identités	Points positifs	Lacunes	189 bits(481)	2e-60	Ajustement de la matrice de composition.	90/90(100%)	90/90(100%)	0/90(0%)																																																																					
Score	Attendre	Méthode	Identités	Points positifs	Lacunes																																																																													
189 bits(481)	2e-60	Ajustement de la matrice de composition.	90/90(100%)	90/90(100%)	0/90(0%)																																																																													
DECISION:	9330 because it leads to the longest ORF and is conserved in homologs, unlike 9450. In addition, the 237bp gap is no longer present.																																																																																	

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																										
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>PhagesDB :</p> <table><thead><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Quartz_15, tail assembly chaperone, 90</td><td>109</td><td>2e-24</td></tr><tr><td>Nucci_15, tail assembly chaperone, 90</td><td>109</td><td>2e-24</td></tr><tr><td>Mandalorian_15, tail assembly chaperone, 90</td><td>109</td><td>2e-24</td></tr><tr><td>Carostasia_Draft_15, function unknown, 50</td><td>109</td><td>2e-24</td></tr><tr><td>Zenitsu_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Shamu_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Saratos_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Sansa_14, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>MCubed_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Glamour_14, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Eleri_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>ColaCorta_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>ChikPic_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Andromedas_15, tail assembly chaperone, 125</td><td>102</td><td>3e-22</td></tr><tr><td>Finny_15, tail assembly chaperone, 51</td><td>102</td><td>3e-22</td></tr><tr><td>Juanyo_15, tail assembly chaperone, 90</td><td>101</td><td>6e-22</td></tr><tr><td>YuuY_15, tail assembly chaperone, 90</td><td>98</td><td>6e-21</td></tr><tr><td>Pikmin_15, tail assembly chaperone, 307</td><td>94</td><td>1e-19</td></tr><tr><td>Pajaza_15, tail assembly chaperone, 307</td><td>94</td><td>1e-19</td></tr><tr><td>Casey_15, tail assembly chaperone, 307</td><td>94</td><td>1e-19</td></tr><tr><td>MonChoix_15, tail assembly chaperone, 123</td><td>89</td><td>5e-18</td></tr><tr><td>FarmerDoug_Draft_16, function unknown, 123</td><td>89</td><td>5e-18</td></tr><tr><td>Superfresh_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr><tr><td>Schnapsidee_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr><tr><td>Robinson_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr><tr><td>Renzie_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr><tr><td>Redfield_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr><tr><td>Raccoon_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr><tr><td>Peppino_16, tail assembly chaperone, 123</td><td>88</td><td>8e-18</td></tr></tbody></table> <p>Query= Carostasia_15 (50 letters)</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>202729: Redfield_16, tail assembly chaperone, 123, 5e-17.8 E=0e-18</p> <p>Color Key for Alignment Scores</p> <p>nr :</p>	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_15, tail assembly chaperone, 90	109	2e-24	Nucci_15, tail assembly chaperone, 90	109	2e-24	Mandalorian_15, tail assembly chaperone, 90	109	2e-24	Carostasia_Draft_15, function unknown, 50	109	2e-24	Zenitsu_15, tail assembly chaperone, 125	102	3e-22	Shamu_15, tail assembly chaperone, 125	102	3e-22	Saratos_15, tail assembly chaperone, 125	102	3e-22	Sansa_14, tail assembly chaperone, 125	102	3e-22	MCubed_15, tail assembly chaperone, 125	102	3e-22	Glamour_14, tail assembly chaperone, 125	102	3e-22	Eleri_15, tail assembly chaperone, 125	102	3e-22	ColaCorta_15, tail assembly chaperone, 125	102	3e-22	ChikPic_15, tail assembly chaperone, 125	102	3e-22	Andromedas_15, tail assembly chaperone, 125	102	3e-22	Finny_15, tail assembly chaperone, 51	102	3e-22	Juanyo_15, tail assembly chaperone, 90	101	6e-22	YuuY_15, tail assembly chaperone, 90	98	6e-21	Pikmin_15, tail assembly chaperone, 307	94	1e-19	Pajaza_15, tail assembly chaperone, 307	94	1e-19	Casey_15, tail assembly chaperone, 307	94	1e-19	MonChoix_15, tail assembly chaperone, 123	89	5e-18	FarmerDoug_Draft_16, function unknown, 123	89	5e-18	Superfresh_16, tail assembly chaperone, 123	88	8e-18	Schnapsidee_16, tail assembly chaperone, 123	88	8e-18	Robinson_16, tail assembly chaperone, 123	88	8e-18	Renzie_16, tail assembly chaperone, 123	88	8e-18	Redfield_16, tail assembly chaperone, 123	88	8e-18	Raccoon_16, tail assembly chaperone, 123	88	8e-18	Peppino_16, tail assembly chaperone, 123	88	8e-18
Sequences producing significant alignments:	Score (bits)	E Value																																																																																									
Quartz_15, tail assembly chaperone, 90	109	2e-24																																																																																									
Nucci_15, tail assembly chaperone, 90	109	2e-24																																																																																									
Mandalorian_15, tail assembly chaperone, 90	109	2e-24																																																																																									
Carostasia_Draft_15, function unknown, 50	109	2e-24																																																																																									
Zenitsu_15, tail assembly chaperone, 125	102	3e-22																																																																																									
Shamu_15, tail assembly chaperone, 125	102	3e-22																																																																																									
Saratos_15, tail assembly chaperone, 125	102	3e-22																																																																																									
Sansa_14, tail assembly chaperone, 125	102	3e-22																																																																																									
MCubed_15, tail assembly chaperone, 125	102	3e-22																																																																																									
Glamour_14, tail assembly chaperone, 125	102	3e-22																																																																																									
Eleri_15, tail assembly chaperone, 125	102	3e-22																																																																																									
ColaCorta_15, tail assembly chaperone, 125	102	3e-22																																																																																									
ChikPic_15, tail assembly chaperone, 125	102	3e-22																																																																																									
Andromedas_15, tail assembly chaperone, 125	102	3e-22																																																																																									
Finny_15, tail assembly chaperone, 51	102	3e-22																																																																																									
Juanyo_15, tail assembly chaperone, 90	101	6e-22																																																																																									
YuuY_15, tail assembly chaperone, 90	98	6e-21																																																																																									
Pikmin_15, tail assembly chaperone, 307	94	1e-19																																																																																									
Pajaza_15, tail assembly chaperone, 307	94	1e-19																																																																																									
Casey_15, tail assembly chaperone, 307	94	1e-19																																																																																									
MonChoix_15, tail assembly chaperone, 123	89	5e-18																																																																																									
FarmerDoug_Draft_16, function unknown, 123	89	5e-18																																																																																									
Superfresh_16, tail assembly chaperone, 123	88	8e-18																																																																																									
Schnapsidee_16, tail assembly chaperone, 123	88	8e-18																																																																																									
Robinson_16, tail assembly chaperone, 123	88	8e-18																																																																																									
Renzie_16, tail assembly chaperone, 123	88	8e-18																																																																																									
Redfield_16, tail assembly chaperone, 123	88	8e-18																																																																																									
Raccoon_16, tail assembly chaperone, 123	88	8e-18																																																																																									
Peppino_16, tail assembly chaperone, 123	88	8e-18																																																																																									



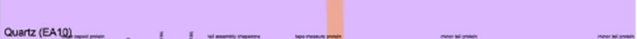

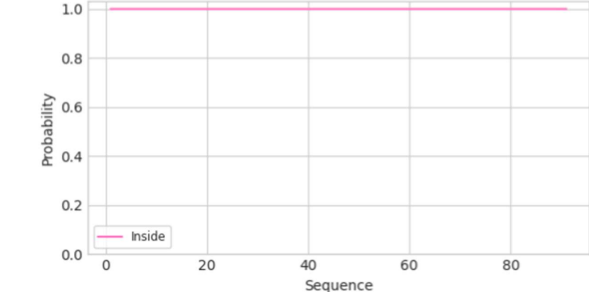
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
✓	tail assembly chaperone [Microbacterium pha...	Microba...	189	189	98%	2e-60	100.00%	90	QDK02408.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	178	178	98%	5e-56	91.11%	90	USH44814.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	178	178	97%	8e-56	92.13%	90	QFP95344.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	173	173	98%	1e-53	87.78%	125	AXH48587.2
✓	tail assembly chaperone [Microbacterium pha...	Microba...	173	173	97%	2e-53	88.76%	125	YP_009623053.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	168	168	97%	4e-49	84.27%	307	YP_009624269.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	155	155	97%	1e-46	78.65%	123	AUX83102.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	155	155	97%	1e-46	78.65%	123	YP_009908654.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	155	155	97%	2e-46	79.78%	123	AUX82728.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	155	155	97%	2e-46	79.78%	123	YP_009623117.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	154	154	97%	4e-46	78.65%	123	YP_009623180.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	154	154	97%	6e-46	78.65%	120	QDM57688.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	154	154	97%	6e-46	78.65%	124	URC17873.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	153	153	97%	1e-45	78.65%	123	AVR56281.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	152	152	97%	3e-45	77.53%	124	WIC89574.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	152	152	97%	4e-45	77.53%	125	Q8Z73249.1
✓	tail assembly chaperone [Microbacterium pha...	Microba...	149	149	97%	6e-44	76.40%	124	YP_009908593.1



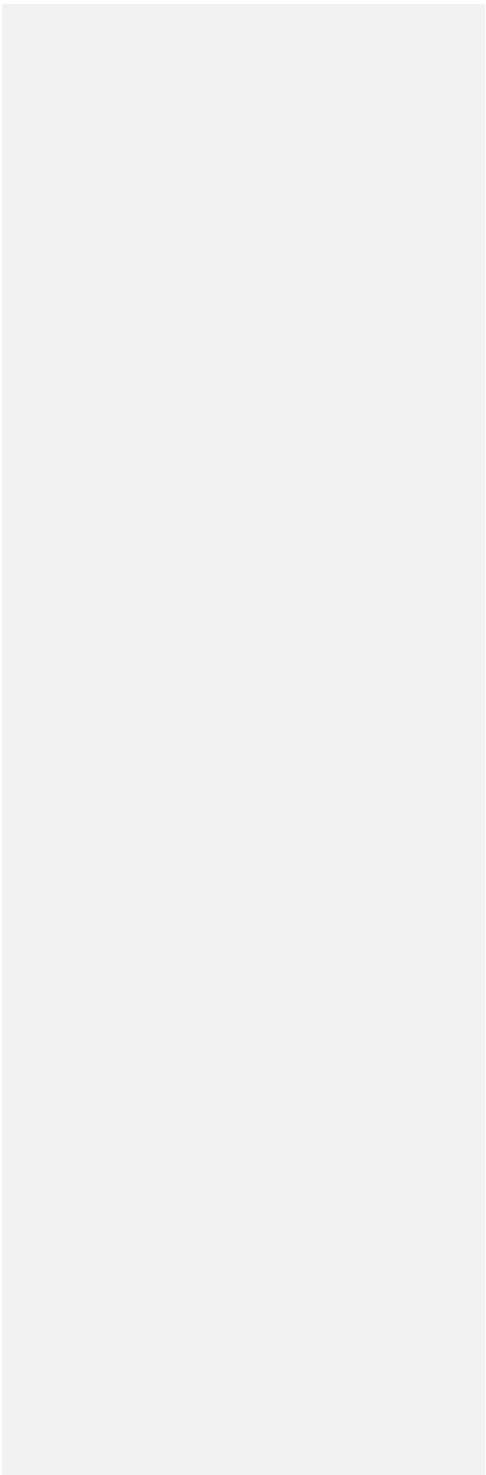
Tous les hits trouvés sont des tail assembly chaperone, avec une e-value <<<10⁻⁴, et un query coverage excellent pour grande partie >98%.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

HHpred trouve des A2L zinc ribbon domain, une protéine arginine kinase activator et une RAd50 zinc hook motif.

	<div><div><div><input type="checkbox"/> 1 PF08792.14 ; A2L_zn_ribbon ; A2L zinc ribbon domain92.660.08225.39022</div><div><input type="checkbox"/> 2 8GQD_H Protein-arginine kinase activator protein; Arginine kinase, Protein quality control, Zinc finger, Enzyme activation, TRA91.70.3829.952.119</div><div><input type="checkbox"/> 3 PF04423.18 ; Rad50_zn_hook ; Rad50 zinc hook motif91.090.2827.29114</div></div></div>
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<p>Chez les phages du même groupe, les gènes 14 et 16 codent respectivement pour une tail assembly chaperone et une tail measure protein, il y aurait donc syntenie.</p> <p>Carostasia_Draft (EA10)</p>  <p>Nuvio (EA10)</p>  <p>Quartz (EA10)</p> 
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p>D'après DeepTMHMM c'est une protéine globulaire.</p> <p>DeepTMHMM - Most Likely Topology Type: Globular</p>  <p>DeepTMHMM - Posterior Probabilities</p> 
Est-ce que la fonction proposée fait partie de liste	Oui.

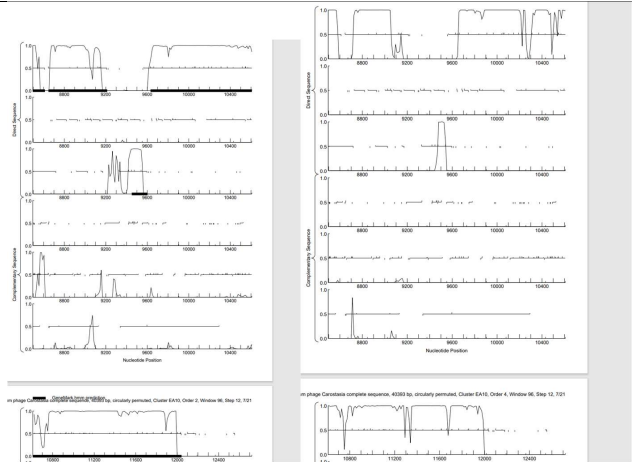
de fonctions approuvées par SEA-PHAGES ?	
DECISION:	<i>Ce gène code pour une tail assembly chaperone.</i>

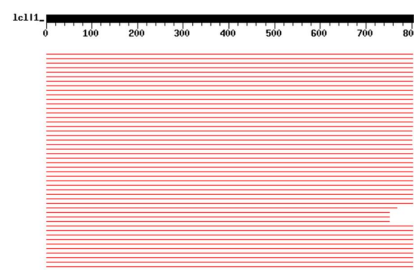


Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	16
Coordonnées du Stop	12042
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 28
Coordonnées du Start retenu	9631
Fonction prédite	Tape measure protein

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	 <p>Potential coding for both even if genemarks gives the start at 9600 instead of 9631.</p>

	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>353188 Quartz_16, tape measure protein, 803. S=1493 E=0.0</p> <p>Color Key for Alignment Scores</p> <p><40 40-50 50-60 60-70 70-80 80-90 90-100</p>  <p>Sequences producing significant alignments:</p> <table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Carostasia_Draft_16, function unknown, 803</td><td>1588</td><td>0.0</td></tr><tr><td>Quartz_16, tape measure protein, 803</td><td>1493</td><td>0.0</td></tr><tr><td>Hucci_16, tape measure protein, 803</td><td>1487</td><td>0.0</td></tr><tr><td>Mandalorian_16, tape measure protein, 803</td><td>1487</td><td>0.0</td></tr><tr><td>YuuY_16, tape measure protein, 803</td><td>1399</td><td>0.0</td></tr><tr><td>Shamu_Draft_17, function unknown, 802</td><td>1126</td><td>0.0</td></tr><tr><td>Saratos_16, tape measure protein, 802</td><td>1126</td><td>0.0</td></tr><tr><td>Glamour_15, tape measure protein, 802</td><td>1126</td><td>0.0</td></tr><tr><td>Eleri_16, tape measure protein, 802</td><td>1126</td><td>0.0</td></tr><tr><td>ColaCorta_16, tape measure protein, 802</td><td>1126</td><td>0.0</td></tr><tr><td>Andromeda_16, tape measure protein, 802</td><td>1126</td><td>0.0</td></tr><tr><td>Zenitsu_16, tape measure protein, 802</td><td>1124</td><td>0.0</td></tr><tr><td>Mcubed_16, tape measure protein, 802</td><td>1124</td><td>0.0</td></tr><tr><td>ChikPic_16, tape measure protein, 802</td><td>1124</td><td>0.0</td></tr><tr><td>Sansa_15, tape measure protein, 802</td><td>1123</td><td>0.0</td></tr><tr><td>Finny_16, tape measure protein, 802</td><td>1122</td><td>0.0</td></tr><tr><td>Jumyo_16, tape measure protein, 806</td><td>1144</td><td>0.0</td></tr><tr><td>SirVictor_17, tape measure protein, 802</td><td>1137</td><td>0.0</td></tr><tr><td>Guetzie_17, tape measure protein, 802</td><td>1137</td><td>0.0</td></tr><tr><td>Lucky3_17, tape measure protein, 802</td><td>1135</td><td>0.0</td></tr><tr><td>Golden_17, tape measure protein, 802</td><td>1135</td><td>0.0</td></tr><tr><td>Koji_17, tape measure protein, 802</td><td>1099</td><td>0.0</td></tr><tr><td>PrincePherus_17, tape measure protein, 802</td><td>1028</td><td>0.0</td></tr><tr><td>BouleyBill_Draft_18, function unknown, 802</td><td>1028</td><td>0.0</td></tr><tr><td>Sinatra_17, tape measure protein, 802</td><td>1025</td><td>0.0</td></tr><tr><td>Pherbot_17, tape measure protein, 802</td><td>1020</td><td>0.0</td></tr><tr><td>Bustleton_17, tape measure protein, 802</td><td>1020</td><td>0.0</td></tr><tr><td>Kauala_17, tape measure protein, 802</td><td>1020</td><td>0.0</td></tr><tr><td>GaeCeo_18, tape measure protein, 802</td><td>1014</td><td>0.0</td></tr><tr><td>ChickenKing_17, tape measure protein, 802</td><td>1013</td><td>0.0</td></tr><tr><td>Cheeto_18, tape measure protein, 802</td><td>1077</td><td>0.0</td></tr><tr><td>Schubert_17, tape measure protein, 798</td><td>1060</td><td>0.0</td></tr><tr><td>Theresita_17, tape measure protein, 802</td><td>1059</td><td>0.0</td></tr><tr><td>WilliamStrong_16, tape measure protein, 810</td><td>924</td><td>0.0</td></tr></tbody></table> <p><i>Strong homologs with the gene</i></p>		Score (bits)	E Value	Carostasia_Draft_16, function unknown, 803	1588	0.0	Quartz_16, tape measure protein, 803	1493	0.0	Hucci_16, tape measure protein, 803	1487	0.0	Mandalorian_16, tape measure protein, 803	1487	0.0	YuuY_16, tape measure protein, 803	1399	0.0	Shamu_Draft_17, function unknown, 802	1126	0.0	Saratos_16, tape measure protein, 802	1126	0.0	Glamour_15, tape measure protein, 802	1126	0.0	Eleri_16, tape measure protein, 802	1126	0.0	ColaCorta_16, tape measure protein, 802	1126	0.0	Andromeda_16, tape measure protein, 802	1126	0.0	Zenitsu_16, tape measure protein, 802	1124	0.0	Mcubed_16, tape measure protein, 802	1124	0.0	ChikPic_16, tape measure protein, 802	1124	0.0	Sansa_15, tape measure protein, 802	1123	0.0	Finny_16, tape measure protein, 802	1122	0.0	Jumyo_16, tape measure protein, 806	1144	0.0	SirVictor_17, tape measure protein, 802	1137	0.0	Guetzie_17, tape measure protein, 802	1137	0.0	Lucky3_17, tape measure protein, 802	1135	0.0	Golden_17, tape measure protein, 802	1135	0.0	Koji_17, tape measure protein, 802	1099	0.0	PrincePherus_17, tape measure protein, 802	1028	0.0	BouleyBill_Draft_18, function unknown, 802	1028	0.0	Sinatra_17, tape measure protein, 802	1025	0.0	Pherbot_17, tape measure protein, 802	1020	0.0	Bustleton_17, tape measure protein, 802	1020	0.0	Kauala_17, tape measure protein, 802	1020	0.0	GaeCeo_18, tape measure protein, 802	1014	0.0	ChickenKing_17, tape measure protein, 802	1013	0.0	Cheeto_18, tape measure protein, 802	1077	0.0	Schubert_17, tape measure protein, 798	1060	0.0	Theresita_17, tape measure protein, 802	1059	0.0	WilliamStrong_16, tape measure protein, 810	924	0.0
	Score (bits)	E Value																																																																																																								
Carostasia_Draft_16, function unknown, 803	1588	0.0																																																																																																								
Quartz_16, tape measure protein, 803	1493	0.0																																																																																																								
Hucci_16, tape measure protein, 803	1487	0.0																																																																																																								
Mandalorian_16, tape measure protein, 803	1487	0.0																																																																																																								
YuuY_16, tape measure protein, 803	1399	0.0																																																																																																								
Shamu_Draft_17, function unknown, 802	1126	0.0																																																																																																								
Saratos_16, tape measure protein, 802	1126	0.0																																																																																																								
Glamour_15, tape measure protein, 802	1126	0.0																																																																																																								
Eleri_16, tape measure protein, 802	1126	0.0																																																																																																								
ColaCorta_16, tape measure protein, 802	1126	0.0																																																																																																								
Andromeda_16, tape measure protein, 802	1126	0.0																																																																																																								
Zenitsu_16, tape measure protein, 802	1124	0.0																																																																																																								
Mcubed_16, tape measure protein, 802	1124	0.0																																																																																																								
ChikPic_16, tape measure protein, 802	1124	0.0																																																																																																								
Sansa_15, tape measure protein, 802	1123	0.0																																																																																																								
Finny_16, tape measure protein, 802	1122	0.0																																																																																																								
Jumyo_16, tape measure protein, 806	1144	0.0																																																																																																								
SirVictor_17, tape measure protein, 802	1137	0.0																																																																																																								
Guetzie_17, tape measure protein, 802	1137	0.0																																																																																																								
Lucky3_17, tape measure protein, 802	1135	0.0																																																																																																								
Golden_17, tape measure protein, 802	1135	0.0																																																																																																								
Koji_17, tape measure protein, 802	1099	0.0																																																																																																								
PrincePherus_17, tape measure protein, 802	1028	0.0																																																																																																								
BouleyBill_Draft_18, function unknown, 802	1028	0.0																																																																																																								
Sinatra_17, tape measure protein, 802	1025	0.0																																																																																																								
Pherbot_17, tape measure protein, 802	1020	0.0																																																																																																								
Bustleton_17, tape measure protein, 802	1020	0.0																																																																																																								
Kauala_17, tape measure protein, 802	1020	0.0																																																																																																								
GaeCeo_18, tape measure protein, 802	1014	0.0																																																																																																								
ChickenKing_17, tape measure protein, 802	1013	0.0																																																																																																								
Cheeto_18, tape measure protein, 802	1077	0.0																																																																																																								
Schubert_17, tape measure protein, 798	1060	0.0																																																																																																								
Theresita_17, tape measure protein, 802	1059	0.0																																																																																																								
WilliamStrong_16, tape measure protein, 810	924	0.0																																																																																																								
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p><i>Gap of 29 therefore acceptable</i></p>																																																																																																									
DECISION:	<p>yes</p>																																																																																																									

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	9631

Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?

Best RBS score: -2.443

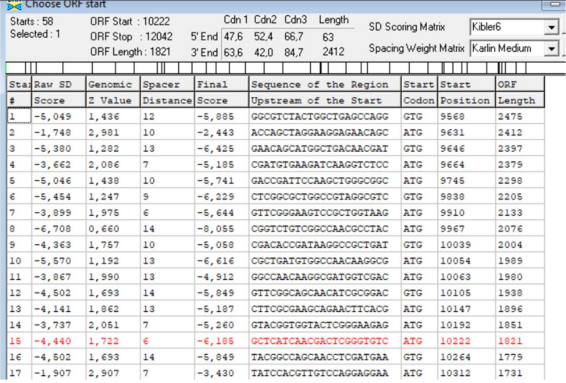
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?

We then have a length of 2412 bp with a gap of 29 with the preceding gene.

Est-ce que le start est conservé chez les homologues voir Starterator ?

We have the same start with starterator

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?

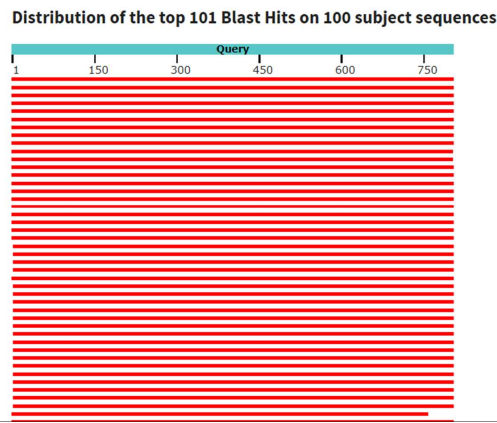


Best RBS score: -2.443

We then have a length of 2412 bp with a gap of 29 with the preceding gene.

Gene: Carostasia_16 Start: 9631, Stop: 12042, Start Num: 4
Candidate Starts for Carostasia_16:
(1, 9568), (2, 9644), (2, 9631), (3, 9646), (6, 9664), (9, 9745), (18, 9838), (21, 9910), (24, 9967), (28, 10039), (30, 10054), (31, 10063), (34, 10105), (35, 10147), (36, 10192), (38, 10222), (42, 10264), (45, 10312), (47, 10372), (51, 10429), (60, 10516), (61, 10546), (62, 10549), (65, 10570), (71, 10627), (75, 10672), (78, 10693), (88, 10762), (99, 10822), (109, 10879), (115, 10924), (116, 10930), (121, 10957), (124, 10984), (126, 11002), (135, 11062), (138, 11098), (139, 11101), (141, 11110), (142, 11122), (148, 11155), (153, 11185), (162, 11350), (168, 11389), (177, 11452), (186, 11539), (192, 11614), (193, 11617), (194, 11665), (200, 11707), (201, 11713), (205, 11770), (219, 11854), (221, 11860), (235, 11938), (236, 11947), (244, 12004), (248, 12031), (249, 12034).

We have the same start with starterator



	tape measure protein [Microbacterium phage Quartz] Sequence ID: UVK59235.1 Length: 803 Number of Matches: 1				
	Range 1: 1 to 803 GenPept Graphics			▼ Next Match ▲ Previous	
	Score	Expect	Method	Identities	Positives
	1532 bits(3966)	0.0	Compositional matrix adjust.	792/803(99%)	799/803(99%)
					Gaps
					0/803(0%)
Query	1	MADNDVKIKVSMGDKVINGLTGIGDAADKTDSKLGGMKGGLAGAGKALVGFSAAAVA			60
Sbjct	1	MADNDVKIKVSMGDKVINGLTGIGDAADKTDSKLGGMKGGLAGAGKALVGFSAAAVA			60
Query	61	AGSALAVGVVGAAYAEQNIIGGIETLFGKSAGKMEQYANDAYKTAGLSANAYMEQVTSFS			120
Sbjct	61	AGSALAVGVVGAAYAEQNIIGGIETLFGKSAGKMEQYANDAYKTAGLSANAYMEQVTSFS			120
Query	121	ASLLQGLGGDTDKAADVANKAMVMSDNANKFGSNIADVQNAVYQGFQNFTHLDNLKLG			180
Sbjct	121	ASLLQGLGGDTDKAADVANKAMVMSDNANKFGSNIADVQNAVYQGFQNFTHLDNLKLG			180
Query	181	YGGTREETMARLINDSGVMGDTFTATASNDEVSFDKIIEGHHVVOEEMGVAGTTAKEASE			240
Sbjct	181	YGGTREETMARLINDSGVMGDTFTATASNDEVSFDKIIEGHHVVOEEMGVAGTTAKEASE			240
Query	241	TISGSIHMLKGSFENLLVGLGSTQDQVAALAGNVDSLETVTINIAPIENLGSNMATLG			300
Sbjct	241	TISGSIHMLKGSFENLLVGLGSTQDQVAALAGNVDSLETVTINIAPIENLGSNMATLG			300
Query	301	PKLGTMMESLVGAVSSAIPALLDAGVGLIGGLVSGITSALPSLITALVPVIGLVQSLAE			360
Sbjct	301	PKLGTMMESLVGAVSSAIPALLDAGVGLIGGLVSGITSALPSLITALVPVIGLVQSLAE			360
Query	361	LLPQLTSAGAEAAALAVGTAEAAPTLIPAIIEGTLGHVQAIIDNLPLLDAGLQLVTGL			420
Sbjct	361	LLPQLTSAGAEAAALAVGTAEAAPTLIPAIIEGTLGHV A+I+NLPLLDAGLQLV GL			420
Query	421	ATGILEAIPVIVEMLPQLIEGIVAFIATGVPHLIEAGVOLFTALVGPLDIITQIVAVLP			480
Sbjct	421	ATGILESIPVIEMLPVLIQGIVSFQITGVPHLIEAGIQLFTALVGPLDIITQIVAVLP			480
Query	481	QIIQSVIGAVGVAPLLVDAGIQLLTALVTALPEIITTVAAALPQIISVLTAVFDAIPL			540
Sbjct	481	QIIQSVIGAVGVAPLLVDAGIQLLTALVTALPEIITTVAAALPQIISVLTAVFDAIPL			540
Query	541	LIDAGIKLFISLIEALPQIINTIIGATPQIIGSVNLATGAIPQIMAGVKLLTSVITNL			600
Sbjct	541	LIDAGIKLFISLIEALPQIINTIIGATPQIIGSVNLATGAIPQIMAGVKLLTSVITNL			600
Query	601	PSIIGTIVGAIPQIIGGIVGALNGIGQIISVGGNIVRGLWDGISGAAGWLFNKITGGFVN			660
Sbjct	601	PSIIGTIVGAIPQIIGGIVGALNGIGQIISVGGNIVRGLWDGISGAAGWLFNKITGGFVN			660
Query	661	DVMGNIGSFFGIASPSKRMDEIGAFLPSGIGVGVVERHEADAIQPTDLNKKVKKEAQQL			720
Sbjct	661	DVMGNIGSFFGIASPSKRMDEIGAFLPSGIGVGVVERHEADAIQPTDLNKKVKKEAQQL			720
Query	721	PTLGFDDQVQTQLTQQIIPQVPMQPSSTPVGLGAAAPQVVISGPLVSAEHSVRDORD			780
Sbjct	721	PTLGFDDQVQTQLTQQIIPQVPMQPSSTPVGLGAAAPQVVISGPLVSAEHSVRDORD			780
Query	781	IRALSTQLKTDMTRELRAQGVMS 803			
Sbjct	781	IRALSTQLKTDMTRELRAQGVMS 803			
DECISION:	Start retained: 9631 because best RBS score, coding potential that corresponds and no breach of principles.				

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec	Listez le meilleur hit Blastp pour chaque source :

une e-value < 10⁻⁴ et une couverture acceptable ?

```
*** PhagesDB :

>Quart_16, tape measure protein, 803
Length = 803

Score = 1493 bits (3866), Expect = 0.0
Identities = 792/803 (98%), Positives = 799/803 (99%)

Query: 1 MADNDVKIKVSHDGDVVINGLTGIGDAADKTDKSLGGHIKGLAGAGKALVGFSAAVA 60
MADNDVKIKVSHDGDVVINGLTGIGDAADKTDKSLGGHIKGLAGAGKALVGFSAAVA
Sbjct: 1 MADNDVKIKVSHDGDVVINGLTGIGDAADKTDKSLGGHIKGLAGAGKALVGFSAAVA 60

Query: 61 AGSALAVGVVGYAAEQNIGGIETLFGKSAGKMEQYANDAYKTAGLSANAYMEQVTSFS 120
AGSALAVGVVGYAAEQNIGGIETLFGKSAGKMEQYANDAYKTAGLSANAYMEQVTSFS
Sbjct: 61 AGSALAVGVVGYAAEQNIGGIETLFGKSAGKMEQYANDAYKTAGLSANAYMEQVTSFS 120

Query: 121 ASLLQGLGGDTOKAADVANKAMVMSDNANKFGSNIADVNAYQGFQKQFTMLDNKLG 180
ASLLQGLGGDTOKAADVANKAMVMSDNANKFGSNIADVNAYQGFQKQFTMLDNKLG
Sbjct: 121 ASLLQGLGGDTOKAADVANKAMVMSDNANKFGSNIADVNAYQGFQKQFTMLDNKLG 180

Query: 181 YGGTREEMARLINDSGVMGDTFTATASNLDVSVFKIIEGIHVQEEHVGAGTTAKEASE 240
YGGTREEMARLINDSGVMGDTFTATASNLDVSVFKIIEGIHVQEEHVGAGTTAKEASE
Sbjct: 181 YGGTREEMARLINDSGVMGDTFTATASNLDVSVFKIIEGIHVQEEHVGAGTTAKEASE 240

Query: 241 TISGSIHMLKGSFENLLVGLSTQDQVAALAGNVIDSLETVITNIAPVIENTLSHATLG 300
TISGSIHMLKGSFENLLVGLSTQDQVAALAGNVIDSLETVITNIAPVIENTLSHATLG
Sbjct: 241 TISGSIHMLKGSFENLLVGLSTQDQVAALAGNVIDSLETVITNIAPVIENTLSHATLG 300

Query: 301 PKLGTWIESLVGAVSSAIPALLDAGVGLIGGLVSGITSALPSLITALVPQVIGLVQSLAE 360
PKLGTWIESLVGAVSSAIPALLDAGVGLIGGLVSGITSALPSLITALVPQVIGLVQSLAE
Sbjct: 301 PKLGTWIESLVGAVSSAIPALLDAGVGLIGGLVSGITSALPSLITALVPQVIGLVQSLAE 360

Query: 361 LLPQLISAGAEAAALAVGIAEAAPTLIPAIIEGILGNVQAIIDNPLLLDAGLQVTL 420
LLPQLISAGAEAAALAVGIAEAAPTLIPAIIEGILGNV A+I+NLPLLLDAGLQV GL
Sbjct: 361 LLPQLISAGAEAAALAVGIAEAAPTLIPAIIEGILGNVAVIENLPLLLDAGLQVGL 420

Query: 421 ATGILEATPVIVHLPQIEGIVAFIATGVPHLIEAGQLFTALVGALPOITQIVAVLP 480
ATGILE+IPVI+HLP LI+GIV+FI TGVPHLIEAG+QLFTALVGALPOITQIVAVLP
Sbjct: 421 ATGILESIPVIEHLPVLIQGISVFIQTGVPHLIEAGQLFTALVGALPOITQIVAVLP 480

Query: 481 QIIQSVIGAVGVAPLLVDAGIQLLTALVTALPEIITTVAAALPQIIDSVLAVFDAIPL 540
QIIQSVIGAVGVAPLLVDAGIQLLTALVTALPEIITTVAAALPQIIDSVLAVFDAIPL
Sbjct: 481 QIIQSVIGAVGVAPLLVDAGIQLLTALVTALPEIITTVAAALPQIIDSVLAVFDAIPL 540

Query: 541 LIDAGIKLFISLIEALPQIINTIIGAIIPQIIGSVLNLTAIPQLINAGVKLLTSVITNL 600
LIDAGIKLFISLIEALPQIINTIIGAIIPQIIGSVLNLTAIPQLINAGVKLLTSVITNL
Sbjct: 541 LIDAGIKLFISLIEALPQIINTIIGAIIPQIIGSVLNLTAIPQLINAGVKLLTSVITNL 600

Query: 601 PSIIIGTIVGAIPQIIGGIVGALNGIGQIISVGGNIVRGLWDIGSAGAGLGNKIGGFVN 660
PSIIIGTIVGAIPQIIGGIVGALNGIGQIISVGGNIVRGLWDIGSAGAGLGNKIGGFVN
Sbjct: 601 PSIIIGTIVGAIPQIIGGIVGALNGIGQIISVGGNIVRGLWDIGSAGAGLGNKIGGFVN 660

Query: 661 DVMGNIGSFFGIASPSKRNIRDEIGAF LPSGIGVGVERHEADAIQPIITDLNKKVMKEAQQ 720
DVMGNIGSFFGIASPSKRNIRDEIGAF LPSGIGVGVERHEADAIQPIITDLNKKVMKEAQQ
Sbjct: 661 DVMGNIGSFFGIASPSKRNIRDEIGAF LPSGIGVGVERHEADAIQPIITDLNKKVMKEAQQ 720

Query: 721 PTLGFDQVQTQLTQQIIPQVPHQPSSTPVGLGAAAPQQVTSIGPLVSAEHSVRD 780
PTLGFDQVQTQLTQQIIPQVPHQPSSTPVGLGAAAPQQVTSIGPLVSAEHSVRD
Sbjct: 721 PTLGFDQVQTQLTQQIIPQVPHQPSSTPVGLGAAAPQQVTSIGPLVSAEHSVRD 780

Query: 781 IRALSTQLKTDHIRELRAQGVMS 803
IRALSTQLKTDHIRELRAQGVMS
Sbjct: 781 IRALSTQLKTDHIRELRAQGVMS 803
```

*** nr :

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Pos. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hace measure system (Microbacterium phage Ques2)	Microbacterium phage Ques2	1532	1532	99%	0.0	98.63%	803	U955625.1
<input checked="" type="checkbox"/>	hace measure system (Microbacterium phage Nucc2)	Microbacterium phage Nucc2	1523	1523	99%	0.0	98.36%	803	Q602400.1

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

UniProt :

Template alignment

☐ 1. Q24L11 TMP_BPPCD Probable tape measure protein OS=Clostridium phage phiCD119 (strain Clostridium difficile/United States/Govind/2006) OX=1283341 PE=3 SV=1
Probability: 99.97%, E-value: 1.2e-21, Score: 229.55, Aligned cols: 130, Identities: 28%, Similarity: 0.47, Template Neff: 12

PDB :

Template alignment | Template 3D Structure | PDBe

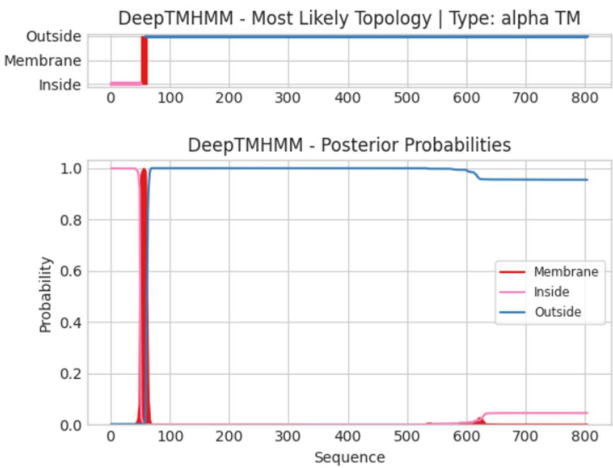
☐ 3. 6V81_BF Tape Measure Protein, gp57; phage tail, tail tip, tape measure protein, VIRAL PROTEIN; 3.7A (Staphylococcus virus 80alpha)
Probability: 99.95%, E-value: 1.6e-18, Score: 206.74, Aligned cols: 126, Identities: 23%, Similarity: 0.426, Template Neff: 10.8

Ce gène est-il situé à côté de gènes de fonction

Proche de la minor tail protein et de la tail assembly chaperone

connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Oui, tape measure protein

tape measure protein	Tape Measure, tmp, tape measure	If you have sir
----------------------	---------------------------------	-----------------

DECISION:

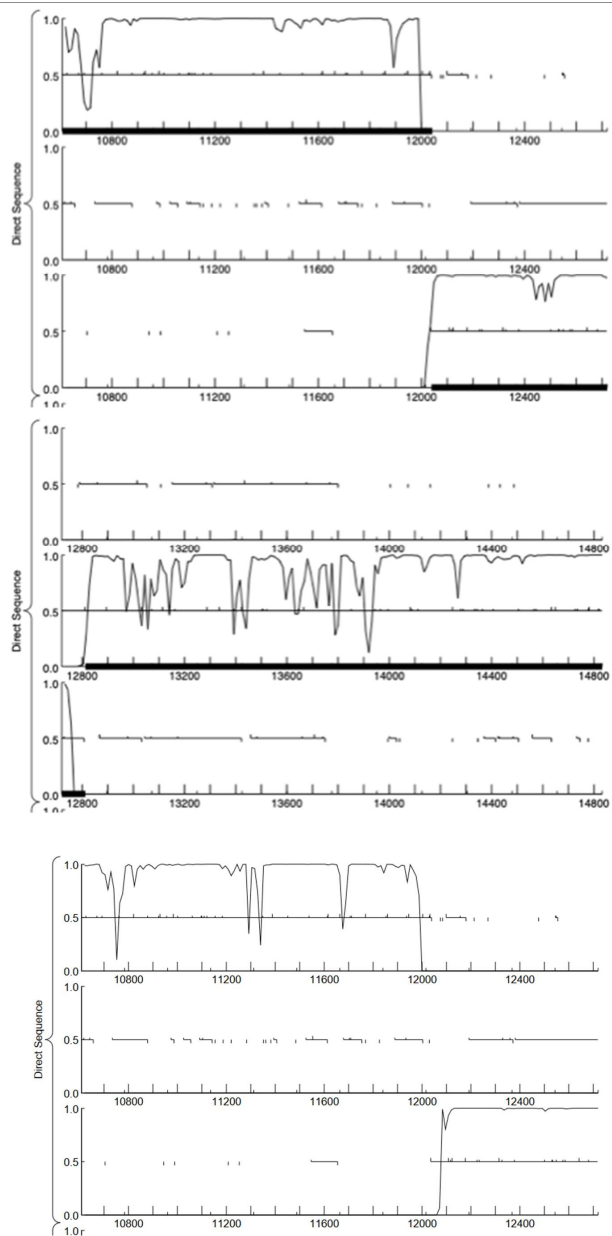
Tape Measure Protein

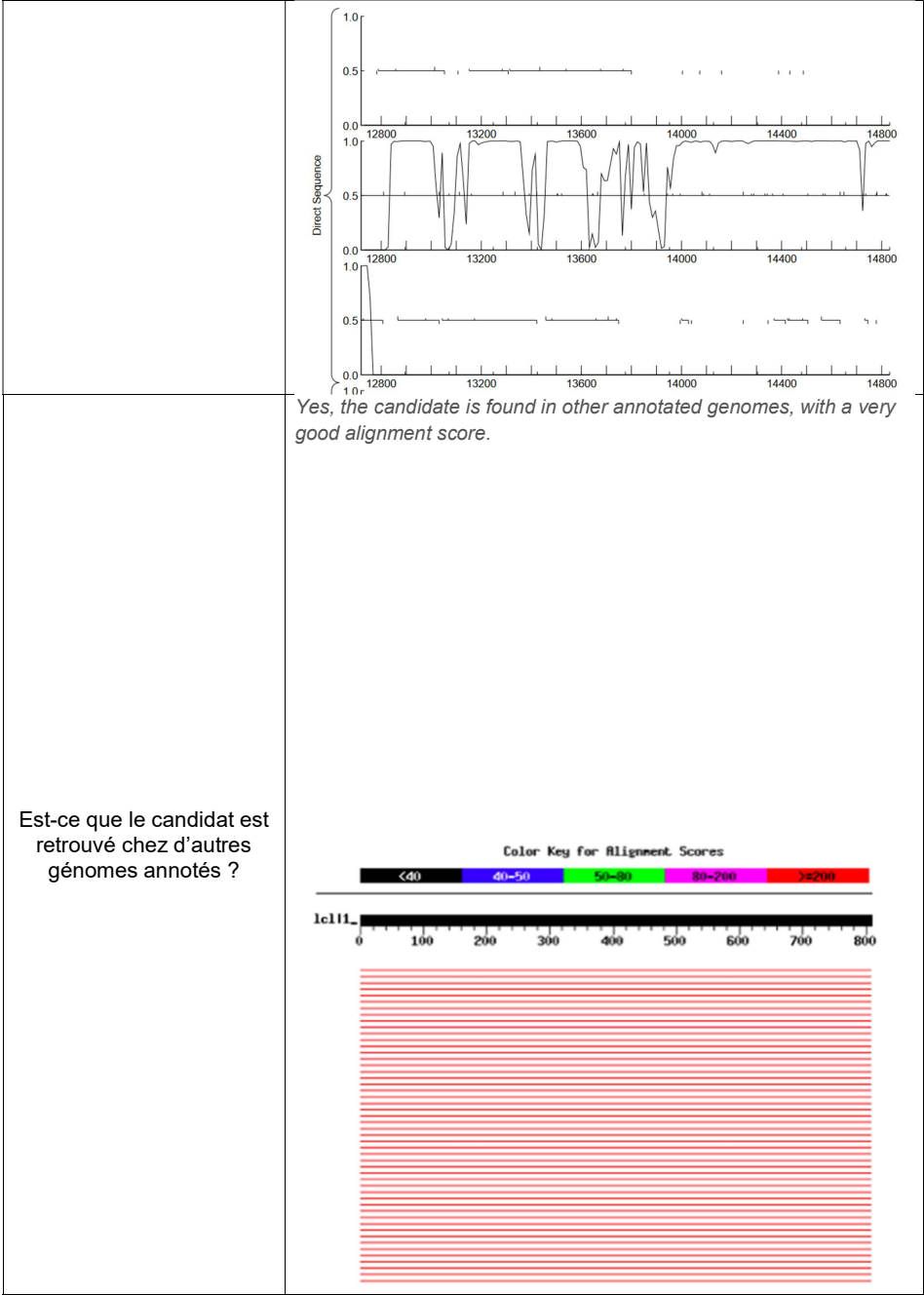
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	17
Coordonnées du Stop	12812
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	ove 4
Coordonnées du Start retenu	12039
Fonction prédite	<i>minor tail protein</i>

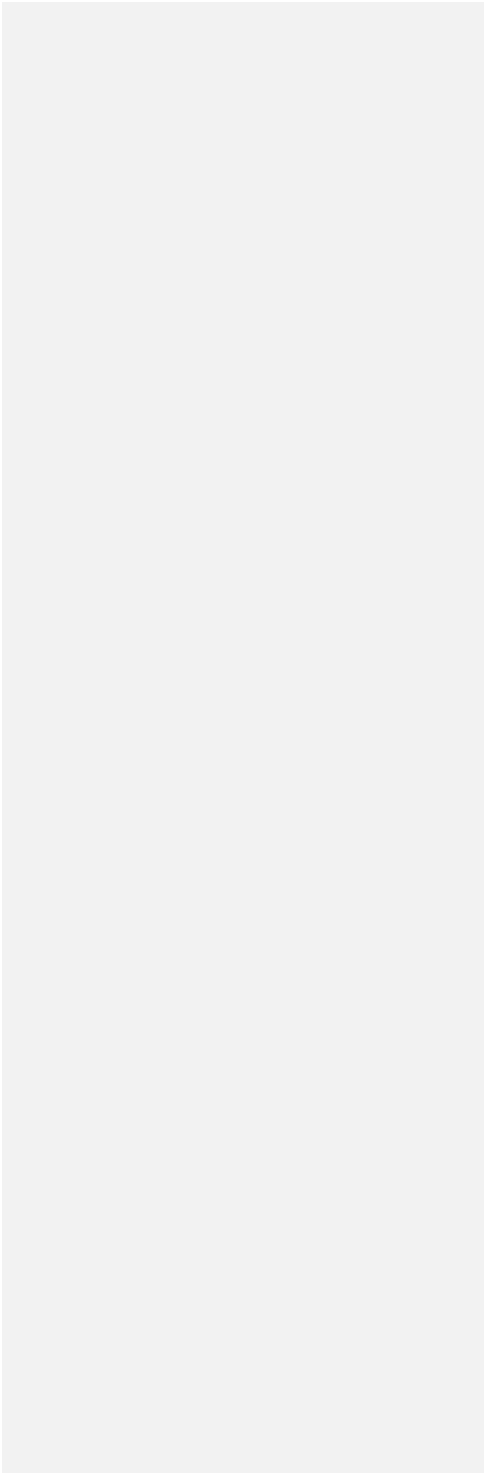
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both.</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>GeneMarks detects a coding potential along the entire length of the ORF.</i>





Est-ce que le candidat est en contradiction avec les principes d'annotation ?	No
DECISION:	Yes



Décision #2 : Quel est le start du gène ?

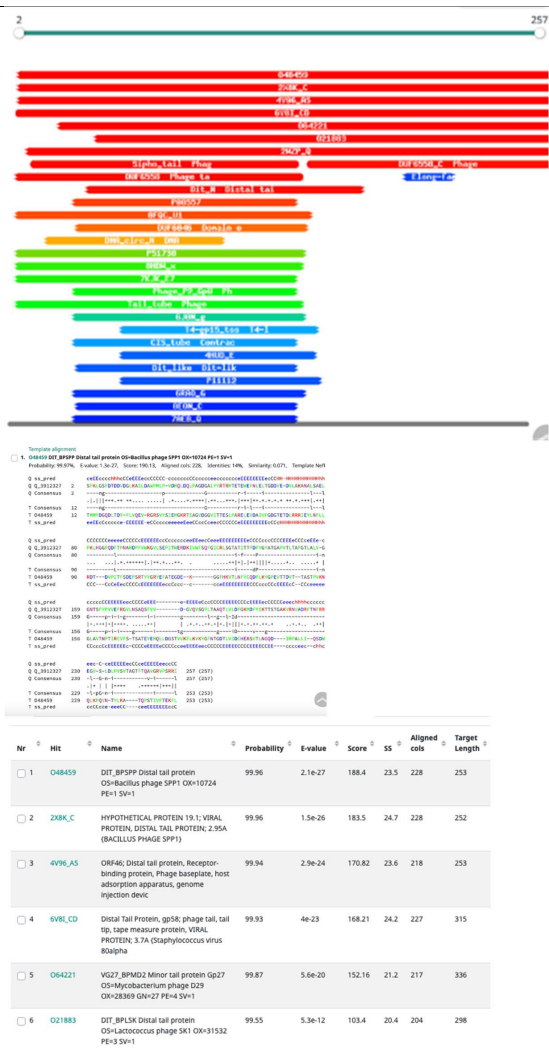
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	12039 by Glimmer and GeneMark
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<i>This start is an ATG associated with a good-scoring RBS. Final score = -2,505.</i>
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<i>Yes, the predicted start leads to the longest ORF= 774bp, no excessive overlap (-4).</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<i>Alone in his group, impossible to conclude by definition.</i>
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p><i>Yes, the start is preserved in other homologs with Blastp.</i></p> <hr/> <pre> >Mandalorian_17, minor tail protein, 257 Length = 257 Score = 514 bits (1325), Expect = e-146 Identities = 254/257 (98%), Positives = 255/257 (99%) Query: 1 MSFKLGSFDTDDVDGLKAILDAMPMLFVDMQLDQLPAGDGALYYTRMTETEWEFNLELT 60 MSFKLGSFDTDDVDGLKAILDAMPMLFVDMQLDQLPAGDGALYYTRMTETEWEFNLELT Sbjct: 1 MSFKLGSFDTDDVDGLKAILDAMPMLFVDMQLDQLPAGDGALYYTRMTETEWEFNLELT 60 Query: 61 GDDVEDVLAKANALSAELNPKLHGCPQDFTPNAMDPPVWKGVLSEPIWTERDKIVNFSQF 120 GDDVEDVLAKANALSAELNPKLHGCPQDFTPNAMDPPVWKGVLSEPIWTERDKIVNFSQF Sbjct: 61 GDDVEDVLAKANALSAELNPKLHGCPQDFTPNAMDPPVWKGVLSEPIWTERDKIVNFSQF 120 Query: 121 GICRLSGTATITTPDPYGYATGAPVTLTAPGTLALVGTGNTSFYPPVEFRGLNSAQSFV 180 GICRLSGTATITTPDPYGYATGAPVTLTAPGTLALV TGNTSFYPPVEFRGLNSAQSFV Sbjct: 121 GICRLSGTATITTPDPYGYATGAPVTLTAPGTLALVSTGNTSFYPPVEFRGLNSAQSFV 180 Query: 181 VDGQVQSGPLTAAQTLVLDFGKMDFYIKTTSTGAKVRNVADRFTNFRLEGVSLDLPVSV 240 VDGQVQSGPLTAAQTLVLDFGKMDFYIKTTSTGAKVRNVADRFTNFRLEGVSLDLPVSV Sbjct: 181 VDGQVQSGPLTAAQTLVLDFGKMDFYIKTTSSGNKVRNVADRFTNFRLEGVSLDLPVSV 240 Query: 241 TAGTFTQAVGRVPSRRI 257 TAGTFTQAVGRVPSRRI Sbjct: 241 TAGTFTQAVGRVPSRRI 257 </pre>

DECISION:	Start retained ATG at 12039bp, confers the largest ORF without excessive overlap. Conserved in homologs, good RBS score.
-----------	--

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																						
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Phage DB :</p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_17, function unknown, 257</td><td>521</td><td>e-148</td></tr><tr><td>Mandalorian_17, minor tail protein, 257</td><td>514</td><td>e-146</td></tr><tr><td>Quartz_17, minor tail protein, 257</td><td>514</td><td>e-145</td></tr><tr><td>Nucci_17, minor tail protein, 257</td><td>513</td><td>e-145</td></tr><tr><td>Yuuy_17, minor tail protein, 257</td><td>410</td><td>e-114</td></tr><tr><td>Thorongil_18, minor tail protein, 257</td><td>405</td><td>e-113</td></tr><tr><td>Thompson_18, minor tail protein, 257</td><td>402</td><td>e-112</td></tr><tr><td>Vispistious_Draft_18, function unknown, 257</td><td>402</td><td>e-112</td></tr><tr><td>ShaiHulud_Draft_18, function unknown, 257</td><td>402</td><td>e-112</td></tr><tr><td>PuppyEggo_18, minor tail protein, 257</td><td>402</td><td>e-112</td></tr><tr><td>AxiPup_18, minor tail protein, 257</td><td>402</td><td>e-112</td></tr><tr><td>Sedgewig_18, minor tail protein, 257</td><td>400</td><td>e-111</td></tr><tr><td>Nattles_18, minor tail protein, 257</td><td>400</td><td>e-111</td></tr><tr><td>Calix_18, minor tail protein, 257</td><td>400</td><td>e-111</td></tr><tr><td>Baines_18, minor tail protein, 257</td><td>400</td><td>e-111</td></tr><tr><td>Alyxandracam_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>SonOfLevi_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Klimt_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Gelo_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Garey24_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Erla_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Chamuel_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>TinSulphur_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>TatarkaPM_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>StirfryIV_Draft_18, function unknown, 257</td><td>399</td><td>e-111</td></tr><tr><td>Stanktossa_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Shee_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Rapheph_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>PhredFlintston_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Phireproof_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>Phiderman_Draft_18, function unknown, 257</td><td>399</td><td>e-111</td></tr><tr><td>Pherferi_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr><tr><td>ParleG_18, minor tail protein, 257</td><td>399</td><td>e-111</td></tr></table> <p>Query: Carostasia_17 (257 letters)</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show details and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <p>0 50 100 150 200 250</p> <p>nr :</p>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_17, function unknown, 257	521	e-148	Mandalorian_17, minor tail protein, 257	514	e-146	Quartz_17, minor tail protein, 257	514	e-145	Nucci_17, minor tail protein, 257	513	e-145	Yuuy_17, minor tail protein, 257	410	e-114	Thorongil_18, minor tail protein, 257	405	e-113	Thompson_18, minor tail protein, 257	402	e-112	Vispistious_Draft_18, function unknown, 257	402	e-112	ShaiHulud_Draft_18, function unknown, 257	402	e-112	PuppyEggo_18, minor tail protein, 257	402	e-112	AxiPup_18, minor tail protein, 257	402	e-112	Sedgewig_18, minor tail protein, 257	400	e-111	Nattles_18, minor tail protein, 257	400	e-111	Calix_18, minor tail protein, 257	400	e-111	Baines_18, minor tail protein, 257	400	e-111	Alyxandracam_18, minor tail protein, 257	399	e-111	SonOfLevi_18, minor tail protein, 257	399	e-111	Klimt_18, minor tail protein, 257	399	e-111	Gelo_18, minor tail protein, 257	399	e-111	Garey24_18, minor tail protein, 257	399	e-111	Erla_18, minor tail protein, 257	399	e-111	Chamuel_18, minor tail protein, 257	399	e-111	TinSulphur_18, minor tail protein, 257	399	e-111	TatarkaPM_18, minor tail protein, 257	399	e-111	StirfryIV_Draft_18, function unknown, 257	399	e-111	Stanktossa_18, minor tail protein, 257	399	e-111	Shee_18, minor tail protein, 257	399	e-111	Rapheph_18, minor tail protein, 257	399	e-111	PhredFlintston_18, minor tail protein, 257	399	e-111	Phireproof_18, minor tail protein, 257	399	e-111	Phiderman_Draft_18, function unknown, 257	399	e-111	Pherferi_18, minor tail protein, 257	399	e-111	ParleG_18, minor tail protein, 257	399	e-111
Sequences producing significant alignments:	Score (bits)	E Value																																																																																																					
Carostasia_Draft_17, function unknown, 257	521	e-148																																																																																																					
Mandalorian_17, minor tail protein, 257	514	e-146																																																																																																					
Quartz_17, minor tail protein, 257	514	e-145																																																																																																					
Nucci_17, minor tail protein, 257	513	e-145																																																																																																					
Yuuy_17, minor tail protein, 257	410	e-114																																																																																																					
Thorongil_18, minor tail protein, 257	405	e-113																																																																																																					
Thompson_18, minor tail protein, 257	402	e-112																																																																																																					
Vispistious_Draft_18, function unknown, 257	402	e-112																																																																																																					
ShaiHulud_Draft_18, function unknown, 257	402	e-112																																																																																																					
PuppyEggo_18, minor tail protein, 257	402	e-112																																																																																																					
AxiPup_18, minor tail protein, 257	402	e-112																																																																																																					
Sedgewig_18, minor tail protein, 257	400	e-111																																																																																																					
Nattles_18, minor tail protein, 257	400	e-111																																																																																																					
Calix_18, minor tail protein, 257	400	e-111																																																																																																					
Baines_18, minor tail protein, 257	400	e-111																																																																																																					
Alyxandracam_18, minor tail protein, 257	399	e-111																																																																																																					
SonOfLevi_18, minor tail protein, 257	399	e-111																																																																																																					
Klimt_18, minor tail protein, 257	399	e-111																																																																																																					
Gelo_18, minor tail protein, 257	399	e-111																																																																																																					
Garey24_18, minor tail protein, 257	399	e-111																																																																																																					
Erla_18, minor tail protein, 257	399	e-111																																																																																																					
Chamuel_18, minor tail protein, 257	399	e-111																																																																																																					
TinSulphur_18, minor tail protein, 257	399	e-111																																																																																																					
TatarkaPM_18, minor tail protein, 257	399	e-111																																																																																																					
StirfryIV_Draft_18, function unknown, 257	399	e-111																																																																																																					
Stanktossa_18, minor tail protein, 257	399	e-111																																																																																																					
Shee_18, minor tail protein, 257	399	e-111																																																																																																					
Rapheph_18, minor tail protein, 257	399	e-111																																																																																																					
PhredFlintston_18, minor tail protein, 257	399	e-111																																																																																																					
Phireproof_18, minor tail protein, 257	399	e-111																																																																																																					
Phiderman_Draft_18, function unknown, 257	399	e-111																																																																																																					
Pherferi_18, minor tail protein, 257	399	e-111																																																																																																					
ParleG_18, minor tail protein, 257	399	e-111																																																																																																					

	<div><div>DescriptionsGraphic SummaryAlignmentsTaxonomy</div><div><div>Never to see the titleClick to show alignmentsShow Conserved Domains</div><div>Alignment Scores<div><div>< 40</div><div>40 - 50</div><div>50 - 80</div><div>80 - 200</div><div>>= 200</div></div></div><div>88 sequences selectedNo putative conserved domains have been detected</div><div><div>Distribution of the top 88 Blast Hits on 88 subject sequences</div></div><table><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Mandelorian)</td><td>Microbacterium phage Mandelorian</td><td>516</td><td>516</td><td>99%</td><td>0.0</td><td>98.83%</td><td>257</td><td>GX013611.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Nucoli)</td><td>Microbacterium phage Nucoli</td><td>515</td><td>515</td><td>99%</td><td>0.0</td><td>98.44%</td><td>257</td><td>GQK02410.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Quartet)</td><td>Microbacterium phage Quartet</td><td>515</td><td>515</td><td>99%</td><td>0.0</td><td>98.44%</td><td>257</td><td>UJ559236.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage YuoY)</td><td>Microbacterium phage YuoY</td><td>411</td><td>411</td><td>99%</td><td>9e-143</td><td>77.52%</td><td>257</td><td>QEP99346.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Thorongil)</td><td>Microbacterium phage Thorongil</td><td>407</td><td>407</td><td>99%</td><td>4e-141</td><td>74.32%</td><td>257</td><td>QEP62419.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Thomassone)</td><td>Microbacterium phage Thomassone</td><td>404</td><td>404</td><td>99%</td><td>3e-140</td><td>73.54%</td><td>257</td><td>UJ552622.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage AutPup)</td><td>Microbacterium phage AutPup</td><td>404</td><td>404</td><td>99%</td><td>4e-140</td><td>73.54%</td><td>257</td><td>ALJ52667.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Sedgewick)</td><td>Microbacterium phage Sedgewick</td><td>402</td><td>402</td><td>99%</td><td>2e-139</td><td>72.76%</td><td>257</td><td>QJ459860.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Batnes)</td><td>Microbacterium phage Batnes</td><td>402</td><td>402</td><td>99%</td><td>4e-139</td><td>73.15%</td><td>257</td><td>ALJ582730.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Alexandracum)</td><td>Microbacterium phage Alexandracum</td><td>401</td><td>401</td><td>99%</td><td>6e-139</td><td>73.15%</td><td>257</td><td>GQF44973.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Martin)</td><td>Microbacterium phage Martin</td><td>401</td><td>401</td><td>99%</td><td>7e-139</td><td>72.37%</td><td>257</td><td>BM020733.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Pinedilla)</td><td>Microbacterium phage Pinedilla</td><td>401</td><td>401</td><td>99%</td><td>1e-138</td><td>72.76%</td><td>257</td><td>AS044829.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Itoai)</td><td>Microbacterium phage Itoai</td><td>401</td><td>401</td><td>99%</td><td>1e-138</td><td>72.37%</td><td>257</td><td>YP_009623182.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage BisRedClifford)</td><td>Microbacterium phage BisRedClifford</td><td>400</td><td>400</td><td>99%</td><td>1e-138</td><td>72.76%</td><td>257</td><td>QKY79261.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Gubbabump)</td><td>Microbacterium phage Gubbabump</td><td>400</td><td>400</td><td>99%</td><td>1e-138</td><td>72.76%</td><td>257</td><td>QOR55953.1</td></tr><tr><td><input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Oktober96)</td><td>Microbacterium phage Oktober96</td><td>400</td><td>400</td><td>99%</td><td>1e-138</td><td>73.15%</td><td>257</td><td>QIN83950.1</td></tr></table></div></div>	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Mandelorian)	Microbacterium phage Mandelorian	516	516	99%	0.0	98.83%	257	GX013611.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Nucoli)	Microbacterium phage Nucoli	515	515	99%	0.0	98.44%	257	GQK02410.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Quartet)	Microbacterium phage Quartet	515	515	99%	0.0	98.44%	257	UJ559236.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage YuoY)	Microbacterium phage YuoY	411	411	99%	9e-143	77.52%	257	QEP99346.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Thorongil)	Microbacterium phage Thorongil	407	407	99%	4e-141	74.32%	257	QEP62419.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Thomassone)	Microbacterium phage Thomassone	404	404	99%	3e-140	73.54%	257	UJ552622.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage AutPup)	Microbacterium phage AutPup	404	404	99%	4e-140	73.54%	257	ALJ52667.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Sedgewick)	Microbacterium phage Sedgewick	402	402	99%	2e-139	72.76%	257	QJ459860.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Batnes)	Microbacterium phage Batnes	402	402	99%	4e-139	73.15%	257	ALJ582730.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Alexandracum)	Microbacterium phage Alexandracum	401	401	99%	6e-139	73.15%	257	GQF44973.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Martin)	Microbacterium phage Martin	401	401	99%	7e-139	72.37%	257	BM020733.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Pinedilla)	Microbacterium phage Pinedilla	401	401	99%	1e-138	72.76%	257	AS044829.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Itoai)	Microbacterium phage Itoai	401	401	99%	1e-138	72.37%	257	YP_009623182.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage BisRedClifford)	Microbacterium phage BisRedClifford	400	400	99%	1e-138	72.76%	257	QKY79261.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Gubbabump)	Microbacterium phage Gubbabump	400	400	99%	1e-138	72.76%	257	QOR55953.1	<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Oktober96)	Microbacterium phage Oktober96	400	400	99%	1e-138	73.15%	257	QIN83950.1
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Mandelorian)	Microbacterium phage Mandelorian	516	516	99%	0.0	98.83%	257	GX013611.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Nucoli)	Microbacterium phage Nucoli	515	515	99%	0.0	98.44%	257	GQK02410.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Quartet)	Microbacterium phage Quartet	515	515	99%	0.0	98.44%	257	UJ559236.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage YuoY)	Microbacterium phage YuoY	411	411	99%	9e-143	77.52%	257	QEP99346.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Thorongil)	Microbacterium phage Thorongil	407	407	99%	4e-141	74.32%	257	QEP62419.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Thomassone)	Microbacterium phage Thomassone	404	404	99%	3e-140	73.54%	257	UJ552622.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage AutPup)	Microbacterium phage AutPup	404	404	99%	4e-140	73.54%	257	ALJ52667.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Sedgewick)	Microbacterium phage Sedgewick	402	402	99%	2e-139	72.76%	257	QJ459860.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Batnes)	Microbacterium phage Batnes	402	402	99%	4e-139	73.15%	257	ALJ582730.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Alexandracum)	Microbacterium phage Alexandracum	401	401	99%	6e-139	73.15%	257	GQF44973.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Martin)	Microbacterium phage Martin	401	401	99%	7e-139	72.37%	257	BM020733.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Pinedilla)	Microbacterium phage Pinedilla	401	401	99%	1e-138	72.76%	257	AS044829.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Itoai)	Microbacterium phage Itoai	401	401	99%	1e-138	72.37%	257	YP_009623182.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage BisRedClifford)	Microbacterium phage BisRedClifford	400	400	99%	1e-138	72.76%	257	QKY79261.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Gubbabump)	Microbacterium phage Gubbabump	400	400	99%	1e-138	72.76%	257	QOR55953.1																																																																																																																																																		
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Oktober96)	Microbacterium phage Oktober96	400	400	99%	1e-138	73.15%	257	QIN83950.1																																																																																																																																																		
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	HHpred trouve des distal tail protein, mais cette fonction n'est pas listée chez SEA-PHAGE. Mais HHPred trouve un subject avec plus de 99% de probabilité pour une minor tail protein.																																																																																																																																																									



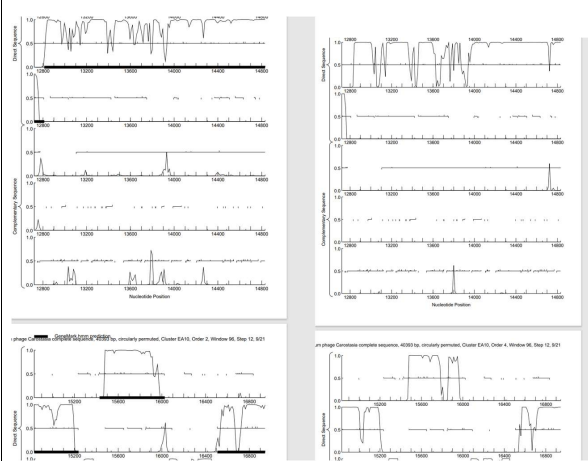
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Gene avant : tape measure protein, gene après : major tail protein

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	18
Coordonnées du Stop	15238
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove 1
Coordonnées du Start retenu	12812
Fonction prédite	<i>minor tail protein.</i>

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	 <p>Coding potential for both, starts well at the start and ends at the stop, but not continuous throughout the orf.</p>

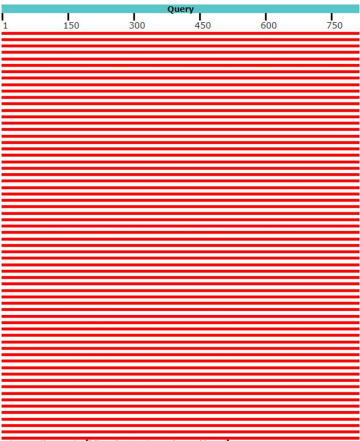
	<p><u>Distribution of 100 Blast Hits on the Query Sequence</u></p> <div><div>88773 ColaCorta_18, minor tail protein, 808. S=1429 E=0.0</div><div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>200</div></div></div><div><div>1c111</div><div><div>0</div><div>100</div><div>200</div><div>300</div><div>400</div><div>500</div><div>600</div><div>700</div><div>800</div></div><div><div></div></div></div><p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p></div>
	<div><div><div>Sequences producing significant alignments:</div><div><div><div>Score</div><div>E</div></div><div><div>(bits)</div><div>Value</div></div></div><div><div>Carostasia_Draft_18, function unknown, 808</div><div>1619</div><div>0.0</div></div><div><div>Quartz_18, minor tail protein, 808</div><div>1617</div><div>0.0</div></div><div><div>Nucci_18, minor tail protein, 808</div><div>1609</div><div>0.0</div></div><div><div>Mandalorian_18, minor tail protein, 808</div><div>1609</div><div>0.0</div></div><div><div>Juanyo_18, minor tail protein, 808</div><div>1494</div><div>0.0</div></div><div><div>YuuV_18, minor tail protein, 808</div><div>1455</div><div>0.0</div></div><div><div>ChikPic_18, minor tail protein, 808</div><div>1432</div><div>0.0</div></div><div><div>Saratos_18, minor tail protein, 808</div><div>1429</div><div>0.0</div></div><div><div>Glamour_17, minor tail protein, 808</div><div>1429</div><div>0.0</div></div><div><div>ColaCorta_18, minor tail protein, 808</div><div>1429</div><div>0.0</div></div><div><div>Andromedas_18, minor tail protein, 808</div><div>1429</div><div>0.0</div></div><div><div>Eleri_18, minor tail protein, 808</div><div>1427</div><div>0.0</div></div><div><div>Zenitsu_18, minor tail protein, 808</div><div>1426</div><div>0.0</div></div><div><div>NCubed_18, minor tail protein, 808</div><div>1426</div><div>0.0</div></div><div><div>Finny_18, minor tail protein, 808</div><div>1426</div><div>0.0</div></div><div><div>Shamu_Draft_19, function unknown, 808</div><div>1424</div><div>0.0</div></div><div><div>Sansa_17, minor tail protein, 808</div><div>1419</div><div>0.0</div></div><div><div>Kavala_19, minor tail protein, 810</div><div>1319</div><div>0.0</div></div><div><div>Theresita_19, minor tail protein, 811</div><div>1295</div><div>0.0</div></div><div><div>Koji_19, minor tail protein, 807</div><div>1288</div><div>0.0</div></div><div><div>BouleyBill_Draft_20, function unknown, 807</div><div>1288</div><div>0.0</div></div><div><div>Lucky3_19, minor tail protein, 807</div><div>1278</div><div>0.0</div></div><div><div>Golden_19, minor tail protein, 807</div><div>1278</div><div>0.0</div></div><div><div>Pherbot_19, minor tail protein, 807</div><div>1276</div><div>0.0</div></div><div><div>Bustleton_19, minor tail protein, 807</div><div>1276</div><div>0.0</div></div><div><div>Sinatra_19, minor tail protein, 807</div><div>1276</div><div>0.0</div></div><div><div>SirVictor_19, minor tail protein, 807</div><div>1275</div><div>0.0</div></div><div><div>Guetzie_19, minor tail protein, 807</div><div>1275</div><div>0.0</div></div><div><div>PrincePhergus_19, minor tail protein, 807</div><div>1273</div><div>0.0</div></div><div><div>Cheeto1_20, minor tail protein, 808</div><div>1266</div><div>0.0</div></div><div><div>ChickenKing_19, minor tail protein, 808</div><div>1259</div><div>0.0</div></div></div><div>Homology with other genes found</div></div>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Overlap -1 therefore acceptable
DECISION:	Yes

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																													
Quel start Glimmer et GeneMark suggèrent-ils ?	12812 for both																																																																																																																																																																																													
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Starts: 50 Selected: 1</div><div>ORF Start : 13667 ORF Stop : 15238 ORF Length : 1572</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 90.2 58.7 53.1 429 3' End 59.3 48.6 87.1 2427</div><div>SD Scoring Matrix: <div>Kabier6</div><div>Explore</div></div><div>Spacing Weight Matrix: <div>Kalin Medium</div><div>Document</div></div></div> <table><tr><th>Stk</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,802</td><td>2,020</td><td>13</td><td>-4,848</td><td>CGTGGAGGGGATAGATCTGT</td><td>GTG</td><td>12893</td><td>2856</td></tr><tr><td>2</td><td>-3,061</td><td>2,367</td><td>10</td><td>-3,756</td><td>GTGTCCCTTCCCGAGGATCTG</td><td>ATG</td><td>12812</td><td>2427</td></tr><tr><td>3</td><td>-6,213</td><td>0,892</td><td>13</td><td>-7,259</td><td>CCTCACCCAGCTCTGTCTCCGC</td><td>ATG</td><td>12896</td><td>2343</td></tr><tr><td>4</td><td>-3,456</td><td>2,163</td><td>17</td><td>-5,456</td><td>CGATGGGGTACTGGCTGGGG</td><td>ATG</td><td>13034</td><td>2205</td></tr><tr><td>5</td><td>-3,952</td><td>1,950</td><td>16</td><td>-5,747</td><td>CTGTGAGAGAGCTCATCCGT</td><td>GTG</td><td>13073</td><td>2166</td></tr><tr><td>6</td><td>-2,509</td><td>2,428</td><td>9</td><td>-3,277</td><td>GTTCCTCCTTCAGAGAGCTG</td><td>ATG</td><td>13115</td><td>2124</td></tr><tr><td>7</td><td>-2,645</td><td>2,551</td><td>8</td><td>-3,891</td><td>GCTGACCTTCAGTGGACCCG</td><td>GTG</td><td>13163</td><td>2076</td></tr><tr><td>8</td><td>-4,853</td><td>1,528</td><td>8</td><td>-6,075</td><td>CACCCAGCTCATCAGCGATGG</td><td>ATG</td><td>13289</td><td>1950</td></tr><tr><td>9</td><td>-2,931</td><td>2,428</td><td>14</td><td>-4,278</td><td>GTGTACTGATCGCGCGGTG</td><td>ATG</td><td>13337</td><td>1902</td></tr><tr><td>10</td><td>-4,775</td><td>1,565</td><td>7</td><td>-6,297</td><td>CCTCAGCTCACACGGGGCACG</td><td>GTG</td><td>13394</td><td>1845</td></tr><tr><td>11</td><td>-3,387</td><td>2,214</td><td>15</td><td>-4,989</td><td>TGCTCTGTACTGTCTGCCAAC</td><td>ATG</td><td>13424</td><td>1815</td></tr><tr><td>12</td><td>-4,463</td><td>1,711</td><td>7</td><td>-5,986</td><td>CGTGGACCGCTGACGAGACC</td><td>GTG</td><td>13505</td><td>1734</td></tr><tr><td>13</td><td>-4,463</td><td>1,711</td><td>10</td><td>-6,158</td><td>TGCGACCTGACGAGACCTG</td><td>GTG</td><td>13508</td><td>1731</td></tr><tr><td>14</td><td>-3,841</td><td>1,993</td><td>16</td><td>-5,457</td><td>GACCTGTGTGTGTGCTCCCTC</td><td>GTG</td><td>13523</td><td>1716</td></tr><tr><td>15</td><td>-5,650</td><td>1,155</td><td>13</td><td>-6,495</td><td>CGGTGAGGGCTACCTGCTGT</td><td>GTG</td><td>13646</td><td>1593</td></tr><tr><td>16</td><td>-5,650</td><td>1,155</td><td>16</td><td>-7,446</td><td>GTGAGCGCTACCTGCTGTGT</td><td>GTG</td><td>13649</td><td>1590</td></tr><tr><td>17</td><td>-6,463</td><td>0,775</td><td>11</td><td>-7,220</td><td>TGTGTGGCCGACGACCTGGC</td><td>ATG</td><td>13667</td><td>1572</td></tr><tr><td>18</td><td>-6,304</td><td>0,849</td><td>10</td><td>-6,999</td><td>CCTCACTGGACCGACAGCTTC</td><td>GTG</td><td>13853</td><td>1386</td></tr><tr><td>19</td><td>-3,800</td><td>2,021</td><td>13</td><td>-4,846</td><td>GTCTCCTGGACGCTCTGCTG</td><td>GTG</td><td>13946</td><td>1293</td></tr></table> <p>The predicted start has the best RBS score: -3.756</p>	Stk	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,802	2,020	13	-4,848	CGTGGAGGGGATAGATCTGT	GTG	12893	2856	2	-3,061	2,367	10	-3,756	GTGTCCCTTCCCGAGGATCTG	ATG	12812	2427	3	-6,213	0,892	13	-7,259	CCTCACCCAGCTCTGTCTCCGC	ATG	12896	2343	4	-3,456	2,163	17	-5,456	CGATGGGGTACTGGCTGGGG	ATG	13034	2205	5	-3,952	1,950	16	-5,747	CTGTGAGAGAGCTCATCCGT	GTG	13073	2166	6	-2,509	2,428	9	-3,277	GTTCCTCCTTCAGAGAGCTG	ATG	13115	2124	7	-2,645	2,551	8	-3,891	GCTGACCTTCAGTGGACCCG	GTG	13163	2076	8	-4,853	1,528	8	-6,075	CACCCAGCTCATCAGCGATGG	ATG	13289	1950	9	-2,931	2,428	14	-4,278	GTGTACTGATCGCGCGGTG	ATG	13337	1902	10	-4,775	1,565	7	-6,297	CCTCAGCTCACACGGGGCACG	GTG	13394	1845	11	-3,387	2,214	15	-4,989	TGCTCTGTACTGTCTGCCAAC	ATG	13424	1815	12	-4,463	1,711	7	-5,986	CGTGGACCGCTGACGAGACC	GTG	13505	1734	13	-4,463	1,711	10	-6,158	TGCGACCTGACGAGACCTG	GTG	13508	1731	14	-3,841	1,993	16	-5,457	GACCTGTGTGTGTGCTCCCTC	GTG	13523	1716	15	-5,650	1,155	13	-6,495	CGGTGAGGGCTACCTGCTGT	GTG	13646	1593	16	-5,650	1,155	16	-7,446	GTGAGCGCTACCTGCTGTGT	GTG	13649	1590	17	-6,463	0,775	11	-7,220	TGTGTGGCCGACGACCTGGC	ATG	13667	1572	18	-6,304	0,849	10	-6,999	CCTCACTGGACCGACAGCTTC	GTG	13853	1386	19	-3,800	2,021	13	-4,846	GTCTCCTGGACGCTCTGCTG	GTG	13946	1293
Stk	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																						
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																						
1	-3,802	2,020	13	-4,848	CGTGGAGGGGATAGATCTGT	GTG	12893	2856																																																																																																																																																																																						
2	-3,061	2,367	10	-3,756	GTGTCCCTTCCCGAGGATCTG	ATG	12812	2427																																																																																																																																																																																						
3	-6,213	0,892	13	-7,259	CCTCACCCAGCTCTGTCTCCGC	ATG	12896	2343																																																																																																																																																																																						
4	-3,456	2,163	17	-5,456	CGATGGGGTACTGGCTGGGG	ATG	13034	2205																																																																																																																																																																																						
5	-3,952	1,950	16	-5,747	CTGTGAGAGAGCTCATCCGT	GTG	13073	2166																																																																																																																																																																																						
6	-2,509	2,428	9	-3,277	GTTCCTCCTTCAGAGAGCTG	ATG	13115	2124																																																																																																																																																																																						
7	-2,645	2,551	8	-3,891	GCTGACCTTCAGTGGACCCG	GTG	13163	2076																																																																																																																																																																																						
8	-4,853	1,528	8	-6,075	CACCCAGCTCATCAGCGATGG	ATG	13289	1950																																																																																																																																																																																						
9	-2,931	2,428	14	-4,278	GTGTACTGATCGCGCGGTG	ATG	13337	1902																																																																																																																																																																																						
10	-4,775	1,565	7	-6,297	CCTCAGCTCACACGGGGCACG	GTG	13394	1845																																																																																																																																																																																						
11	-3,387	2,214	15	-4,989	TGCTCTGTACTGTCTGCCAAC	ATG	13424	1815																																																																																																																																																																																						
12	-4,463	1,711	7	-5,986	CGTGGACCGCTGACGAGACC	GTG	13505	1734																																																																																																																																																																																						
13	-4,463	1,711	10	-6,158	TGCGACCTGACGAGACCTG	GTG	13508	1731																																																																																																																																																																																						
14	-3,841	1,993	16	-5,457	GACCTGTGTGTGTGCTCCCTC	GTG	13523	1716																																																																																																																																																																																						
15	-5,650	1,155	13	-6,495	CGGTGAGGGCTACCTGCTGT	GTG	13646	1593																																																																																																																																																																																						
16	-5,650	1,155	16	-7,446	GTGAGCGCTACCTGCTGTGT	GTG	13649	1590																																																																																																																																																																																						
17	-6,463	0,775	11	-7,220	TGTGTGGCCGACGACCTGGC	ATG	13667	1572																																																																																																																																																																																						
18	-6,304	0,849	10	-6,999	CCTCACTGGACCGACAGCTTC	GTG	13853	1386																																																																																																																																																																																						
19	-3,800	2,021	13	-4,846	GTCTCCTGGACGCTCTGCTG	GTG	13946	1293																																																																																																																																																																																						
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	With the start at 12812, the orf is 2427 bp with an overlap of -1.																																																																																																																																																																																													
Est-ce que le start est conservé chez les homologues voir Starterator ?	<div>Gene: Carostasia_18 Start: 12812, Stop: 15238, Start Num: 2</div> <div>Candidate Starts for Carostasia_18:</div> <div>Identical start with starterator</div>																																																																																																																																																																																													

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?

Distribution of the top 104 Blast Hits on 100 subject sequences



minor tail protein [Microbacterium phage Nucci]

Sequence ID: [QDK02411.1](#) Length: 808 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 808 [GenPept](#) [Graphics](#)

[View Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1634 bits(4231)	0.0	Compositional matrix adjust.	800/808(99%)	806/808(99%)	0/808(0%)
Query 1	MAEYSRVTSRGWITGENDPDVATQQLLRNDSFVGHENIRPTDSSGKORTAIAINVGSGT	60			
Subject 1	MAEYSRVTSRGWITGENDPDVATQQLLRNDSFVGHENIRPTDSSGKORTAIAINVGSGT	60			
Query 61	SATAVASRIGTALRMIEAGVSEKTSIRVANDTLFPSSGVHGTWYKGLDRQNPVLS	120			
Subject 61	SATAVASRIGTALRMIEAGVSEKTSIRVANDTLFPSSGVHGTWYKGLDRNPVLS	120			
Query 121	TRGGAEPLFHIAVNTSANGPQWYRFYSASGTQLSEMDTLLQADENYHIGAVDQLTL	180			
Subject 121	TRGGAEPLFHIAVNTSANGPQWYRFYSASGTQLSEMDTLLQADENYHIGAVDQLTL	180			
Query 181	GTUSVYTVSLTTGTVESRSCTAANPNAACANILEVGVPDEYSRATVDETVVAPFVGN	240			
Subject 181	GTUSVYTVSLTTGTVESRSCTAANPNAACANILEVGVPDEYSRATVDETVVAPFVGN	240			
Query 241	AGNALRIIRLANGALEATQADHTVSGRIAPSTATLPVVAQTRAMPQWGSOTRQITIN	300			
Subject 241	AGNALRIIRLANGALEATQADHTVSGRIAPSTATLPVVAQTRAMPQWGSOTRQITIN	300			
Query 301	QGGATVRYRTSAILTTNSANKNASQIGTEPHIAHIQVEVTLTATDSFVDDILLSTAPPV	360			
Subject 301	QGGATVRYRTSAILTTNSANKNASQIGTEPHIAHIQVEVTLTATDSFVDDILLSTAPPV	360			
Query 361	PPLPSTRVQVPSLQPLLVPTGGVLLQDTLLSCVTROTDSHISTLTFSLSLADSKATQ	420			
Subject 361	PPLPSTRVQVPSLQPLLVPTGGVLLQDTLLSCVTROTDSHISTLTFSLSLADSKATQ	420			
Query 421	TEAEPVVFVKGRRHVARIGITTTKSRNTHVEVYECRMVYDLYAGQIDAGTHTTTAAGAH	480			
Subject 421	TEAEPVVFVKGRRHVARIGITTTKSRNTHVEVYECRMVYDLYAGQIDAGTHTTTAAGAH	480			
Query 481	TSLLADSGVYGVQVPTVEVLGHEHSGTVLQVVKQIASVVGDELVDVNHKFMHLSQGG	540			
Subject 481	TSLLADSGVYGVQVPTVEVLGHEHSGTVLQVVKQIASVVGDELVDVNHKFMHLSQGG	540			
Query 541	RDRGTYFSEYKIGSSARKREDTILVTRYGRHAEGLTIAPAHNGVPPVEDFTWTHVEVE	600			
Subject 541	RDRGTYFSEYKIGSSARKREDTILVTRYGRHAEGLTIAPAHNGVPPVEDFTWTHVEVE	600			
Query 601	STYDFRSQHTPQNLRFQAFIVGRSQPSISEYEYAGLVORVAEVDREVLDFVDFVIDE	660			
Subject 601	STYDFRSQHTPQNLRFQAFIVGRSQPSISEYEYAGLVORVAEVDREVLDFVDFVIDE	660			
Query 661	DYDQSVKSRVQLDIDNADLRKSKITLSMKLRSLASSDOSTPGALTGGQITDTRDINPF	720			
Subject 661	DYDQSVKSRVQLDIDNADLRKSKITLSMKLRSLASSDOSTPGALTGGQITDTRDINPF	720			
Query 721	NILLVNSRGDNGHNAHAGSQVVEGATGRYSYAFGANIGTLEQTVASDRENFIFSAQV	780			
Subject 721	NILLVNSRGDNGHNAHAGSQVVEGATGRYSYAFGANIGTLEQTVASDRENFIFSAQV	780			
Query 781	DADSESSKVEVTFQYTDGTSETQLLEL	808			
Subject 781	DADSESSKVEVTFQYTDGTSETQLLEL	808			

DECISION:

Start retained at 12812 because homologous, good RBS score, and it corresponds to the coding potential for the 2 genemarks.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Rationnelle

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :

>Quartz_18, minor tail protein, 808
Length = 808

Score = 1617 bits (4186), Expect = 0.0
Identities = 806/808 (99%), Positives = 807/808 (99%)

Query: 1 MAEYSRVTSRGDWTGEWDPAVTQLLRMDSFVGNENNRPTDSSGKDRTAIAINVGSGT 60
MAEYSRVTSRGDWTGEWDPAVTQLLRMDSFVGNENNRPTDSSGKDRTAIAINVGSGT
Sbjct: 1 MAEYSRVTSRGDWTGEWDPAVTQLLRMDSFVGNENNRPTDSSGKDRTAIAINVGSGT 60

Query: 61 SAIVASRWGTALRMNEAGVSEKTSIRVANDGTLFPSSGVSVMGTYKGLDRQWNPVLS 120
SAIVASRWGTALRMNEAGVSEKTSIRVANDGTLFPSSGVSVMGTYKGLDRQWNPVLS
Sbjct: 61 SAVAVASRWGTALRMNEAGVSEKTSIRVANDGTLFPSSGVSVMGTYKGLDRQWNPVLS 120

Query: 121 TRGGAENPLFHIAVNTSANGPQVNYRFYSASGTQLISEWMDTLQADEWYHIGAVMDLTL 180
TRGGAENPLFHIAVNTSANGPQVNYRFYSASGTQLISEWMDTLQADEWYHIGAVMDLTL
Sbjct: 121 TRGGAENPLFHIAVNTSANGPQVNYRFYSASGTQLISEWMDTLQADEWYHIGAVMDLTL 180

Query: 181 GTHSVYVTSLTGTGVSRSAGTAANNNAECTANLEVGYPDPPEYSRATVDETWWAPFVGN 240
GTHSVY VSLTTGTGVSRSAGTAANNNAECTANLEVGYPDPPEYSRATVDETWWAPFVGN
Sbjct: 181 GTHSVYCVSLTGTGVSRSAGTAANNNAECTANLEVGYPDPPEYSRATVDETWWAPFVGN 240

Query: 241 AGQWALNRRLANGALEATQADTNTVSGRIAPRSTATLPVVAQTRAMPQWGSQDTPQITIN 300
AGQWALNRRLANGALEATQADTNTVSGRIAPRSTATLPVVAQTRAMPQWGSQDTPQITIN
Sbjct: 241 AGQWALNRRLANGALEATQADTNTVSGRIAPRSTATLPVVAQTRAMPQWGSQDTPQITIN 300

Query: 301 GQGATVRYRTSANLTTWSAHKNASQIGTEPHAAHIQYEVTLTATDSFVDDILLSTAPPVT 360
GQGATVRYRTSANLTTWSAHKNASQIGTEPHAAHIQYEVTLTATDSFVDDILLSTAPPVT
Sbjct: 301 GQGATVRYRTSANLTTWSAHKNASQIGTEPHAAHIQYEVTLTATDSFVDDILLSTAPPVT 360

Query: 361 PPLPSTRVQPFSLDPLLVPTGGGVLQDTLLSCVTRDTSNESTLTFSLSLADSKATQ 420
PPLPSTRVQPFSLDPLLVPTGGGVLQDTLLSCVTRDTSNESTLTFSLSLADSKATQ
Sbjct: 361 PPLPSTRVQPFSLDPLLVPTGGGVLQDTLLSCVTRDTSNESTLTFSLSLADSKATQ 420

Query: 421 IEAEMPVVFGRHYVARGITTTKSRNTHEVEVYCERNWYDLLYAGQIDAQTHTTTAAGAH 480
IEAEMPVVFGRHYVARGITTTKSRNTHEVEVYCERNWYDLLYAGQIDAQTHTTTAAGAH
Sbjct: 421 IEAEMPVVFGRHYVARGITTTKSRNTHEVEVYCERNWYDLLYAGQIDAQTHTTTAAGAH 480

Query: 481 TSLLADSGHYVGQVDPTEVLGHENESGTVLGVLKQIASVYGGDLVFDVNNKFVHLLSQGG 540
TSLLADSGHYVGQVDPTEVLGHENESGTVLGVLKQIASVYGGDLVFDVNNKFVHLLSQGG
Sbjct: 481 TSLLADSGHYVGQVDPTEVLGHENESGTVLGVLKQIASVYGGDLVFDVNNKFVHLLSQGG 540

Query: 541 RDRGTFSYEKGISSARKREDTTNLVTRIYGRNAEGLTIAPANNQVPYVEDFTHTNEVRE 600
RDRGTFSYEKGISSARKREDTTNLVTRIYGRNAEGLTIAPANNQVPYVEDFTHTNEVRE
Sbjct: 541 RDRGTFSYEKGISSARKREDTTNLVTRIYGRNAEGLTIAPANNQVPYVEDFTHTNEVRE 600

Query: 601 STYDFRSGMTQAMRLFLQAFIVGRSQPSISYEYEVAGLVDRVAEVDREVLDFVFNDE 660
STYDFRSGMTQAMRLFLQAFIVGRSQPSISYEYEVAGLVDRVAEVDREVLDFVFNDE
Sbjct: 601 STYDFRSGMTQAMRLFLQAFIVGRSQPSISYEYEVAGLVDRVAEVDREVLDFVFNDE 660

Query: 661 DYDQSVKSRVVQLDIDWDLRKSKITLSNKLRLASSDDSDPGALTGGQIDTRDINPF 720
DYDQSVKSRVVQLDIDWDLRKSKITLSNKLRLASSDDSDPGALTGGQIDTRDINPF
Sbjct: 661 DYDQSVKSRVVQLDIDWDLRKSKITLSNKLRLASSDDSDPGALTGGQIDTRDINPF 720

Query: 721 NLLVNSRGDNGMAHJAGSGVQVVEGGATGRYSFAGANGTLEQTVASDNRENFIISAQV 780
NLLVNSRGDNGMAHJAGSGVQVVEGGATGRYSFAGANGTLEQTVASDNRENFIISAQV
Sbjct: 721 NLLVNSRGDNGMAHJAGSGVQVVEGGATGRYSFAGANGTLEQTVASDNRENFIISAQV 780

Query: 781 DADSESSVKVEVTFQYDGTSETQILEL 808
DADSESSVKVEVTFQYDGTSETQILEL
Sbjct: 781 DADSESSVKVEVTFQYDGTSETQILEL 808

*** nr :

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Quartz)	Microbacterium phage Quartz	1642	1642	99%	0.0	99.75%	808	U959237.1
<input checked="" type="checkbox"/> minor tail protein (Microbacterium phage Nucoi)	Microbacterium phage Nucoi	1634	1634	99%	0.0	99.01%	808	Q959241.1

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<p>UniProt :</p> <p>Template alignment</p> <p><input type="checkbox"/> 1. Q0PDK6 FIBER_BP5PP Tail spike protein OS=Bacillus phage SPP1 OX=10724 GN=21 PE=1 SV=1 Probability: 99.94%, E-value: 1.2e-23, Score: 266.28, Aligned cols: 316, Identities: 15%, Similarity: 0.154, Template Neff: 9.1</p> <p>PDB :</p> <p>Template alignment Template 3D Structure PDBe</p> <p><input type="checkbox"/> 2. 6V81_CE Tail-Associated Lysin, gp59: phage tail, tail tip, tape measure protein, VIRAL PROTEIN; 3.7A (Staphylococcus virus 80alpha) Probability: 99.93%, E-value: 2.1e-23, Score: 249.67, Aligned cols: 309, Identities: 14%, Similarity: 0.08, Template Neff: 11.4</p> <p>Pfam :</p> <p>Template alignment CDD</p> <p><input type="checkbox"/> 4. PF06605.15 : Prophage_tail ; Prophage endopeptidase tail Probability: 99.69%, E-value: 5.3e-15, Score: 156.06, Aligned cols: 234, Identities: 15%, Similarity: 0.142, Template Neff: 12</p> <p>CD :</p> <p>Template alignment</p> <p><input type="checkbox"/> cd00152 PTX; Pentraxins are plasma proteins characterized by their pentameric discoid assembly and their Ca2+ dependent ligand binding. 37, such as Serum amyloid P component (SAP) and C-reactive Protein (CRP), which are cytokine-inducible acute-phase proteins implicated in innate immunity. Probability: 98.69%, E-value: 0.0000016, Score: 85.56, Aligned cols: 190, Identities: 9%, Similarity: -0.001, Template Neff: 13.9</p>		
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Proche de la tape measure protein et de la tail assembly chaperone		
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	Non		
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<div>Oui, minor tail protein</div> <table><tr><td><div>minor tail protein</div></td><td><div>tail fiber-like protein, collagen-like, glycine rich</div></td></tr></table>	<div>minor tail protein</div>	<div>tail fiber-like protein, collagen-like, glycine rich</div>
<div>minor tail protein</div>	<div>tail fiber-like protein, collagen-like, glycine rich</div>		
DECISION:	Minor tail protein		

Student Gene Annotation Worksheet

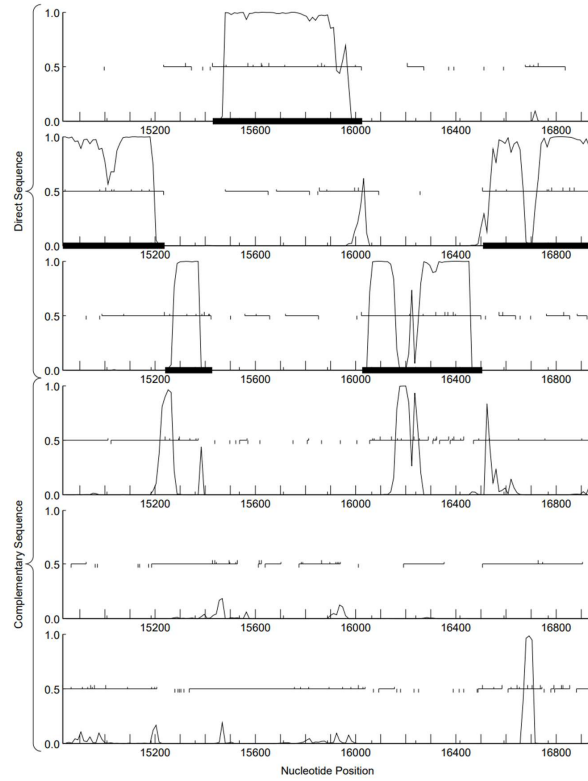
Carosmatique-gene-19

Basic Phage Information	
Nom du Phage	carostasia
Gène #	19
Coordonnées du Stop	15428
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 1
Coordonnées du Start retenu	15240
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	

GenMark hmm prediction
erium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 9/21



Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																												
Quel start Glimmer et GeneMark suggèrent-ils ?	15240																																																																																																												
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Start associé à un RBS</div> <div><div>Starts: 10 Selected: 1</div><div>ORF Start : 15261 ORF Stop : 15428 ORF Length: 168</div><div>5' End 58.6 3' End 50.4</div><div>Cdn1 75.9 Cdn2 69.0 Cdn3 87</div><div>Length 351</div><div>SD Scoring Matrix</div><div>SD Scoring Matrix Kibler6</div><div>Explore</div><div>Spacing Weight Matrix</div><div>Spacing Weight Matrix Karlin Medium</div><div>Document</div></div> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,697</td><td>2,069</td><td>11</td><td>-4,454</td><td>ITCAACCTGCTGGTGAACAGCC</td><td>GTG</td><td>14991</td><td>438</td></tr><tr><td>2</td><td>-5,344</td><td>1,298</td><td>12</td><td>-6,180</td><td>GGCCGGTACAGCTTGGCGTTGC</td><td>GTG</td><td>15078</td><td>351</td></tr><tr><td>3</td><td>-2,279</td><td>2,733</td><td>11</td><td>-3,036</td><td>ACCCAGATTCTGGAGCTGTGAC</td><td>ATG</td><td>15240</td><td>189</td></tr><tr><td>4</td><td>-7,061</td><td>0,495</td><td>11</td><td>-7,818</td><td>CATGCCCTCCCGGCCCCACC</td><td>GTG</td><td>15261</td><td>168</td></tr><tr><td>5</td><td>-5,412</td><td>1,267</td><td>10</td><td>-6,107</td><td>CCTCCAGACTACCGATGCCGAC</td><td>GTG</td><td>15330</td><td>99</td></tr><tr><td>6</td><td>-4,654</td><td>1,622</td><td>10</td><td>-5,348</td><td>CCTGATCCTCCAGCTGGCACC</td><td>GTG</td><td>15366</td><td>63</td></tr><tr><td>7</td><td>-4,283</td><td>1,795</td><td>11</td><td>-5,040</td><td>GGGCGCGGCTGGGTTCCCAAC</td><td>GTG</td><td>15390</td><td>39</td></tr><tr><td>8</td><td>-5,046</td><td>1,438</td><td>10</td><td>-5,741</td><td>CTGGGTTCCCAAGTGAACGAG</td><td>ATG</td><td>15399</td><td>30</td></tr><tr><td>9</td><td>-3,800</td><td>2,021</td><td>11</td><td>-4,557</td><td>CGAGATGCCCTGGACCCCTGGG</td><td>GTG</td><td>15417</td><td>12</td></tr><tr><td>10</td><td>-3,800</td><td>2,021</td><td>14</td><td>-5,147</td><td>GATGCCCTGGACCGCTGGGGTG</td><td>GTG</td><td>15420</td><td>9</td></tr></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,697	2,069	11	-4,454	ITCAACCTGCTGGTGAACAGCC	GTG	14991	438	2	-5,344	1,298	12	-6,180	GGCCGGTACAGCTTGGCGTTGC	GTG	15078	351	3	-2,279	2,733	11	-3,036	ACCCAGATTCTGGAGCTGTGAC	ATG	15240	189	4	-7,061	0,495	11	-7,818	CATGCCCTCCCGGCCCCACC	GTG	15261	168	5	-5,412	1,267	10	-6,107	CCTCCAGACTACCGATGCCGAC	GTG	15330	99	6	-4,654	1,622	10	-5,348	CCTGATCCTCCAGCTGGCACC	GTG	15366	63	7	-4,283	1,795	11	-5,040	GGGCGCGGCTGGGTTCCCAAC	GTG	15390	39	8	-5,046	1,438	10	-5,741	CTGGGTTCCCAAGTGAACGAG	ATG	15399	30	9	-3,800	2,021	11	-4,557	CGAGATGCCCTGGACCCCTGGG	GTG	15417	12	10	-3,800	2,021	14	-5,147	GATGCCCTGGACCGCTGGGGTG	GTG	15420	9
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																					
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																					
1	-3,697	2,069	11	-4,454	ITCAACCTGCTGGTGAACAGCC	GTG	14991	438																																																																																																					
2	-5,344	1,298	12	-6,180	GGCCGGTACAGCTTGGCGTTGC	GTG	15078	351																																																																																																					
3	-2,279	2,733	11	-3,036	ACCCAGATTCTGGAGCTGTGAC	ATG	15240	189																																																																																																					
4	-7,061	0,495	11	-7,818	CATGCCCTCCCGGCCCCACC	GTG	15261	168																																																																																																					
5	-5,412	1,267	10	-6,107	CCTCCAGACTACCGATGCCGAC	GTG	15330	99																																																																																																					
6	-4,654	1,622	10	-5,348	CCTGATCCTCCAGCTGGCACC	GTG	15366	63																																																																																																					
7	-4,283	1,795	11	-5,040	GGGCGCGGCTGGGTTCCCAAC	GTG	15390	39																																																																																																					
8	-5,046	1,438	10	-5,741	CTGGGTTCCCAAGTGAACGAG	ATG	15399	30																																																																																																					
9	-3,800	2,021	11	-4,557	CGAGATGCCCTGGACCCCTGGG	GTG	15417	12																																																																																																					
10	-3,800	2,021	14	-5,147	GATGCCCTGGACCGCTGGGGTG	GTG	15420	9																																																																																																					
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Meilleur score : -3.036 Plus long ORF avec un ATG. Les autres orf (1,2 et 3) conduisent à un chevauchement excessif avec le gène precedent																																																																																																												
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																																													
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div><div>Download ▾ GenPept Graphics</div><div>Next ▾ Previous ◀ Descriptions</div></div> <p>hypothetical protein SEA_NUCCL_19 [Microbacterium phage Nucleo]</p> <p>Sequence ID: QDK02412.1 Length: 62 Number of Matches: 1</p> <p>See 1 more title(s) ▾ See all identical Proteins(PG)</p> <div><div>Range 1: 1 to 62</div><div>GenPept Graphics</div><div>Next Match ▾ Previous Match</div></div> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>124 bits(310)</td><td>3e-35</td><td>Compositional matrix adjust.</td><td>62/62(100%)</td><td>62/62(100%)</td><td>0/62(0%)</td></tr></table> <p>Query 1 NPVRATVRNPFSKQTLTIKVSQTTDAQVEVTDLLIQAQTVGTGQVPRVTERPMTAGV 68</p> <p>NPVRATVRNPFSKQTLTIKVSQTTDAQVEVTDLLIQAQTVGTGQVPRVTERPMTAGV 68</p> <p>Seqct 1 NPVRATVRNPFSKQTLTIKVSQTTDAQVEVTDLLIQAQTVGTGQVPRVTERPMTAGV 68</p> <p>Query 61 VS 62</p> <p>VS 62</p> <p>Seqct 61 VS 62</p>	Score	Expect	Method	Identities	Positives	Gaps	124 bits(310)	3e-35	Compositional matrix adjust.	62/62(100%)	62/62(100%)	0/62(0%)																																																																																																
Score	Expect	Method	Identities	Positives	Gaps																																																																																																								
124 bits(310)	3e-35	Compositional matrix adjust.	62/62(100%)	62/62(100%)	0/62(0%)																																																																																																								
DECISION:	Le start conservé est celui à la position 15240 car il n'est pas en contradiction avec les principes d'annotation et Il est conservé																																																																																																												

Download ▾ GenPept Graphics

Next ▾ Previous ◀ Descriptions

Range 1: 1 to 62

GenPept Graphics

Next Match ▾ Previous Match

	chez d'autres homologues. De plus, le RBS associé au start présente le meilleur score et le plus long ORF avec un ATG.
--	--

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																						
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Phages DB :</p> <table><thead><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Nucci_19, function unknown, 62</td><td>127</td><td>1e-29</td></tr><tr><td>Mandalorian_19, function unknown, 62</td><td>127</td><td>1e-29</td></tr><tr><td>Carostasia_Draft_19, function unknown, 62</td><td>127</td><td>1e-29</td></tr><tr><td>Quartz_19, function unknown, 62</td><td>125</td><td>4e-29</td></tr><tr><td>Yuuy_19, function unknown, 62</td><td>113</td><td>1e-25</td></tr><tr><td>Juanyo_19, function unknown, 62</td><td>112</td><td>3e-25</td></tr><tr><td>Pikmin_19, function unknown, 62</td><td>108</td><td>4e-24</td></tr><tr><td>Pajaza_19, function unknown, 62</td><td>108</td><td>4e-24</td></tr><tr><td>Casey_19, function unknown, 62</td><td>108</td><td>4e-24</td></tr><tr><td>Theresita_20, function unknown, 62</td><td>108</td><td>5e-24</td></tr><tr><td>SirVictor_20, function unknown, 62</td><td>108</td><td>5e-24</td></tr><tr><td>Lucky3_20, function unknown, 62</td><td>108</td><td>5e-24</td></tr><tr><td>Guetzle_20, function unknown, 62</td><td>108</td><td>5e-24</td></tr><tr><td>Golden_20, function unknown, 62</td><td>108</td><td>5e-24</td></tr><tr><td>ChickenKing_20, function unknown, 62</td><td>103</td><td>1e-22</td></tr><tr><td>Sinatra_20, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>PrincePhergus_20, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>Pherbot_20, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>Bustleton_20, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>BouleyBill_Draft_21, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>Koji_20, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>GaeCeo_21, function unknown, 62</td><td>102</td><td>3e-22</td></tr><tr><td>Cheetol_21, function unknown, 62</td><td>102</td><td>4e-22</td></tr><tr><td>Schubert_20, function unknown, 63</td><td>101</td><td>7e-22</td></tr><tr><td>Zenitsu_19, function unknown, 62</td><td>100</td><td>2e-21</td></tr><tr><td>Shamu_19, function unknown, 62</td><td>100</td><td>2e-21</td></tr><tr><td>Sansa_18, function unknown, 62</td><td>100</td><td>2e-21</td></tr><tr><td>MCubed_19, function unknown, 62</td><td>100</td><td>2e-21</td></tr><tr><td>Finny_19, function unknown, 62</td><td>100</td><td>2e-21</td></tr><tr><td>chikPic_19, function unknown, 62</td><td>100</td><td>2e-21</td></tr><tr><td>Saratos_19, function unknown, 62</td><td>99</td><td>3e-21</td></tr><tr><td>Glamour_18, function unknown, 62</td><td>99</td><td>3e-21</td></tr><tr><td>Eleri_19, function unknown, 62</td><td>99</td><td>3e-21</td></tr></tbody></table> <p>Query= Carostasia_19 (62 letters)</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>nr :</p>	Sequences producing significant alignments:	Score (bits)	E Value	Nucci_19, function unknown, 62	127	1e-29	Mandalorian_19, function unknown, 62	127	1e-29	Carostasia_Draft_19, function unknown, 62	127	1e-29	Quartz_19, function unknown, 62	125	4e-29	Yuuy_19, function unknown, 62	113	1e-25	Juanyo_19, function unknown, 62	112	3e-25	Pikmin_19, function unknown, 62	108	4e-24	Pajaza_19, function unknown, 62	108	4e-24	Casey_19, function unknown, 62	108	4e-24	Theresita_20, function unknown, 62	108	5e-24	SirVictor_20, function unknown, 62	108	5e-24	Lucky3_20, function unknown, 62	108	5e-24	Guetzle_20, function unknown, 62	108	5e-24	Golden_20, function unknown, 62	108	5e-24	ChickenKing_20, function unknown, 62	103	1e-22	Sinatra_20, function unknown, 62	102	3e-22	PrincePhergus_20, function unknown, 62	102	3e-22	Pherbot_20, function unknown, 62	102	3e-22	Bustleton_20, function unknown, 62	102	3e-22	BouleyBill_Draft_21, function unknown, 62	102	3e-22	Koji_20, function unknown, 62	102	3e-22	GaeCeo_21, function unknown, 62	102	3e-22	Cheetol_21, function unknown, 62	102	4e-22	Schubert_20, function unknown, 63	101	7e-22	Zenitsu_19, function unknown, 62	100	2e-21	Shamu_19, function unknown, 62	100	2e-21	Sansa_18, function unknown, 62	100	2e-21	MCubed_19, function unknown, 62	100	2e-21	Finny_19, function unknown, 62	100	2e-21	chikPic_19, function unknown, 62	100	2e-21	Saratos_19, function unknown, 62	99	3e-21	Glamour_18, function unknown, 62	99	3e-21	Eleri_19, function unknown, 62	99	3e-21
Sequences producing significant alignments:	Score (bits)	E Value																																																																																																					
Nucci_19, function unknown, 62	127	1e-29																																																																																																					
Mandalorian_19, function unknown, 62	127	1e-29																																																																																																					
Carostasia_Draft_19, function unknown, 62	127	1e-29																																																																																																					
Quartz_19, function unknown, 62	125	4e-29																																																																																																					
Yuuy_19, function unknown, 62	113	1e-25																																																																																																					
Juanyo_19, function unknown, 62	112	3e-25																																																																																																					
Pikmin_19, function unknown, 62	108	4e-24																																																																																																					
Pajaza_19, function unknown, 62	108	4e-24																																																																																																					
Casey_19, function unknown, 62	108	4e-24																																																																																																					
Theresita_20, function unknown, 62	108	5e-24																																																																																																					
SirVictor_20, function unknown, 62	108	5e-24																																																																																																					
Lucky3_20, function unknown, 62	108	5e-24																																																																																																					
Guetzle_20, function unknown, 62	108	5e-24																																																																																																					
Golden_20, function unknown, 62	108	5e-24																																																																																																					
ChickenKing_20, function unknown, 62	103	1e-22																																																																																																					
Sinatra_20, function unknown, 62	102	3e-22																																																																																																					
PrincePhergus_20, function unknown, 62	102	3e-22																																																																																																					
Pherbot_20, function unknown, 62	102	3e-22																																																																																																					
Bustleton_20, function unknown, 62	102	3e-22																																																																																																					
BouleyBill_Draft_21, function unknown, 62	102	3e-22																																																																																																					
Koji_20, function unknown, 62	102	3e-22																																																																																																					
GaeCeo_21, function unknown, 62	102	3e-22																																																																																																					
Cheetol_21, function unknown, 62	102	4e-22																																																																																																					
Schubert_20, function unknown, 63	101	7e-22																																																																																																					
Zenitsu_19, function unknown, 62	100	2e-21																																																																																																					
Shamu_19, function unknown, 62	100	2e-21																																																																																																					
Sansa_18, function unknown, 62	100	2e-21																																																																																																					
MCubed_19, function unknown, 62	100	2e-21																																																																																																					
Finny_19, function unknown, 62	100	2e-21																																																																																																					
chikPic_19, function unknown, 62	100	2e-21																																																																																																					
Saratos_19, function unknown, 62	99	3e-21																																																																																																					
Glamour_18, function unknown, 62	99	3e-21																																																																																																					
Eleri_19, function unknown, 62	99	3e-21																																																																																																					

hypothetical protein PRD_0CCE008_21
[Mycobacterium pho...]

No putative conserved domains have been detected

Distribution of the top 100 Blast Hits on 100 subject sequences

Query

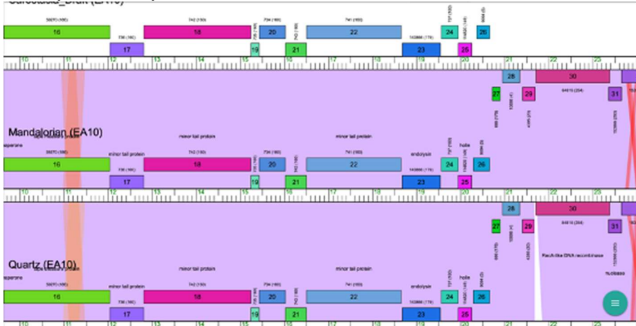
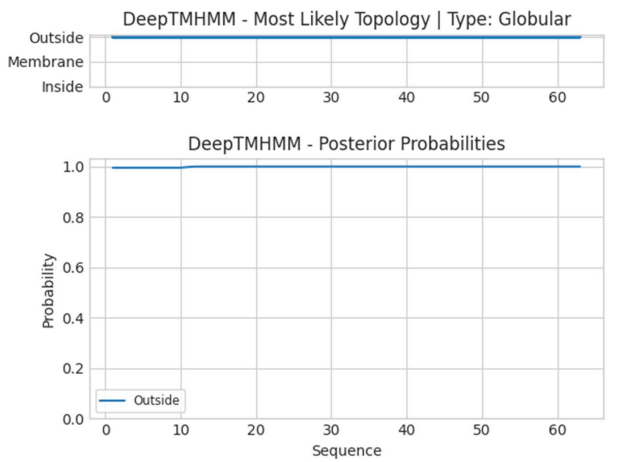
1 10 20 30 40 50 60

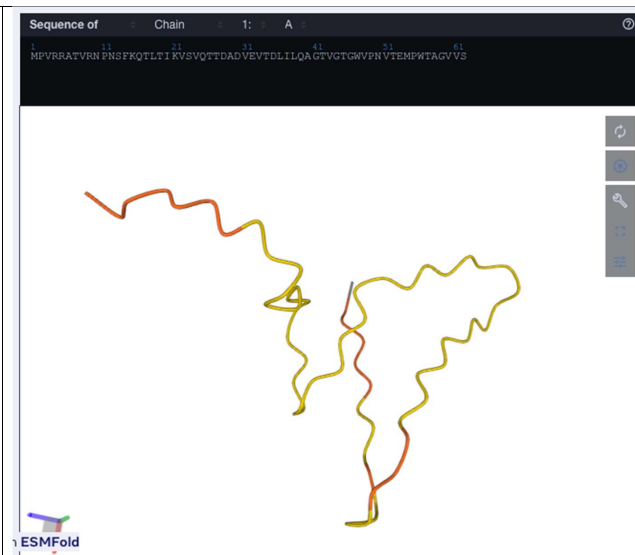
Query cover : au-dessus de 90% pour beaucoup.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?



HHpred trouve deux homologues avec une probabilité >85%.
Ce sont des tail fiber proteins.

<p>Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?</p>	<p>Le gène en aval est une minor tail protein, et celui en amont code pour une protéine de fonction inconnue.</p> 
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>Non, ce serait une protéine globulaire.</p> 
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>OUI.</p>
<p>Structure</p>	<p>D'après ESMfold la protéine est complètement déstructurée, il y a donc peu de chances qu'elle corresponde à une tail fiber protein. Il pourrait s'agir d'un résidu de tail fiber protein.</p>



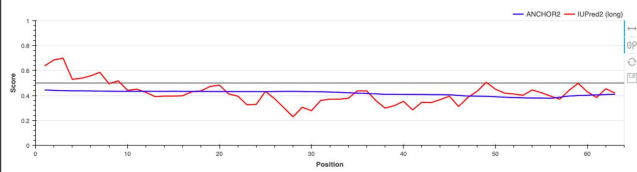
Le gène 20 montre aussi une possible homologie avec une tail fiber protein d'après sa séquence. Mais sa structure est prédite très déstructurée par ESMFold.



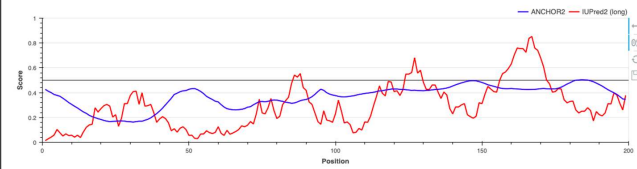
Une prédiction structurale en dimère des protéines des gènes 19 et 20 est aussi très déstructurée.

Néanmoins, d'après IUPred ce résultat serait dû au manque de données chez alpha fold plutôt qu'une vraie déstructuration.

IUPred 19 :



IUPred20 :



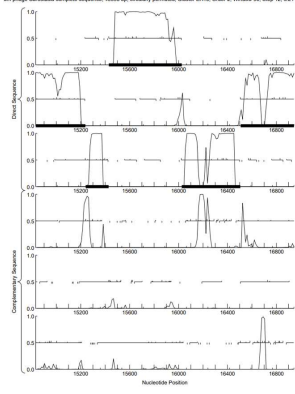
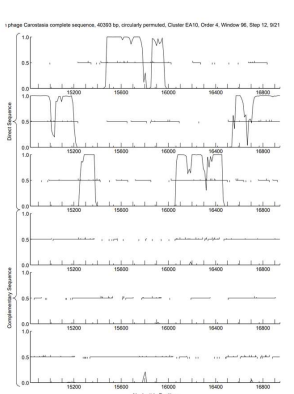
DECISION:

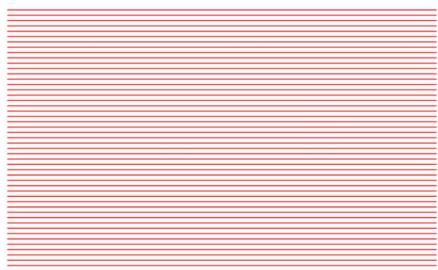
Fonction inconnue, elle pourrait être un reliquat de tail fiber protein.

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	20
Coordonnées du Stop	16026
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 1
Coordonnées du Start retenu	15430
Fonction prédite	Minor tail protein

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	<div><div><p>vir phage Carostasia complete sequence, 40303 bp, circularly permuted, Cluster E410, Order 2, Window 96, Step 12, 921</p></div><div><p>phage Carostasia complete sequence, 40303 bp, circularly permuted, Cluster E410, Order 4, Window 96, Step 12, 921</p></div></div> <p>Potentiel codant correspond</p>

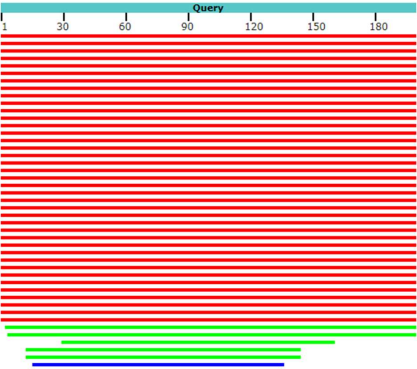
	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <div><p>Mouse-over to show define and scores. Click to show alignments</p><p>Color Key for Alignment Scores</p><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>200</div></div><p>1c111</p><div><div>0</div><div>50</div><div>100</div><div>150</div></div></div>																																																															
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div>Sequences producing significant alignments:</div><table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Nucci_20, function unknown, 198</td><td>415</td><td>e-116</td></tr><tr><td>Mandalorian_20, function unknown, 198</td><td>415</td><td>e-116</td></tr><tr><td>Carostasia_Draft_20, function unknown, 198</td><td>415</td><td>e-116</td></tr><tr><td>Quartz_20, function unknown, 198</td><td>414</td><td>e-116</td></tr><tr><td>Juanyo_20, function unknown, 198</td><td>374</td><td>e-104</td></tr><tr><td>YuuY_20, function unknown, 198</td><td>365</td><td>e-101</td></tr><tr><td>ChickenKing_21, function unknown, 199</td><td>328</td><td>1e-90</td></tr><tr><td>Sinatra_21, function unknown, 199</td><td>328</td><td>4e-90</td></tr><tr><td>PrincePhergus_21, function unknown, 199</td><td>328</td><td>4e-90</td></tr><tr><td>Pherbot_21, function unknown, 199</td><td>328</td><td>4e-90</td></tr><tr><td>Bustleton_21, function unknown, 199</td><td>328</td><td>4e-90</td></tr><tr><td>BouleyBill_Draft_22, function unknown, 199</td><td>328</td><td>4e-90</td></tr><tr><td>Koji_21, function unknown, 199</td><td>322</td><td>1e-88</td></tr><tr><td>Kavala_21, function unknown, 199</td><td>322</td><td>4e-88</td></tr><tr><td>Theresita_21, function unknown, 198</td><td>320</td><td>1e-87</td></tr><tr><td>Zenitsu_20, function unknown, 198</td><td>319</td><td>2e-87</td></tr><tr><td>Lucky3_21, function unknown, 199</td><td>319</td><td>2e-87</td></tr><tr><td>Golden_21, function unknown, 199</td><td>319</td><td>2e-87</td></tr><tr><td>MCubed_20, function unknown, 198</td><td>319</td><td>3e-87</td></tr><tr><td>Finny_20, function unknown, 198</td><td>318</td><td>5e-87</td></tr></tbody></table><p>Homologue retrouvé dans ce gène</p></div>		Score (bits)	E Value	Nucci_20, function unknown, 198	415	e-116	Mandalorian_20, function unknown, 198	415	e-116	Carostasia_Draft_20, function unknown, 198	415	e-116	Quartz_20, function unknown, 198	414	e-116	Juanyo_20, function unknown, 198	374	e-104	YuuY_20, function unknown, 198	365	e-101	ChickenKing_21, function unknown, 199	328	1e-90	Sinatra_21, function unknown, 199	328	4e-90	PrincePhergus_21, function unknown, 199	328	4e-90	Pherbot_21, function unknown, 199	328	4e-90	Bustleton_21, function unknown, 199	328	4e-90	BouleyBill_Draft_22, function unknown, 199	328	4e-90	Koji_21, function unknown, 199	322	1e-88	Kavala_21, function unknown, 199	322	4e-88	Theresita_21, function unknown, 198	320	1e-87	Zenitsu_20, function unknown, 198	319	2e-87	Lucky3_21, function unknown, 199	319	2e-87	Golden_21, function unknown, 199	319	2e-87	MCubed_20, function unknown, 198	319	3e-87	Finny_20, function unknown, 198	318	5e-87
	Score (bits)	E Value																																																														
Nucci_20, function unknown, 198	415	e-116																																																														
Mandalorian_20, function unknown, 198	415	e-116																																																														
Carostasia_Draft_20, function unknown, 198	415	e-116																																																														
Quartz_20, function unknown, 198	414	e-116																																																														
Juanyo_20, function unknown, 198	374	e-104																																																														
YuuY_20, function unknown, 198	365	e-101																																																														
ChickenKing_21, function unknown, 199	328	1e-90																																																														
Sinatra_21, function unknown, 199	328	4e-90																																																														
PrincePhergus_21, function unknown, 199	328	4e-90																																																														
Pherbot_21, function unknown, 199	328	4e-90																																																														
Bustleton_21, function unknown, 199	328	4e-90																																																														
BouleyBill_Draft_22, function unknown, 199	328	4e-90																																																														
Koji_21, function unknown, 199	322	1e-88																																																														
Kavala_21, function unknown, 199	322	4e-88																																																														
Theresita_21, function unknown, 198	320	1e-87																																																														
Zenitsu_20, function unknown, 198	319	2e-87																																																														
Lucky3_21, function unknown, 199	319	2e-87																																																														
Golden_21, function unknown, 199	319	2e-87																																																														
MCubed_20, function unknown, 198	319	3e-87																																																														
Finny_20, function unknown, 198	318	5e-87																																																														
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de conflit, juste un gap de 1 donc acceptable																																																															
DECISION:	Oui																																																															

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																
Quel start Glimmer et GeneMark suggèrent-ils ?	15430 pour les 2																																																																																																																																																
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Stats: 14 ORF Start : 15860 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler6 Selected: 1 ORF Stop : 16026 5' End 53.3 60.0 93.3 45 Spacing Weight Matrix Karlin Medium ORF Length: 177 3' End 65.2 49.5 81.5 552</div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,861</td><td>1,593</td><td>11</td><td>-4,618</td><td>ACCGCTGGGGTGGTGAAGTGAC</td><td>ATG</td><td>15430</td><td>597</td></tr><tr><td>2</td><td>-7,368</td><td>0,351</td><td>11</td><td>-8,125</td><td>CGCTCTCGGCGCGTCCAGCGCG</td><td>GTG</td><td>15475</td><td>552</td></tr><tr><td>3</td><td>-3,967</td><td>1,943</td><td>10</td><td>-4,661</td><td>CGCGTCCAGCGCGTGTCTGGGT</td><td>GTG</td><td>15484</td><td>543</td></tr><tr><td>4</td><td>-3,967</td><td>1,943</td><td>16</td><td>-5,763</td><td>CAAGCGCGTGTCTGGGTGTGGCC</td><td>TTG</td><td>15490</td><td>537</td></tr><tr><td>5</td><td>-4,827</td><td>1,540</td><td>8</td><td>-6,049</td><td>GAGCGTGTCTCTGTGGGCACT</td><td>ATG</td><td>15571</td><td>456</td></tr><tr><td>6</td><td>-5,918</td><td>1,030</td><td>7</td><td>-7,441</td><td>TATGCTGGGCGACTGGGCGCTG</td><td>GTG</td><td>15592</td><td>435</td></tr><tr><td>7</td><td>-4,585</td><td>1,654</td><td>17</td><td>-6,585</td><td>CAAGTGGGCGTACATCAACGGC</td><td>ATG</td><td>15625</td><td>402</td></tr><tr><td>8</td><td>-4,735</td><td>1,584</td><td>6</td><td>-6,479</td><td>GTGGCGCTACATCAACGGCATG</td><td>GTG</td><td>15628</td><td>399</td></tr><tr><td>9</td><td>-4,283</td><td>1,795</td><td>5</td><td>-6,283</td><td>GGCGGACTAGCGGACTGGGTA</td><td>ATG</td><td>15655</td><td>372</td></tr><tr><td>10</td><td>-4,394</td><td>1,743</td><td>13</td><td>-5,440</td><td>GTACCGCACGGGTCCGCAAGG</td><td>GTG</td><td>15718</td><td>309</td></tr><tr><td>11</td><td>-3,288</td><td>2,261</td><td>7</td><td>-4,811</td><td>CCAGGACCAAGCTCTGGACCTG</td><td>GTG</td><td>15850</td><td>177</td></tr><tr><td>12</td><td>-3,741</td><td>2,049</td><td>6</td><td>-5,485</td><td>GGACCTGGTGAATGATGGTATC</td><td>ATG</td><td>15865</td><td>162</td></tr><tr><td>13</td><td>-3,741</td><td>2,049</td><td>18</td><td>-6,042</td><td>TGATGGTATCATGTGGGGATC</td><td>GTG</td><td>15877</td><td>150</td></tr><tr><td>14</td><td>-3,079</td><td>2,358</td><td>13</td><td>-4,125</td><td>CTGGCAAGAGGACTGGGCAGAT</td><td>GTG</td><td>16000</td><td>27</td></tr></table> <p>Start prédit correspond au meilleur score RBS : -4.618. Ce start correspond à un ATG.</p>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,861	1,593	11	-4,618	ACCGCTGGGGTGGTGAAGTGAC	ATG	15430	597	2	-7,368	0,351	11	-8,125	CGCTCTCGGCGCGTCCAGCGCG	GTG	15475	552	3	-3,967	1,943	10	-4,661	CGCGTCCAGCGCGTGTCTGGGT	GTG	15484	543	4	-3,967	1,943	16	-5,763	CAAGCGCGTGTCTGGGTGTGGCC	TTG	15490	537	5	-4,827	1,540	8	-6,049	GAGCGTGTCTCTGTGGGCACT	ATG	15571	456	6	-5,918	1,030	7	-7,441	TATGCTGGGCGACTGGGCGCTG	GTG	15592	435	7	-4,585	1,654	17	-6,585	CAAGTGGGCGTACATCAACGGC	ATG	15625	402	8	-4,735	1,584	6	-6,479	GTGGCGCTACATCAACGGCATG	GTG	15628	399	9	-4,283	1,795	5	-6,283	GGCGGACTAGCGGACTGGGTA	ATG	15655	372	10	-4,394	1,743	13	-5,440	GTACCGCACGGGTCCGCAAGG	GTG	15718	309	11	-3,288	2,261	7	-4,811	CCAGGACCAAGCTCTGGACCTG	GTG	15850	177	12	-3,741	2,049	6	-5,485	GGACCTGGTGAATGATGGTATC	ATG	15865	162	13	-3,741	2,049	18	-6,042	TGATGGTATCATGTGGGGATC	GTG	15877	150	14	-3,079	2,358	13	-4,125	CTGGCAAGAGGACTGGGCAGAT	GTG	16000	27
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																									
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																									
1	-3,861	1,593	11	-4,618	ACCGCTGGGGTGGTGAAGTGAC	ATG	15430	597																																																																																																																																									
2	-7,368	0,351	11	-8,125	CGCTCTCGGCGCGTCCAGCGCG	GTG	15475	552																																																																																																																																									
3	-3,967	1,943	10	-4,661	CGCGTCCAGCGCGTGTCTGGGT	GTG	15484	543																																																																																																																																									
4	-3,967	1,943	16	-5,763	CAAGCGCGTGTCTGGGTGTGGCC	TTG	15490	537																																																																																																																																									
5	-4,827	1,540	8	-6,049	GAGCGTGTCTCTGTGGGCACT	ATG	15571	456																																																																																																																																									
6	-5,918	1,030	7	-7,441	TATGCTGGGCGACTGGGCGCTG	GTG	15592	435																																																																																																																																									
7	-4,585	1,654	17	-6,585	CAAGTGGGCGTACATCAACGGC	ATG	15625	402																																																																																																																																									
8	-4,735	1,584	6	-6,479	GTGGCGCTACATCAACGGCATG	GTG	15628	399																																																																																																																																									
9	-4,283	1,795	5	-6,283	GGCGGACTAGCGGACTGGGTA	ATG	15655	372																																																																																																																																									
10	-4,394	1,743	13	-5,440	GTACCGCACGGGTCCGCAAGG	GTG	15718	309																																																																																																																																									
11	-3,288	2,261	7	-4,811	CCAGGACCAAGCTCTGGACCTG	GTG	15850	177																																																																																																																																									
12	-3,741	2,049	6	-5,485	GGACCTGGTGAATGATGGTATC	ATG	15865	162																																																																																																																																									
13	-3,741	2,049	18	-6,042	TGATGGTATCATGTGGGGATC	GTG	15877	150																																																																																																																																									
14	-3,079	2,358	13	-4,125	CTGGCAAGAGGACTGGGCAGAT	GTG	16000	27																																																																																																																																									
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	L'orf avec le start retenu est de 597p, la plus longue																																																																																																																																																
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Pas de track</p> <p>Gene: Carostasia_20 Start: 15430, Stop: 16026, Start Num: 8</p> <p>Candidate Starts for Carostasia_20:</p> <p>(Start: 8 @15430 has 137 MA's), (12, 15475), (13, 15484), (14, 15490), (20, 15571), (23, 15592), (25, 15625), (26, 15628), (28, 15655), (36, 15718), (43, 15850), (44, 15865), (45, 15877), (50, 16000),</p> <p>Oui, il est conservé</p>																																																																																																																																																

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?

Distribution of the top 45 Blast Hits on 45 subject sequences



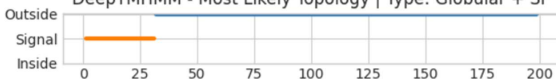
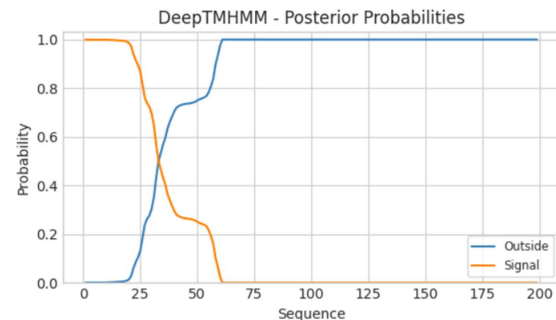
Download GenPept Graphics	
hypothetical protein SEA_NUCCI_20 [Microbacterium phage Nucci]	
Sequence ID: QDK02413.1 Length: 198 Number of Matches: 1	
See 1 more title(s) See all Identical Proteins(IPG)	
Range 1: 1 to 198 GenPept Graphics Next Match Previous Match	
Score	Expect Method Identities Positives Gaps
401 bits(1030)	9e-141 Compositional matrix adjust. 198/198(100%) 198/198(100%) 0/198(0%)
Query 1	MAFIRTFGALRPSKPVLGVALRTAEAEAKVTDLHPGSSLSFSAHPVIGDLALVAAPKN 60
Subjct 1	MAFIRTFGALRPSKPVLGVALRTAEAEAKVTDLHPGSSLSFSAHPVIGDLALVAAPKN 60
Query 61	RYINGHVQADYATWMSDEEDVSPYRGDLVPTGPQTVEMGLLYLGEISSHQVFDGYHTT 120
Subjct 61	RYINGHVQADYATWMSDEEDVSPYRGDLVPTGPQTVEMGLLYLGEISSHQVFDGYHTT 120
Query 121	TAGAGVPHLTARQQRLDLVTDGINSALVATRGINTDPGSHARTDTGTVTESHAEGKSA 180
Subjct 121	TAGAGVPHLTARQQRLDLVTDGINSALVATRGINTDPGSHARTDTGTVTESHAEGKSA 180
Query 181	VVAHMEHNAQVLTONGGM 198
Subjct 181	VVAHMEHNAQVLTONGGM 198
Download GenPept Graphics	
hypothetical protein SEA_QUARTZ_20 [Microbacterium phage Quartz]	
Sequence ID: UVK59239.1 Length: 198 Number of Matches: 1	
Range 1: 1 to 198 GenPept Graphics Next Match Previous Match	
Score	Expect Method Identities Positives Gaps
400 bits(1027)	3e-140 Compositional matrix adjust. 197/198(99%) 198/198(100%) 0/198(0%)
Query 1	MAFIRTFGALRPSKPVLGVALRTAEAEAKVTDLHPGSSLSFSAHPVIGDLALVAAPKN 60
Subjct 1	MAFIRTFGALRPSKPVLGVALRTAEAEAKVTDLHPGSSLSFSAHPVIGDLALVAAPKN 60
Query 61	RYINGHVQADYATWMSDEEDVSPYRGDLVPTGPQTVEMGLLYLGEISSHQVFDGYHTT 120
Subjct 61	RYINGHVQADYATWMSDEEDVSPYRGDLVPTGPQTVEMGLLYLGEISSHQVFDGYHTT 120
Query 121	TAGAGVPHLTARQQRLDLVTDGINSALVATRGINTDPGSHARTDTGTVTESHAEGKSA 180
Subjct 121	TAGAGVPHLTARQQRLDLVTDGINSALVATRGINTDPGSHARTDTGTVTESHAEGKSA 180
Query 181	VVAHMEHNAQVLTONGGM 198
Subjct 181	VVAHMEHNAQVLTONGGM 198

DECISION:

Start 15430. Il est en accord avec genemark et glimmer, il conduit à l'orf le plus long, il est globalement bien conservé chez les homologues et son RBS est associé à un bon score

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																	
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>>Nucci_20, function unknown, 198 Length = 198</p> <p>Score = 415 bits (1067), Expect = e-116 Identities = 198/198 (100%), Positives = 198/198 (100%)</p> <p>Query: 1 MAFIRTFGALRPSKPVLGVALRITAEDAEKVTDIQLNPGSSLSFNAPMVGDLALVAAPKW 60 MAFIRTFGALRPSKPVLGVALRITAEDAEKVTDIQLNPGSSLSFNAPMVGDLALVAAPKW Sbjct: 1 MAFIRTFGALRPSKPVLGVALRITAEDAEKVTDIQLNPGSSLSFNAPMVGDLALVAAPKW 60</p> <p>Query: 61 RYINGMVQADYATWMSDEDVASPYRGDLYTPGPQTVEWGLLYLGEISSHQVFDGYDHTT 120 RYINGMVQADYATWMSDEDVASPYRGDLYTPGPQTVEWGLLYLGEISSHQVFDGYDHTT Sbjct: 61 RYINGMVQADYATWMSDEDVASPYRGDLYTPGPQTVEWGLLYLGEISSHQVFDGYDHTT 120</p> <p>Query: 121 TAGAGVTPHLTARQDQRLDLVTDGIMSIAIVATRGHIDPGSNARTDGTGTVTSHAEGWSA 180 TAGAGVTPHLTARQDQRLDLVTDGIMSIAIVATRGHIDPGSNARTDGTGTVTSHAEGWSA Sbjct: 121 TAGAGVTPHLTARQDQRLDLVTDGIMSIAIVATRGHIDPGSNARTDGTGTVTSHAEGWSA 180</p> <p>Query: 181 VWAHEDHADVLTDHGGW 198 VWAHEDHADVLTDHGGW Sbjct: 181 VWAHEDHADVLTDHGGW 198</p> <p>*** nr :</p> <table><tr><th>select all</th><th>44 sequences selected</th><th>GenPept</th><th>Graphics</th><th>Distance tree of results</th><th>Multiple alignment</th><th>MSA Viewer</th></tr><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_NUCCI_20 (Microbacterium phage Nucci)</td><td>Microbacterium phage Nucci</td><td>401</td><td>401</td><td>99%</td><td>1e-140</td><td>100.00%</td><td>198</td><td>GD952413.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_QUARTZ_20 (Microbacterium phage Quartz)</td><td>Microbacterium phage Quartz</td><td>400</td><td>400</td><td>99%</td><td>3e-140</td><td>99.49%</td><td>198</td><td>U9559239.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_JUANVO_20 (Microbacterium phage Juanvo)</td><td>Microbacterium phage Juanvo</td><td>361</td><td>361</td><td>99%</td><td>6e-125</td><td>87.88%</td><td>198</td><td>U9844770.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_YUUY_20 (Microbacterium phage Yuuy)</td><td>Microbacterium phage Yuuy</td><td>353</td><td>353</td><td>99%</td><td>7e-122</td><td>84.85%</td><td>198</td><td>QEP5489.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_CHICKENKING_21 (Microbacterium phage ChickenKing)</td><td>Microbacterium phage ChickenKing</td><td>321</td><td>321</td><td>99%</td><td>5e-109</td><td>78.28%</td><td>198</td><td>QEP5489.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_PHERBOT_21 (Microbacterium phage Pherbot)</td><td>Microbacterium phage Pherbot</td><td>319</td><td>319</td><td>99%</td><td>2e-108</td><td>77.78%</td><td>198</td><td>GBZ73072.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_PRINCEPHARUS_21 (Microbacterium phage PrincePharus)</td><td>Microbacterium phage PrincePharus</td><td>319</td><td>319</td><td>99%</td><td>3e-108</td><td>77.27%</td><td>198</td><td>GBZ72890.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein ED483_wd21 (Microbacterium phage Koli)</td><td>Microbacterium phage Koli</td><td>315</td><td>315</td><td>99%</td><td>1e-106</td><td>76.26%</td><td>198</td><td>YF_009624202.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_KAUALA_21 (Microbacterium phage Kauala)</td><td>Microbacterium phage Kauala</td><td>315</td><td>315</td><td>99%</td><td>1e-106</td><td>75.76%</td><td>198</td><td>GNL31033.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_MCUBED_20 (Microbacterium phage MCUbed)</td><td>Microbacterium phage MCUbed</td><td>310</td><td>310</td><td>99%</td><td>9e-105</td><td>73.74%</td><td>198</td><td>QGN50680.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein ED482_wd21 (Microbacterium phage Golden)</td><td>Microbacterium phage Golden</td><td>310</td><td>310</td><td>99%</td><td>9e-105</td><td>73.76%</td><td>198</td><td>YF_009624162.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_ZENITSU_20 (Microbacterium phage Zenitsu)</td><td>Microbacterium phage Zenitsu</td><td>310</td><td>310</td><td>99%</td><td>1e-104</td><td>73.74%</td><td>198</td><td>W88693821.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_FENNY_20 (Microbacterium phage Fenry)</td><td>Microbacterium phage Fenry</td><td>310</td><td>310</td><td>99%</td><td>1e-104</td><td>73.23%</td><td>198</td><td>QGF15749.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein ED484_wd20 (Microbacterium phage Dael)</td><td>Microbacterium phage Dael</td><td>309</td><td>309</td><td>99%</td><td>2e-104</td><td>73.23%</td><td>198</td><td>YF_009623058.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_SIRVICTOR_21 (Microbacterium phage SirVictor)</td><td>Microbacterium phage SirVictor</td><td>308</td><td>308</td><td>99%</td><td>8e-104</td><td>75.25%</td><td>198</td><td>W88693964.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_THERSULT_21 (Microbacterium phage Thersult)</td><td>Microbacterium phage Thersult</td><td>308</td><td>308</td><td>99%</td><td>9e-104</td><td>72.73%</td><td>198</td><td>GBZ73212.1</td></tr></table>	select all	44 sequences selected	GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer		Description	Scientific Name	Max Score	Total Score	Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/>	hypothetical protein SEA_NUCCI_20 (Microbacterium phage Nucci)	Microbacterium phage Nucci	401	401	99%	1e-140	100.00%	198	GD952413.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_QUARTZ_20 (Microbacterium phage Quartz)	Microbacterium phage Quartz	400	400	99%	3e-140	99.49%	198	U9559239.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_JUANVO_20 (Microbacterium phage Juanvo)	Microbacterium phage Juanvo	361	361	99%	6e-125	87.88%	198	U9844770.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_YUUY_20 (Microbacterium phage Yuuy)	Microbacterium phage Yuuy	353	353	99%	7e-122	84.85%	198	QEP5489.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_CHICKENKING_21 (Microbacterium phage ChickenKing)	Microbacterium phage ChickenKing	321	321	99%	5e-109	78.28%	198	QEP5489.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_PHERBOT_21 (Microbacterium phage Pherbot)	Microbacterium phage Pherbot	319	319	99%	2e-108	77.78%	198	GBZ73072.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_PRINCEPHARUS_21 (Microbacterium phage PrincePharus)	Microbacterium phage PrincePharus	319	319	99%	3e-108	77.27%	198	GBZ72890.1	<input checked="" type="checkbox"/>	hypothetical protein ED483_wd21 (Microbacterium phage Koli)	Microbacterium phage Koli	315	315	99%	1e-106	76.26%	198	YF_009624202.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_KAUALA_21 (Microbacterium phage Kauala)	Microbacterium phage Kauala	315	315	99%	1e-106	75.76%	198	GNL31033.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_MCUBED_20 (Microbacterium phage MCUbed)	Microbacterium phage MCUbed	310	310	99%	9e-105	73.74%	198	QGN50680.1	<input checked="" type="checkbox"/>	hypothetical protein ED482_wd21 (Microbacterium phage Golden)	Microbacterium phage Golden	310	310	99%	9e-105	73.76%	198	YF_009624162.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_ZENITSU_20 (Microbacterium phage Zenitsu)	Microbacterium phage Zenitsu	310	310	99%	1e-104	73.74%	198	W88693821.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_FENNY_20 (Microbacterium phage Fenry)	Microbacterium phage Fenry	310	310	99%	1e-104	73.23%	198	QGF15749.1	<input checked="" type="checkbox"/>	hypothetical protein ED484_wd20 (Microbacterium phage Dael)	Microbacterium phage Dael	309	309	99%	2e-104	73.23%	198	YF_009623058.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_SIRVICTOR_21 (Microbacterium phage SirVictor)	Microbacterium phage SirVictor	308	308	99%	8e-104	75.25%	198	W88693964.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_THERSULT_21 (Microbacterium phage Thersult)	Microbacterium phage Thersult	308	308	99%	9e-104	72.73%	198	GBZ73212.1
select all	44 sequences selected	GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer																																																																																																																																																																												
	Description	Scientific Name	Max Score	Total Score	Cover	E value	Per. Ident	Acc. Len	Accession																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_NUCCI_20 (Microbacterium phage Nucci)	Microbacterium phage Nucci	401	401	99%	1e-140	100.00%	198	GD952413.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_QUARTZ_20 (Microbacterium phage Quartz)	Microbacterium phage Quartz	400	400	99%	3e-140	99.49%	198	U9559239.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_JUANVO_20 (Microbacterium phage Juanvo)	Microbacterium phage Juanvo	361	361	99%	6e-125	87.88%	198	U9844770.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_YUUY_20 (Microbacterium phage Yuuy)	Microbacterium phage Yuuy	353	353	99%	7e-122	84.85%	198	QEP5489.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_CHICKENKING_21 (Microbacterium phage ChickenKing)	Microbacterium phage ChickenKing	321	321	99%	5e-109	78.28%	198	QEP5489.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_PHERBOT_21 (Microbacterium phage Pherbot)	Microbacterium phage Pherbot	319	319	99%	2e-108	77.78%	198	GBZ73072.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_PRINCEPHARUS_21 (Microbacterium phage PrincePharus)	Microbacterium phage PrincePharus	319	319	99%	3e-108	77.27%	198	GBZ72890.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein ED483_wd21 (Microbacterium phage Koli)	Microbacterium phage Koli	315	315	99%	1e-106	76.26%	198	YF_009624202.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_KAUALA_21 (Microbacterium phage Kauala)	Microbacterium phage Kauala	315	315	99%	1e-106	75.76%	198	GNL31033.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_MCUBED_20 (Microbacterium phage MCUbed)	Microbacterium phage MCUbed	310	310	99%	9e-105	73.74%	198	QGN50680.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein ED482_wd21 (Microbacterium phage Golden)	Microbacterium phage Golden	310	310	99%	9e-105	73.76%	198	YF_009624162.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_ZENITSU_20 (Microbacterium phage Zenitsu)	Microbacterium phage Zenitsu	310	310	99%	1e-104	73.74%	198	W88693821.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_FENNY_20 (Microbacterium phage Fenry)	Microbacterium phage Fenry	310	310	99%	1e-104	73.23%	198	QGF15749.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein ED484_wd20 (Microbacterium phage Dael)	Microbacterium phage Dael	309	309	99%	2e-104	73.23%	198	YF_009623058.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_SIRVICTOR_21 (Microbacterium phage SirVictor)	Microbacterium phage SirVictor	308	308	99%	8e-104	75.25%	198	W88693964.1																																																																																																																																																																									
<input checked="" type="checkbox"/>	hypothetical protein SEA_THERSULT_21 (Microbacterium phage Thersult)	Microbacterium phage Thersult	308	308	99%	9e-104	72.73%	198	GBZ73212.1																																																																																																																																																																									
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<p>UniProt : (seul hit satisfaisant)</p> <p>Template alignment</p> <p><input type="checkbox"/> 1.ABE283 VPN7_BPPHE Tail fiber protein O5=Enterococcus phage phiEF24C OX=442493 GN=EPF_031 PE=1 SV=1 Probability: 91.72%, E-value: 0.54, Score: 54.04, Aligned cols: 51, Identities: 14%, Similarity: 0.321, Template Neff: 7.9</p> <p>Visualisation</p>																																																																																																																																																																																	

	<table><tr><th>N°</th><th>Frapper</th><th>Nom</th><th>Probabilité</th><th>Valeur E</th><th>Score</th><th>SS</th><th>Cols alignés</th><th>Longueur cible</th></tr><tr><td><input type="checkbox"/> 1</td><td>A8E283</td><td>VPN7_BPPHE Protéine des fibres de queue OS=Enterococcus phage phiEF24C OX=442493 GN=EFP_031 PE=1 SV=1</td><td>91.72</td><td>0,54</td><td>54.04</td><td>4.7</td><td>51</td><td>1825</td></tr><tr><td><input type="checkbox"/> 2</td><td>PF07902.15</td><td>; GP58 : protéine de type gp58</td><td>54.03</td><td>41</td><td>35.33</td><td>4</td><td>50</td><td>580</td></tr><tr><td><input type="checkbox"/> 3</td><td>4DOQ_A</td><td>HYALURONATE LYASE ; LYASE, MODULE DE LIAISON DES HYALURONANES, DES HYDRATES DE CARBONE, CBM, PL FAMILLE 8, PL8, HYL ; HET : EDO ; 1,2A (ST</td><td>52,7</td><td>130</td><td>21,99</td><td>5,3</td><td>51</td><td>169</td></tr></table>	N°	Frapper	Nom	Probabilité	Valeur E	Score	SS	Cols alignés	Longueur cible	<input type="checkbox"/> 1	A8E283	VPN7_BPPHE Protéine des fibres de queue OS=Enterococcus phage phiEF24C OX=442493 GN=EFP_031 PE=1 SV=1	91.72	0,54	54.04	4.7	51	1825	<input type="checkbox"/> 2	PF07902.15	; GP58 : protéine de type gp58	54.03	41	35.33	4	50	580	<input type="checkbox"/> 3	4DOQ_A	HYALURONATE LYASE ; LYASE, MODULE DE LIAISON DES HYALURONANES, DES HYDRATES DE CARBONE, CBM, PL FAMILLE 8, PL8, HYL ; HET : EDO ; 1,2A (ST	52,7	130	21,99	5,3	51	169
N°	Frapper	Nom	Probabilité	Valeur E	Score	SS	Cols alignés	Longueur cible																													
<input type="checkbox"/> 1	A8E283	VPN7_BPPHE Protéine des fibres de queue OS=Enterococcus phage phiEF24C OX=442493 GN=EFP_031 PE=1 SV=1	91.72	0,54	54.04	4.7	51	1825																													
<input type="checkbox"/> 2	PF07902.15	; GP58 : protéine de type gp58	54.03	41	35.33	4	50	580																													
<input type="checkbox"/> 3	4DOQ_A	HYALURONATE LYASE ; LYASE, MODULE DE LIAISON DES HYALURONANES, DES HYDRATES DE CARBONE, CBM, PL FAMILLE 8, PL8, HYL ; HET : EDO ; 1,2A (ST	52,7	130	21,99	5,3	51	169																													
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Proche des tail proteins																																				
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div><p>DeepTMHMM - Most Likely Topology Type: Globular + SP</p><p>DeepTMHMM - Posterior Probabilities</p></div>																																				
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<div><div>Oui, Minor Tail Protein</div><table><tr><td>minor tail protein</td><td>tail fiber-like protein, collagen-like, glycine rich</td><td>If you have significant hits to either collagen-like or glycine-rich proteins, and are in the syntenic region of minor tail proteins, you can call them minor tail proteins.</td></tr></table></div>	minor tail protein	tail fiber-like protein, collagen-like, glycine rich	If you have significant hits to either collagen-like or glycine-rich proteins, and are in the syntenic region of minor tail proteins, you can call them minor tail proteins.																																	
minor tail protein	tail fiber-like protein, collagen-like, glycine rich	If you have significant hits to either collagen-like or glycine-rich proteins, and are in the syntenic region of minor tail proteins, you can call them minor tail proteins.																																			
DECISION:	NKF																																				

Student Gene Annotation Worksheet

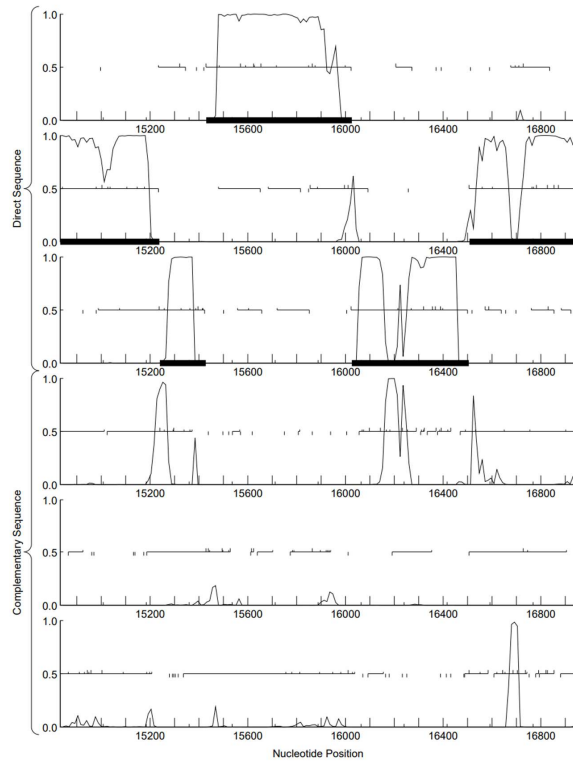
Carosmatique-gene-21

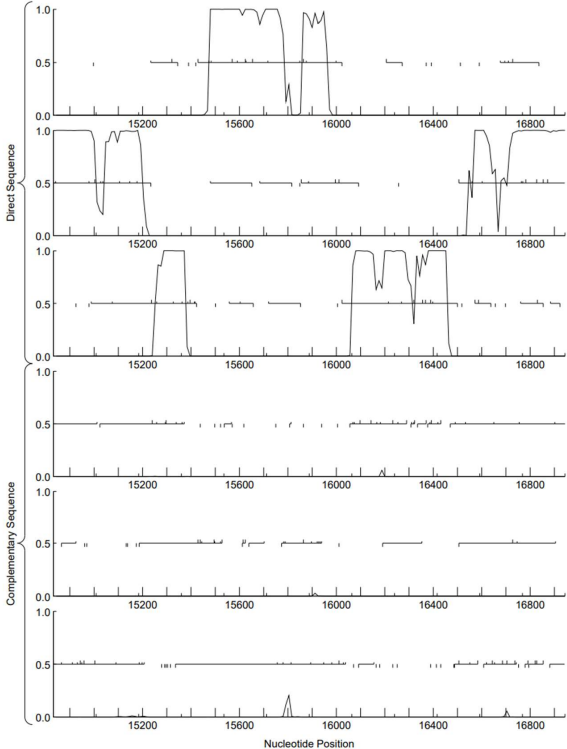
Basic Phage Information	
Nom du Phage	carostasia
Gène #	21
Coordonnées du Stop	16505
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Chev 1
Coordonnées du Start retenu	16026
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

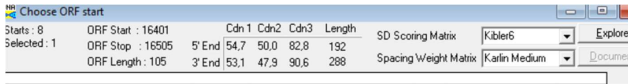
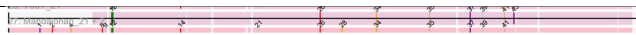
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Cadre +3</i>

terium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 9/21



	<p>acterium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 9/21</p> 																														
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Nucci_21, function unknown, 159</td><td>313</td><td>8e-86</td></tr><tr><td>Mandalorian_21, function unknown, 159</td><td>313</td><td>8e-86</td></tr><tr><td>Carostasia_Draft_21, function unknown, 159</td><td>313</td><td>8e-86</td></tr><tr><td>Quartz_21, function unknown, 159</td><td>306</td><td>1e-83</td></tr><tr><td>Juanyo_21, function unknown, 159</td><td>282</td><td>3e-76</td></tr><tr><td>SirVictor_22, function unknown, 158</td><td>259</td><td>1e-69</td></tr><tr><td>Guertzie_22, function unknown, 158</td><td>252</td><td>1e-69</td></tr><tr><td>Finny_21, function unknown, 158</td><td>249</td><td>1e-66</td></tr><tr><td>Yuuy_21, function unknown, 158</td><td>248</td><td>3e-66</td></tr></table>	Sequences producing significant alignments:	Score (bits)	E Value	Nucci_21, function unknown, 159	313	8e-86	Mandalorian_21, function unknown, 159	313	8e-86	Carostasia_Draft_21, function unknown, 159	313	8e-86	Quartz_21, function unknown, 159	306	1e-83	Juanyo_21, function unknown, 159	282	3e-76	SirVictor_22, function unknown, 158	259	1e-69	Guertzie_22, function unknown, 158	252	1e-69	Finny_21, function unknown, 158	249	1e-66	Yuuy_21, function unknown, 158	248	3e-66
Sequences producing significant alignments:	Score (bits)	E Value																													
Nucci_21, function unknown, 159	313	8e-86																													
Mandalorian_21, function unknown, 159	313	8e-86																													
Carostasia_Draft_21, function unknown, 159	313	8e-86																													
Quartz_21, function unknown, 159	306	1e-83																													
Juanyo_21, function unknown, 159	282	3e-76																													
SirVictor_22, function unknown, 158	259	1e-69																													
Guertzie_22, function unknown, 158	252	1e-69																													
Finny_21, function unknown, 158	249	1e-66																													
Yuuy_21, function unknown, 158	248	3e-66																													
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non																														
DECISION:	Oui																														

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																										
Quel start Glimmer et GeneMark suggèrent-ils ?	16026																																																																																										
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div></div> <table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,633</td><td>2,567</td><td>11</td><td>-3,390</td><td>TTACTGATCATGGAGGCTGGTA</td><td>ATG</td><td>16026</td><td>480</td></tr><tr><td>2</td><td>-6,028</td><td>0,978</td><td>9</td><td>-6,803</td><td>GTACTGGAACAATGCTGCTAC</td><td>GTG</td><td>16218</td><td>288</td></tr><tr><td>3</td><td>-6,193</td><td>0,901</td><td>13</td><td>-7,238</td><td>GGTACGCTTGACCTCGGCTG</td><td>GTG</td><td>16272</td><td>234</td></tr><tr><td>4</td><td>-7,056</td><td>0,497</td><td>7</td><td>-8,579</td><td>CCAGTATTACACCTCGGCAAC</td><td>ATG</td><td>16323</td><td>183</td></tr><tr><td>5</td><td>-3,638</td><td>2,097</td><td>12</td><td>-4,474</td><td>CGGGCGACCGGGAAGCACCG</td><td>ATG</td><td>16359</td><td>147</td></tr><tr><td>6</td><td>-3,201</td><td>2,302</td><td>14</td><td>-4,548</td><td>GAAGCACCGGATGGGCTGGC</td><td>ATG</td><td>16371</td><td>135</td></tr><tr><td>7</td><td>-5,689</td><td>1,137</td><td>12</td><td>-6,525</td><td>CATGATCTACGCTGCCACGGTC</td><td>ATG</td><td>16392</td><td>114</td></tr><tr><td>8</td><td>-3,990</td><td>1,932</td><td>8</td><td>-5,212</td><td>CGCTGCCACGGTCATGAAGCC</td><td>GTG</td><td>16401</td><td>105</td></tr></table>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,633	2,567	11	-3,390	TTACTGATCATGGAGGCTGGTA	ATG	16026	480	2	-6,028	0,978	9	-6,803	GTACTGGAACAATGCTGCTAC	GTG	16218	288	3	-6,193	0,901	13	-7,238	GGTACGCTTGACCTCGGCTG	GTG	16272	234	4	-7,056	0,497	7	-8,579	CCAGTATTACACCTCGGCAAC	ATG	16323	183	5	-3,638	2,097	12	-4,474	CGGGCGACCGGGAAGCACCG	ATG	16359	147	6	-3,201	2,302	14	-4,548	GAAGCACCGGATGGGCTGGC	ATG	16371	135	7	-5,689	1,137	12	-6,525	CATGATCTACGCTGCCACGGTC	ATG	16392	114	8	-3,990	1,932	8	-5,212	CGCTGCCACGGTCATGAAGCC	GTG	16401	105
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																			
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																			
1	-2,633	2,567	11	-3,390	TTACTGATCATGGAGGCTGGTA	ATG	16026	480																																																																																			
2	-6,028	0,978	9	-6,803	GTACTGGAACAATGCTGCTAC	GTG	16218	288																																																																																			
3	-6,193	0,901	13	-7,238	GGTACGCTTGACCTCGGCTG	GTG	16272	234																																																																																			
4	-7,056	0,497	7	-8,579	CCAGTATTACACCTCGGCAAC	ATG	16323	183																																																																																			
5	-3,638	2,097	12	-4,474	CGGGCGACCGGGAAGCACCG	ATG	16359	147																																																																																			
6	-3,201	2,302	14	-4,548	GAAGCACCGGATGGGCTGGC	ATG	16371	135																																																																																			
7	-5,689	1,137	12	-6,525	CATGATCTACGCTGCCACGGTC	ATG	16392	114																																																																																			
8	-3,990	1,932	8	-5,212	CGCTGCCACGGTCATGAAGCC	GTG	16401	105																																																																																			
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Oui, le start donne l'orf la plus longue Meilleur Score : -3.390																																																																																										
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																											
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Oui >Nucci_21, function unknown, 159 Length = 159 Score = 313 bits (803), Expect = 8e-86 Identities = 159/159 (100%), Positives = 159/159 (100%) Query: 1 MALQTPPTPAYGPNTFTSLAELASILLANDEYLEDKIDNIAPPAPSVTKAAITLASGYSN 60 MALQTPPTPAYGPNTFTSLAELASILLANDEYLEDKIDNIAPPAPSVTKAAITLASGYSN Sbjct: 1 MALQTPPTPAYGPNTFTSLAELASILLANDEYLEDKIDNIAPPAPSVTKAAITLASGYSN 60 Query: 61 NASYVSSEAIKIGNLVTLDLGLVNCPSAIGGNQYVTLGTHPAGYRPTGKHRHGVGMIIYAA 120 NASYVSSEAIKIGNLVTLDLGLVNCPSAIGGNQYVTLGTHPAGYRPTGKHRHGVGMIIYAA Sbjct: 61 NASYVSSEAIKIGNLVTLDLGLVNCPSAIGGNQYVTLGTHPAGYRPTGKHRHGVGMIIYAA 120 Query: 121 TVMKPVQLRLNTDGSINFLSAESVSGASYLFASAITFAI 159 TVMKPVQLRLNTDGSINFLSAESVSGASYLFASAITFAI Sbjct: 121 TVMKPVQLRLNTDGSINFLSAESVSGASYLFASAITFAI 159																																																																																										
DECISION:	Le start conservé est celui à la position 16026 car il n'est pas en contradiction avec les principes d'annotation et Il est conservé chez d'autres homologues. De																																																																																										

	plus, le RBS associé au start présente le meilleur score et le plus long ORF avec un ATG
--	--

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

*** *PhagesDB* :

Sequence producing significant alignment:		
NC_012121.1	Function unknown, 159	112
NC_012121.1	Function unknown, 159	111
NC_012121.1	Function unknown, 159	110
NC_012121.1	Function unknown, 159	109
NC_012121.1	Function unknown, 159	108
NC_012121.1	Function unknown, 159	107
NC_012121.1	Function unknown, 159	106
NC_012121.1	Function unknown, 159	105
NC_012121.1	Function unknown, 159	104
NC_012121.1	Function unknown, 159	103
NC_012121.1	Function unknown, 159	102
NC_012121.1	Function unknown, 159	101
NC_012121.1	Function unknown, 159	100
NC_012121.1	Function unknown, 159	99
NC_012121.1	Function unknown, 159	98
NC_012121.1	Function unknown, 159	97
NC_012121.1	Function unknown, 159	96
NC_012121.1	Function unknown, 159	95
NC_012121.1	Function unknown, 159	94
NC_012121.1	Function unknown, 159	93
NC_012121.1	Function unknown, 159	92
NC_012121.1	Function unknown, 159	91
NC_012121.1	Function unknown, 159	90
NC_012121.1	Function unknown, 159	89
NC_012121.1	Function unknown, 159	88
NC_012121.1	Function unknown, 159	87
NC_012121.1	Function unknown, 159	86
NC_012121.1	Function unknown, 159	85
NC_012121.1	Function unknown, 159	84
NC_012121.1	Function unknown, 159	83
NC_012121.1	Function unknown, 159	82
NC_012121.1	Function unknown, 159	81
NC_012121.1	Function unknown, 159	80
NC_012121.1	Function unknown, 159	79
NC_012121.1	Function unknown, 159	78
NC_012121.1	Function unknown, 159	77
NC_012121.1	Function unknown, 159	76
NC_012121.1	Function unknown, 159	75
NC_012121.1	Function unknown, 159	74
NC_012121.1	Function unknown, 159	73
NC_012121.1	Function unknown, 159	72
NC_012121.1	Function unknown, 159	71
NC_012121.1	Function unknown, 159	70
NC_012121.1	Function unknown, 159	69
NC_012121.1	Function unknown, 159	68
NC_012121.1	Function unknown, 159	67
NC_012121.1	Function unknown, 159	66
NC_012121.1	Function unknown, 159	65
NC_012121.1	Function unknown, 159	64
NC_012121.1	Function unknown, 159	63
NC_012121.1	Function unknown, 159	62
NC_012121.1	Function unknown, 159	61
NC_012121.1	Function unknown, 159	60
NC_012121.1	Function unknown, 159	59
NC_012121.1	Function unknown, 159	58
NC_012121.1	Function unknown, 159	57
NC_012121.1	Function unknown, 159	56
NC_012121.1	Function unknown, 159	55
NC_012121.1	Function unknown, 159	54
NC_012121.1	Function unknown, 159	53
NC_012121.1	Function unknown, 159	52
NC_012121.1	Function unknown, 159	51
NC_012121.1	Function unknown, 159	50
NC_012121.1	Function unknown, 159	49
NC_012121.1	Function unknown, 159	48
NC_012121.1	Function unknown, 159	47
NC_012121.1	Function unknown, 159	46
NC_012121.1	Function unknown, 159	45
NC_012121.1	Function unknown, 159	44
NC_012121.1	Function unknown, 159	43
NC_012121.1	Function unknown, 159	42
NC_012121.1	Function unknown, 159	41
NC_012121.1	Function unknown, 159	40
NC_012121.1	Function unknown, 159	39
NC_012121.1	Function unknown, 159	38
NC_012121.1	Function unknown, 159	37
NC_012121.1	Function unknown, 159	36
NC_012121.1	Function unknown, 159	35
NC_012121.1	Function unknown, 159	34
NC_012121.1	Function unknown, 159	33
NC_012121.1	Function unknown, 159	32
NC_012121.1	Function unknown, 159	31
NC_012121.1	Function unknown, 159	30
NC_012121.1	Function unknown, 159	29
NC_012121.1	Function unknown, 159	28
NC_012121.1	Function unknown, 159	27
NC_012121.1	Function unknown, 159	26
NC_012121.1	Function unknown, 159	25
NC_012121.1	Function unknown, 159	24
NC_012121.1	Function unknown, 159	23
NC_012121.1	Function unknown, 159	22
NC_012121.1	Function unknown, 159	21
NC_012121.1	Function unknown, 159	20
NC_012121.1	Function unknown, 159	19
NC_012121.1	Function unknown, 159	18
NC_012121.1	Function unknown, 159	17
NC_012121.1	Function unknown, 159	16
NC_012121.1	Function unknown, 159	15
NC_012121.1	Function unknown, 159	14
NC_012121.1	Function unknown, 159	13
NC_012121.1	Function unknown, 159	12
NC_012121.1	Function unknown, 159	11
NC_012121.1	Function unknown, 159	10
NC_012121.1	Function unknown, 159	9
NC_012121.1	Function unknown, 159	8
NC_012121.1	Function unknown, 159	7
NC_012121.1	Function unknown, 159	6
NC_012121.1	Function unknown, 159	5
NC_012121.1	Function unknown, 159	4
NC_012121.1	Function unknown, 159	3
NC_012121.1	Function unknown, 159	2
NC_012121.1	Function unknown, 159	1

(q#: s#) : [alignment] ; e-value

Query: NC_012121.1
(159 letters)

Distribution of 100 Blast Hits on the Query Sequence

Mean score to show define and scores. Click to show alignments

Color Key for Alignment Scores

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

100 0 100 250

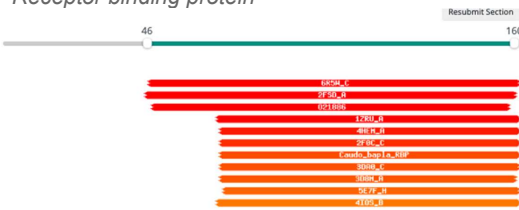
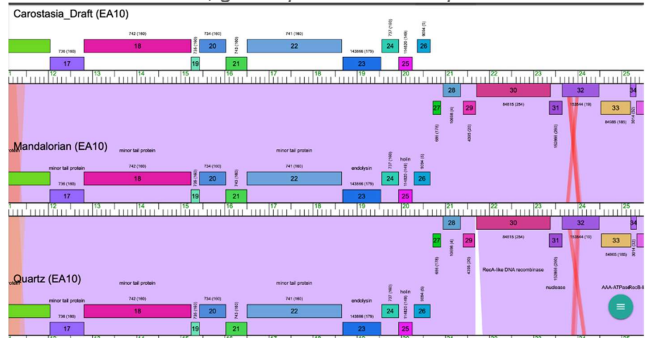
100 0 100 250

100 0 100 250

100 0 100 250

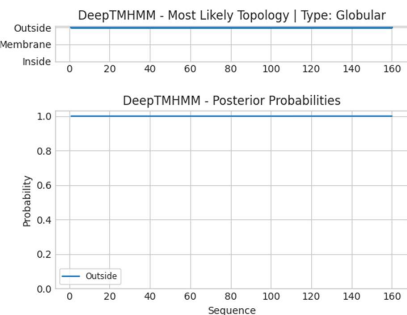
100 0 100 250

1

	Que des hits pour les deux DB de protéines à la fonction inconnue/putatives, la query cover est très satisfaisante et la e value est $<<10^{-4}$.																																													
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHpred avec une proba $\geq 90\%$ et une couverture acceptable ?	<p>HHpred trouve plusieurs subjects avec une probabilité de plus de 90%. Parmi les fonctions on a :</p> <ul style="list-style-type: none">- Gp15 protein, receptor binding protein- putative baseplate protein- Receptor binding protein  <p>Hitlist</p> <p>Show 25 Entries Search:</p> <table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/></td><td>1</td><td>6RSW_C Gp15 protein; Listeria, homotrimeric, receptor binding protein, Bacteriophage, VIRAL PROTEIN; HET: 1PE, ACT; 1.7A (Liste</td><td>98.83</td><td>1.5e-7</td><td>68.8</td><td>10.5</td><td>108</td><td>173</td></tr><tr><td><input type="checkbox"/></td><td>2</td><td>2FSD_A putative baseplate protein; Phage, Lactococcus lactis, receptor binding protein, head domain, VIRAL PROTEIN; 2.3A (unide</td><td>97.29</td><td>0.016</td><td>41.96</td><td>10.2</td><td>108</td><td>142</td></tr><tr><td><input type="checkbox"/></td><td>3</td><td>O21886 RBP, BPLSK Receptor binding protein OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1</td><td>95.63</td><td>1.3</td><td>35.78</td><td>12.1</td><td>101</td><td>264</td></tr><tr><td><input type="checkbox"/></td><td>4</td><td>1ZRJ_A lactophage p2 receptor binding protein; 3 domains: BETA BARREL, BETA PRISM, BETA BARREL, Structural Genomics, Structural</td><td>95</td><td>1.9</td><td>34.68</td><td>11.4</td><td>85</td><td>264</td></tr></table>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/>	1	6RSW_C Gp15 protein; Listeria, homotrimeric, receptor binding protein, Bacteriophage, VIRAL PROTEIN; HET: 1PE, ACT; 1.7A (Liste	98.83	1.5e-7	68.8	10.5	108	173	<input type="checkbox"/>	2	2FSD_A putative baseplate protein; Phage, Lactococcus lactis, receptor binding protein, head domain, VIRAL PROTEIN; 2.3A (unide	97.29	0.016	41.96	10.2	108	142	<input type="checkbox"/>	3	O21886 RBP, BPLSK Receptor binding protein OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1	95.63	1.3	35.78	12.1	101	264	<input type="checkbox"/>	4	1ZRJ_A lactophage p2 receptor binding protein; 3 domains: BETA BARREL, BETA PRISM, BETA BARREL, Structural Genomics, Structural	95	1.9	34.68	11.4	85	264
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																						
<input type="checkbox"/>	1	6RSW_C Gp15 protein; Listeria, homotrimeric, receptor binding protein, Bacteriophage, VIRAL PROTEIN; HET: 1PE, ACT; 1.7A (Liste	98.83	1.5e-7	68.8	10.5	108	173																																						
<input type="checkbox"/>	2	2FSD_A putative baseplate protein; Phage, Lactococcus lactis, receptor binding protein, head domain, VIRAL PROTEIN; 2.3A (unide	97.29	0.016	41.96	10.2	108	142																																						
<input type="checkbox"/>	3	O21886 RBP, BPLSK Receptor binding protein OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1	95.63	1.3	35.78	12.1	101	264																																						
<input type="checkbox"/>	4	1ZRJ_A lactophage p2 receptor binding protein; 3 domains: BETA BARREL, BETA PRISM, BETA BARREL, Structural Genomics, Structural	95	1.9	34.68	11.4	85	264																																						
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<p>Gene avant inconnu, gene après : minor tail protein</p> 																																													

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

Non, deepTMHMMla prédit comme globulaire.

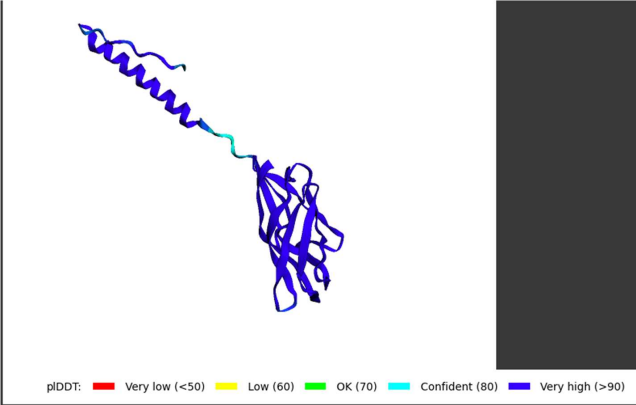


Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Oui.

AlphaFold

Structure :

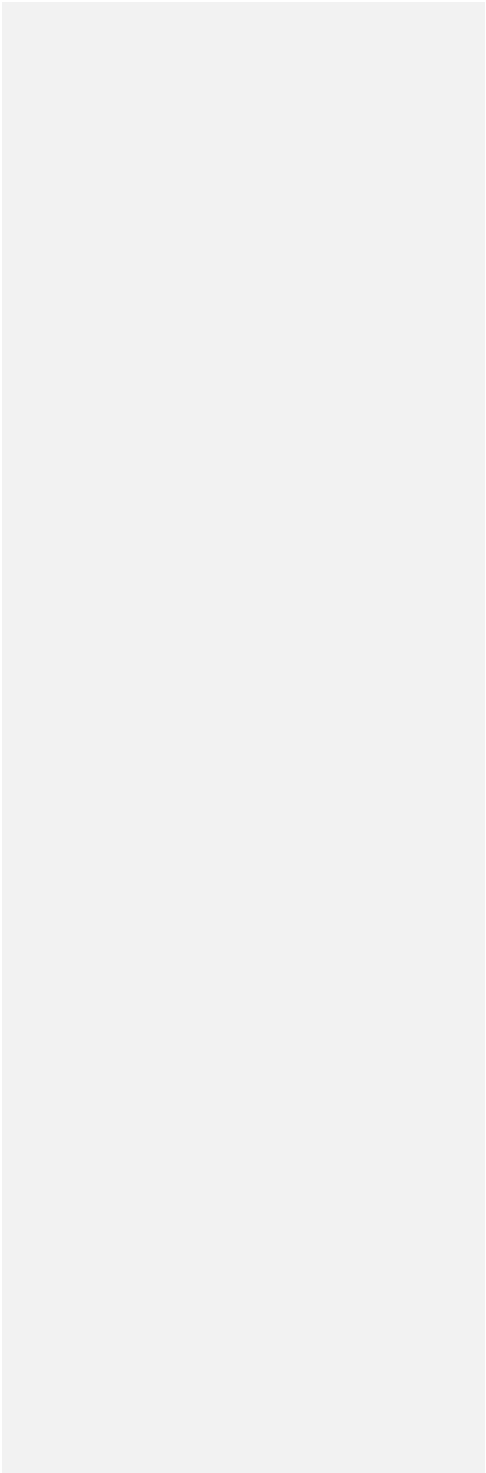


Recherche SeekFold :

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
AF-A04482F2Z04-F1.model	Uncharacterized protein	Enterobacteriaceae	1.00	21.6	6.37e-7	1-160	100%
AF-A04482F2Z04-F1.model	OLP2479 domain-contain.	Enterobacteriaceae	1.00	20.9	3.70e-5	1-160	100%
AF-Q200U2-F1.model_v4	ArgL4880	Acetivibrio thalassae	0.23	10.9	1.01e+0	1-160	100%
AF-Q200U2-F1.model_v4	Mitochondrial rRNA speci...	Danio rerio	0.21	14.5	1.17e+0	1-160	100%
AF-Q200V9-F1.model_v4	Ochrobactrum 700 protein	Caula bella-Abuotoba Group	0.18	15.7	3.30e+0	1-160	100%
AF-Q13941-F1.model_v4	Mitochondrial rRNA speci...	Salmonella enterica serovar	0.16	8	1.00e+0	1-160	100%
AF-Q200U2-F1.model_v4	Heat shock protein, putative	Enterobacteriaceae	0.15	18.7	2.08e+0	1-160	100%

Aucune homologie avec une protéine à la fonction connue.

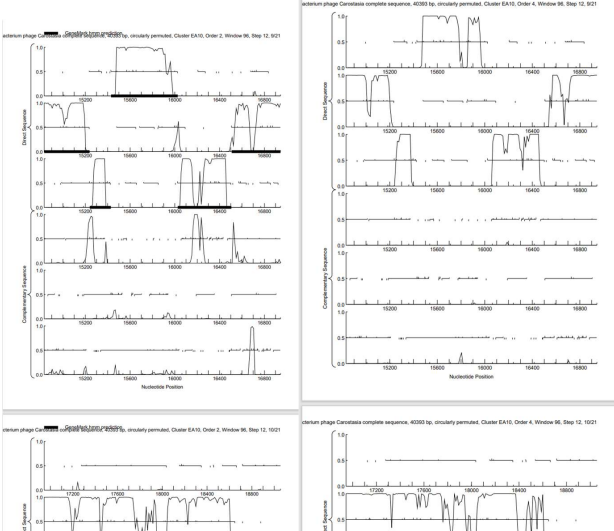
DECISION:	<i>Fonction inconnue.</i>
------------------	---------------------------



Student Gene Annotation Worksheet

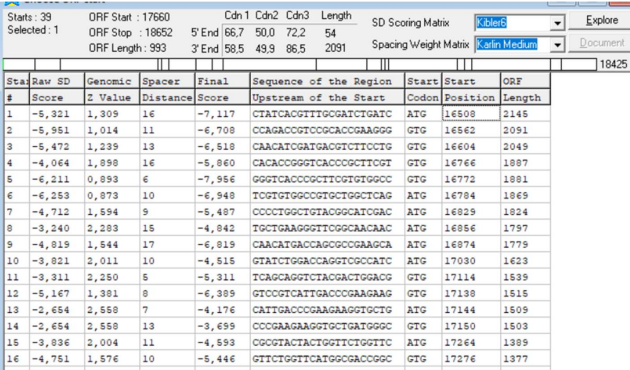
Basic Phage Information	
Nom du Phage	carostasia
Gène #	22
Coordonnées du Stop	18652
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	gap 2
Coordonnées du Start retenu	16508
Fonction prédite	Minor tail protein

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
<p>Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?</p>	<p>Yes both</p>
<p>Y-a-t-il des éléments supportant un potentiel codant ?</p>	

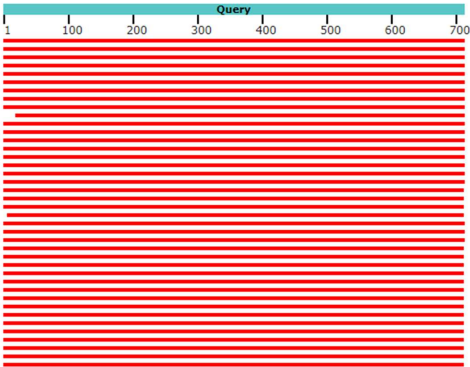
<div>Distribution of 100 Blast Hits on the Query Sequence</div> <div>Mouse-over to show define and scores. Click to show alignments</div> <div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>200</div></div><div><div>1c111</div><div><div>0</div><div>100</div><div>200</div><div>300</div><div>400</div><div>500</div><div>600</div><div>700</div></div><div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div></div></div>	
---	--

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	16508
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div></div> <p>Pas meilleur score RBS : -7.117, ORF la plus longue 2145, le meilleur est -3.699 ORF de 1503pb mais celui-ci conduit à une ORF moins longue et donc à un gap plus important.</p>
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	Avec le start en 16508, orf de 2145 pb Gap +2 . L'orf le plus long conduit à un gap excessif.
Est-ce que le start est conservé chez les homologues voir Starterator ?	Gene: Carostasia_22 Start: 16508, Stop: 18652, Start Num: 12 Candidate Starts for Carostasia_22: (Start: 12 @16508 has 102 MA's), (Start: 16 @16562 has 1 MA's), (20, 16604), (31, 16766), (32, 16772), (33, 16784), (35, 16829), (36, 16856), (39, 16874), (46, 17030), (50, 17114), (52, 17138), (53, 17144), (54, 17150), (56, 17264), (57, 17276), (58, 17288), (65, 17420), (66, 17438), (71, 17513), (73, 17543), (76, 17570), (77, 17573), (80, 17660), (83, 17783), (88, 17882), (96, 18017), (97, 18029), (99, 18077), (103, 18098), (105, 18131), (107, 18140), (110, 18209), (111, 18215), (117, 18299), (118, 18311), (125, 18440), (130, 18506), (132, 18545), Même start conservé pas de track

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?

Distribution of the top 100 Blast Hits on 100 subject sequence:



Download ▾ GenPeot Graphics	
minor tail protein [Microbacterium phage Quartz]	
Sequence ID: UWK59241.1 Length: 714 Number of Matches: 1	
Range 1: 1 to 714 GenPeot Graphics ▼ NCBI:BLAST ▲ EMBL	
Score	Expect Method Identical Positives Query
1464 bits(3790) 0.0	Compositional matrix adjust. 710/714(99%) 712/714(99%) 10/714(0%)
Query 1	NR01VALMHANQTVRTTEGVLDKYGNIDDOVFLVGVLLGTGLENLKHITVEAARAGQPS 60
Subject 1	NR01VALMHANQTVRTTEGVLDKYGNIDDOVFLVGVLLGTGLENLKHITVEAARAGQPS 60
Query 61	IRHMLTIQCFSSAANKAIGTQPLRVAVALAQHETTYATYPHLVGEDMAEGFGHHTSA 120
Subject 61	IRHMLTIQCFSSAANKAIGTQPLRVAVALAQHETTYATYPHLVGEDMAEGFGHHTSA 120
Query 121	EARMGYRALGDHARSFGKXLSAALPAATEGIFSIIGENILDYAAFAYLDQVADHTYDFA 180
Subject 121	EARMGYRALGDHARSFGKXLSAALPAATEGIFSIIGENILDYAAFAYLDQVADHTYDFA 180
Query 181	NSGSAAPGFLAPRFHQDYQVNTVSVIDPKXVLCVPAYQWNSIHVPTNLEPOYSPSAPY 240
Subject 181	NSGSAAPGFLAPRFHQDYQVNTVSVIDPKXVLCVPAYQWNSIHVPTNLEPOYSPSAPY 240
Query 241	RHSGAYYWFHFNATGVWSSHSQDKEPAEMQTHASMTYRDIETISPTLLIGYAMA 300
Subject 241	RHSGAYYWFHFNATGVWSSHSQDKEPAEMQTHASMTYRDIETISPTLLIGYAMA 300
Query 301	TANHVTSSGHEKXSLGKPYFTRYGRAAADPVGHNQDPTTQVQHVHVP1QVDSAG 360
Subject 301	TANHVTSSGHEKXSLGKPYFTRYGRAAADPVGHNQDPTTQVQHVHVP1QVDSAG 360
Query 361	RWYKQDVHMLTLEVLQDQPSATIHDDCASTGTLDITYYRTGAATHRSSOPLNPPA 420
Subject 361	RWYKQDVHMLTLEVLQDQPSATIHDDCASTGTLDITYYRTGAATHRSSOPLNPPA 420
Query 421	YQGFVIVAGAGRLALTRITNGGGEFHQGRQFQPSAGAGVYHSFEATVDQAGVLRSLNG 480
Subject 421	YQGFVIVAGAGRLALTRITNGGGEFHQGRQFQPSAGAGVYHSFEATVDQAGVLRSLNG 480
Query 481	TTLATQVNSAPOTSSSPSARAVGLRVGRHARVYVYRSTENSVPLILEYVDPSTPGQ 540
Subject 481	TTLATQVNSAPOTSSSPSARAVGLRVGRHARVYVYRSTENSVPLILEYVDPSTPGQ 540
Query 541	GRLSHSGAAMFDHVRWGDAMVNPPEAVNVQCSFTFTNVRIPRTNVTDAFHFVRNT 600
Subject 541	GRLSHSGAAMFDHVRWGDAMVNPPEAVNVQCSFTFTNVRIPRTNVTDAFHFVRNT 600
Query 601	DVEERTTRTQASADNEFTHLKGIGEPAGQTKAVRIIPQDIDQWSTAYLCPKGTLSH 660
Subject 601	DVEERTTRTQASADNEFTHLKGIGEPAGQTKAVRIIPQDIDQWSTAYLCPKGTLSH 660
Query 661	YADAEYQHCROIASSEMLDGIJAWRLQEDVRFHETKGGRLDPDQRIQLA 714
Subject 661	YADAEYQHCROIASSEMLDGIJAWRLQEDVRFHETKGGRLDPDQRIQLA 714

Il correspond bien au start retenu chez les homologues.

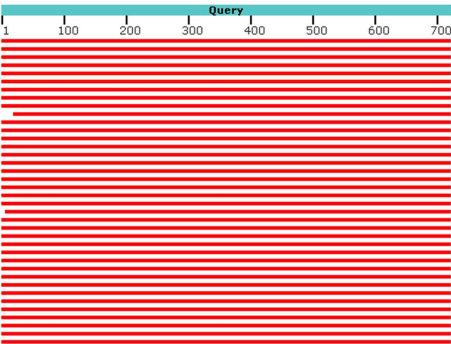
DECISION:

16508. Il est retenu chez les homologues, il conduit à l'orf le plus long. Il est prédit par les deux programmes.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																				
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <div><div><div><40</div><div>40-50</div><div>50-100</div><div>100-200</div><div>>200</div></div><div>1000</div></div> <p>Sequences producing significant alignments:</p> <table><thead><tr><th></th><th>Score</th><th>E</th></tr><tr><th></th><th>(bits)</th><th>Value</th></tr></thead><tbody><tr><td>Nucci_22, minor tail protein, 714</td><td>1507</td><td>0.0</td></tr><tr><td>Mandalorian_22, minor tail protein, 714</td><td>1507</td><td>0.0</td></tr><tr><td>Carostasia_Draft_22, function unknown, 714</td><td>1507</td><td>0.0</td></tr><tr><td>Quartz_22, minor tail protein, 714</td><td>1498</td><td>0.0</td></tr><tr><td>Juanyo_22, minor tail protein, 713</td><td>1388</td><td>0.0</td></tr><tr><td>Yuuy_22, minor tail protein, 714</td><td>1360</td><td>0.0</td></tr><tr><td>MCubed_22, minor tail protein, 713</td><td>1306</td><td>0.0</td></tr><tr><td>Saratos_22, minor tail protein, 713</td><td>1303</td><td>0.0</td></tr><tr><td>Sansa_21, minor tail protein, 713</td><td>1303</td><td>0.0</td></tr><tr><td>ChikPic_22, minor tail protein, 713</td><td>1303</td><td>0.0</td></tr><tr><td>Zenitsu_22, minor tail protein, 713</td><td>1301</td><td>0.0</td></tr><tr><td>Shamu_22, minor tail protein, 715</td><td>1301</td><td>0.0</td></tr><tr><td>Glamour_21, minor tail protein, 713</td><td>1301</td><td>0.0</td></tr><tr><td>Eleri_22, minor tail protein, 713</td><td>1301</td><td>0.0</td></tr><tr><td>ColaCorta_22, minor tail protein, 713</td><td>1301</td><td>0.0</td></tr><tr><td>Andromedas_22, minor tail protein, 713</td><td>1301</td><td>0.0</td></tr><tr><td>>Nucci_22, minor tail protein, 714</td><td></td><td></td></tr><tr><td>Length = 714</td><td></td><td></td></tr><tr><td>Score = 1507 bits (3902), Expect = 0.0</td><td></td><td></td></tr><tr><td>Identities = 714/714 (100%), Positives = 714/714 (100%)</td><td></td><td></td></tr><tr><td>Query: 1 HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS 60</td><td></td><td></td></tr><tr><td> HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS</td><td></td><td></td></tr><tr><td>Sbjct: 1 HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS 60</td><td></td><td></td></tr><tr><td>Query: 61 IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA 120</td><td></td><td></td></tr><tr><td> IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA</td><td></td><td></td></tr><tr><td>Sbjct: 61 IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA 120</td><td></td><td></td></tr><p>*** nr : (q#: s#) : [alignment] ; e-value :</p></tbody></table>		Score	E		(bits)	Value	Nucci_22, minor tail protein, 714	1507	0.0	Mandalorian_22, minor tail protein, 714	1507	0.0	Carostasia_Draft_22, function unknown, 714	1507	0.0	Quartz_22, minor tail protein, 714	1498	0.0	Juanyo_22, minor tail protein, 713	1388	0.0	Yuuy_22, minor tail protein, 714	1360	0.0	MCubed_22, minor tail protein, 713	1306	0.0	Saratos_22, minor tail protein, 713	1303	0.0	Sansa_21, minor tail protein, 713	1303	0.0	ChikPic_22, minor tail protein, 713	1303	0.0	Zenitsu_22, minor tail protein, 713	1301	0.0	Shamu_22, minor tail protein, 715	1301	0.0	Glamour_21, minor tail protein, 713	1301	0.0	Eleri_22, minor tail protein, 713	1301	0.0	ColaCorta_22, minor tail protein, 713	1301	0.0	Andromedas_22, minor tail protein, 713	1301	0.0	>Nucci_22, minor tail protein, 714			Length = 714			Score = 1507 bits (3902), Expect = 0.0			Identities = 714/714 (100%), Positives = 714/714 (100%)			Query: 1 HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS 60			HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS			Sbjct: 1 HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS 60			Query: 61 IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA 120			IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA			Sbjct: 61 IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA 120		
	Score	E																																																																																			
	(bits)	Value																																																																																			
Nucci_22, minor tail protein, 714	1507	0.0																																																																																			
Mandalorian_22, minor tail protein, 714	1507	0.0																																																																																			
Carostasia_Draft_22, function unknown, 714	1507	0.0																																																																																			
Quartz_22, minor tail protein, 714	1498	0.0																																																																																			
Juanyo_22, minor tail protein, 713	1388	0.0																																																																																			
Yuuy_22, minor tail protein, 714	1360	0.0																																																																																			
MCubed_22, minor tail protein, 713	1306	0.0																																																																																			
Saratos_22, minor tail protein, 713	1303	0.0																																																																																			
Sansa_21, minor tail protein, 713	1303	0.0																																																																																			
ChikPic_22, minor tail protein, 713	1303	0.0																																																																																			
Zenitsu_22, minor tail protein, 713	1301	0.0																																																																																			
Shamu_22, minor tail protein, 715	1301	0.0																																																																																			
Glamour_21, minor tail protein, 713	1301	0.0																																																																																			
Eleri_22, minor tail protein, 713	1301	0.0																																																																																			
ColaCorta_22, minor tail protein, 713	1301	0.0																																																																																			
Andromedas_22, minor tail protein, 713	1301	0.0																																																																																			
>Nucci_22, minor tail protein, 714																																																																																					
Length = 714																																																																																					
Score = 1507 bits (3902), Expect = 0.0																																																																																					
Identities = 714/714 (100%), Positives = 714/714 (100%)																																																																																					
Query: 1 HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS 60																																																																																					
HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS																																																																																					
Sbjct: 1 HSNRVALHHAQTVRTGVLKYGDNIDVFLVGYLLGTGGLLENLNIHNTVEAARAKWPS 60																																																																																					
Query: 61 IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA 120																																																																																					
IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA																																																																																					
Sbjct: 61 IRNHLTIQCFSSAANKAIGTPGHPLRVAVLAQMETIYATYPMLYIGIDHDAEGFGNNHTSA 120																																																																																					

Distribution of the top 100 Blast Hits on 100 subject sequences:



select all 100 sequences selected		GenPept	Graphics	Distance tree of results	Multiple alignment	MSA View		
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage Nucci]	Microbacterium phage Nucci	1472	1472	99%	0.0	100.00%	714	GD952415.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage Quantz]	Microbacterium phage Quantz	1464	1464	99%	0.0	99.44%	714	LD955241.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage Juarez]	Microbacterium phage Juarez	1367	1367	99%	0.0	90.34%	713	LDH44772.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage YauT]	Microbacterium phage YauT	1332	1332	99%	0.0	89.60%	714	GF99361.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage MCubed]	Microbacterium phage MCubed	1280	1280	99%	0.0	84.87%	713	GD93882.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage Sarsaa]	Microbacterium phage Sarsaa	1276	1276	99%	0.0	84.45%	713	ADH46934.2
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage Santos]	Microbacterium phage Santos	1274	1274	99%	0.0	84.59%	713	LDG78019.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage ChikPic]	Microbacterium phage ChikPic	1274	1274	99%	0.0	84.59%	713	VNH95916.1
<input checked="" type="checkbox"/> minor tail protein [Microbacterium phage Elex]	Microbacterium phage Elex	1273	1273	99%	0.0	84.45%	713	XP_008620388.1

minor tail protein [Microbacterium phage Nucci]

Sequence ID: [QDK02415.1](#) Length: 714 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 714		GenPept	Graphics	Next Match	Previous Match
Score	Expect	Method	Identities	Positives	Gaps
1472 bits(3812)	0.0	Compositional matrix adjust.	714/714(100%)	714/714(100%)	0/714(0%)
Query 1	MGNRVALMHHANQTVRTGVLQKYGDNIDDFLVGYLLGTGLENLNIHNTVEAARAKWPS 60				
Sbjct 1	MGNRVALMHHANQTVRTGVLQKYGDNIDDFLVGYLLGTGLENLNIHNTVEAARAKWPS 60				
Query 61	IRRWLTIQCFSSAAMKAIQTGPHPLRVAVLAQWETIYATYPWLYGIDWDAGFGNNMTSA 120				
Sbjct 61	IRRWLTIQCFSSAAMKAIQTGPHPLRVAVLAQWETIYATYPWLYGIDWDAGFGNNMTSA 120				
Query 121	EAMAGYRALGDHARSFGKKLSAALPAATEGNFSIGGENWLDYAAFAGYLDQVAIMTYDFA 180				
Sbjct 121	EAMAGYRALGDHARSFGKKLSAALPAATEGNFSIGGENWLDYAAFAGYLDQVAIMTYDFA 180				

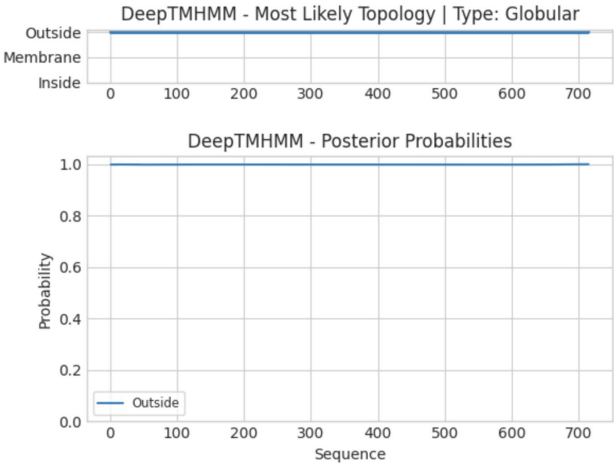
Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Number of Hits: 250
Query MSA diversity (Neff): 9.46861

Visualization



Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<i>non</i>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<i>non</i>  <p><i>Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebeius par ex).</i></p>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Oui</i> .
DECISION:	<i>Minor tail protein</i>

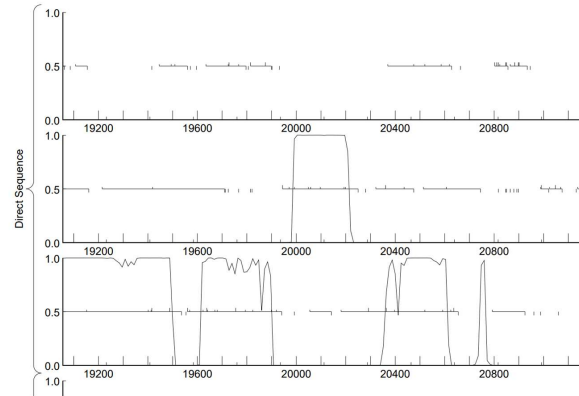
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	23
Coordonnées du Stop	19541
Direction (For/Rev)	FOR
Gap/chevauchement avec un autre gène	Gap 25
Coordonnées du Start retenu	18678
Fonction prédite	endolysine, protease M15 domain

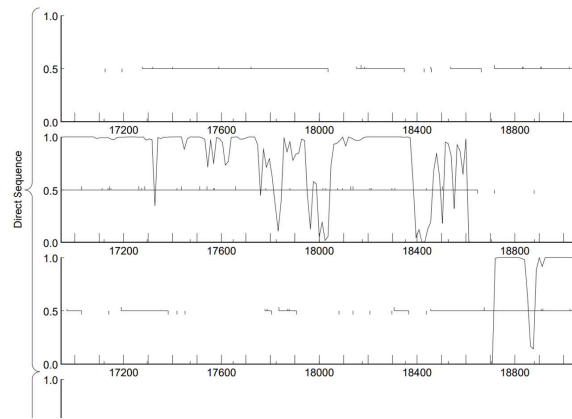
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Cadre +3</i>

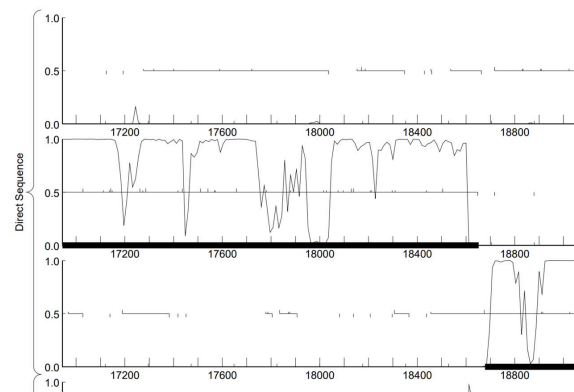
erium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 11/21

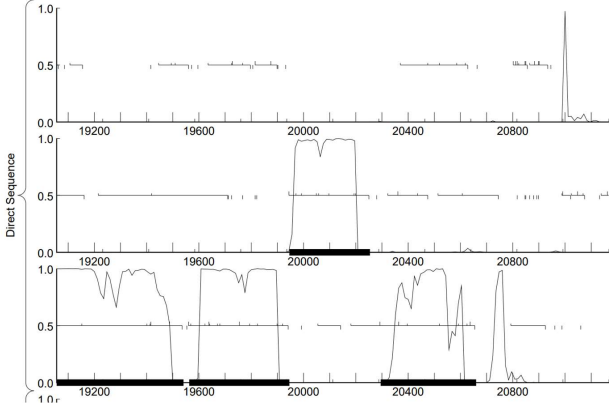


erium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 10/21



erium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 10/21



																																					
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Nucci_23, lysin A, 287</td><td>615</td><td>e-176</td></tr><tr><td>Mandalorian_23, endolysin, 287</td><td>615</td><td>e-176</td></tr><tr><td>Carostasia_Draft_23, function unknown, 287</td><td>615</td><td>e-176</td></tr><tr><td>Quartz_23, endolysin, 287</td><td>611</td><td>e-175</td></tr><tr><td>YuuY_23, lysin A, 287</td><td>581</td><td>e-166</td></tr><tr><td>Saratos_23, endolysin, 260</td><td>368</td><td>e-101</td></tr><tr><td>Zenitsu_23, endolysin, 260</td><td>367</td><td>e-101</td></tr><tr><td>Shamu_Draft_24, function unknown, 260</td><td>367</td><td>e-101</td></tr><tr><td>Glamour_22, lysin A, 260</td><td>367</td><td>e-101</td></tr><tr><td>Eleri_23, lysin A, 260</td><td>367</td><td>e-101</td></tr><tr><td>Colafacta_23, lysin A, 260</td><td>367</td><td>e-101</td></tr></tbody></table>		Score (bits)	E Value	Nucci_23, lysin A, 287	615	e-176	Mandalorian_23, endolysin, 287	615	e-176	Carostasia_Draft_23, function unknown, 287	615	e-176	Quartz_23, endolysin, 287	611	e-175	YuuY_23, lysin A, 287	581	e-166	Saratos_23, endolysin, 260	368	e-101	Zenitsu_23, endolysin, 260	367	e-101	Shamu_Draft_24, function unknown, 260	367	e-101	Glamour_22, lysin A, 260	367	e-101	Eleri_23, lysin A, 260	367	e-101	Colafacta_23, lysin A, 260	367	e-101
	Score (bits)	E Value																																			
Nucci_23, lysin A, 287	615	e-176																																			
Mandalorian_23, endolysin, 287	615	e-176																																			
Carostasia_Draft_23, function unknown, 287	615	e-176																																			
Quartz_23, endolysin, 287	611	e-175																																			
YuuY_23, lysin A, 287	581	e-166																																			
Saratos_23, endolysin, 260	368	e-101																																			
Zenitsu_23, endolysin, 260	367	e-101																																			
Shamu_Draft_24, function unknown, 260	367	e-101																																			
Glamour_22, lysin A, 260	367	e-101																																			
Eleri_23, lysin A, 260	367	e-101																																			
Colafacta_23, lysin A, 260	367	e-101																																			
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Non</i>																																				
DECISION:	<i>Oui</i>																																				

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																												
Quel start Glimmer et GeneMark suggèrent-ils ?	18678																																																																																																												
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>OUI</div> <div><div>Starts : 10 Selected : 1</div><div>ORF Start : 19404 ORF Stop : 19541 ORF Length : 138</div><div>Cdn1 Cdn2 Cdn3 Length 45.2 79.5 61.6 219</div><div>SD Scoring Matrix Spacing Weight Matrix</div><div>Kibler6 Karlin Medium</div><div>Explore Document</div></div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-5,224</td><td>1,355</td><td>11</td><td>-5,981</td><td>TGGGTGAGCAAGGCATACCTGT</td><td>GTG</td><td>18459</td><td>1083</td></tr><tr><td>2</td><td>-1,236</td><td>3,221</td><td>11</td><td>-1,593</td><td>CGGCACACTAAGGAGAACGAA</td><td>ATG</td><td>18678</td><td>864</td></tr><tr><td>3</td><td>-3,808</td><td>2,017</td><td>6</td><td>-5,553</td><td>CAACCACTCCGGAGAGGTGGT</td><td>GTG</td><td>18912</td><td>630</td></tr><tr><td>4</td><td>-3,861</td><td>1,993</td><td>9</td><td>-4,635</td><td>CTCCCGGACGGTGGTGGGCC</td><td>GTG</td><td>18918</td><td>624</td></tr><tr><td>5</td><td>-2,994</td><td>2,398</td><td>11</td><td>-3,751</td><td>CTTCAGGGGCTGGAGCGTGGC</td><td>GTG</td><td>19029</td><td>513</td></tr><tr><td>6</td><td>-5,508</td><td>1,222</td><td>11</td><td>-6,265</td><td>GGCTACCAAGACTTCTCTGGC</td><td>GTG</td><td>19155</td><td>387</td></tr><tr><td>7</td><td>-3,420</td><td>2,105</td><td>10</td><td>-4,315</td><td>CTGGGGCTACAAGGCAACGAC</td><td>GTG</td><td>19404</td><td>138</td></tr><tr><td>8</td><td>-5,369</td><td>1,287</td><td>13</td><td>-6,414</td><td>GGGCAACGACGTGCTGGTCCG</td><td>GTG</td><td>19416</td><td>126</td></tr><tr><td>9</td><td>-5,369</td><td>1,287</td><td>16</td><td>-7,165</td><td>CAACGACGTGCTGGTCCGGTG</td><td>ATG</td><td>19419</td><td>123</td></tr><tr><td>10</td><td>-4,975</td><td>1,471</td><td>9</td><td>-5,750</td><td>CAACGACACCCCGGTCCGGTG</td><td>ATG</td><td>19491</td><td>51</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-5,224	1,355	11	-5,981	TGGGTGAGCAAGGCATACCTGT	GTG	18459	1083	2	-1,236	3,221	11	-1,593	CGGCACACTAAGGAGAACGAA	ATG	18678	864	3	-3,808	2,017	6	-5,553	CAACCACTCCGGAGAGGTGGT	GTG	18912	630	4	-3,861	1,993	9	-4,635	CTCCCGGACGGTGGTGGGCC	GTG	18918	624	5	-2,994	2,398	11	-3,751	CTTCAGGGGCTGGAGCGTGGC	GTG	19029	513	6	-5,508	1,222	11	-6,265	GGCTACCAAGACTTCTCTGGC	GTG	19155	387	7	-3,420	2,105	10	-4,315	CTGGGGCTACAAGGCAACGAC	GTG	19404	138	8	-5,369	1,287	13	-6,414	GGGCAACGACGTGCTGGTCCG	GTG	19416	126	9	-5,369	1,287	16	-7,165	CAACGACGTGCTGGTCCGGTG	ATG	19419	123	10	-4,975	1,471	9	-5,750	CAACGACACCCCGGTCCGGTG	ATG	19491	51
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																					
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																					
1	-5,224	1,355	11	-5,981	TGGGTGAGCAAGGCATACCTGT	GTG	18459	1083																																																																																																					
2	-1,236	3,221	11	-1,593	CGGCACACTAAGGAGAACGAA	ATG	18678	864																																																																																																					
3	-3,808	2,017	6	-5,553	CAACCACTCCGGAGAGGTGGT	GTG	18912	630																																																																																																					
4	-3,861	1,993	9	-4,635	CTCCCGGACGGTGGTGGGCC	GTG	18918	624																																																																																																					
5	-2,994	2,398	11	-3,751	CTTCAGGGGCTGGAGCGTGGC	GTG	19029	513																																																																																																					
6	-5,508	1,222	11	-6,265	GGCTACCAAGACTTCTCTGGC	GTG	19155	387																																																																																																					
7	-3,420	2,105	10	-4,315	CTGGGGCTACAAGGCAACGAC	GTG	19404	138																																																																																																					
8	-5,369	1,287	13	-6,414	GGGCAACGACGTGCTGGTCCG	GTG	19416	126																																																																																																					
9	-5,369	1,287	16	-7,165	CAACGACGTGCTGGTCCGGTG	ATG	19419	123																																																																																																					
10	-4,975	1,471	9	-5,750	CAACGACACCCCGGTCCGGTG	ATG	19491	51																																																																																																					
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start ne conduit pas à l'ORF la plus longue, cependant le start qui conduit à l'ORF le plus long est un GTG situé en position 18459. Celui-ci conduit à un chevauchement de 192pb avec le gène précédent.																																																																																																												
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																																													
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div>>Nucci_23, lysin A, 287 Length = 287</div> <div>Score = 615 bits (1585), Expect = e-176 Identities = 287/287 (100%), Positives = 287/287 (100%)</div> <div>Query: 1 MATLKNHPGYNLRDDAAAFDAYEAKYGKRTVNSAGRTKAEQQNLINRW DAGGAANRPPY 60 MATLKNHPGYNLRDDAAAFDAYEAKYGKRTVNSAGRTKAEQQNLINRW DAGGAANRPPY Sbjct: 1 MATLKNHPGYNLRDDAAAFDAYEAKYGKRTVNSAGRTKAEQQNLINRW DAGGAANRPPY 60</div> <div></div>																																																																																																												
DECISION:	Le start conservé est 18678 car il n'est pas en contraction avec les principes d'annotation et il est conservé chez d'autres homologues.																																																																																																												

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

*** PhagesDB :

Sequences producing significant alignments:	Score (bits)	E Value
Nucci_23, lysin A, 287	615	e-176
Hendalocaris_23, endolysin, 287	615	e-176
Carostasia_Draft_23, function unknown, 287	615	e-176
Quartz_23, endolysin, 287	611	e-175
Yaut_23, lysin A, 287	581	e-166
Saratou_23, endolysin, 260	268	e-101
Senitus_23, endolysin, 260	267	e-101
Shamo_23, endolysin, 260	267	e-101
Glammur_22, lysin A, 260	267	e-101
Eleri_23, lysin A, 260	267	e-101
ColaCorta_23, lysin A, 260	267	e-101
Chloric_23, endolysin, 260	267	e-101
Andromedas_23, lysin A, 260	267	e-101
Pinsy_23, lysin A, 260	267	e-101
Mcubed_23, lysin A, 260	267	e-101
Sanna_22, lysin A, 260	257	1e-98
Juanyo_23, endolysin, 257	245	6e-95
Count_3, endolysin, 241	230	2e-66
Eden_20, endolysin, 256	227	3e-62
Gack_Draft_20, function unknown, 256	232	7e-61
Franklin22_21, endolysin, 256	230	3e-60
Trisemit_27, lysin A, 312	228	2e-59
Zada_24, lysin A, 268	225	8e-59
Robinson_24, lysin A, 268	225	8e-59
Alyxandrew_24, lysin A, 268	225	8e-59
WildOut_24, lysin A, 268	224	2e-58
Strachdee_24, lysin A, 268	224	2e-58
Solitrelin_Draft_24, function unknown, 268	224	2e-58
Riyhil_24, lysin A, 268	224	2e-58
Raptor_24, lysin A, 268	224	2e-58
Phriends_24, lysin A, 268	224	2e-58
Phireproof_24, endolysin, 268	224	2e-58

Image generated by NCBI (287 letters)

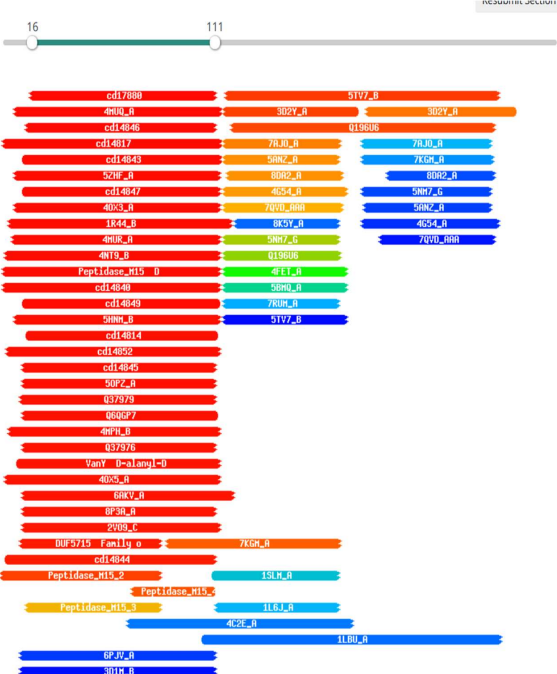
Distribution of 181 Blast Hits on the Query Sequence

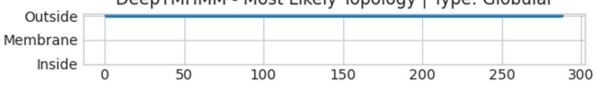
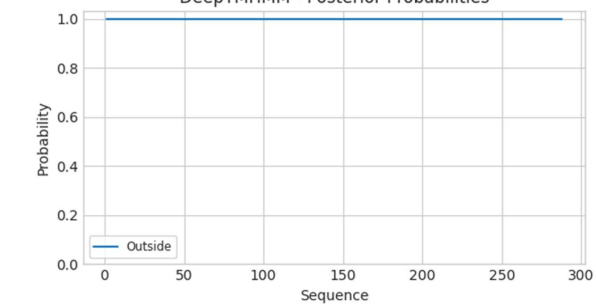
*** nr :

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Isin A [Microbacterium phage Nuori]	Microbacterium phage Nuori	588	588	0%	150.000	287	62032818.1	
<input checked="" type="checkbox"/> endolysin [Microbacterium phage Quarte]	Microbacterium phage Quarte	585	585	99%	0%	98.00%	287	U959262.1
<input checked="" type="checkbox"/> Isin A [Microbacterium phage Tauf]	Microbacterium phage Tauf	538	538	99%	0%	98.10%	287	GF575362.1
<input checked="" type="checkbox"/> Isin A [Microbacterium phage Finny]	Microbacterium phage Finny	335	335	87%	5e-112	70.50%	280	GF0715762.1
<input checked="" type="checkbox"/> endolysin [Microbacterium phage Sarcate]	Microbacterium phage Sarcate	335	335	87%	5e-112	70.50%	280	UG070885.1
<input checked="" type="checkbox"/> endolysin [Microbacterium phage Elaci]	Microbacterium phage Elaci	335	335	87%	5e-112	70.50%	280	VF_00952286.1
<input checked="" type="checkbox"/> Isin A [Microbacterium phage MC-Jared]	Microbacterium phage MC-Jared	335	335	87%	5e-112	70.50%	280	GU503961.1
<input checked="" type="checkbox"/> Isin A [Microbacterium phage Seme]	Microbacterium phage Seme	327	327	87%	8e-109	68.87%	280	AK060555.1
<input checked="" type="checkbox"/> endolysin [Microbacterium phage Juireal]	Microbacterium phage Juireal	315	315	64%	3e-104	84.49%	257	U5248772.1
<input checked="" type="checkbox"/> endolysin-binding protein [Microbacterium sp. XT11]	Microbacterium sp. XT11	254	254	65%	3e-40	58.98%	280	WP_06720134.1
<input checked="" type="checkbox"/> endolysin-binding protein [Microbacterium sp. BG28]	Microbacterium sp. BG28	251	251	96%	8e-79	51.77%	273	WP_32043296.1
<input checked="" type="checkbox"/> endolysin [Microbacterium phage pBAC2003]	Microbacterium phage pBAC2003	233	233	99%	3e-48	48.84%	241	U915355.1
<input checked="" type="checkbox"/> endolysin [Microbacterium phage Court]	Microbacterium phage Court	228	228	99%	3e-48	65.54%	241	AK060326.1
<input checked="" type="checkbox"/> Isin A [Microbacterium phage Robesart]	Microbacterium phage Robesart	214	214	88%	1e-64	47.08%	258	AF058504.1

Image generated by NCBI (287 letters)

Distribution of the top 180 Blast Hits on 180 subject sequences

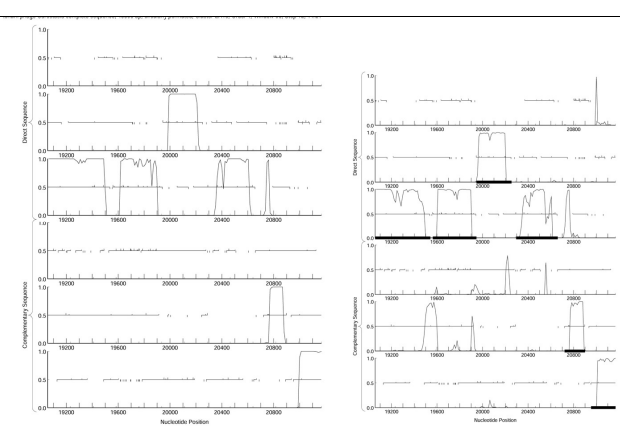
	<p>Les hits trouvés sont des lysine et endolysines, avec un query cover >85% pour une grande partie, et une e value <<<10⁻⁴. Le meilleur hit est une lysine A chez le phage Nucci.</p>
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?</p>	<p>Le meilleur hit est une D-Ala-D-Ala_dipeptidase; D-Ala-D-Ala Dipeptidase avec une probabilité de 99,21%. Des hits similaires sont trouvés avec une probabilité de plus de 90%.</p> 

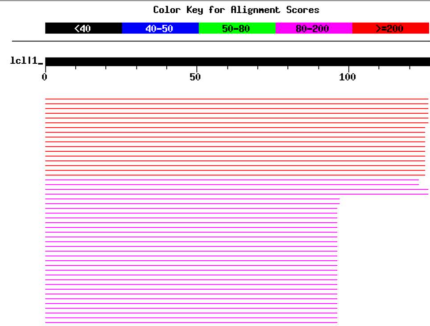
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p><i>Non, elle est prédite comme globulaire.</i></p> <p>DeepTMHMM - Most Likely Topology Type: Globular</p>  <p>DeepTMHMM - Posterior Probabilities</p> 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<p><i>Oui</i></p>
DECISION:	<p><i>Ce gène code pour une endolysin, protease M15 domain</i></p>

Student Gene Annotation Worksheet


Basic Phage Information	
Nom du Phage	Carostasia
Gène #	24
Coordonnées du Stop	19946
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	gap 21
Coordonnées du Start retenu	19563
Fonction prédite	membrane protein

Décision #1 : Est-ce un gène ?

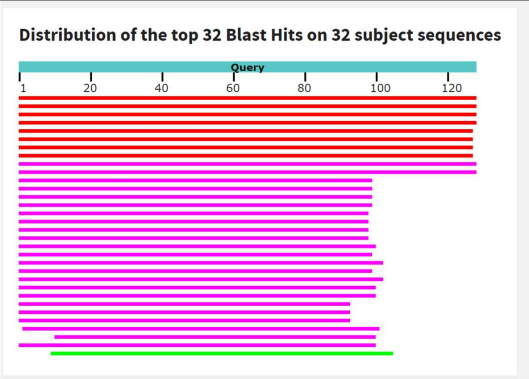
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	 <p>Potentiel codant observé chez les 2</p>

	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <div><div>377374 Sansa_23, function unknown, 127..S= 206 E=1e-53</div><div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-60</div><div>60-200</div><div>2000</div></div></div><p>Sequences producing significant alignments:</p><table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Nucci_24, function unknown, 127</td><td>244</td><td>7e-65</td></tr><tr><td>Mandalorian_24, function unknown, 127</td><td>244</td><td>7e-65</td></tr><tr><td>Carostasia_Draft_24, function unknown, 127</td><td>244</td><td>7e-65</td></tr><tr><td>Quartz_24, function unknown, 127</td><td>240</td><td>1e-63</td></tr><tr><td>Juanyo_24, function unknown, 127</td><td>216</td><td>2e-56</td></tr><tr><td>YuuY_24, function unknown, 127</td><td>213</td><td>1e-55</td></tr><tr><td>Zenitsu_24, function unknown, 127</td><td>202</td><td>1e-54</td></tr><tr><td>MCubed_24, function unknown, 127</td><td>202</td><td>3e-54</td></tr><tr><td>Finny_24, function unknown, 127</td><td>202</td><td>3e-54</td></tr><tr><td>Sansa_23, function unknown, 127</td><td>206</td><td>1e-53</td></tr><tr><td>Shamu_Draft_25, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>Saratos_24, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>Glamour_23, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>Eleri_24, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>ColaCorta_24, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>ChikPic_24, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>Andromedas_24, function unknown, 127</td><td>206</td><td>2e-53</td></tr><tr><td>JasonD Draft 29, function unknown, 137</td><td>146</td><td>2e-35</td></tr></tbody></table><p>Homologue du gène chez d'autres phages</p></div>		Score (bits)	E Value	Nucci_24, function unknown, 127	244	7e-65	Mandalorian_24, function unknown, 127	244	7e-65	Carostasia_Draft_24, function unknown, 127	244	7e-65	Quartz_24, function unknown, 127	240	1e-63	Juanyo_24, function unknown, 127	216	2e-56	YuuY_24, function unknown, 127	213	1e-55	Zenitsu_24, function unknown, 127	202	1e-54	MCubed_24, function unknown, 127	202	3e-54	Finny_24, function unknown, 127	202	3e-54	Sansa_23, function unknown, 127	206	1e-53	Shamu_Draft_25, function unknown, 127	206	2e-53	Saratos_24, function unknown, 127	206	2e-53	Glamour_23, function unknown, 127	206	2e-53	Eleri_24, function unknown, 127	206	2e-53	ColaCorta_24, function unknown, 127	206	2e-53	ChikPic_24, function unknown, 127	206	2e-53	Andromedas_24, function unknown, 127	206	2e-53	JasonD Draft 29, function unknown, 137	146	2e-35
	Score (bits)	E Value																																																								
Nucci_24, function unknown, 127	244	7e-65																																																								
Mandalorian_24, function unknown, 127	244	7e-65																																																								
Carostasia_Draft_24, function unknown, 127	244	7e-65																																																								
Quartz_24, function unknown, 127	240	1e-63																																																								
Juanyo_24, function unknown, 127	216	2e-56																																																								
YuuY_24, function unknown, 127	213	1e-55																																																								
Zenitsu_24, function unknown, 127	202	1e-54																																																								
MCubed_24, function unknown, 127	202	3e-54																																																								
Finny_24, function unknown, 127	202	3e-54																																																								
Sansa_23, function unknown, 127	206	1e-53																																																								
Shamu_Draft_25, function unknown, 127	206	2e-53																																																								
Saratos_24, function unknown, 127	206	2e-53																																																								
Glamour_23, function unknown, 127	206	2e-53																																																								
Eleri_24, function unknown, 127	206	2e-53																																																								
ColaCorta_24, function unknown, 127	206	2e-53																																																								
ChikPic_24, function unknown, 127	206	2e-53																																																								
Andromedas_24, function unknown, 127	206	2e-53																																																								
JasonD Draft 29, function unknown, 137	146	2e-35																																																								
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non																																																									
DECISION:	Oui																																																									

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																														
Quel start Glimmer et GeneMark suggèrent-ils ?	19563 pour les 2																																																																																																																														
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,934</td><td>2,427</td><td>10</td><td>-3,628</td><td>ACCACACCTGGAGGGGTAGATA</td><td>ATG</td><td>19563</td><td>384</td></tr><tr><td>2</td><td>-4,829</td><td>1,539</td><td>17</td><td>-6,829</td><td>GGAGGGGTAGATAATGCTGGAA</td><td>GTG</td><td>19572</td><td>375</td></tr><tr><td>3</td><td>-7,532</td><td>0,274</td><td>10</td><td>-8,227</td><td>ACTCTTCGTGCGCGTCATTACG</td><td>GTG</td><td>19626</td><td>321</td></tr><tr><td>4</td><td>-3,254</td><td>2,277</td><td>9</td><td>-4,029</td><td>CATTACGGTGTACGGGGGTGTC</td><td>ATG</td><td>19641</td><td>306</td></tr><tr><td>5</td><td>-3,254</td><td>2,277</td><td>12</td><td>-4,090</td><td>TACGGTGTACGGGGGTGTCATG</td><td>GTG</td><td>19644</td><td>303</td></tr><tr><td>6</td><td>-3,662</td><td>2,086</td><td>13</td><td>-4,708</td><td>GATCAACAAGGTCAAGGCCAAC</td><td>GTG</td><td>19674</td><td>273</td></tr><tr><td>7</td><td>-2,955</td><td>2,416</td><td>7</td><td>-4,478</td><td>GGTCAAGGCCACGTGGAGAG</td><td>GTG</td><td>19683</td><td>264</td></tr><tr><td>8</td><td>-7,189</td><td>0,435</td><td>11</td><td>-7,946</td><td>CGATCGGCTCACCCTCGCGTC</td><td>ATG</td><td>19758</td><td>189</td></tr><tr><td>9</td><td>-6,213</td><td>0,892</td><td>13</td><td>-7,259</td><td>TCAGGACGAGCTCATCTGGAC</td><td>GTG</td><td>19797</td><td>150</td></tr><tr><td>10</td><td>-3,867</td><td>1,990</td><td>16</td><td>-5,663</td><td>GCTGAAGGCGCATGACGCTGAA</td><td>GTG</td><td>19827</td><td>120</td></tr><tr><td>11</td><td>-6,832</td><td>0,602</td><td>11</td><td>-7,588</td><td>TGCTCCGACCATCCCGGTA</td><td>GTG</td><td>19902</td><td>45</td></tr><tr><td>12</td><td>-6,663</td><td>0,681</td><td>14</td><td>-8,009</td><td>AGTGAAGCGGTTCCACGCCCTC</td><td>GTG</td><td>19923</td><td>24</td></tr></table> <p>Le start prédit à le meilleur score de RBS : -3.628. Ce start correspond à un ATG</p>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,934	2,427	10	-3,628	ACCACACCTGGAGGGGTAGATA	ATG	19563	384	2	-4,829	1,539	17	-6,829	GGAGGGGTAGATAATGCTGGAA	GTG	19572	375	3	-7,532	0,274	10	-8,227	ACTCTTCGTGCGCGTCATTACG	GTG	19626	321	4	-3,254	2,277	9	-4,029	CATTACGGTGTACGGGGGTGTC	ATG	19641	306	5	-3,254	2,277	12	-4,090	TACGGTGTACGGGGGTGTCATG	GTG	19644	303	6	-3,662	2,086	13	-4,708	GATCAACAAGGTCAAGGCCAAC	GTG	19674	273	7	-2,955	2,416	7	-4,478	GGTCAAGGCCACGTGGAGAG	GTG	19683	264	8	-7,189	0,435	11	-7,946	CGATCGGCTCACCCTCGCGTC	ATG	19758	189	9	-6,213	0,892	13	-7,259	TCAGGACGAGCTCATCTGGAC	GTG	19797	150	10	-3,867	1,990	16	-5,663	GCTGAAGGCGCATGACGCTGAA	GTG	19827	120	11	-6,832	0,602	11	-7,588	TGCTCCGACCATCCCGGTA	GTG	19902	45	12	-6,663	0,681	14	-8,009	AGTGAAGCGGTTCCACGCCCTC	GTG	19923	24
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																							
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																							
1	-2,934	2,427	10	-3,628	ACCACACCTGGAGGGGTAGATA	ATG	19563	384																																																																																																																							
2	-4,829	1,539	17	-6,829	GGAGGGGTAGATAATGCTGGAA	GTG	19572	375																																																																																																																							
3	-7,532	0,274	10	-8,227	ACTCTTCGTGCGCGTCATTACG	GTG	19626	321																																																																																																																							
4	-3,254	2,277	9	-4,029	CATTACGGTGTACGGGGGTGTC	ATG	19641	306																																																																																																																							
5	-3,254	2,277	12	-4,090	TACGGTGTACGGGGGTGTCATG	GTG	19644	303																																																																																																																							
6	-3,662	2,086	13	-4,708	GATCAACAAGGTCAAGGCCAAC	GTG	19674	273																																																																																																																							
7	-2,955	2,416	7	-4,478	GGTCAAGGCCACGTGGAGAG	GTG	19683	264																																																																																																																							
8	-7,189	0,435	11	-7,946	CGATCGGCTCACCCTCGCGTC	ATG	19758	189																																																																																																																							
9	-6,213	0,892	13	-7,259	TCAGGACGAGCTCATCTGGAC	GTG	19797	150																																																																																																																							
10	-3,867	1,990	16	-5,663	GCTGAAGGCGCATGACGCTGAA	GTG	19827	120																																																																																																																							
11	-6,832	0,602	11	-7,588	TGCTCCGACCATCCCGGTA	GTG	19902	45																																																																																																																							
12	-6,663	0,681	14	-8,009	AGTGAAGCGGTTCCACGCCCTC	GTG	19923	24																																																																																																																							
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>A partir du start prédit qui à le meilleur score de RBS on a une ORF de 384 pb, et il s'agit de l'orf le plus long</p>																																																																																																																														
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Gene: Carostasia_24 Start: 19563, Stop: 19946, Start Num: 11 Candidate Starts for Carostasia_24: (Start: 11 @19563 has 44 MA's), (13, 19572), (22, 19626), (24, 19641), (25, 19644), (30, 19674), (31, 19683), (34, 19758), (37, 19797), (40, 19827), (49, 19902), (51, 19923),</p> 																																																																																																																														

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?



membrane protein [Microbacterium phage Nucci]

Sequence ID: [QDK02417.1](#) Length: 127 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 127 [GenPept](#) [Graphics](#) [Next Match](#) [Prev](#)

Score	Expect	Method	Identities	Positives	Gaps
249 bits(636)	5e-83	Compositional matrix adjust.	127/127(100%)	127/127(100%)	0/127(0%)
Query 1	MLEVIERIGAVLTPLFVAVITVYGGVMVAKINKVKANVEKVSQDIITNHGSKNIGDAIDR				60
Sbjct 1	MLEVIERIGAVLTPLFVAVITVYGGVMVAKINKVKANVEKVSQDIITNHGSKNIGDAIDR				60
Query 61	LTTRVMVISDNQQQLILDVKALKAHDAEVDTRLNIDIKSSQDTKDAVSDTIRVVKPFHAL				120
Sbjct 61	LTTRVMVISDNQQQLILDVKALKAHDAEVDTRLNIDIKSSQDTKDAVSDTIRVVKPFHAL				120
Query 121	VDKLGK 127				
Sbjct 121	VDKLGK 127				

[Download](#) [GenPept](#) [Graphics](#)

membrane protein [Microbacterium phage Quartz]

Sequence ID: [UVK59243.1](#) Length: 127 Number of Matches: 1

Range 1: 1 to 127 [GenPept](#) [Graphics](#) [Next Match](#) [Prev](#)

Score	Expect	Method	Identities	Positives	Gaps
244 bits(624)	4e-81	Compositional matrix adjust.	125/127(98%)	126/127(99%)	0/127(0%)
Query 1	MLEVIERIGAVLTPLFVAVITVYGGVMVAKINKVKANVEKVSQDIITNHGSKNIGDAIDR				60
Sbjct 1	MLEVIERIGAVLTPLFVAVITVYGGVMVAKINKVKANVEKVSQDIITNHGSKNIGDAIDR				60
Query 61	LTTRVMVISDNQQQLILDVKALKAHDAEVDTRLNIDIKSSQDTKDAVSDTIRVVKPFHAL				120
Sbjct 61	LTTRVMVISDNQQQLILDVKALKAHDAEVDTRLNIDIKSSQDTKDAVSDTIRVVKPFHAL				120
Query 121	VDKLGK 127				
Sbjct 121	VDKLGK 127				

DECISION:

Start 19563 retenu est celui proposé, meilleur score rbs, correspond au règles, bon potentiel codant

Décision #3 : Quelle est la fonction de la protéine putative ?

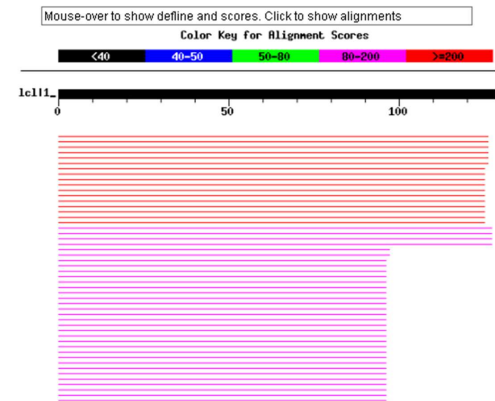
Collection des éléments de réponse

Rationnelle

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :
(q#: s#) : [alignment] ; e-value :
[Distribution of 100 Blast Hits on the Query Sequence](#)



Sequences producing significant alignments:		Score (bits)	E Value
Nucci_24, function unknown, 127		244	7e-65
Mandalorian_24, function unknown, 127		244	7e-65
Carostasia_Draft_24, function unknown, 127		244	7e-65
Quartz_24, function unknown, 127		240	1e-63
Juanyo_24, function unknown, 127		216	2e-56
YuuY_24, function unknown, 127		213	1e-55
Zenitsu_24, function unknown, 127		209	1e-54
HCubed_24, function unknown, 127		209	3e-54
Finny_24, function unknown, 127		209	3e-54
Sansa_23, function unknown, 127		206	1e-53
Shamu_24, function unknown, 127		206	2e-53
Saratos_24, function unknown, 127		206	2e-53
Glamour_23, function unknown, 127		206	2e-53
Eleri_24, function unknown, 127		206	2e-53
ColaCorta_24, function unknown, 127		206	2e-53
ChikPic_24, function unknown, 127		206	2e-53
Andromedas_24, function unknown, 127		206	2e-53
JasonD_Draft_29, function unknown, 137		148	5e-36
Acosta_Draft_26, function unknown, 137		145	3e-35

>Mandalorian_24, function unknown, 127
Length = 127

Score = 244 bits (622), Expect = 7e-65
Identities = 127/127 (100%), Positives = 127/127 (100%)

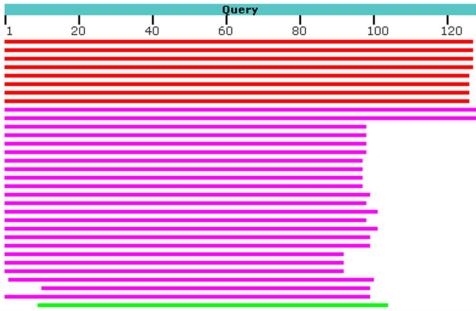
Query: 1	MLEVIERIGAVLTPLFVAVITVYGGVMHAKINKVKANVEKVQSDIITNHGSKNIGDAIDR 60
	MLEVIERIGAVLTPLFVAVITVYGGVMHAKINKVKANVEKVQSDIITNHGSKNIGDAIDR
Sbjct: 1	MLEVIERIGAVLTPLFVAVITVYGGVMHAKINKVKANVEKVQSDIITNHGSKNIGDAIDR 60

Query: 61	LTTRVHVISDNQDQLILDVKALKAHDAEVDTRLDNIDKSSQDTKDAVSDTIRVVKPFHAL 120
	LTTRVHVISDNQDQLILDVKALKAHDAEVDTRLDNIDKSSQDTKDAVSDTIRVVKPFHAL
Sbjct: 61	LTTRVHVISDNQDQLILDVKALKAHDAEVDTRLDNIDKSSQDTKDAVSDTIRVVKPFHAL 120

Query: 121	VDKLGK 127
	VDKLGK
Sbjct: 121	VDKLGK 127

*** nr :
(q#: s#) : [alignment] ; e-value :

Distribution of the top 32 Blast Hits on 32 subject sequences



select all 32 sequences selected									
		GenPept	Graphics	Distance tree of results		Multiple alignment		MSA Views	
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Nucci]	Microbacterium phage Nucci	249	249	99%	6e-83	100.00%	127	GQK02417.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Quetz]	Microbacterium phage Quetz	244	244	99%	6e-81	96.43%	127	LV650243.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Juarez]	Microbacterium phage Juarez	220	220	99%	2e-71	86.83%	127	U0H44774.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage YauY]	Microbacterium phage YauY	217	217	99%	3e-70	86.04%	127	GFF596383.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Zentou]	Microbacterium phage Zentou	213	213	98%	2e-68	84.92%	127	W0803826.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Finny]	Microbacterium phage Finny	212	212	98%	4e-68	84.13%	127	GQF16753.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Sassa]	Microbacterium phage Sassa	209	209	98%	3e-67	83.33%	127	G0148506.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Eliot]	Microbacterium phage Eliot	209	209	98%	3e-67	83.33%	127	JF_00672962.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Jesso]	Microbacterium phage Jesso	148	148	100%	1e-42	56.30%	137	Q0014496.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage McGibbon]	Microbacterium phage McGibbon	147	147	100%	1e-42	56.30%	137	JF_00990802.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Kaula]	Microbacterium phage Kaula	122	122	76%	7e-33	60.00%	128	P0631037.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Chertoff]	Microbacterium phage Chertoff	120	120	76%	7e-32	63.27%	130	U0334892.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Chertoff]	Microbacterium phage Chertoff	120	120	76%	1e-31	55.19%	128	G0272076.1
<input checked="" type="checkbox"/>	membrane protein [Microbacterium phage Chickadee]	Microbacterium phage Chickadee	110	110	76%	3e-31	62.24%	131	GQ008605.1

membrane protein [Microbacterium phage Nucci]

Sequence ID: [QDK02417.1](#) Length: 127 Number of Matches: 1

[See 1 more title\(s\)](#) [See all identical Proteins \(IPG\)](#)

Range 1: 1 to 127 [GenPept](#) [Graphics](#)

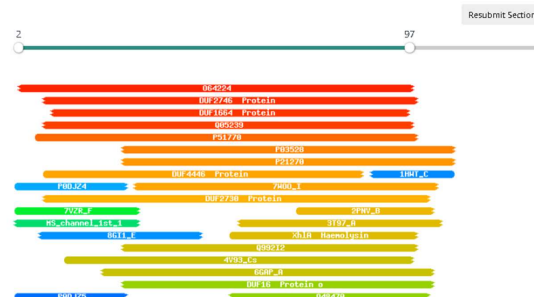
[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
249 bits(636)	5e-83	Compositional matrix adjust.	127/127(100%)	127/127(100%)	0/127(0%)
Query	1	MLEVIERIGAVLTPLFVAVITVYGGVHVAKI NKVKANVEKVQSDIITNHGSKNIGDAIDR			60
Sbjct	1	MLEVIERIGAVLTPLFVAVITVYGGVHVAKI NKVKANVEKVQSDIITNHGSKNIGDAIDR			60
Query	61	LTTRVWVISDWQDQLLDVKALKAHDAEVDTRLDNI DKSSQDTKDVS DTRVVKP PHAL			120
Sbjct	61	LTTRVWVISDWQDQLLDVKALKAHDAEVDTRLDNI DKSSQDTKDVS DTRVVKP PHAL			120
Query	121	VDKLGK			127
Sbjct	121	VDKLGK			127

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Number of Hits: **250**
Query MSA diversity (Neff): **2.08253**
Detected sequence features: **•Transmembrane segment(s)**

Visualization



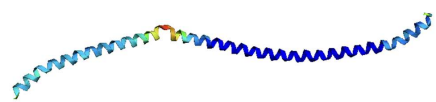
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cals	Target Length
1	Q64224	VG30_BPMDD2 Gene 30 protein O5=Mycobacterium phage D29 OX=28369 GN=30 PE=4 SV=1	96.89	0.046	40.43	9.5	71	88
2	PF10874.12	; DUF2746; Protein of unknown function (DUF2746)	96.1	0.39	36.39	10.4	86	101
3	PF07899.16	; DUF1664; Protein of unknown function (DUF1664)	95.48	0.42	33.47	7.9	84	122
4	Q05239	VG30_BPMLS5 Gene 30 protein O5=Mycobacterium phage L5 OX=31757 GN=30 PE=4 SV=1	94.64	2.3	32.29	10.2	80	101
5	P51770	SPAN1_BPP2 Probable spanin, inner membrane subunit O5=Escherichia phage P2 OX=10679 GN=lysB PE=3 SV=1	91.08	9.1	28.24	8.7	93	141
6	P03528	SIGMA1_P6OXD Outer capsid protein sigma-1 O5=Reovirus type 3 (strain Dearling) OX=10886 GN=S1 PE=1 SV=3	88.89	4.7	37.95	7.3	75	455

1. **O64224** VG30_BPMD2 Gene 30 protein OS=Mycobacterium phage D29 OX=28369 GN=30 PE=1 SV=1
Probability: 96.89%, E-value: 0.046, Score: 40.43, Aligned cols: 71, Identities: 24%, Similarity: 0.351, Template Neff: 5.5

[illegible]

	<p>Prediction of UNNAMED</p> <table><tr><td>ID</td><td>UNNAMED</td><td></td><td></td><td></td></tr><tr><td>FT</td><td>TOPO_DOM</td><td>1</td><td>11</td><td>NON CYTOPLASMIC.</td></tr><tr><td>FT</td><td>TRANSMEM</td><td>12</td><td>31</td><td></td></tr><tr><td>FT</td><td>TOPO_DOM</td><td>32</td><td>128</td><td>CYTOPLASMIC.</td></tr><tr><td>//</td><td></td><td></td><td></td><td></td></tr></table> <p>Phobius posterior probabilities for UNNAMED</p> <p>DeepTMHMM - Most Likely Topology Type: alpha TM</p> <p>DeepTMHMM - Posterior Probabilities</p> <p><i>Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebeius par ex).</i></p>	ID	UNNAMED				FT	TOPO_DOM	1	11	NON CYTOPLASMIC.	FT	TRANSMEM	12	31		FT	TOPO_DOM	32	128	CYTOPLASMIC.	//				
ID	UNNAMED																									
FT	TOPO_DOM	1	11	NON CYTOPLASMIC.																						
FT	TRANSMEM	12	31																							
FT	TOPO_DOM	32	128	CYTOPLASMIC.																						
//																										
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>Non</p> <p>Répondez Oui ou Non. Une fois que vous avez pris une décision quant-à-la fonction du gène, vérifiez la liste des</p>																									

	fonctions officielles de SEA-PHAGES (SEA-PHAGES Official Function List) pour vous assurer que vous suivez les directives de nommage des fonctions. Les fonctions qui ne figurent pas sur la liste approuvée doivent être soigneusement examinées pour être approuvées.
DECISION:	membrane protein (peut être pinholine car 1 domaine TM et située entre une holine et une endolysine)



pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

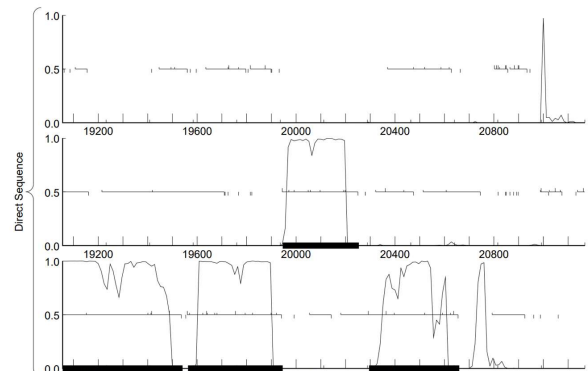
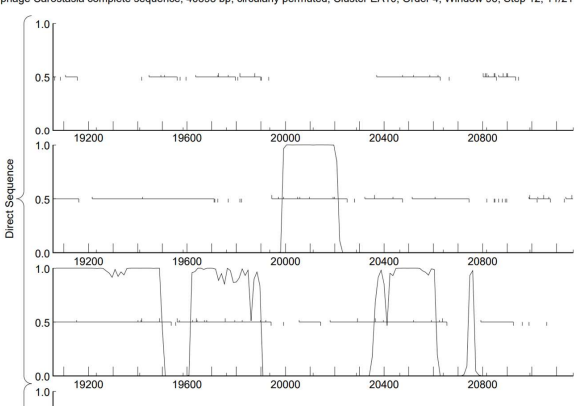
Student Gene Annotation Worksheet

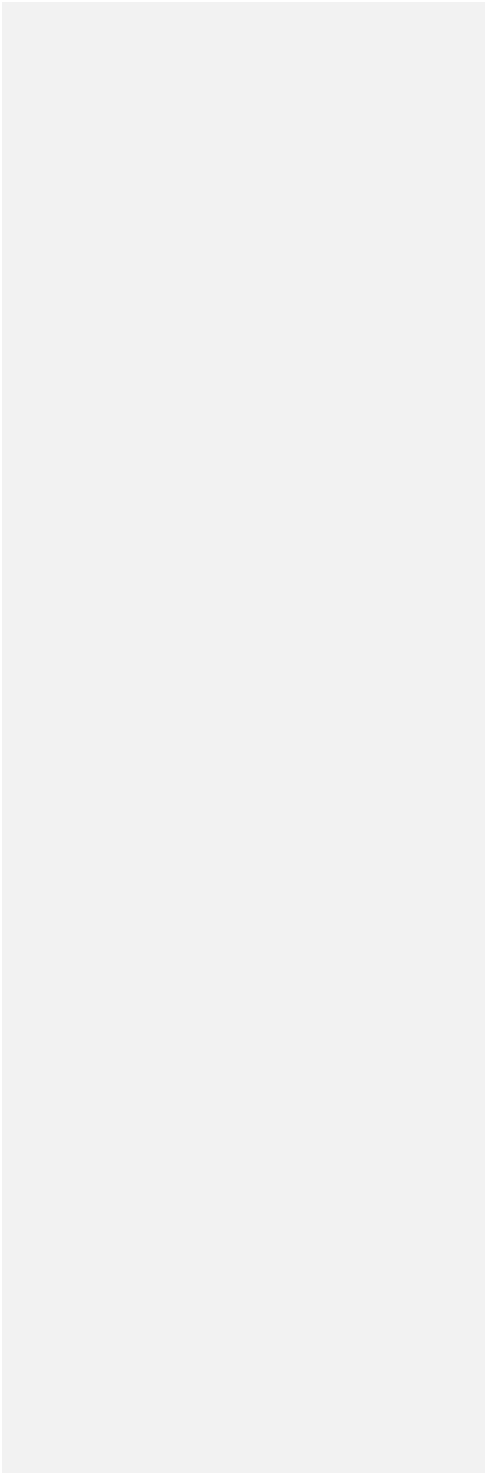
Carosmatique-gene-25

Basic Phage Information	
Nom du Phage	carostasia
Gène #	25
Coordonnées du Stop	20254
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Chev 1
Coordonnées du Start retenu	19946
Fonction prédite	holin

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	

	<div><div>Cadre +2</div><div><div>GeneMark hm prediction</div><div>erium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 11/21</div><div></div><div>erium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 11/21</div><div></div></div></div>																											
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div>Sequences producing significant alignments:</div><div><table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_25, holin, 102</td><td>194</td><td>5e-50</td></tr><tr><td>Nucci_25, holin, 102</td><td>194</td><td>5e-50</td></tr><tr><td>Mandalorian_25, holin, 102</td><td>194</td><td>5e-50</td></tr><tr><td>Carostasia_Draft_25, function unknown, 102</td><td>194</td><td>5e-50</td></tr><tr><td>Theresita_26, holin, 102</td><td>164</td><td>9e-41</td></tr><tr><td>SirVictor_26, function unknown, 97</td><td>163</td><td>2e-40</td></tr><tr><td>Guetzie_26, function unknown, 97</td><td>163</td><td>2e-40</td></tr><tr><td>Lucky3_26, holin, 97</td><td>159</td><td>2e-39</td></tr></table></div></div>		Score (bits)	E Value	Quartz_25, holin, 102	194	5e-50	Nucci_25, holin, 102	194	5e-50	Mandalorian_25, holin, 102	194	5e-50	Carostasia_Draft_25, function unknown, 102	194	5e-50	Theresita_26, holin, 102	164	9e-41	SirVictor_26, function unknown, 97	163	2e-40	Guetzie_26, function unknown, 97	163	2e-40	Lucky3_26, holin, 97	159	2e-39
	Score (bits)	E Value																										
Quartz_25, holin, 102	194	5e-50																										
Nucci_25, holin, 102	194	5e-50																										
Mandalorian_25, holin, 102	194	5e-50																										
Carostasia_Draft_25, function unknown, 102	194	5e-50																										
Theresita_26, holin, 102	164	9e-41																										
SirVictor_26, function unknown, 97	163	2e-40																										
Guetzie_26, function unknown, 97	163	2e-40																										
Lucky3_26, holin, 97	159	2e-39																										
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<div>non</div>																											
DECISION:	<div>oui</div>																											



Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																			
Quel start Glimmer et GeneMark suggèrent-ils ?	19946																																																																																																			
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Oui</div> <div>Score : -4.802</div> <div><div>Stats: 9 Selected: 1</div><div>ORF Start : 20099 ORF Stop : 20254 ORF Length: 156</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 44.4 55.6 77.8 27 3' End 71.3 44.7 90.4 282</div><div>SD Scoring Matrix Spacing Weight Matrix</div><div>Kibler6 Kalin Medium</div><div>Explore Document</div></div> <table><thead><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr></thead><tbody><tr><td>1</td><td>-3,590</td><td>2,124</td><td>8</td><td>-4,802</td><td>TGGACAAACTGAAAGGCAAGTA</td><td>ATG</td><td>19946</td><td>309</td></tr><tr><td>2</td><td>-5,712</td><td>1,126</td><td>10</td><td>-6,407</td><td>TACCATGCAACGCTGGCCACC</td><td>GTG</td><td>19973</td><td>282</td></tr><tr><td>3</td><td>-6,359</td><td>0,823</td><td>15</td><td>-7,961</td><td>CGTGGCGGCTGTCTATGCACTG</td><td>GTG</td><td>19994</td><td>261</td></tr><tr><td>4</td><td>-6,089</td><td>0,950</td><td>10</td><td>-6,784</td><td>GCTGTGGCGGCTGTCTATGCACTG</td><td>GTG</td><td>20051</td><td>204</td></tr><tr><td>5</td><td>-6,253</td><td>0,973</td><td>10</td><td>-6,948</td><td>GCTGTGGCGGCTGTCTATGCACTG</td><td>GTG</td><td>20060</td><td>195</td></tr><tr><td>6</td><td>-5,691</td><td>1,136</td><td>9</td><td>-6,466</td><td>GGCGGTGTCTGTCTGTGTGGCC</td><td>TTC</td><td>20066</td><td>189</td></tr><tr><td>7</td><td>-4,663</td><td>1,617</td><td>7</td><td>-6,186</td><td>CGGCGGCTGTCTGTGTGGCCACG</td><td>GTG</td><td>20099</td><td>156</td></tr><tr><td>8</td><td>-5,202</td><td>1,365</td><td>12</td><td>-6,038</td><td>GCGGGGCATCGGTGGCAACACC</td><td>GTG</td><td>20195</td><td>60</td></tr><tr><td>9</td><td>-5,276</td><td>1,330</td><td>15</td><td>-6,878</td><td>CATGGTGGCAACCGTGGTCTC</td><td>GTG</td><td>20201</td><td>54</td></tr></tbody></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,590	2,124	8	-4,802	TGGACAAACTGAAAGGCAAGTA	ATG	19946	309	2	-5,712	1,126	10	-6,407	TACCATGCAACGCTGGCCACC	GTG	19973	282	3	-6,359	0,823	15	-7,961	CGTGGCGGCTGTCTATGCACTG	GTG	19994	261	4	-6,089	0,950	10	-6,784	GCTGTGGCGGCTGTCTATGCACTG	GTG	20051	204	5	-6,253	0,973	10	-6,948	GCTGTGGCGGCTGTCTATGCACTG	GTG	20060	195	6	-5,691	1,136	9	-6,466	GGCGGTGTCTGTCTGTGTGGCC	TTC	20066	189	7	-4,663	1,617	7	-6,186	CGGCGGCTGTCTGTGTGGCCACG	GTG	20099	156	8	-5,202	1,365	12	-6,038	GCGGGGCATCGGTGGCAACACC	GTG	20195	60	9	-5,276	1,330	15	-6,878	CATGGTGGCAACCGTGGTCTC	GTG	20201	54
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																												
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																												
1	-3,590	2,124	8	-4,802	TGGACAAACTGAAAGGCAAGTA	ATG	19946	309																																																																																												
2	-5,712	1,126	10	-6,407	TACCATGCAACGCTGGCCACC	GTG	19973	282																																																																																												
3	-6,359	0,823	15	-7,961	CGTGGCGGCTGTCTATGCACTG	GTG	19994	261																																																																																												
4	-6,089	0,950	10	-6,784	GCTGTGGCGGCTGTCTATGCACTG	GTG	20051	204																																																																																												
5	-6,253	0,973	10	-6,948	GCTGTGGCGGCTGTCTATGCACTG	GTG	20060	195																																																																																												
6	-5,691	1,136	9	-6,466	GGCGGTGTCTGTCTGTGTGGCC	TTC	20066	189																																																																																												
7	-4,663	1,617	7	-6,186	CGGCGGCTGTCTGTGTGGCCACG	GTG	20099	156																																																																																												
8	-5,202	1,365	12	-6,038	GCGGGGCATCGGTGGCAACACC	GTG	20195	60																																																																																												
9	-5,276	1,330	15	-6,878	CATGGTGGCAACCGTGGTCTC	GTG	20201	54																																																																																												
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<div>Score le plus bas et ORF le plus long.</div> <div></div>																																																																																																			
Est-ce que le start est conservé chez les homologues voir Starterator ?	<div></div>																																																																																																			
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div><div>holin [microbactérie phage Nucci]</div><div>ID de séquence : QDK02418.1 Longueur : 102 Nombre de correspondances : 1</div><div>Voir 2 titres supplémentaires Voir toutes les protéines identiques (IPG)</div><div>Plage 1 : 1 à 102 graphiques GenPept Prochain match Match précédent</div><table><thead><tr><th>Score</th><th>Attendre</th><th>Méthode</th><th>Identités</th><th>Points positifs</th><th>Lacunes</th></tr></thead><tbody><tr><td>193 bits(491)</td><td>1e-61</td><td>Ajustement de la matrice de composition.</td><td>102/102(100%)</td><td>102/102(100%)</td><td>0/102(0%)</td></tr></tbody></table><div>Requête 1 MDTIATLATVPAVIALVTLFKDLGLPSKLSPLLAVALGVALSVLGALSLGTVANWYETVS 60 MDTIATLATVPAVIALVTLFKDLGLPSKLSPLLAVALGVALSVLGALSLGTVANWYETVS Sbjct 1 MDTIATLATVPAVIALVTLFKDLGLPSKLSPLLAVALGVALSVLGALSLGTVANWYETVS 60</div><div>Requête 61 LGVILGLSAAGLYDGARAIGNTVVVEEDPATKPLLRDLNN 102 LGVILGLSAAGLYDGARAIGNTVVVEEDPATKPLLRDLNN Sbjct 61 LGVILGLSAAGLYDGARAIGNTVVVEEDPATKPLLRDLNN 102</div></div> <div><div>Télécharger GenPept Graphique</div><div><div>holin [phage microbactérien Theresita]</div><div>ID de séquence : QBZ73216.1 Longueur : 102 Nombre de correspondances : 1</div><div>Plage 1 : 1 à 101 graphiques GenPept Prochain match Match précédent</div><table><thead><tr><th>Score</th><th>Attendre</th><th>Méthode</th><th>Identités</th><th>Points positifs</th><th>Lacunes</th></tr></thead><tbody><tr><td>162 bits(409)</td><td>4e-49</td><td>Ajustement de la matrice de composition.</td><td>88/101(87%)</td><td>90/101(89%)</td><td>1/101(0%)</td></tr></tbody></table><div>Requête 1 MDTIATLATVPAVIALVTLFKDLGLPSKLSPLLAVALGVALSVLGALSLGTVANWYETVS 60 MDTIATLATVPAVIALVTLFKDLGLPSKLSPLLAVALGVALSVLGALSLGTVANWYETVS Sbjct 1 MDTIATLATVPAVIALVTLFKDLGLPSKLSPLLAVALGVALSVLGALSLGTVANWYETVS 60</div><div>Requête 61 LGVILGLSAAGLYDGARAIGNTVVVEEDPATKPLLRDL 100 LGVILGLSAAGLYDGARAIGNTVVVEEDPATKPLLRDL Sbjct 61 LGVILGLSAAGLYDGARAIGNTVVVEEDPATKPLLRDL 101</div></div></div>	Score	Attendre	Méthode	Identités	Points positifs	Lacunes	193 bits(491)	1e-61	Ajustement de la matrice de composition.	102/102(100%)	102/102(100%)	0/102(0%)	Score	Attendre	Méthode	Identités	Points positifs	Lacunes	162 bits(409)	4e-49	Ajustement de la matrice de composition.	88/101(87%)	90/101(89%)	1/101(0%)																																																																											
Score	Attendre	Méthode	Identités	Points positifs	Lacunes																																																																																															
193 bits(491)	1e-61	Ajustement de la matrice de composition.	102/102(100%)	102/102(100%)	0/102(0%)																																																																																															
Score	Attendre	Méthode	Identités	Points positifs	Lacunes																																																																																															
162 bits(409)	4e-49	Ajustement de la matrice de composition.	88/101(87%)	90/101(89%)	1/101(0%)																																																																																															

DECISION:	<p>Le start conservé est 19946 car il n'est pas en contraction avec les principes d'annotation, il est conservé chez les homologues et le RBS associé au start présente le meilleur score et le plus long ORF avec un ATG</p>
-----------	---

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Rationnelle

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

*** *PhagesDB* :

Sequences producing significant alignments:	Score	(bits) Value
Quarta_25_bolin, 102	181	5e-50
Bucci_25_bolin, 102	181	5e-50
Mandelorian_25_bolin, 102	181	5e-50
Cercoaria_25b_25, function unknown, 102	181	5e-50
Therania_24_bolin, 102	164	9e-41
SLVIRUS_24, function unknown, 97	163	2e-40
Quarta_24, function unknown, 97	162	2e-40
Lucky1_24_bolin, 97	153	2e-39
Quarta_24_bolin, 97	153	2e-39
Vuvf_25_bolin, 97	154	2e-38
SLVIRUS_25_bolin, 97	153	2e-38
PrincoPhagus_24_bolin, 97	155	2e-38
Pharos_24_bolin, 97	155	2e-38
Kavala_24_bolin, 97	155	2e-38
Buhalman_24_bolin, 97	155	2e-38
Quarta_24, function unknown, 97	155	2e-38
SLVIRUS_24b_27, function unknown, 97	155	2e-38
Koji_24_bolin, 97	153	2e-38
SLVIRUS_24, function unknown, 94	152	4e-37
Senitis_25_bolin, 90	144	2e-35
Shaw_25_bolin, 90	144	2e-35
Seratos_25_bolin, 90	141	2e-35
Mimod_25_bolin, 90	144	2e-35
Glanow_24_bolin, 90	144	2e-35
SLVIRUS_25_bolin, 90	144	2e-35
SLVIRUS_24b_29, 90	144	2e-35
ChikPir_25_bolin, 90	143	2e-35
Autodroma_24_bolin, 90	144	2e-35
Sanea_24_bolin, 90	137	1e-32

Minimum to show details and scores. Click to show alignments
 Color Key for Alignment Scores

*** *PhagesDB* :

Sequences producing significant alignments:	Score	(bits) Value
Quarta_25_bolin, 102	181	5e-50
Bucci_25_bolin, 102	181	5e-50

*** nr :

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	holin (Microbacterium phage Nucl)	Microbacterium phage Nucl	193	193	99%	1e-61	100.00%	102	Q052418.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Theresia)	Microbacterium phage Theresia	162	162	97%	5e-49	87.13%	102	Q8Z732.1
<input checked="" type="checkbox"/>	membrane protein (Microbacterium phage StVictor)	Microbacterium phage StVictor	160	160	91%	1e-48	89.89%	97	VNM87799.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Golden)	Microbacterium phage Golden	157	157	90%	2e-47	88.94%	97	YP_009524167.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage XauC)	Microbacterium phage XauC	154	154	91%	3e-46	92.39%	97	Q2P9334.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Ksaula)	Microbacterium phage Ksaula	153	153	91%	7e-46	85.97%	97	Q8L3138.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage PhineoTheraps)	Microbacterium phage PhineoTheraps	153	153	91%	9e-46	85.57%	97	Q8Z72995.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Koi)	Microbacterium phage Koi	152	152	91%	2e-45	84.54%	97	YP_009624225.1
<input checked="" type="checkbox"/>	membrane protein (Microbacterium phage Juarez)	Microbacterium phage Juarez	150	150	92%	1e-44	86.73%	96	U8564775.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage MCubed)	Microbacterium phage MCubed	143	143	80%	6e-42	89.16%	90	Q003665.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Elari)	Microbacterium phage Elari	142	142	80%	1e-41	89.16%	90	YP_009623053.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Sarsa)	Microbacterium phage Sarsa	133	133	80%	5e-38	85.54%	90	AX148597.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Palasa)	Microbacterium phage Palasa	119	119	86%	2e-32	67.74%	99	AVJ51016.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Platin)	Microbacterium phage Platin	117	117	86%	1e-31	68.67%	99	YP_009524280.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Finny)	Microbacterium phage Finny	116	116	80%	2e-31	66.86%	74	Q2P15794.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Schubert)	Microbacterium phage Schubert	114	114	91%	2e-30	63.64%	101	YP_009518958.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Merlot)	Microbacterium phage Merlot	105	105	99%	1e-26	55.96%	113	AK003241.1
<input checked="" type="checkbox"/>	holin (Microbacterium phage Jemsa)	Microbacterium phage Jemsa	105	105	94%	1e-26	58.65%	112	QX144897.1

Pour les deux DB, les meilleurs hits sont des holin de phage. NuclAvec un excellent query cover (>90%) et des e-values <<10⁻⁴.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHpred avec une proba >= 90% et une couverture acceptable ?

Les meilleurs hits (probabilité >99%) sont des holins, ils proviennent de Uniprot.

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	PF00946.15	: Phage_holin_5_1 : Bacteriophage A118-like holin, H01118	99.76	5.8e-17	102.89	11.8	81	93
2	Q37977	HOLIN_BP450 Anisholin OS=Listeria phage A500 GN=40522 GN-Pol PE=3 SV=1	99.72	3.7e-16	99.73	11.2	81	96
3	PF16079.9	: Phage_holin_5_2 : Phage holin family HolA, in holin superfamily Y	99.6	1.8e-14	86.33	7.8	64	65

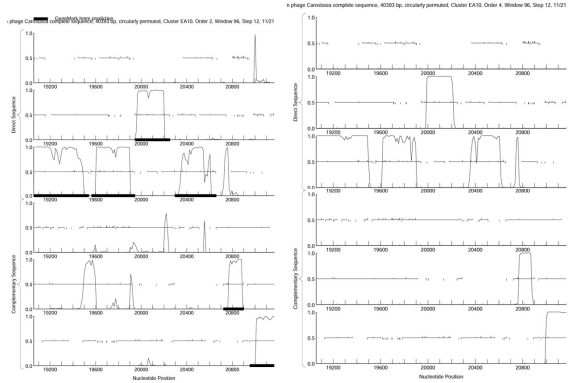
Residues Section

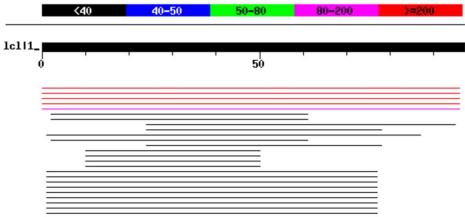
<p>Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?</p>	<div data-bbox="399 459 917 728"> </div> <p>Aucune synténie observée".</p>
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<div data-bbox="399 817 829 1131"> </div> <div data-bbox="399 1142 829 1534"> <pre> ID UNWRPED 1 35 FT SIGNAL 1 2 FT REGION 1 2 N-REGION FT REGION 3 18 H-REGION FT REGION 11 35 C-REGION FT TOP2_DON 16 38 NON CYTOPLASMIC FT TRANSDON 31 156 FT TOP2_DON 57 182 CYTOPLASMIC // </pre> </div>
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>Oui.</p>
<p>DECISION:</p>	<p>Ce gène code pour une holin.</p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	26
Coordonnées du Stop	20660
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	gap 40
Coordonnées du Start retenu	20295
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	

	<p>Distribution of 26 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p>  <p>Sequences producing significant alignments:</p> <table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Nucci_26, function unknown, 121</td><td>213</td><td>1e-55</td></tr><tr><td>Mandalorian_26, function unknown, 121</td><td>213</td><td>1e-55</td></tr><tr><td>Carostasia_Draft_26, function unknown, 97</td><td>213</td><td>1e-55</td></tr><tr><td>Quartz_26, function unknown, 121</td><td>208</td><td>4e-54</td></tr><tr><td>YuuY_26, function unknown, 121</td><td>186</td><td>2e-47</td></tr><tr><td>McGalleon_3, function unknown, 113</td><td>34</td><td>0.11</td></tr></tbody></table> <p><i>Gène retrouvé chez d'autres espèce mais peu</i></p>		Score (bits)	E Value	Nucci_26, function unknown, 121	213	1e-55	Mandalorian_26, function unknown, 121	213	1e-55	Carostasia_Draft_26, function unknown, 97	213	1e-55	Quartz_26, function unknown, 121	208	4e-54	YuuY_26, function unknown, 121	186	2e-47	McGalleon_3, function unknown, 113	34	0.11
	Score (bits)	E Value																				
Nucci_26, function unknown, 121	213	1e-55																				
Mandalorian_26, function unknown, 121	213	1e-55																				
Carostasia_Draft_26, function unknown, 97	213	1e-55																				
Quartz_26, function unknown, 121	208	4e-54																				
YuuY_26, function unknown, 121	186	2e-47																				
McGalleon_3, function unknown, 113	34	0.11																				
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p><i>On a 2 start possible qui donne dans tous les cas un grands gap de 113pb ou de 40.</i></p>																					
DECISION:	<p><i>Oui</i></p>																					

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																										
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Glimmer : 20367</i> <i>GeneMark : 20295</i>																																																																																										
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,766</td><td>2,037</td><td>16</td><td>-5,562</td><td>TACGACGGGGGGGGGCGCATCG</td><td>GTG</td><td>20184</td><td>477</td></tr><tr><td>2</td><td>-3,322</td><td>2,245</td><td>11</td><td>-4,079</td><td>ACACACAGTTAGGAATCATACA</td><td>ATG</td><td>20295</td><td>366</td></tr><tr><td>3</td><td>-3,298</td><td>2,261</td><td>14</td><td>-4,635</td><td>CGACTTCTGGACGTACCGATGG</td><td>ATG</td><td>20367</td><td>294</td></tr><tr><td>4</td><td>-7,368</td><td>0,351</td><td>12</td><td>-8,204</td><td>GATCGCGCGCGCTCTTCTGGAC</td><td>GTG</td><td>20400</td><td>261</td></tr><tr><td>5</td><td>-4,819</td><td>1,544</td><td>11</td><td>-5,576</td><td>CACGCTCTACAGACGTGGGGT</td><td>GTG</td><td>20523</td><td>138</td></tr><tr><td>6</td><td>-6,011</td><td>0,986</td><td>9</td><td>-6,786</td><td>CCGCTTCCGCTGCCGTGCTTCG</td><td>GTG</td><td>20595</td><td>66</td></tr><tr><td>7</td><td>-3,967</td><td>1,943</td><td>9</td><td>-4,741</td><td>CCGCTGGGGAGCCCGTGCCTGG</td><td>GTG</td><td>20628</td><td>33</td></tr><tr><td>8</td><td>-4,283</td><td>1,795</td><td>14</td><td>-5,630</td><td>CGGTGCTGGGTGAACCTCAAG</td><td>ATG</td><td>20640</td><td>21</td></tr></table> <p><i>meilleur score rbs : -4.079 pour le start prédit par genemark qui correspond au gap le plus faible (41). Ce start correspond à un ATG</i></p>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,766	2,037	16	-5,562	TACGACGGGGGGGGGCGCATCG	GTG	20184	477	2	-3,322	2,245	11	-4,079	ACACACAGTTAGGAATCATACA	ATG	20295	366	3	-3,298	2,261	14	-4,635	CGACTTCTGGACGTACCGATGG	ATG	20367	294	4	-7,368	0,351	12	-8,204	GATCGCGCGCGCTCTTCTGGAC	GTG	20400	261	5	-4,819	1,544	11	-5,576	CACGCTCTACAGACGTGGGGT	GTG	20523	138	6	-6,011	0,986	9	-6,786	CCGCTTCCGCTGCCGTGCTTCG	GTG	20595	66	7	-3,967	1,943	9	-4,741	CCGCTGGGGAGCCCGTGCCTGG	GTG	20628	33	8	-4,283	1,795	14	-5,630	CGGTGCTGGGTGAACCTCAAG	ATG	20640	21
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																			
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																			
1	-3,766	2,037	16	-5,562	TACGACGGGGGGGGGCGCATCG	GTG	20184	477																																																																																			
2	-3,322	2,245	11	-4,079	ACACACAGTTAGGAATCATACA	ATG	20295	366																																																																																			
3	-3,298	2,261	14	-4,635	CGACTTCTGGACGTACCGATGG	ATG	20367	294																																																																																			
4	-7,368	0,351	12	-8,204	GATCGCGCGCGCTCTTCTGGAC	GTG	20400	261																																																																																			
5	-4,819	1,544	11	-5,576	CACGCTCTACAGACGTGGGGT	GTG	20523	138																																																																																			
6	-6,011	0,986	9	-6,786	CCGCTTCCGCTGCCGTGCTTCG	GTG	20595	66																																																																																			
7	-3,967	1,943	9	-4,741	CCGCTGGGGAGCCCGTGCCTGG	GTG	20628	33																																																																																			
8	-4,283	1,795	14	-5,630	CGGTGCTGGGTGAACCTCAAG	ATG	20640	21																																																																																			
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p><i>L'ORF la plus longue en prenant donc le meilleur score RBS est donc de 366pb pour un gap de 41 avec le gène précédent. Pour le 2^e start prédit (glimmer) on a une ORF de 294pb pour 113pb de gap avec un moins bon score rbs</i></p>																																																																																										

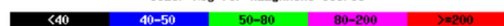
Est-ce que le start est conservé chez les homologues voir Starterator ?

</

Distribution of 28 Blast Hits on the Query Sequence

Mouse-over to show define and scores. Click to show alignments

Color Key for Alignment Scores



Sequences producing significant alignments:	Score (bits)	E Value
Nucci_26, function unknown, 121	267	6e-72
Mandalorian_26, function unknown, 121	267	6e-72
Quartz_26, function unknown, 121	260	9e-70
YuuY_26, function unknown, 121	238	4e-63
Carostasia_Draft_26, function unknown, 97	213	1e-55
McGalleon_3, function unknown, 113	40	0.002
Jenos_3, function unknown, 113	40	0.002
Endor_3, function unknown, 114	37	0.022
Endor_28, function unknown, 137	35	0.065
Acosta_Draft_28, function unknown, 137	34	0.11
Jason0_Draft_31, function unknown, 137	32	0.42
Jenos_29, function unknown, 136	31	0.94
Syleon_28, minor tail protein, 583	30	1.6
Sephiroth_28, minor tail protein, 583	30	1.6
Octobien14_28, minor tail protein, 583	30	1.6

>Nucci_26, function unknown, 121
Length = 121

Score = 267 bits (683), Expect = 6e-72
Identities = 121/121 (100%), Positives = 121/121 (100%)

Query: 1 HAGYATPSGSCGQQSATDFNTYRHMGGTIRARLLDVSLAPDPNPNHGWGQASIRIDLVG 60
HAGYATPSGSCGQQSATDFNTYRHMGGTIRARLLDVSLAPDPNPNHGWGQASIRIDLVG
Sbjct: 1 HAGYATPSGSCGQQSATDFNTYRHMGGTIRARLLDVSLAPDPNPNHGWGQASIRIDLVG 60

Query: 61 RDGKNKASNTLYKTGWLPAYKTFVFNVAQEVFRFCRASVPSYPHGAGAHVNFKNEFSW 120
RDGKNKASNTLYKTGWLPAYKTFVFNVAQEVFRFCRASVPSYPHGAGAHVNFKNEFSW
Sbjct: 61 RDGKNKASNTLYKTGWLPAYKTFVFNVAQEVFRFCRASVPSYPHGAGAHVNFKNEFSW 120

Query: 121 G 121
G
Sbjct: 121 G 121

>Mandalorian_26, function unknown, 121
Length = 121

Score = 267 bits (683), Expect = 6e-72
Identities = 121/121 (100%), Positives = 121/121 (100%)

Query: 1 HAGYATPSGSCGQQSATDFNTYRHMGGTIRARLLDVSLAPDPNPNHGWGQASIRIDLVG 60
HAGYATPSGSCGQQSATDFNTYRHMGGTIRARLLDVSLAPDPNPNHGWGQASIRIDLVG
Sbjct: 1 HAGYATPSGSCGQQSATDFNTYRHMGGTIRARLLDVSLAPDPNPNHGWGQASIRIDLVG 60

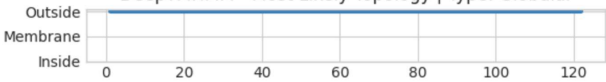
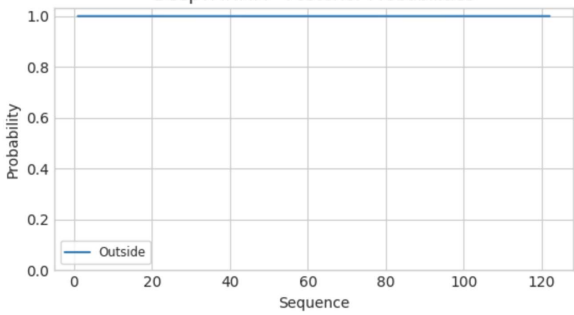
Query: 61 RDGKNKASNTLYKTGWLPAYKTFVFNVAQEVFRFCRASVPSYPHGAGAHVNFKNEFSW 120
RDGKNKASNTLYKTGWLPAYKTFVFNVAQEVFRFCRASVPSYPHGAGAHVNFKNEFSW
Sbjct: 61 RDGKNKASNTLYKTGWLPAYKTFVFNVAQEVFRFCRASVPSYPHGAGAHVNFKNEFSW 120

Query: 121 G 121
G
Sbjct: 121 G 121

*** nr :

(q#: s#) : [alignment] ; e-value :

	<div><div>No putative conserved domains have been detected</div><div><div>Distribution of the top 5 Blast Hits on 5 subject sequences</div></div><table><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max. Score</th><th>Total Score</th><th>Query Cover</th><th>E-value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td>✓</td><td>hypothetical protein SEA_NUCCI_26 [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>250</td><td>250</td><td>99%</td><td>2e-83</td><td>100.00%</td><td>121</td><td>QD02419.1</td></tr><tr><td>✓</td><td>hypothetical protein SEA_QUART_36 [Microbacterium phage Quart]</td><td>Microbacterium phage Quart</td><td>243</td><td>243</td><td>99%</td><td>1e-80</td><td>97.52%</td><td>121</td><td>U055545.1</td></tr><tr><td>✓</td><td>hypothetical protein SEA_TU07_36 [Microbacterium phage Tu07]</td><td>Microbacterium phage Tu07</td><td>222</td><td>222</td><td>99%</td><td>3e-72</td><td>89.26%</td><td>121</td><td>CEP55355.1</td></tr><tr><td>✓</td><td>hypothetical protein [Microbacterium sp.]</td><td>Microbacterium sp.</td><td>55.8</td><td>55.8</td><td>72%</td><td>6e-07</td><td>35.48%</td><td>101</td><td>MF0333887.1</td></tr><tr><td>✓</td><td>hypothetical protein [Microbacterium sp. CFSP-0734]</td><td>Microbacterium sp. CFSP-0734</td><td>45.8</td><td>45.8</td><td>84%</td><td>0.005</td><td>28.85%</td><td>125</td><td>WP_102014993.1</td></tr></table><div>hypothetical protein SEA_NUCCI_26 [Microbacterium phage Nucci] Sequence ID: QD02419.1 Length: 121 Number of Matches: 1 See 1 more title(s) See all Identical Proteins (16)</div><div>Range 1: 1 to 121 GenPept Graphics ▼ Next Match ▲ Previous Match</div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>250 bits(638)</td><td>2e-83</td><td>Compositional matrix adjust.</td><td>121/121(100%)</td><td>121/121(100%)</td><td>0/121(0%)</td></tr></table><div><div>Query 1</div><div>MAGYATPSGSGGQGSATDFWTYRWGGTIRARLLDVS LAPDPHYNWQGWQASRIDLVG 60</div><div>MAGYATPSGSGGQGSATDFWTYRWGGTIRARLLDVS LAPDPHYNWQGWQASRIDLVG 60</div><div>Sbjct 1</div><div>MAGYATPSGSGGQGSATDFWTYRWGGTIRARLLDVS LAPDPHYNWQGWQASRIDLVG 60</div></div><div><div>Query 61</div><div>RDGKINKASNTLYKTGWLPAYKTFVFMVVAQEVRFRCRASVPSPYWGAGAWNFKMF FSW 120</div><div>RDGKINKASNTLYKTGWLPAYKTFVFMVVAQEVRFRCRASVPSPYWGAGAWNFKMF FSW 120</div><div>Sbjct 61</div><div>RDGKINKASNTLYKTGWLPAYKTFVFMVVAQEVRFRCRASVPSPYWGAGAWNFKMF FSW 120</div></div><div><div>Query 121</div><div>G 121</div><div>G 121</div><div>Sbjct 121</div><div>G 121</div></div></div>		Description	Scientific Name	Max. Score	Total Score	Query Cover	E-value	Per. Ident	Acc. Len	Accession	✓	hypothetical protein SEA_NUCCI_26 [Microbacterium phage Nucci]	Microbacterium phage Nucci	250	250	99%	2e-83	100.00%	121	QD02419.1	✓	hypothetical protein SEA_QUART_36 [Microbacterium phage Quart]	Microbacterium phage Quart	243	243	99%	1e-80	97.52%	121	U055545.1	✓	hypothetical protein SEA_TU07_36 [Microbacterium phage Tu07]	Microbacterium phage Tu07	222	222	99%	3e-72	89.26%	121	CEP55355.1	✓	hypothetical protein [Microbacterium sp.]	Microbacterium sp.	55.8	55.8	72%	6e-07	35.48%	101	MF0333887.1	✓	hypothetical protein [Microbacterium sp. CFSP-0734]	Microbacterium sp. CFSP-0734	45.8	45.8	84%	0.005	28.85%	125	WP_102014993.1	Score	Expect	Method	Identities	Positives	Gaps	250 bits(638)	2e-83	Compositional matrix adjust.	121/121(100%)	121/121(100%)	0/121(0%)
	Description	Scientific Name	Max. Score	Total Score	Query Cover	E-value	Per. Ident	Acc. Len	Accession																																																																
✓	hypothetical protein SEA_NUCCI_26 [Microbacterium phage Nucci]	Microbacterium phage Nucci	250	250	99%	2e-83	100.00%	121	QD02419.1																																																																
✓	hypothetical protein SEA_QUART_36 [Microbacterium phage Quart]	Microbacterium phage Quart	243	243	99%	1e-80	97.52%	121	U055545.1																																																																
✓	hypothetical protein SEA_TU07_36 [Microbacterium phage Tu07]	Microbacterium phage Tu07	222	222	99%	3e-72	89.26%	121	CEP55355.1																																																																
✓	hypothetical protein [Microbacterium sp.]	Microbacterium sp.	55.8	55.8	72%	6e-07	35.48%	101	MF0333887.1																																																																
✓	hypothetical protein [Microbacterium sp. CFSP-0734]	Microbacterium sp. CFSP-0734	45.8	45.8	84%	0.005	28.85%	125	WP_102014993.1																																																																
Score	Expect	Method	Identities	Positives	Gaps																																																																				
250 bits(638)	2e-83	Compositional matrix adjust.	121/121(100%)	121/121(100%)	0/121(0%)																																																																				
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<div>Number of Hits: 18 Query MSA diversity (Neff): 4.47829</div> <div>Visualization</div>  <table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/></td><td>1</td><td>PF11454.12 ; DUF3016 ; Protein of unknown function (DUF3016)</td><td>50.32</td><td>140</td><td>23.71</td><td>5.5</td><td>42</td><td>134</td></tr><tr><td><input type="checkbox"/></td><td>2</td><td>PF14186.10 ; Aida_C2 ; Axin Interactor dorsalisation-associated protein, C-terminal</td><td>49.62</td><td>260</td><td>22.75</td><td>7.6</td><td>66</td><td>143</td></tr><tr><td><input type="checkbox"/></td><td>3</td><td>4C47_B INNER MEMBRANE LIPOPROTEIN; CELL ADHESION, BACTERIAL ADHESION, MEMBRANE TRAFFICKING, MEMBRANE INSERTION, AUTOTRANSPORT,</td><td>45.27</td><td>360</td><td>23.39</td><td>7.4</td><td>91</td><td>207</td></tr></table> <div>Pas de correspondance pertinente</div>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/>	1	PF11454.12 ; DUF3016 ; Protein of unknown function (DUF3016)	50.32	140	23.71	5.5	42	134	<input type="checkbox"/>	2	PF14186.10 ; Aida_C2 ; Axin Interactor dorsalisation-associated protein, C-terminal	49.62	260	22.75	7.6	66	143	<input type="checkbox"/>	3	4C47_B INNER MEMBRANE LIPOPROTEIN; CELL ADHESION, BACTERIAL ADHESION, MEMBRANE TRAFFICKING, MEMBRANE INSERTION, AUTOTRANSPORT,	45.27	360	23.39	7.4	91	207																																				
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																																	
<input type="checkbox"/>	1	PF11454.12 ; DUF3016 ; Protein of unknown function (DUF3016)	50.32	140	23.71	5.5	42	134																																																																	
<input type="checkbox"/>	2	PF14186.10 ; Aida_C2 ; Axin Interactor dorsalisation-associated protein, C-terminal	49.62	260	22.75	7.6	66	143																																																																	
<input type="checkbox"/>	3	4C47_B INNER MEMBRANE LIPOPROTEIN; CELL ADHESION, BACTERIAL ADHESION, MEMBRANE TRAFFICKING, MEMBRANE INSERTION, AUTOTRANSPORT,	45.27	360	23.39	7.4	91	207																																																																	
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une	<div>Non</div> <div>étroitement apparenté (astuce: utilisez Phamerator) et reportez la fonction des gènes trouvés de chaque côté du gène candidat</div>																																																																								

forte conservation de l'ordre des gènes ?	<i>dans le phage le plus étroitement apparenté. Si la réponse est NON, entrez "Aucune synténie observée".</i>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div>Non</div> <div><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p></div> <div><i>Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebeius par ex).</i></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Non</i>
DECISION:	<i>NFK</i>

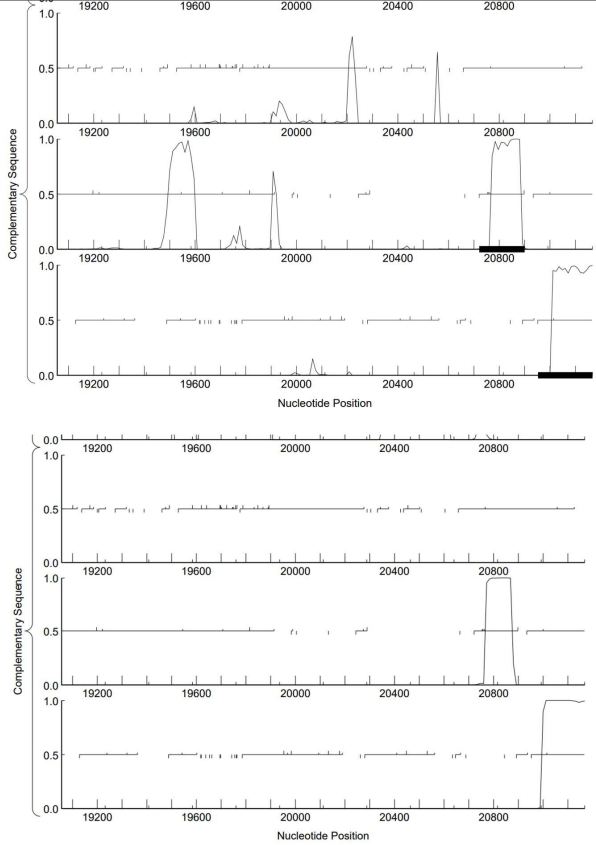
Student Gene Annotation Worksheet

Carosmatique-gene-27

Basic Phage Information	
Nom du Phage	carostasia
Gène #	27
Coordonnées du Stop	20721
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	gap 52
Coordonnées du Start retenu	20900
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH

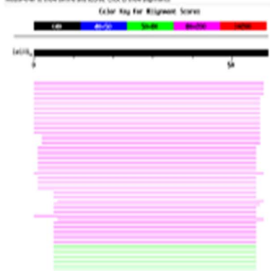
<p>Y-a-t-il des éléments supportant un potentiel codant ?</p>	 <p>Cadre °2</p>																					
<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<div><div>Sequences producing significant alignments:</div><table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>Nucci_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>Mandalorian_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>Carostasia Draft_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>YuuY_27, function unknown, 77</td><td>110</td><td>9e-25</td></tr><tr><td>ChikPic_27, function unknown, 58</td><td>110</td><td>1e-24</td></tr></table></div>		Score (bits)	E Value	Quartz_27, function unknown, 59	117	1e-26	Nucci_27, function unknown, 59	117	1e-26	Mandalorian_27, function unknown, 59	117	1e-26	Carostasia Draft_27, function unknown, 59	117	1e-26	YuuY_27, function unknown, 77	110	9e-25	ChikPic_27, function unknown, 58	110	1e-24
	Score (bits)	E Value																				
Quartz_27, function unknown, 59	117	1e-26																				
Nucci_27, function unknown, 59	117	1e-26																				
Mandalorian_27, function unknown, 59	117	1e-26																				
Carostasia Draft_27, function unknown, 59	117	1e-26																				
YuuY_27, function unknown, 77	110	9e-25																				
ChikPic_27, function unknown, 58	110	1e-24																				
<p>Est-ce que le candidat est en contradiction avec les principes d'annotation ?</p>	<p><i>non</i></p>																					
<p>DECISION:</p>	<p><i>oui</i></p>																					

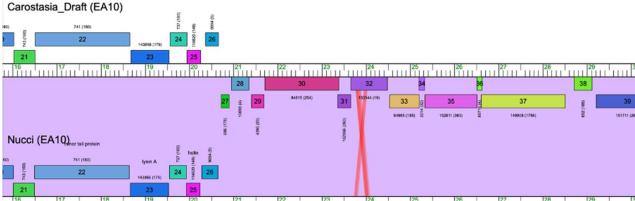

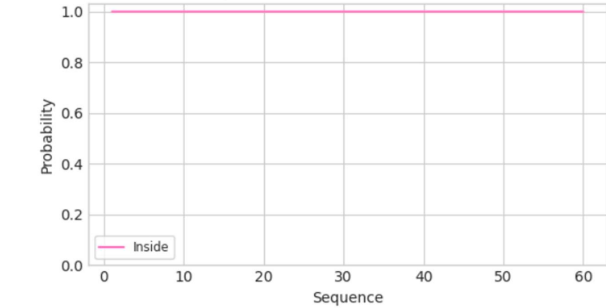
Décision #2 : Quel est le start du gène ?

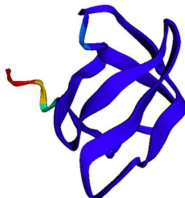
Collection des éléments de réponse	Rationnelle																																																
Quel start Glimmer et GeneMark suggèrent-ils ?	20900																																																
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Starts : 4 ORF Start : 20900 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kib Selected : 1 ORF Stop : 20721 5' End 46,7 62,2 66,7 135 Spacing Weight Matrix Kar ORF Length : 180 3' End 56,5 67,4 91,3 138</div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th></tr><tr><td>1</td><td>-1,748</td><td>2,981</td><td>10</td><td>-2,443</td><td>CAACCAACCGAAGGAGGOCATC</td><td>ATG</td><td>20900</td></tr><tr><td>2</td><td>-3,912</td><td>1,969</td><td>10</td><td>-4,606</td><td>GCGTCTCGAGACGGACCTGGGT</td><td>GTG</td><td>20765</td></tr><tr><td>3</td><td>-4,283</td><td>1,795</td><td>10</td><td>-4,977</td><td>CGAGACGGACCTGGGTGTGGCC</td><td>GTG</td><td>20759</td></tr><tr><td>4</td><td>-4,283</td><td>1,795</td><td>13</td><td>-5,329</td><td>GACGGACCTGGGTGTGGCCGTG</td><td>GTG</td><td>20756</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	1	-1,748	2,981	10	-2,443	CAACCAACCGAAGGAGGOCATC	ATG	20900	2	-3,912	1,969	10	-4,606	GCGTCTCGAGACGGACCTGGGT	GTG	20765	3	-4,283	1,795	10	-4,977	CGAGACGGACCTGGGTGTGGCC	GTG	20759	4	-4,283	1,795	13	-5,329	GACGGACCTGGGTGTGGCCGTG	GTG	20756
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start																																										
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position																																										
1	-1,748	2,981	10	-2,443	CAACCAACCGAAGGAGGOCATC	ATG	20900																																										
2	-3,912	1,969	10	-4,606	GCGTCTCGAGACGGACCTGGGT	GTG	20765																																										
3	-4,283	1,795	10	-4,977	CGAGACGGACCTGGGTGTGGCC	GTG	20759																																										
4	-4,283	1,795	13	-5,329	GACGGACCTGGGTGTGGCCGTG	GTG	20756																																										
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Meilleur score et ORF le plus long																																																
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																	
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>1c111 0 50</p> <p>>Quartz_27, function unknown, 59 Length = 59</p> <p>Score = 117 bits (293), Expect = 1e-26 Identities = 59/59 (100%), Positives = 59/59 (100%)</p> <p>Query: 1 MASKGIKAEDLTGQVIIDPEGNEATVIRIRRIDHQRGRLETLGVAVVQLSDRFPLAP 59 MASKGIKAEDLTGQVIIDPEGNEATVIRIRRIDHQRGRLETLGVAVVQLSDRFPLAP Sbjct: 1 MASKGIKAEDLTGQVIIDPEGNEATVIRIRRIDHQRGRLETLGVAVVQLSDRFPLAP 59</p>																																																
DECISION:	Le start conservé est 20900 car il n'est pas en contraction avec les principes d'annotation et Il est conservé chez																																																

	d'autres homologues et il donne l'ORF le plus long avec le meilleur score RBS
--	---

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																																								
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** <i>PhagesDB</i> :</p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>Nucci_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>Mandalorian_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>Carostasia_Draft_27, function unknown, 59</td><td>117</td><td>1e-26</td></tr><tr><td>YuuY_27, function unknown, 77</td><td>110</td><td>1e-24</td></tr><tr><td>ChikPic_27, function unknown, 58</td><td>110</td><td>1e-24</td></tr><tr><td>Zenitsu_27, function unknown, 58</td><td>108</td><td>4e-24</td></tr><tr><td>Shamu_27, function unknown, 58</td><td>108</td><td>4e-24</td></tr><tr><td>Glamour_26, function unknown, 58</td><td>108</td><td>4e-24</td></tr><tr><td>Eleri_27, function unknown, 58</td><td>108</td><td>4e-24</td></tr><tr><td>ColaCorta_27, function unknown, 58</td><td>108</td><td>4e-24</td></tr><tr><td>Andromedas_27, function unknown, 58</td><td>108</td><td>4e-24</td></tr><tr><td>Juanyo_27, function unknown, 58</td><td>108</td><td>6e-24</td></tr><tr><td>Sansa_26, function unknown, 58</td><td>107</td><td>8e-24</td></tr><tr><td>MCubed_27, function unknown, 57</td><td>104</td><td>7e-23</td></tr><tr><td>Finny_27, function unknown, 57</td><td>104</td><td>7e-23</td></tr><tr><td>Saratos_27, function unknown, 57</td><td>103</td><td>2e-22</td></tr><tr><td>Koji_27, function unknown, 60</td><td>91</td><td>1e-18</td></tr><tr><td>BouleyBill_Draft_28, function unknown, 60</td><td>91</td><td>1e-18</td></tr><tr><td>Kauala_27, function unknown, 60</td><td>90</td><td>2e-18</td></tr><tr><td>Sinatra_27, function unknown, 60</td><td>90</td><td>2e-18</td></tr><tr><td>PrincePhergus_27, function unknown, 65</td><td>90</td><td>2e-18</td></tr><tr><td>Pherbot_27, function unknown, 60</td><td>90</td><td>2e-18</td></tr></table> <p>Mouse-over to show details and scores. Click to show alignment.</p> <p>Color Key for Alignment Scores</p>  <p>Sequences producing significant alignments:</p> <table><tr><th>Source</th><th>E Value</th></tr><tr><td>Quartz_27, function unknown, 59</td><td>117</td></tr><tr><td>Nucci_27, function unknown, 59</td><td>117</td></tr><tr><td>Mandalorian_27, function unknown, 59</td><td>117</td></tr><tr><td>Carostasia_Draft_27, function unknown, 59</td><td>117</td></tr><tr><td>YuuY_27, function unknown, 77</td><td>110</td></tr><tr><td>ChikPic_27, function unknown, 58</td><td>110</td></tr><tr><td>Zenitsu_27, function unknown, 58</td><td>108</td></tr><tr><td>Shamu_27, function unknown, 58</td><td>108</td></tr><tr><td>Glamour_26, function unknown, 58</td><td>108</td></tr><tr><td>Eleri_27, function unknown, 58</td><td>108</td></tr><tr><td>ColaCorta_27, function unknown, 58</td><td>108</td></tr><tr><td>Andromedas_27, function unknown, 58</td><td>108</td></tr><tr><td>Juanyo_27, function unknown, 58</td><td>108</td></tr><tr><td>Sansa_26, function unknown, 58</td><td>107</td></tr><tr><td>MCubed_27, function unknown, 57</td><td>104</td></tr><tr><td>Finny_27, function unknown, 57</td><td>104</td></tr><tr><td>Saratos_27, function unknown, 57</td><td>103</td></tr><tr><td>Koji_27, function unknown, 60</td><td>91</td></tr><tr><td>BouleyBill_Draft_28, function unknown, 60</td><td>91</td></tr><tr><td>Kauala_27, function unknown, 60</td><td>90</td></tr><tr><td>Sinatra_27, function unknown, 60</td><td>90</td></tr><tr><td>PrincePhergus_27, function unknown, 65</td><td>90</td></tr><tr><td>Pherbot_27, function unknown, 60</td><td>90</td></tr></table> <p>*** <i>nr</i> :</p> <p>(q#: s#) : [alignment] ; e-value :</p>	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_27, function unknown, 59	117	1e-26	Nucci_27, function unknown, 59	117	1e-26	Mandalorian_27, function unknown, 59	117	1e-26	Carostasia_Draft_27, function unknown, 59	117	1e-26	YuuY_27, function unknown, 77	110	1e-24	ChikPic_27, function unknown, 58	110	1e-24	Zenitsu_27, function unknown, 58	108	4e-24	Shamu_27, function unknown, 58	108	4e-24	Glamour_26, function unknown, 58	108	4e-24	Eleri_27, function unknown, 58	108	4e-24	ColaCorta_27, function unknown, 58	108	4e-24	Andromedas_27, function unknown, 58	108	4e-24	Juanyo_27, function unknown, 58	108	6e-24	Sansa_26, function unknown, 58	107	8e-24	MCubed_27, function unknown, 57	104	7e-23	Finny_27, function unknown, 57	104	7e-23	Saratos_27, function unknown, 57	103	2e-22	Koji_27, function unknown, 60	91	1e-18	BouleyBill_Draft_28, function unknown, 60	91	1e-18	Kauala_27, function unknown, 60	90	2e-18	Sinatra_27, function unknown, 60	90	2e-18	PrincePhergus_27, function unknown, 65	90	2e-18	Pherbot_27, function unknown, 60	90	2e-18	Source	E Value	Quartz_27, function unknown, 59	117	Nucci_27, function unknown, 59	117	Mandalorian_27, function unknown, 59	117	Carostasia_Draft_27, function unknown, 59	117	YuuY_27, function unknown, 77	110	ChikPic_27, function unknown, 58	110	Zenitsu_27, function unknown, 58	108	Shamu_27, function unknown, 58	108	Glamour_26, function unknown, 58	108	Eleri_27, function unknown, 58	108	ColaCorta_27, function unknown, 58	108	Andromedas_27, function unknown, 58	108	Juanyo_27, function unknown, 58	108	Sansa_26, function unknown, 58	107	MCubed_27, function unknown, 57	104	Finny_27, function unknown, 57	104	Saratos_27, function unknown, 57	103	Koji_27, function unknown, 60	91	BouleyBill_Draft_28, function unknown, 60	91	Kauala_27, function unknown, 60	90	Sinatra_27, function unknown, 60	90	PrincePhergus_27, function unknown, 65	90	Pherbot_27, function unknown, 60	90
Sequences producing significant alignments:	Score (bits)	E Value																																																																																																																							
Quartz_27, function unknown, 59	117	1e-26																																																																																																																							
Nucci_27, function unknown, 59	117	1e-26																																																																																																																							
Mandalorian_27, function unknown, 59	117	1e-26																																																																																																																							
Carostasia_Draft_27, function unknown, 59	117	1e-26																																																																																																																							
YuuY_27, function unknown, 77	110	1e-24																																																																																																																							
ChikPic_27, function unknown, 58	110	1e-24																																																																																																																							
Zenitsu_27, function unknown, 58	108	4e-24																																																																																																																							
Shamu_27, function unknown, 58	108	4e-24																																																																																																																							
Glamour_26, function unknown, 58	108	4e-24																																																																																																																							
Eleri_27, function unknown, 58	108	4e-24																																																																																																																							
ColaCorta_27, function unknown, 58	108	4e-24																																																																																																																							
Andromedas_27, function unknown, 58	108	4e-24																																																																																																																							
Juanyo_27, function unknown, 58	108	6e-24																																																																																																																							
Sansa_26, function unknown, 58	107	8e-24																																																																																																																							
MCubed_27, function unknown, 57	104	7e-23																																																																																																																							
Finny_27, function unknown, 57	104	7e-23																																																																																																																							
Saratos_27, function unknown, 57	103	2e-22																																																																																																																							
Koji_27, function unknown, 60	91	1e-18																																																																																																																							
BouleyBill_Draft_28, function unknown, 60	91	1e-18																																																																																																																							
Kauala_27, function unknown, 60	90	2e-18																																																																																																																							
Sinatra_27, function unknown, 60	90	2e-18																																																																																																																							
PrincePhergus_27, function unknown, 65	90	2e-18																																																																																																																							
Pherbot_27, function unknown, 60	90	2e-18																																																																																																																							
Source	E Value																																																																																																																								
Quartz_27, function unknown, 59	117																																																																																																																								
Nucci_27, function unknown, 59	117																																																																																																																								
Mandalorian_27, function unknown, 59	117																																																																																																																								
Carostasia_Draft_27, function unknown, 59	117																																																																																																																								
YuuY_27, function unknown, 77	110																																																																																																																								
ChikPic_27, function unknown, 58	110																																																																																																																								
Zenitsu_27, function unknown, 58	108																																																																																																																								
Shamu_27, function unknown, 58	108																																																																																																																								
Glamour_26, function unknown, 58	108																																																																																																																								
Eleri_27, function unknown, 58	108																																																																																																																								
ColaCorta_27, function unknown, 58	108																																																																																																																								
Andromedas_27, function unknown, 58	108																																																																																																																								
Juanyo_27, function unknown, 58	108																																																																																																																								
Sansa_26, function unknown, 58	107																																																																																																																								
MCubed_27, function unknown, 57	104																																																																																																																								
Finny_27, function unknown, 57	104																																																																																																																								
Saratos_27, function unknown, 57	103																																																																																																																								
Koji_27, function unknown, 60	91																																																																																																																								
BouleyBill_Draft_28, function unknown, 60	91																																																																																																																								
Kauala_27, function unknown, 60	90																																																																																																																								
Sinatra_27, function unknown, 60	90																																																																																																																								
PrincePhergus_27, function unknown, 65	90																																																																																																																								
Pherbot_27, function unknown, 60	90																																																																																																																								

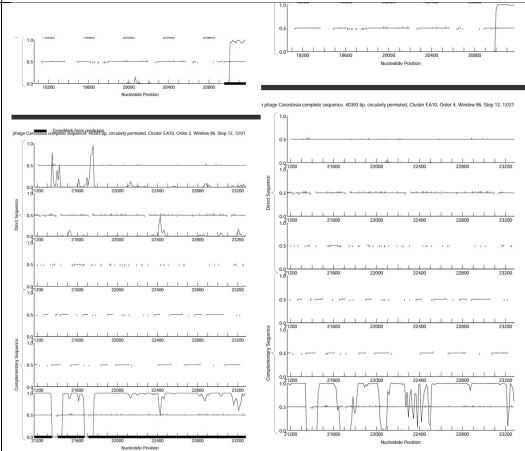
	<table><thead><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr></thead><tbody><tr><td>1</td><td>PF16138.9</td><td>; DUF4846; Domain of unknown function (4846)</td><td>79.7</td><td>8.5</td><td>28.33</td><td>3.8</td><td>31</td><td>240</td></tr><tr><td>2</td><td>PF08741.14</td><td>; YwhD; YwhD family</td><td>74.97</td><td>5.5</td><td>29.65</td><td>1.8</td><td>12</td><td>162</td></tr><tr><td>3</td><td>PF18337.5</td><td>; Tudor_RapA; RapA N-terminal Tudor like domain</td><td>71.4</td><td>29</td><td>20.43</td><td>3.9</td><td>26</td><td>62</td></tr><tr><td>4</td><td>PF13228.10</td><td>; DUF4037; Domain of unknown function (DUF4037)</td><td>60.69</td><td>14</td><td>21.29</td><td>1.2</td><td>21</td><td>100</td></tr><tr><td>5</td><td>cd07267</td><td>ThT_Oxygenase_N; N-terminal domain of 2,4,5-trihydroxytoluene (ThT) oxygenase. This subfamily contains the N-terminal, n</td><td>58.32</td><td>35</td><td>17.61</td><td>2.3</td><td>13</td><td>113</td></tr><tr><td>6</td><td>2ZW5_B</td><td>Bleomycin acetyltransferase; dimer; two domains, Transferase; HET: COA; 2-4A (Streptomyces verticillus)</td><td>57.74</td><td>29</td><td>21.98</td><td>2.3</td><td>14</td><td>301</td></tr></tbody></table>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	1	PF16138.9	; DUF4846; Domain of unknown function (4846)	79.7	8.5	28.33	3.8	31	240	2	PF08741.14	; YwhD; YwhD family	74.97	5.5	29.65	1.8	12	162	3	PF18337.5	; Tudor_RapA; RapA N-terminal Tudor like domain	71.4	29	20.43	3.9	26	62	4	PF13228.10	; DUF4037; Domain of unknown function (DUF4037)	60.69	14	21.29	1.2	21	100	5	cd07267	ThT_Oxygenase_N; N-terminal domain of 2,4,5-trihydroxytoluene (ThT) oxygenase. This subfamily contains the N-terminal, n	58.32	35	17.61	2.3	13	113	6	2ZW5_B	Bleomycin acetyltransferase; dimer; two domains, Transferase; HET: COA; 2-4A (Streptomyces verticillus)	57.74	29	21.98	2.3	14	301
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																								
1	PF16138.9	; DUF4846; Domain of unknown function (4846)	79.7	8.5	28.33	3.8	31	240																																																								
2	PF08741.14	; YwhD; YwhD family	74.97	5.5	29.65	1.8	12	162																																																								
3	PF18337.5	; Tudor_RapA; RapA N-terminal Tudor like domain	71.4	29	20.43	3.9	26	62																																																								
4	PF13228.10	; DUF4037; Domain of unknown function (DUF4037)	60.69	14	21.29	1.2	21	100																																																								
5	cd07267	ThT_Oxygenase_N; N-terminal domain of 2,4,5-trihydroxytoluene (ThT) oxygenase. This subfamily contains the N-terminal, n	58.32	35	17.61	2.3	13	113																																																								
6	2ZW5_B	Bleomycin acetyltransferase; dimer; two domains, Transferase; HET: COA; 2-4A (Streptomyces verticillus)	57.74	29	21.98	2.3	14	301																																																								
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<p><i>Pas de synténie observées avec les gènes voisins directs (fonction inconnue) mais les gènes voisins un peu plus lointains sont des AAA-ATPase, des recombinaisons et des exonucléases.</i></p> 																																																															
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p><i>Non, elle est prédite globulaire par deepTMHMM</i></p> <p>DeepTMHMM - Most Likely Topology Type: Globular</p>  <p>DeepTMHMM - Posterior Probabilities</p> 																																																															
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<p><i>Oui.</i></p>																																																															
AlphaFold	<p><i>Structure :</i></p>																																																															

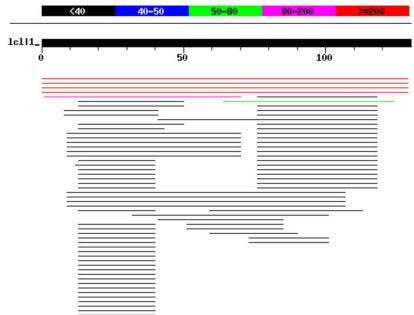
	<div><p>pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)</p><p>FoldSeek:</p><div><div>AFDB-PROTEOME 486 hits</div><div><div>GRAPHICAL</div><div>NUMERIC</div></div><table><tr><th>Target</th><th>Description</th><th>Scientific Name</th><th>Prob.</th><th>Seq. Id.</th><th>E-Value</th><th>Position in query</th><th>Alignment</th></tr><tr><td>AF-X8EVU6-F1-model_v4</td><td>Uncharacterized protein</td><td>Microbacterium uberans sp.</td><td>1.00</td><td>17.4</td><td>4.22e-2</td><td><div><div></div></div></td><td><div><div></div></div></td></tr><tr><td>AF-A0A0PYC9A8-F1-model_v4</td><td>Cs11pGS3100 protein</td><td>Oryza sativa Japonica Group</td><td>1.00</td><td>8.1</td><td>3.32e-2</td><td><div><div></div></div></td><td><div><div></div></div></td></tr><tr><td>AF-A0A0H0G0R8-F1-model_v4</td><td>DNA helicase</td><td>Klebsiella pneumoniae subsp.</td><td>0.99</td><td>16.1</td><td>6.44e-2</td><td><div><div></div></div></td><td><div><div></div></div></td></tr><tr><td>AF-Q329Y9-F1-model_v4</td><td>DNA helicase</td><td>Shigella dysenteriae Sd197</td><td>0.98</td><td>14.2</td><td>1.17e-1</td><td><div><div></div></div></td><td><div><div></div></div></td></tr><tr><td>AF-Q35311-F1-model_v4</td><td>DNA helicase II</td><td>Salmonella enterica subsp.</td><td>0.98</td><td>11.4</td><td>8.69e-2</td><td><div><div></div></div></td><td><div><div></div></div></td></tr></table></div></div> <p><i>La recherche par foldseek montre une homologie structurale avec des DNA helicase avec une probabilité de aux alentours de 98%. Ce sont des hélicases de bactéries.</i></p>	Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment	AF-X8EVU6-F1-model_v4	Uncharacterized protein	Microbacterium uberans sp.	1.00	17.4	4.22e-2	<div><div></div></div>	<div><div></div></div>	AF-A0A0PYC9A8-F1-model_v4	Cs11pGS3100 protein	Oryza sativa Japonica Group	1.00	8.1	3.32e-2	<div><div></div></div>	<div><div></div></div>	AF-A0A0H0G0R8-F1-model_v4	DNA helicase	Klebsiella pneumoniae subsp.	0.99	16.1	6.44e-2	<div><div></div></div>	<div><div></div></div>	AF-Q329Y9-F1-model_v4	DNA helicase	Shigella dysenteriae Sd197	0.98	14.2	1.17e-1	<div><div></div></div>	<div><div></div></div>	AF-Q35311-F1-model_v4	DNA helicase II	Salmonella enterica subsp.	0.98	11.4	8.69e-2	<div><div></div></div>	<div><div></div></div>
Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment																																										
AF-X8EVU6-F1-model_v4	Uncharacterized protein	Microbacterium uberans sp.	1.00	17.4	4.22e-2	<div><div></div></div>	<div><div></div></div>																																										
AF-A0A0PYC9A8-F1-model_v4	Cs11pGS3100 protein	Oryza sativa Japonica Group	1.00	8.1	3.32e-2	<div><div></div></div>	<div><div></div></div>																																										
AF-A0A0H0G0R8-F1-model_v4	DNA helicase	Klebsiella pneumoniae subsp.	0.99	16.1	6.44e-2	<div><div></div></div>	<div><div></div></div>																																										
AF-Q329Y9-F1-model_v4	DNA helicase	Shigella dysenteriae Sd197	0.98	14.2	1.17e-1	<div><div></div></div>	<div><div></div></div>																																										
AF-Q35311-F1-model_v4	DNA helicase II	Salmonella enterica subsp.	0.98	11.4	8.69e-2	<div><div></div></div>	<div><div></div></div>																																										
DECISION:	<i>NFK</i>																																																

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	28
Coordonnées du Stop	20953
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	Gap 60
Coordonnées du Start retenu	21345
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	 <p>Potentiel codant identique chez les 2</p>

	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p>  <p>Sequences producing significant alignments:</p> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Nucci_28, function unknown, 130</td><td>278</td><td>4e-75</td></tr><tr><td>Mandalorian_28, function unknown, 130</td><td>278</td><td>4e-75</td></tr><tr><td>Carostasia_Draft_28, function unknown, 130</td><td>278</td><td>4e-75</td></tr><tr><td>Quartz_28, function unknown, 130</td><td>276</td><td>1e-74</td></tr><tr><td>Juanyo_28, function unknown, 132</td><td>100</td><td>9e-22</td></tr><tr><td>Gingerbug_58, function unknown, 181</td><td>76</td><td>3e-14</td></tr><tr><td>Shaobing_82, function unknown, 115</td><td>31</td><td>0.92</td></tr></table>		Score (bits)	E Value	Nucci_28, function unknown, 130	278	4e-75	Mandalorian_28, function unknown, 130	278	4e-75	Carostasia_Draft_28, function unknown, 130	278	4e-75	Quartz_28, function unknown, 130	276	1e-74	Juanyo_28, function unknown, 132	100	9e-22	Gingerbug_58, function unknown, 181	76	3e-14	Shaobing_82, function unknown, 115	31	0.92
	Score (bits)	E Value																							
Nucci_28, function unknown, 130	278	4e-75																							
Mandalorian_28, function unknown, 130	278	4e-75																							
Carostasia_Draft_28, function unknown, 130	278	4e-75																							
Quartz_28, function unknown, 130	276	1e-74																							
Juanyo_28, function unknown, 132	100	9e-22																							
Gingerbug_58, function unknown, 181	76	3e-14																							
Shaobing_82, function unknown, 115	31	0.92																							
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p><i>Chev de -39 mais pas le choix</i></p>																								
DECISION:	<p><i>Oui</i></p>																								

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																				
Quel start Glimmer et GeneMark suggèrent-ils ?	Même start proposé à 21345pb																																				
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,748</td><td>2,981</td><td>10</td><td>-2,443</td><td>ACACCACTTCAAGGAGCATCCC</td><td>ATG</td><td>21345</td><td>393</td></tr><tr><td>2</td><td>-5,382</td><td>1,281</td><td>12</td><td>-6,218</td><td>GCTGCTCCCCGGCACTACTCTC</td><td>GTG</td><td>21018</td><td>66</td></tr></table> <p>correspond au meilleur RBS : -2.443</p>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,748	2,981	10	-2,443	ACACCACTTCAAGGAGCATCCC	ATG	21345	393	2	-5,382	1,281	12	-6,218	GCTGCTCCCCGGCACTACTCTC	GTG	21018	66
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																													
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																													
1	-1,748	2,981	10	-2,443	ACACCACTTCAAGGAGCATCCC	ATG	21345	393																													
2	-5,382	1,281	12	-6,218	GCTGCTCCCCGGCACTACTCTC	GTG	21018	66																													
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	L'orf la plus longue est donc de 393pb donc un chev -39 qui dépasse 30 mais on ne peut pas supprimer le gène sinon trop grand gap																																				
Est-ce que le start est conservé chez les homologues voir Starterator ?	Pas trouvé sur starterator																																				
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>hypothetical protein SEA_NUCCI_28 [Microbacterium phage Nucci]</p> <p>Sequence ID: QDK02421.1 Length: 130 Number of Matches: 1</p> <p>See 1 more title(s) ▼ See all Identical Proteins (PG)</p> <p>Range 1: 1 to 130 GenPept Graphics ▼ Next Match ▲ Previous</p> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>266 bits(679)</td><td>2e-89</td><td>Compositional matrix adjust.</td><td>130/130(100%)</td><td>130/130(100%)</td><td>0/130(0%)</td></tr></table> <p>Query 1 MNITPEAIAHFVEWLNITVQPPYYDLNRKGTTKEVNAYADYLTEHGLTERDDTTYPSTW 60</p> <p>Sbjct 1 MNITPEAIAHFVEWLNITVQPPYYDLNRKGTTKEVNAYADYLTEHGLTERDDTTYPSTW 60</p> <p>Query 61 IPAVEAGTTVYAEWAVGPAQRSGWVDPQATNLKENTIVGELLPGTYLVGLPKGGITTL 120</p> <p>Sbjct 61 IPAVEAGTTVYAEWAVGPAQRSGWVDPQATNLKENTIVGELLPGTYLVGLPKGGITTL 120</p> <p>Query 121 RPGQFFLKVA 130</p> <p>Sbjct 121 RPGQFFLKVA 130</p> <p>Download ▼ GenPept Graphics</p> <p>hypothetical protein SEA_QUARTZ_28 [Microbacterium phage Quartz]</p> <p>Sequence ID: UVK59247.1 Length: 130 Number of Matches: 1</p> <p>Range 1: 1 to 130 GenPept Graphics ▼ Next Match ▲ Previous</p> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>265 bits(676)</td><td>6e-89</td><td>Compositional matrix adjust.</td><td>129/130(99%)</td><td>130/130(100%)</td><td>0/130(0%)</td></tr></table> <p>Query 1 MNITPEAIAHFVEWLNITVQPPYYDLNRKGTTKEVNAYADYLTEHGLTERDDTTYPSTW 60</p> <p>Sbjct 1 MNITPEAIAHFVEWLNITVQPPYYDLNRKGTTKEVNAYADYLTEHGLTERDDTTYPSTW 60</p> <p>Query 61 IPAVEAGTTVYAEWAVGPAQRSGWVDPQATNLKENTIVGELLPGTYLVGLPKGGITTL 120</p> <p>Sbjct 61 IPAVEAGTTVYAEWAVGPAQRSGWVDPQATNLKENTIVGELLPGTYLVGLPKGGITTL 120</p> <p>Query 121 RPGQFFLKVA 130</p> <p>Sbjct 121 RPGQFFLKVA 130</p>	Score	Expect	Method	Identities	Positives	Gaps	266 bits(679)	2e-89	Compositional matrix adjust.	130/130(100%)	130/130(100%)	0/130(0%)	Score	Expect	Method	Identities	Positives	Gaps	265 bits(676)	6e-89	Compositional matrix adjust.	129/130(99%)	130/130(100%)	0/130(0%)												
Score	Expect	Method	Identities	Positives	Gaps																																
266 bits(679)	2e-89	Compositional matrix adjust.	130/130(100%)	130/130(100%)	0/130(0%)																																
Score	Expect	Method	Identities	Positives	Gaps																																
265 bits(676)	6e-89	Compositional matrix adjust.	129/130(99%)	130/130(100%)	0/130(0%)																																
DECISION:	21345 car trouvé chez les homologues. ORF le plus long et meilleur score RBS																																				

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																									
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <div><div>Mouse-over to show define and scores. Click to show alignments</div><div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>=200</div></div></div><div><div>1c111_</div><div><div>0</div><div>50</div><div>100</div></div></div><div><div>Sequences producing significant alignments:</div><div><table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Nucci_28, function unknown, 130</td><td>278</td><td>4e-75</td></tr><tr><td>Mandalorian_28, function unknown, 130</td><td>278</td><td>4e-75</td></tr><tr><td>Carostasia_Draft_28, function unknown, 130</td><td>278</td><td>4e-75</td></tr><tr><td>Quartz_28, function unknown, 130</td><td>275</td><td>1e-74</td></tr><tr><td>Juanyo_28, function unknown, 132</td><td>100</td><td>1e-21</td></tr><tr><td>Gingerbug_58, function unknown, 181</td><td>76</td><td>3e-14</td></tr><tr><td>Shaobing_82, function unknown, 115</td><td>31</td><td>0.95</td></tr><tr><td>Niklas_83, function unknown, 91</td><td>31</td><td>0.95</td></tr><tr><td>Sombrero_12, minor tail protein, 511</td><td>30</td><td>1.6</td></tr><tr><td>Scarlett_82, function unknown, 114</td><td>30</td><td>1.6</td></tr><tr><td>Oscar_82, function unknown, 111</td><td>30</td><td>1.6</td></tr><tr><td>Jellybones_12, minor tail protein, 511</td><td>30</td><td>1.6</td></tr><tr><td>FelixAlejandro_13, minor tail protein, 511</td><td>30</td><td>1.6</td></tr><tr><td>Sidious_17, tape measure protein, 2157</td><td>30</td><td>2.1</td></tr><tr><td>Validus_86, function unknown, 120</td><td>30</td><td>2.8</td></tr><tr><td>Sam12_12, hydrolase, 511</td><td>30</td><td>2.8</td></tr><tr><td>RemRem_13, minor tail protein, 511</td><td>30</td><td>2.8</td></tr><tr><td>Poland_Draft_15, function unknown, 511</td><td>30</td><td>2.8</td></tr></table></div></div></div>		Score (bits)	E Value	Nucci_28, function unknown, 130	278	4e-75	Mandalorian_28, function unknown, 130	278	4e-75	Carostasia_Draft_28, function unknown, 130	278	4e-75	Quartz_28, function unknown, 130	275	1e-74	Juanyo_28, function unknown, 132	100	1e-21	Gingerbug_58, function unknown, 181	76	3e-14	Shaobing_82, function unknown, 115	31	0.95	Niklas_83, function unknown, 91	31	0.95	Sombrero_12, minor tail protein, 511	30	1.6	Scarlett_82, function unknown, 114	30	1.6	Oscar_82, function unknown, 111	30	1.6	Jellybones_12, minor tail protein, 511	30	1.6	FelixAlejandro_13, minor tail protein, 511	30	1.6	Sidious_17, tape measure protein, 2157	30	2.1	Validus_86, function unknown, 120	30	2.8	Sam12_12, hydrolase, 511	30	2.8	RemRem_13, minor tail protein, 511	30	2.8	Poland_Draft_15, function unknown, 511	30	2.8
	Score (bits)	E Value																																																								
Nucci_28, function unknown, 130	278	4e-75																																																								
Mandalorian_28, function unknown, 130	278	4e-75																																																								
Carostasia_Draft_28, function unknown, 130	278	4e-75																																																								
Quartz_28, function unknown, 130	275	1e-74																																																								
Juanyo_28, function unknown, 132	100	1e-21																																																								
Gingerbug_58, function unknown, 181	76	3e-14																																																								
Shaobing_82, function unknown, 115	31	0.95																																																								
Niklas_83, function unknown, 91	31	0.95																																																								
Sombrero_12, minor tail protein, 511	30	1.6																																																								
Scarlett_82, function unknown, 114	30	1.6																																																								
Oscar_82, function unknown, 111	30	1.6																																																								
Jellybones_12, minor tail protein, 511	30	1.6																																																								
FelixAlejandro_13, minor tail protein, 511	30	1.6																																																								
Sidious_17, tape measure protein, 2157	30	2.1																																																								
Validus_86, function unknown, 120	30	2.8																																																								
Sam12_12, hydrolase, 511	30	2.8																																																								
RemRem_13, minor tail protein, 511	30	2.8																																																								
Poland_Draft_15, function unknown, 511	30	2.8																																																								

>Nucci_28, function unknown, 130
Length = 130

Score = 278 bits (710), Expect = 4e-75
Identities = 130/130 (100%), Positives = 130/130 (100%)

Query: 1 MNITPEAIAHFVEHLNTVQPPYYDLNRKGWTTKEVNAYADYLTEHGLIERDDTTYPSWTS 60
Sbjct: 1 MNITPEAIAHFVEHLNTVQPPYYDLNRKGWTTKEVNAYADYLTEHGLIERDDTTYPSWTS 60

Query: 61 IPAVEAGTTVYAEHAWGPAQRGSGWVDPQATNLKEWTIVGELLPGTYLVGLPKGGTITL 120
IPAVEAGTTVYAEHAWGPAQRGSGWVDPQATNLKEWTIVGELLPGTYLVGLPKGGTITL 120
Sbjct: 61 IPAVEAGTTVYAEHAWGPAQRGSGWVDPQATNLKEWTIVGELLPGTYLVGLPKGGTITL 120

Query: 121 RPKQFFLKVA 130
RPKQFFLKVA
Sbjct: 121 RPKQFFLKVA 130

*** nr :
(q#: s#) : [alignment] ; e-value :

No putative conserved domains have been detected

Distribution of the top 4 Blast Hits on 4 subject sequences

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession
<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_28 [Microbacterium phage Nucci]	Microbacterium phage Nucci	265	265	99%	2e-69	100.00%	130	ODK02421.1
<input checked="" type="checkbox"/> hypothetical protein SEA_OUHAT_28 [Microbacterium phage Ouhat]	Microbacterium phage Ouhat	265	265	99%	6e-69	99.29%	130	LO459247.1
<input checked="" type="checkbox"/> hypothetical protein SEA_JUANYO_28 [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	97.8	97.8	45%	6e-23	71.67%	130	US44778.1
<input checked="" type="checkbox"/> hypothetical protein SEA_OINGERSVU_28 [Microbacterium phage Oingervu]	Microbacterium phage Oingervu	73.6	73.6	48%	7e-13	61.90%	181	U061392.1

hypothetical protein SEA_NUCCI_28 [Microbacterium phage Nucci]
Sequence ID: **ODK02421.1** Length: 130 Number of Matches: 1
[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 130 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
266 bits(679)	2e-69	Compositional matrix adjust.	130/130(100%)	130/130(100%)	0/130(0%)

Query 1 MNITPEAIAHFVEHLNTVQPPYYDLNRKGWTTKEVNAYADYLTEHGLIERDDTTYPSWTS 60
Sbjct 1 MNITPEAIAHFVEHLNTVQPPYYDLNRKGWTTKEVNAYADYLTEHGLIERDDTTYPSWTS 60

Query 61 IPAVEAGTTVYAEHAWGPAQRGSGWVDPQATNLKEWTIVGELLPGTYLVGLPKGGTITL 120
IPAVEAGTTVYAEHAWGPAQRGSGWVDPQATNLKEWTIVGELLPGTYLVGLPKGGTITL 120
Sbjct 61 IPAVEAGTTVYAEHAWGPAQRGSGWVDPQATNLKEWTIVGELLPGTYLVGLPKGGTITL 120

Query 121 RPKQFFLKVA 130
RPKQFFLKVA
Sbjct 121 RPKQFFLKVA 130

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Number of Hits: 250
Query MSA diversity (Neff): 1.24194
Note: your query alignment consists of only 3 sequence(s).

Visualization

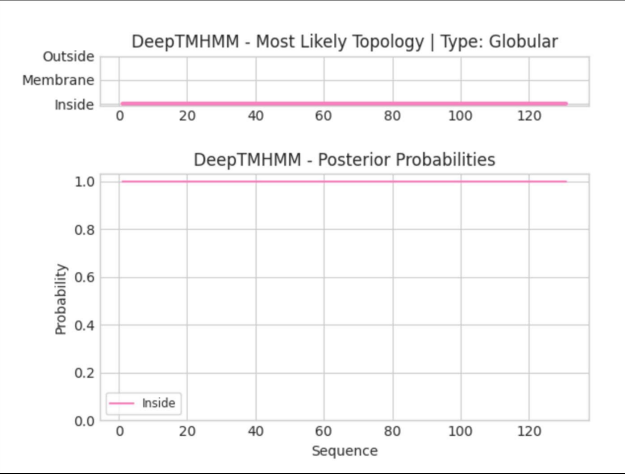
Resubmit Section

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	8B9L_B	RNA-binding protein E3; E3L_PRR, Vaccinia Virus, VIRAL PROTEIN-TRANSFERASE complex(Vaccinia virus WR)	91.49	1.2	25.41	3.8	32	62
2	2L54_A	Double-stranded RNA-specific adenosine deaminase; HYDROLASE; NMR (Homo sapiens) SCOP: a 4.5.19	87.89	3.2	23.39	3.7	33	63
3	3EYL_A	Z-DNA-binding protein 1; Alternative splicing, DNA-binding, Polymorphism, DNA Binding Protein-Z-DNA COMPLEX, DNA Binding	87.76	2.3	27.31	3.4	32	72

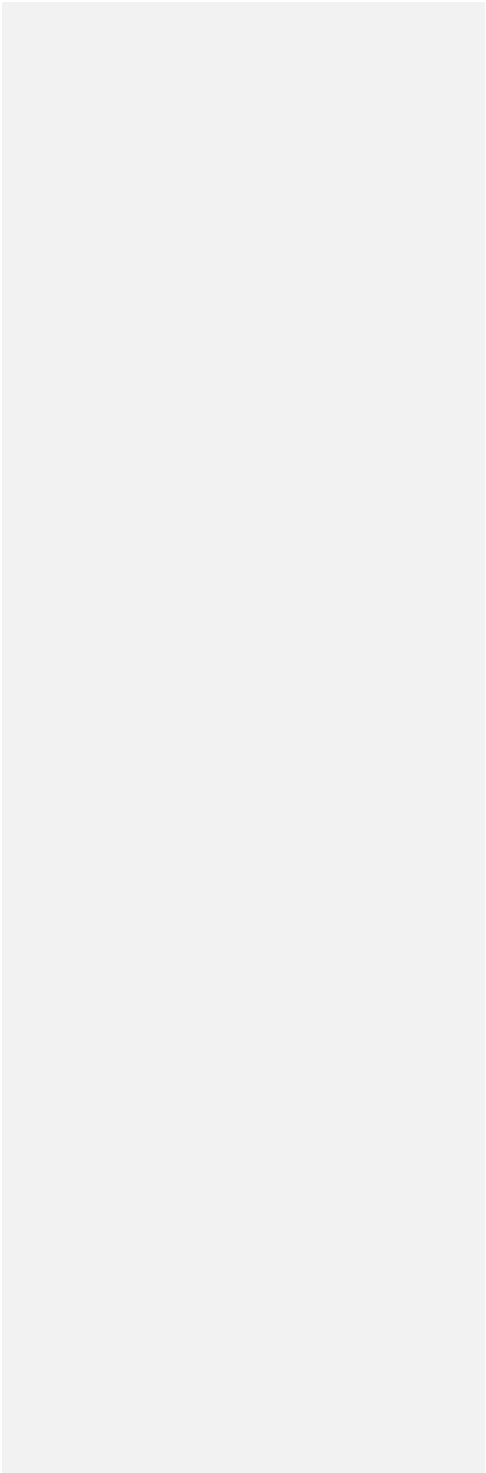
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Aucune synténie observée

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



	<i>Non</i>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui</i>
DECISION:	<i>NFK</i>



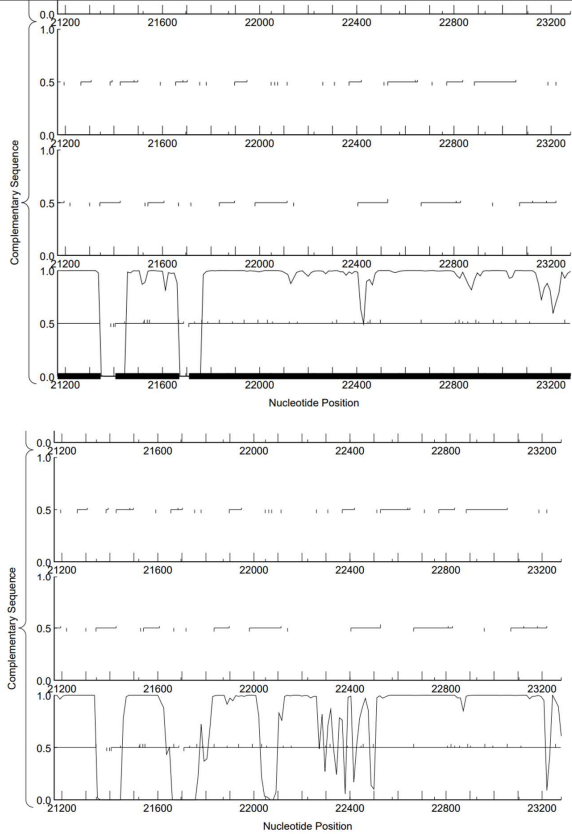
Student Gene Annotation Worksheet

Carosmatique-gene-29

Basic Phage Information	
Nom du Phage	carostasia
Gène #	29
Coordonnées du Stop	21406
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	gap 39
Coordonnées du Start retenu	21669
Fonction prédite	membrane protein

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>YES BOTH</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<i>Cadre+3</i>

																						
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_29, function unknown, 94</td><td>200</td><td>1e-51</td></tr><tr><td>Nucci_29, function unknown, 87</td><td>183</td><td>1e-46</td></tr><tr><td>Mandalorian_29, function unknown, 87</td><td>183</td><td>1e-46</td></tr><tr><td>Quartz_29, function unknown, 87</td><td>166</td><td>1e-41</td></tr><tr><td>YuuY_30, function unknown, 63</td><td>79</td><td>4e-15</td></tr><tr><td>Schubert_30, function unknown, 93</td><td>59</td><td>3e-09</td></tr></table>		Score (bits)	E Value	Carostasia_Draft_29, function unknown, 94	200	1e-51	Nucci_29, function unknown, 87	183	1e-46	Mandalorian_29, function unknown, 87	183	1e-46	Quartz_29, function unknown, 87	166	1e-41	YuuY_30, function unknown, 63	79	4e-15	Schubert_30, function unknown, 93	59	3e-09
	Score (bits)	E Value																				
Carostasia_Draft_29, function unknown, 94	200	1e-51																				
Nucci_29, function unknown, 87	183	1e-46																				
Mandalorian_29, function unknown, 87	183	1e-46																				
Quartz_29, function unknown, 87	166	1e-41																				
YuuY_30, function unknown, 63	79	4e-15																				
Schubert_30, function unknown, 93	59	3e-09																				
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Non</i>																					
DECISION:	<i>oui</i>																					

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																												
Quel start Glimmer et GeneMark suggèrent-ils ?	21690 / 21669																																																																																																												
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Start glimmer : GTG, de score -7.436 Start genemark : ATG, de score -4.699</div> <div><div>Starts: 10 Selected: 1</div><div>ORF Start : 21669 ORF Stop : 21406 ORF Length: 264</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 33.3 100.0 33.3 9 3' End 80.0 60.0 20.0 30</div><div>SD Scoring Matrix Kibler6 Spacing Weight Matrix Karlin Medium</div><div>Explore Document</div></div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,983</td><td>1,467</td><td>10</td><td>-5,678</td><td>ACCTGTCTAGAGACTAGGGC</td><td>TTG</td><td>21705</td><td>300</td></tr><tr><td>2</td><td>-5,503</td><td>1,224</td><td>13</td><td>-6,548</td><td>AGAAGACTAGGGCTTGTGCCCC</td><td>TTG</td><td>21696</td><td>291</td></tr><tr><td>3</td><td>-6,089</td><td>0,950</td><td>14</td><td>-7,436</td><td>CTAGGGCTTGTGCCCTTGCOC</td><td>GTG</td><td>21690</td><td>285</td></tr><tr><td>4</td><td>-3,925</td><td>1,963</td><td>9</td><td>-4,699</td><td>GTTGCTGTGGTACGGTTCTTC</td><td>ATG</td><td>21669</td><td>264</td></tr><tr><td>5</td><td>-4,025</td><td>1,916</td><td>16</td><td>-5,821</td><td>TACCCAGGCGACGATCAACGTC</td><td>ATG</td><td>21549</td><td>144</td></tr><tr><td>6</td><td>-4,547</td><td>1,672</td><td>7</td><td>-6,070</td><td>GACGATCAACGTGATGGTCTG</td><td>ATG</td><td>21540</td><td>135</td></tr><tr><td>7</td><td>-5,571</td><td>1,192</td><td>14</td><td>-6,917</td><td>CATGGTCTGATGCCCTCTCTG</td><td>ATG</td><td>21528</td><td>123</td></tr><tr><td>8</td><td>-5,571</td><td>1,192</td><td>5</td><td>-7,571</td><td>GTCTCTGATGCCCTCTCTGATG</td><td>GTG</td><td>21525</td><td>120</td></tr><tr><td>9</td><td>-3,604</td><td>2,113</td><td>16</td><td>-5,400</td><td>GGACGAGTTCTGGAGGCGCTG</td><td>GTG</td><td>21447</td><td>42</td></tr><tr><td>10</td><td>-5,202</td><td>1,365</td><td>15</td><td>-6,804</td><td>CCAGATCGGTGAATATTGGGGG</td><td>TTG</td><td>21414</td><td>9</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,983	1,467	10	-5,678	ACCTGTCTAGAGACTAGGGC	TTG	21705	300	2	-5,503	1,224	13	-6,548	AGAAGACTAGGGCTTGTGCCCC	TTG	21696	291	3	-6,089	0,950	14	-7,436	CTAGGGCTTGTGCCCTTGCOC	GTG	21690	285	4	-3,925	1,963	9	-4,699	GTTGCTGTGGTACGGTTCTTC	ATG	21669	264	5	-4,025	1,916	16	-5,821	TACCCAGGCGACGATCAACGTC	ATG	21549	144	6	-4,547	1,672	7	-6,070	GACGATCAACGTGATGGTCTG	ATG	21540	135	7	-5,571	1,192	14	-6,917	CATGGTCTGATGCCCTCTCTG	ATG	21528	123	8	-5,571	1,192	5	-7,571	GTCTCTGATGCCCTCTCTGATG	GTG	21525	120	9	-3,604	2,113	16	-5,400	GGACGAGTTCTGGAGGCGCTG	GTG	21447	42	10	-5,202	1,365	15	-6,804	CCAGATCGGTGAATATTGGGGG	TTG	21414	9
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																					
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																					
1	-4,983	1,467	10	-5,678	ACCTGTCTAGAGACTAGGGC	TTG	21705	300																																																																																																					
2	-5,503	1,224	13	-6,548	AGAAGACTAGGGCTTGTGCCCC	TTG	21696	291																																																																																																					
3	-6,089	0,950	14	-7,436	CTAGGGCTTGTGCCCTTGCOC	GTG	21690	285																																																																																																					
4	-3,925	1,963	9	-4,699	GTTGCTGTGGTACGGTTCTTC	ATG	21669	264																																																																																																					
5	-4,025	1,916	16	-5,821	TACCCAGGCGACGATCAACGTC	ATG	21549	144																																																																																																					
6	-4,547	1,672	7	-6,070	GACGATCAACGTGATGGTCTG	ATG	21540	135																																																																																																					
7	-5,571	1,192	14	-6,917	CATGGTCTGATGCCCTCTCTG	ATG	21528	123																																																																																																					
8	-5,571	1,192	5	-7,571	GTCTCTGATGCCCTCTCTGATG	GTG	21525	120																																																																																																					
9	-3,604	2,113	16	-5,400	GGACGAGTTCTGGAGGCGCTG	GTG	21447	42																																																																																																					
10	-5,202	1,365	15	-6,804	CCAGATCGGTGAATATTGGGGG	TTG	21414	9																																																																																																					
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start ATG ne donne pas l'ORF le plus long Mais Meilleur score RBS																																																																																																												
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																																													
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Avec le start 21669 :																																																																																																												

	<div><div><div><div>Download</div><div>GenPept</div><div>Graphics</div></div><div>▼ Next ▲ Previous ◀ Descriptions</div></div><div><div>membrane protein [Microbacterium phage Nucci]</div><div>Sequence ID: QDK02422.1 Length: 87 Number of Matches: 1</div><div>See 1 more title(s) ▼ See all Identical Proteins (PG)</div></div><div><div>Range 1: 1 to 87 GenPept Graphics</div><div>▼ Next Match ▲ Previous Match</div><div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>176 bits(446)</td><td>4e-55</td><td>Compositional matrix adjust.</td><td>87/87(100%)</td><td>87/87(100%)</td><td>0/87(0%)</td></tr></table><div>Query 8 HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA 67</div><div>HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA</div><div>Subject 1 HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA 68</div><div>Query 68 YMGRYMEVLRLVGLIQIGEWMLL 94</div><div>YMGRYMEVLRLVGLIQIGEWMLL</div><div>Subject 61 YMGRYMEVLRLVGLIQIGEWMLL 87</div></div><div><div>Related Information</div><div>Identical Proteins - Identical proteins to QDK02422.1</div></div></div><div><div><div>Download</div><div>GenPept</div><div>Graphics</div></div><div>▼ Next ▲ Previous ◀ Descriptions</div></div><div><div>membrane protein [Microbacterium phage Quartz]</div><div>Sequence ID: UVK59248.1 Length: 87 Number of Matches: 1</div></div><div><div>Range 1: 1 to 87 GenPept Graphics</div><div>▼ Next Match ▲ Previous Match</div><div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>159 bits(403)</td><td>1e-48</td><td>Compositional matrix adjust.</td><td>79/87(91%)</td><td>81/87(93%)</td><td>0/87(0%)</td></tr></table><div>Query 8 HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA 67</div><div>H T L R T S E D L V A A V G A D T E P E Q E K A T P T A T I N W N V L P R L U N V F G V W A L C A A</div><div>Subject 1 R K T T L E T R P T S L E D L V A A V G A D T E P E Q E K A T P T E A T I N W N V L P R L U N V F G V W A L C A A 68</div><div>Query 68 YMGRYMEVLRLVGLIQIGEWMLL 94</div><div>YMGRYMEVLRLVGLIQIGEWMLL</div><div>Subject 61 YMGRYMEVLRLVGLIQIGEWMLL 87</div></div></div><div><p>Ce start n'est pas conservé chez les homologues</p></div><div><p>Avec le start 21669 :</p></div><div><div><div>Download</div><div>GenPept</div><div>Graphics</div></div><div>▼ Next ▲ Previous ◀ Descriptions</div></div><div><div>membrane protein [Microbacterium phage Nucci]</div><div>Sequence ID: QDK02422.1 Length: 87 Number of Matches: 1</div><div>See 1 more title(s) ▼ See all Identical Proteins (PG)</div></div><div><div>Range 1: 1 to 87 GenPept Graphics</div><div>▼ Next Match ▲ Previous Match</div><div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>176 bits(447)</td><td>2e-55</td><td>Compositional matrix adjust.</td><td>87/87(100%)</td><td>87/87(100%)</td><td>0/87(0%)</td></tr></table><div>Query 1 HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA 68</div><div>HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA</div><div>Subject 1 HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA 68</div><div>Query 61 YMGRYMEVLRLVGLIQIGEWMLL 87</div><div>YMGRYMEVLRLVGLIQIGEWMLL</div><div>Subject 61 YMGRYMEVLRLVGLIQIGEWMLL 87</div></div><div><div>Related Information</div><div>Identical Proteins - Identical proteins to QDK02422.1</div></div></div><div><div><div>Download</div><div>GenPept</div><div>Graphics</div></div><div>▼ Next ▲ Previous ◀ Descriptions</div></div><div><div>membrane protein [Microbacterium phage Quartz]</div><div>Sequence ID: UVK59248.1 Length: 87 Number of Matches: 1</div></div><div><div>Range 1: 1 to 87 GenPept Graphics</div><div>▼ Next Match ▲ Previous Match</div><div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>160 bits(404)</td><td>7e-49</td><td>Compositional matrix adjust.</td><td>79/87(91%)</td><td>81/87(93%)</td><td>0/87(0%)</td></tr></table><div>Query 1 HTTLTPRRTTTSFEDLVAAVGADTEPEQEKATPTQATINWNVLPRLUNVFGVWALCAA 68</div><div>H T L R T S E D L V A A V G A D T E P E Q E K A T P T A T I N W N V L P R L U N V F G V W A L C A A</div><div>Subject 1 R K T T L E T R P T S L E D L V A A V G A D T E P E Q E K A T P T E A T I N W N V L P R L U N V F G V W A L C A A 68</div><div>Query 61 YMGRYMEVLRLVGLIQIGEWMLL 87</div><div>YMGRYMEVLRLVGLIQIGEWMLL</div><div>Subject 61 YMGRYMEVLRLVGLIQIGEWMLL 87</div></div></div><div><p>Le start conservé est donc celui à la position 21669</p></div></div>	Score	Expect	Method	Identities	Positives	Gaps	176 bits(446)	4e-55	Compositional matrix adjust.	87/87(100%)	87/87(100%)	0/87(0%)	Score	Expect	Method	Identities	Positives	Gaps	159 bits(403)	1e-48	Compositional matrix adjust.	79/87(91%)	81/87(93%)	0/87(0%)	Score	Expect	Method	Identities	Positives	Gaps	176 bits(447)	2e-55	Compositional matrix adjust.	87/87(100%)	87/87(100%)	0/87(0%)	Score	Expect	Method	Identities	Positives	Gaps	160 bits(404)	7e-49	Compositional matrix adjust.	79/87(91%)	81/87(93%)	0/87(0%)
Score	Expect	Method	Identities	Positives	Gaps																																												
176 bits(446)	4e-55	Compositional matrix adjust.	87/87(100%)	87/87(100%)	0/87(0%)																																												
Score	Expect	Method	Identities	Positives	Gaps																																												
159 bits(403)	1e-48	Compositional matrix adjust.	79/87(91%)	81/87(93%)	0/87(0%)																																												
Score	Expect	Method	Identities	Positives	Gaps																																												
176 bits(447)	2e-55	Compositional matrix adjust.	87/87(100%)	87/87(100%)	0/87(0%)																																												
Score	Expect	Method	Identities	Positives	Gaps																																												
160 bits(404)	7e-49	Compositional matrix adjust.	79/87(91%)	81/87(93%)	0/87(0%)																																												
DECISION:	<p>Le start conservé est le 21669 car il a un meilleur score, il correspond à un ATG et il est conservé chez les homologues.</p>																																																

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP	Listez le meilleur hit Blastp pour chaque source :

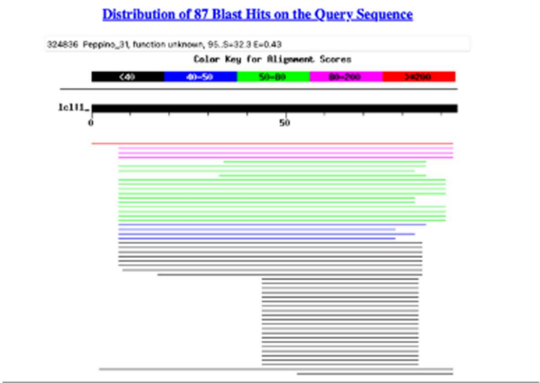
(phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

*** *PhagesDB* :

Sequences producing significant alignments:

	Score (bits)	E Value
Nucci_29, function unknown, 87	183	1e-46
dandalorian_29, function unknown, 87	183	1e-46
zarostasia_Draft_29, function unknown, 94	183	1e-46
quartz_29, function unknown, 87	166	1e-41
ruuY_30, function unknown, 63	79	4e-15
Schubert_30, function unknown, 93	59	3e-09
Koji_30, function unknown, 91	56	3e-08
Juanyo_29, function unknown, 64	56	3e-08

Le meilleur hit est une protéine à fonction inconnue chez le phage Nucci, avec une e-value= 1e-46.



*** *nr* :

☒ select all 11 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Nucci)	Microbacterium phage Nucci	176	176	98%	2e-55	100.00%	87	QDK02452.1
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Quartz)	Microbacterium phage Quartz	160	160	98%	8e-49	90.80%	87	UJX59248.1
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage YuuY)	Microbacterium phage YuuY	73.9	73.9	67%	6e-15	62.71%	63	QFF59365.1
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Juanyo)	Microbacterium phage Juanyo	52.0	52.0	61%	3e-08	40.74%	64	UJH44779.1

Descriptions **Graphics Summary** Alignments Taxonomy

☒ Show Conserved Domains

Alignment Scores: 0-40 40-50 50-60 60-70 70-80 80-90 90-100

No putative conserved domains have been detected

Distribution of the top 11 Blast Hits on 11 subject sequences

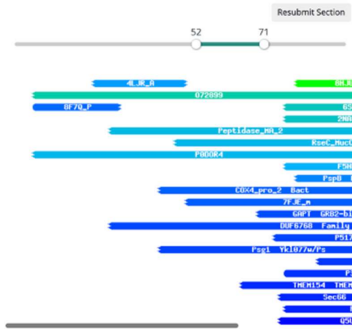
Le meilleur hit est une protéine de membrane du phage Nucci, avec une e-value= 2e-55, et une query cover = 100%

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un cribre HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Pas de correspondance pertinente.

No	Hit	Name	Probability	E-value	Score	SS	Aligned cdd	Target Length
1	BHDL_X	SUBUNIT X, RC-LH1 core complex, PHOTOSYNTHESIS, HET: RC_L, MG2, BPH, DGA, PKD, HEM, KGD, 2.8A (Roseiflexus cambialis DS)	74.98	15	23.02	3.3	20	32
2	659K_X	Cytoskeleton protein beta2, Ribosome, nascent chain in complex with SecA, RIBOSOME, HET: MG, 3.1A (Escherichia coli)	64.37	53	22.57	4.5	27	57
3	2NAB_A	Cytokine receptor common subunit beta1, transmembrane helix, MEMBRANE PROTEIN, NMR (phosphatidyl)	64.35	43	21.12	3.9	25	44

Visualization



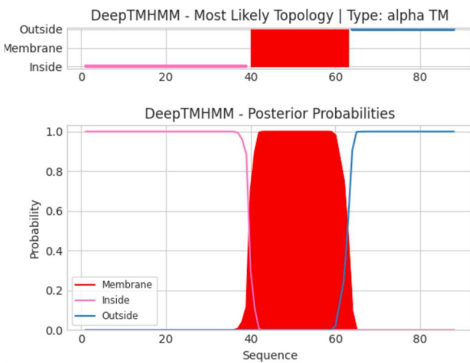
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

"Aucune synténie observée".



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

Oui, c'est cohérent avec l'homologie avec une protéine membranaire.



Phobius prediction

Prediction of UNNAMED

ID	UNNAMED			
FT	TOPO_DOM	1	46	CYTOPLASMIC.
FT	TRANSMEM	47	70	
FT	TOPO_DOM	71	94	NON CYTOPLASMIC.
//				

Phobius posterior probabilities for UNNAMED

transmembrane — cytoplasmic — non cytoplasmic — signal peptide

Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Non.

AlphaFold

Structure :

pLDDT: ■ Very low (<50) ■ Low (60) ■ OK (70) ■ Confident (80) ■ Very high (>90)

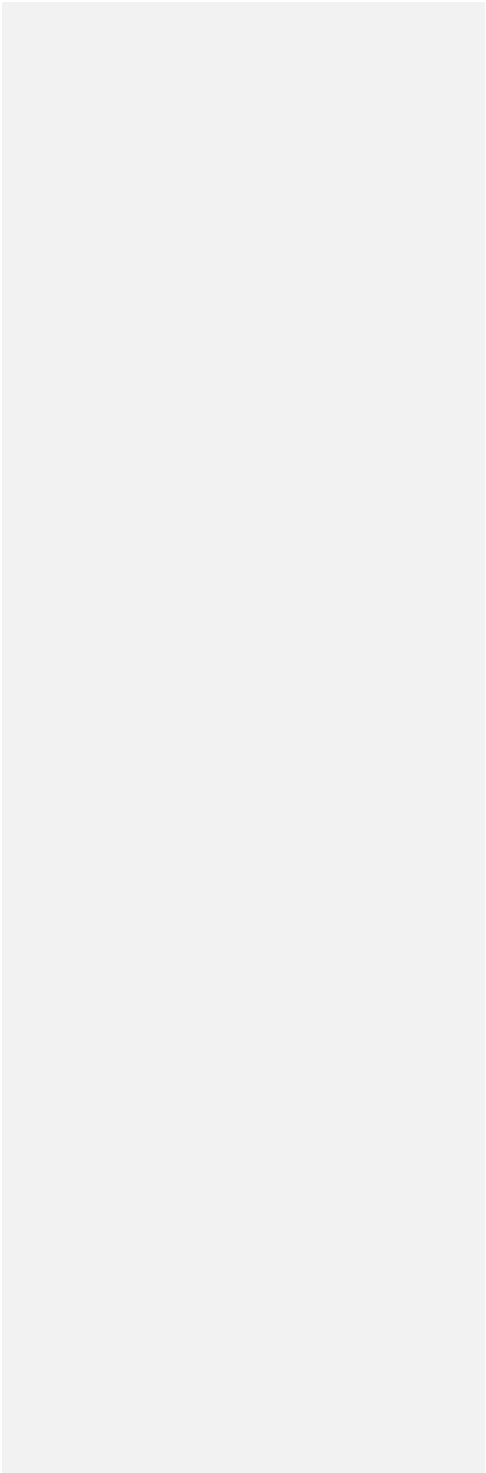
La structure est fiable au niveau de la plus grande hélice.

FoldSeek :

AFDB-PROTEOME 2 hits							GRAPHICAL	NUMERIC
Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment	
AF-ANANCOLUP-F1-mod	Uncharacterized protein	Stigmatopora japonica	0.12	16	3.63e+0			
AF-ANANHTPAZ-F1-mod	Uncharacterized protein	Oryzias latipes	0.09	16	9.08e+0			
AFDB-SWISSPROT 0 hits							GRAPHICAL	NUMERIC
Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment	
AFDB50 3 hits							GRAPHICAL	NUMERIC
Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment	
AF-ANAGQETZ-F1-mod	Protein	Lentulus lobatus	0.28	14.8	9.77e+0			
AF-ULAZ02-F1-mod	Uncharacterized protein	Pyronota unipunctatus CB	0.15	21.3	9.12e+0			
AF-ANAKRQNH-F1-mod	Uncharacterized protein	Stigmatopora sp. Y02.A104	0.13	16.1	9.77e+0			

Aucun hit pertinent, seulement des unknown proteins.

DECISION:	<i>Ce gène code pour une protéine de membrane.</i>

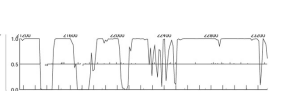
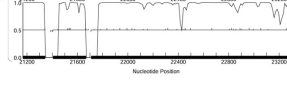


Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	30
Coordonnées du Stop	21709
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	ove 25
Coordonnées du Start retenu	23385
Fonction prédite	DNA primase/helicase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both

<p>Y-a-t-il des éléments supportant un potentiel codant ?</p>	<div><p>Page Coordinates: 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 13371</p><p>Page Coordinates complete sequence: 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 13371</p><p>Direct Sequence</p><p>Complementary Sequence</p></div> <p>Potentiel codant pour les 2</p>																																																																					
	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <p>0-40 40-50 50-60 60-70 70-80 80-90 90-100</p> <p>1c111</p> <p>Sequences producing significant alignments:</p> <table><tr><th></th><th>Score</th><th>E</th></tr><tr><th></th><th>(bits)</th><th>Value</th></tr><tr><td>Nucci_30, function unknown, 558</td><td>1134</td><td>0.0</td></tr><tr><td>Carostasia_Draft_30, function unknown, 558</td><td>1134</td><td>0.0</td></tr><tr><td>Mandalorian_30, RecA-like DNA recombinase, 558</td><td>1132</td><td>0.0</td></tr><tr><td>Quartz_31, DNA primase/helicase, 558</td><td>1123</td><td>0.0</td></tr><tr><td>Juanyo_31, RecA-like DNA recombinase, 558</td><td>1083</td><td>0.0</td></tr><tr><td>Yuuy_32, RecA-like DNA recombinase, 559</td><td>888</td><td>0.0</td></tr><tr><td>Williamstrong_32, DNA recombinase, 555</td><td>846</td><td>0.0</td></tr><tr><td>Sansa_31, DNA recombinase, 559</td><td>811</td><td>0.0</td></tr><tr><td>Mcubed_32, RecA-like DNA recombinase, 560</td><td>810</td><td>0.0</td></tr><tr><td>Fimny_32, RecA-like DNA recombinase, 560</td><td>810</td><td>0.0</td></tr><tr><td>Zenitsu_32, DNA primase/helicase, 560</td><td>809</td><td>0.0</td></tr><tr><td>Shamu_Draft_33, function unknown, 560</td><td>802</td><td>0.0</td></tr><tr><td>Glamour_31, RecA-like DNA recombinase, 560</td><td>802</td><td>0.0</td></tr><tr><td>ColaCorta_32, DNA recombinase, 560</td><td>802</td><td>0.0</td></tr><tr><td>Andromedas_32, DNA recombinase, 560</td><td>802</td><td>0.0</td></tr><tr><td>Saratos_32, RecA-like DNA recombinase, 560</td><td>808</td><td>0.0</td></tr><tr><td>Eleri_32, DNA recombinase, 560</td><td>808</td><td>0.0</td></tr><tr><td>ChikPic_32, DNA primase/helicase, 560</td><td>808</td><td>0.0</td></tr><tr><td>Nebulous_31, DNA primase/helicase, 559</td><td>805</td><td>0.0</td></tr><tr><td>Clock_Draft_31, function unknown, 559</td><td>805</td><td>0.0</td></tr><tr><td>Neferthena_31, DNA recombinase, 560</td><td>800</td><td>0.0</td></tr></table> <p>On a bien d'autres phages qui possèdent le gène</p> <p>Gène suffisamment long, même sens que les voisins, cependant en fonction du start, gap de 44 ou chevauchement de 24</p>		Score	E		(bits)	Value	Nucci_30, function unknown, 558	1134	0.0	Carostasia_Draft_30, function unknown, 558	1134	0.0	Mandalorian_30, RecA-like DNA recombinase, 558	1132	0.0	Quartz_31, DNA primase/helicase, 558	1123	0.0	Juanyo_31, RecA-like DNA recombinase, 558	1083	0.0	Yuuy_32, RecA-like DNA recombinase, 559	888	0.0	Williamstrong_32, DNA recombinase, 555	846	0.0	Sansa_31, DNA recombinase, 559	811	0.0	Mcubed_32, RecA-like DNA recombinase, 560	810	0.0	Fimny_32, RecA-like DNA recombinase, 560	810	0.0	Zenitsu_32, DNA primase/helicase, 560	809	0.0	Shamu_Draft_33, function unknown, 560	802	0.0	Glamour_31, RecA-like DNA recombinase, 560	802	0.0	ColaCorta_32, DNA recombinase, 560	802	0.0	Andromedas_32, DNA recombinase, 560	802	0.0	Saratos_32, RecA-like DNA recombinase, 560	808	0.0	Eleri_32, DNA recombinase, 560	808	0.0	ChikPic_32, DNA primase/helicase, 560	808	0.0	Nebulous_31, DNA primase/helicase, 559	805	0.0	Clock_Draft_31, function unknown, 559	805	0.0	Neferthena_31, DNA recombinase, 560	800	0.0
	Score	E																																																																				
	(bits)	Value																																																																				
Nucci_30, function unknown, 558	1134	0.0																																																																				
Carostasia_Draft_30, function unknown, 558	1134	0.0																																																																				
Mandalorian_30, RecA-like DNA recombinase, 558	1132	0.0																																																																				
Quartz_31, DNA primase/helicase, 558	1123	0.0																																																																				
Juanyo_31, RecA-like DNA recombinase, 558	1083	0.0																																																																				
Yuuy_32, RecA-like DNA recombinase, 559	888	0.0																																																																				
Williamstrong_32, DNA recombinase, 555	846	0.0																																																																				
Sansa_31, DNA recombinase, 559	811	0.0																																																																				
Mcubed_32, RecA-like DNA recombinase, 560	810	0.0																																																																				
Fimny_32, RecA-like DNA recombinase, 560	810	0.0																																																																				
Zenitsu_32, DNA primase/helicase, 560	809	0.0																																																																				
Shamu_Draft_33, function unknown, 560	802	0.0																																																																				
Glamour_31, RecA-like DNA recombinase, 560	802	0.0																																																																				
ColaCorta_32, DNA recombinase, 560	802	0.0																																																																				
Andromedas_32, DNA recombinase, 560	802	0.0																																																																				
Saratos_32, RecA-like DNA recombinase, 560	808	0.0																																																																				
Eleri_32, DNA recombinase, 560	808	0.0																																																																				
ChikPic_32, DNA primase/helicase, 560	808	0.0																																																																				
Nebulous_31, DNA primase/helicase, 559	805	0.0																																																																				
Clock_Draft_31, function unknown, 559	805	0.0																																																																				
Neferthena_31, DNA recombinase, 560	800	0.0																																																																				
<p>Est-ce que le candidat est en contradiction avec les principes d'annotation ?</p>																																																																						
<p>DECISION:</p>	<p>Oui</p>																																																																					

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																																																																																																																																																		
Quel start Glimmer et GeneMark suggèrent-ils ?	Glimmer : 23385 GeneMark : 23316																																																																																																																																																																																																																																																																																																																																																																																		
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Seq</th><th>Raw ID</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Score</th><th>Sequence of the Start</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>2 Value</th><th>Distance</th><th>Score</th><th>Position</th><th>Position</th><th>Position</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,597</td><td>3,052</td><td>12</td><td>-2,433</td><td>CGAAGCTCAGGAAGTACAGAG</td><td>GTG</td><td>23385</td><td>1477</td><td></td></tr><tr><td>2</td><td>-5,691</td><td>1,141</td><td>16</td><td>-7,477</td><td>CCCTCTTGACCTAATCCACAGG</td><td>GTG</td><td>23352</td><td>1444</td><td></td></tr><tr><td>3</td><td>-4,648</td><td>1,415</td><td>15</td><td>-6,270</td><td>GCACCTCGCGGCTCCCGGCTCAC</td><td>GTG</td><td>23322</td><td>1414</td><td></td></tr><tr><td>4</td><td>-4,975</td><td>1,471</td><td>12</td><td>-5,411</td><td>CGGCTCGCGGCTCCCGGCTCAC</td><td>ATG</td><td>23316</td><td>1405</td><td></td></tr><tr><td>5</td><td>-3,867</td><td>1,990</td><td>6</td><td>-5,411</td><td>GGAGCGCTTCGCGGAGGCGCT</td><td>ATG</td><td>23259</td><td>1351</td><td></td></tr><tr><td>6</td><td>-5,791</td><td>1,089</td><td>13</td><td>-6,837</td><td>CCCGCGGAGCGTCCCGCTCGG</td><td>GTG</td><td>23115</td><td>1407</td><td></td></tr><tr><td>7</td><td>-4,636</td><td>1,407</td><td>11</td><td>-5,443</td><td>GGAGCGCTCAGGCGCTTCATC</td><td>ATG</td><td>23059</td><td>1350</td><td></td></tr><tr><td>8</td><td>-1,951</td><td>2,037</td><td>13</td><td>-2,596</td><td>CTCTTCAGAGAGCGCTTCCTC</td><td>GTG</td><td>22950</td><td>1290</td><td></td></tr><tr><td>9</td><td>-4,894</td><td>1,743</td><td>12</td><td>-5,230</td><td>CGATCGGAGCGGCTTCATCGGC</td><td>ATG</td><td>22945</td><td>1257</td><td></td></tr><tr><td>10</td><td>-6,213</td><td>0,892</td><td>10</td><td>-6,908</td><td>CGACATCAGCGAGCTTCCTCCG</td><td>GTG</td><td>22905</td><td>1197</td><td></td></tr><tr><td>11</td><td>-5,392</td><td>1,291</td><td>6</td><td>-7,127</td><td>CTCTTCGCGGCTCCCGGCGAC</td><td>ATG</td><td>22893</td><td>1195</td><td></td></tr><tr><td>12</td><td>-4,653</td><td>1,417</td><td>7</td><td>-6,136</td><td>CGGAGCGCTTCCTCGGCGCG</td><td>GTG</td><td>22840</td><td>1152</td><td></td></tr><tr><td>13</td><td>-5,184</td><td>1,373</td><td>11</td><td>-5,941</td><td>CTGCTCTCTCGAGCGGCGCG</td><td>GTG</td><td>22839</td><td>1131</td><td></td></tr><tr><td>14</td><td>-3,482</td><td>2,184</td><td>11</td><td>-4,209</td><td>CGGCGCGCTTCGAGCGCTCGG</td><td>ATG</td><td>22824</td><td>1116</td><td></td></tr><tr><td>15</td><td>-5,931</td><td>1,024</td><td>13</td><td>-6,977</td><td>CGTTCGATCTCTCTCTCTCTCT</td><td>ATG</td><td>22809</td><td>1101</td><td></td></tr><tr><td>16</td><td>-3,800</td><td>2,021</td><td>10</td><td>-4,495</td><td>AGAGCTGGGCTCGACGAGAC</td><td>TTG</td><td>22490</td><td>972</td><td></td></tr><tr><td>17</td><td>-5,508</td><td>1,222</td><td>14</td><td>-6,955</td><td>GGACGAGCGCTTCGCTTCAGG</td><td>ATG</td><td>22468</td><td>960</td><td></td></tr><tr><td>18</td><td>-3,857</td><td>1,994</td><td>5</td><td>-5,857</td><td>GGAGTCATTGAAGCTCGGAG</td><td>ATG</td><td>22400</td><td>792</td><td></td></tr><tr><td>19</td><td>-4,975</td><td>1,471</td><td>14</td><td>-6,771</td><td>CGGCGCGCTCATCGGACGCGCT</td><td>ATG</td><td>22448</td><td>780</td><td></td></tr><tr><td>20</td><td>-5,951</td><td>1,014</td><td>17</td><td>-7,951</td><td>CATCGGAGCGCTTCATCTCTCG</td><td>GTG</td><td>22449</td><td>741</td><td></td></tr><tr><td>21</td><td>-3,899</td><td>1,975</td><td>16</td><td>-5,495</td><td>CTCTCTGTAGAGCGCGCGAG</td><td>GTG</td><td>22392</td><td>694</td><td></td></tr><tr><td>22</td><td>-3,654</td><td>2,089</td><td>10</td><td>-4,349</td><td>CGGAGCGCGCTCGGCGCTCTCC</td><td>ATG</td><td>22320</td><td>612</td><td></td></tr><tr><td>23</td><td>-5,309</td><td>1,315</td><td>14</td><td>-6,455</td><td>GTTCATCGGAGCGCTTAGCGC</td><td>GTG</td><td>22302</td><td>594</td><td></td></tr><tr><td>24</td><td>-3,345</td><td>2,234</td><td>11</td><td>-4,102</td><td>GGGCTCTCTCGGAGCGCGCG</td><td>GTG</td><td>22158</td><td>450</td><td></td></tr><tr><td>25</td><td>-6,988</td><td>0,829</td><td>11</td><td>-7,745</td><td>CTCTGACCGCTCTCTCTCGCT</td><td>GTG</td><td>22125</td><td>417</td><td></td></tr><tr><td>26</td><td>-4,259</td><td>1,787</td><td>13</td><td>-5,345</td><td>CACATAGCAGCGCTTCGAGCTG</td><td>GTG</td><td>22056</td><td>340</td><td></td></tr><tr><td>27</td><td>-3,497</td><td>2,049</td><td>14</td><td>-5,459</td><td>TGACCTCTGCTCATTCAGAC</td><td>TTG</td><td>22041</td><td>333</td><td></td></tr><tr><td>28</td><td>-6,034</td><td>0,975</td><td>13</td><td>-7,080</td><td>GGTCTCATTCAGCGCTTCCTCC</td><td>ATG</td><td>22035</td><td>327</td><td></td></tr><tr><td>29</td><td>-4,259</td><td>1,787</td><td>7</td><td>-5,822</td><td>CACGCTCTCGAGCGAGAGGAC</td><td>ATG</td><td>21996</td><td>288</td><td></td></tr><tr><td>30</td><td>-5,205</td><td>1,364</td><td>11</td><td>-5,962</td><td>CAGAGAGCGCATCTCGCGCATC</td><td>ATG</td><td>21939</td><td>231</td><td></td></tr><tr><td>31</td><td>-3,240</td><td>2,289</td><td>10</td><td>-4,594</td><td>GCTCTTCAGAGCGCTTCAGAC</td><td>GTG</td><td>21891</td><td>189</td><td></td></tr><tr><td>32</td><td>-4,705</td><td>1,557</td><td>14</td><td>-6,052</td><td>CGAGCGCTGTGCGAGCTTATC</td><td>ATG</td><td>21837</td><td>129</td><td></td></tr><tr><td>33</td><td>-5,296</td><td>1,321</td><td>8</td><td>-6,518</td><td>GAGGAGCTCTCGGCTCGGC</td><td>GTG</td><td>21789</td><td>81</td><td></td></tr><tr><td>34</td><td>-4,291</td><td>1,791</td><td>9</td><td>-5,066</td><td>CGAGCTAAGCGCGGAGACAC</td><td>ATG</td><td>21765</td><td>57</td><td></td></tr><tr><td>35</td><td>-4,414</td><td>1,793</td><td>5</td><td>-6,416</td><td>CATTAGCGCGGCTTCAGAGATC</td><td>GTG</td><td>21735</td><td>27</td><td></td></tr></table> <p>Meilleur score RBS correspond au start prédit par glimmer -2.433 le start prédit par GeneMark est le 4° meilleur score</p>	Seq	Raw ID	Genomic	Spacer	Final	Score	Sequence of the Start	Start	Start	ORF	#	Score	2 Value	Distance	Score	Position	Position	Position	Position	Length	1	-1,597	3,052	12	-2,433	CGAAGCTCAGGAAGTACAGAG	GTG	23385	1477		2	-5,691	1,141	16	-7,477	CCCTCTTGACCTAATCCACAGG	GTG	23352	1444		3	-4,648	1,415	15	-6,270	GCACCTCGCGGCTCCCGGCTCAC	GTG	23322	1414		4	-4,975	1,471	12	-5,411	CGGCTCGCGGCTCCCGGCTCAC	ATG	23316	1405		5	-3,867	1,990	6	-5,411	GGAGCGCTTCGCGGAGGCGCT	ATG	23259	1351		6	-5,791	1,089	13	-6,837	CCCGCGGAGCGTCCCGCTCGG	GTG	23115	1407		7	-4,636	1,407	11	-5,443	GGAGCGCTCAGGCGCTTCATC	ATG	23059	1350		8	-1,951	2,037	13	-2,596	CTCTTCAGAGAGCGCTTCCTC	GTG	22950	1290		9	-4,894	1,743	12	-5,230	CGATCGGAGCGGCTTCATCGGC	ATG	22945	1257		10	-6,213	0,892	10	-6,908	CGACATCAGCGAGCTTCCTCCG	GTG	22905	1197		11	-5,392	1,291	6	-7,127	CTCTTCGCGGCTCCCGGCGAC	ATG	22893	1195		12	-4,653	1,417	7	-6,136	CGGAGCGCTTCCTCGGCGCG	GTG	22840	1152		13	-5,184	1,373	11	-5,941	CTGCTCTCTCGAGCGGCGCG	GTG	22839	1131		14	-3,482	2,184	11	-4,209	CGGCGCGCTTCGAGCGCTCGG	ATG	22824	1116		15	-5,931	1,024	13	-6,977	CGTTCGATCTCTCTCTCTCTCT	ATG	22809	1101		16	-3,800	2,021	10	-4,495	AGAGCTGGGCTCGACGAGAC	TTG	22490	972		17	-5,508	1,222	14	-6,955	GGACGAGCGCTTCGCTTCAGG	ATG	22468	960		18	-3,857	1,994	5	-5,857	GGAGTCATTGAAGCTCGGAG	ATG	22400	792		19	-4,975	1,471	14	-6,771	CGGCGCGCTCATCGGACGCGCT	ATG	22448	780		20	-5,951	1,014	17	-7,951	CATCGGAGCGCTTCATCTCTCG	GTG	22449	741		21	-3,899	1,975	16	-5,495	CTCTCTGTAGAGCGCGCGAG	GTG	22392	694		22	-3,654	2,089	10	-4,349	CGGAGCGCGCTCGGCGCTCTCC	ATG	22320	612		23	-5,309	1,315	14	-6,455	GTTCATCGGAGCGCTTAGCGC	GTG	22302	594		24	-3,345	2,234	11	-4,102	GGGCTCTCTCGGAGCGCGCG	GTG	22158	450		25	-6,988	0,829	11	-7,745	CTCTGACCGCTCTCTCTCGCT	GTG	22125	417		26	-4,259	1,787	13	-5,345	CACATAGCAGCGCTTCGAGCTG	GTG	22056	340		27	-3,497	2,049	14	-5,459	TGACCTCTGCTCATTCAGAC	TTG	22041	333		28	-6,034	0,975	13	-7,080	GGTCTCATTCAGCGCTTCCTCC	ATG	22035	327		29	-4,259	1,787	7	-5,822	CACGCTCTCGAGCGAGAGGAC	ATG	21996	288		30	-5,205	1,364	11	-5,962	CAGAGAGCGCATCTCGCGCATC	ATG	21939	231		31	-3,240	2,289	10	-4,594	GCTCTTCAGAGCGCTTCAGAC	GTG	21891	189		32	-4,705	1,557	14	-6,052	CGAGCGCTGTGCGAGCTTATC	ATG	21837	129		33	-5,296	1,321	8	-6,518	GAGGAGCTCTCGGCTCGGC	GTG	21789	81		34	-4,291	1,791	9	-5,066	CGAGCTAAGCGCGGAGACAC	ATG	21765	57		35	-4,414	1,793	5	-6,416	CATTAGCGCGGCTTCAGAGATC	GTG	21735	27	
Seq	Raw ID	Genomic	Spacer	Final	Score	Sequence of the Start	Start	Start	ORF																																																																																																																																																																																																																																																																																																																																																																										
#	Score	2 Value	Distance	Score	Position	Position	Position	Position	Length																																																																																																																																																																																																																																																																																																																																																																										
1	-1,597	3,052	12	-2,433	CGAAGCTCAGGAAGTACAGAG	GTG	23385	1477																																																																																																																																																																																																																																																																																																																																																																											
2	-5,691	1,141	16	-7,477	CCCTCTTGACCTAATCCACAGG	GTG	23352	1444																																																																																																																																																																																																																																																																																																																																																																											
3	-4,648	1,415	15	-6,270	GCACCTCGCGGCTCCCGGCTCAC	GTG	23322	1414																																																																																																																																																																																																																																																																																																																																																																											
4	-4,975	1,471	12	-5,411	CGGCTCGCGGCTCCCGGCTCAC	ATG	23316	1405																																																																																																																																																																																																																																																																																																																																																																											
5	-3,867	1,990	6	-5,411	GGAGCGCTTCGCGGAGGCGCT	ATG	23259	1351																																																																																																																																																																																																																																																																																																																																																																											
6	-5,791	1,089	13	-6,837	CCCGCGGAGCGTCCCGCTCGG	GTG	23115	1407																																																																																																																																																																																																																																																																																																																																																																											
7	-4,636	1,407	11	-5,443	GGAGCGCTCAGGCGCTTCATC	ATG	23059	1350																																																																																																																																																																																																																																																																																																																																																																											
8	-1,951	2,037	13	-2,596	CTCTTCAGAGAGCGCTTCCTC	GTG	22950	1290																																																																																																																																																																																																																																																																																																																																																																											
9	-4,894	1,743	12	-5,230	CGATCGGAGCGGCTTCATCGGC	ATG	22945	1257																																																																																																																																																																																																																																																																																																																																																																											
10	-6,213	0,892	10	-6,908	CGACATCAGCGAGCTTCCTCCG	GTG	22905	1197																																																																																																																																																																																																																																																																																																																																																																											
11	-5,392	1,291	6	-7,127	CTCTTCGCGGCTCCCGGCGAC	ATG	22893	1195																																																																																																																																																																																																																																																																																																																																																																											
12	-4,653	1,417	7	-6,136	CGGAGCGCTTCCTCGGCGCG	GTG	22840	1152																																																																																																																																																																																																																																																																																																																																																																											
13	-5,184	1,373	11	-5,941	CTGCTCTCTCGAGCGGCGCG	GTG	22839	1131																																																																																																																																																																																																																																																																																																																																																																											
14	-3,482	2,184	11	-4,209	CGGCGCGCTTCGAGCGCTCGG	ATG	22824	1116																																																																																																																																																																																																																																																																																																																																																																											
15	-5,931	1,024	13	-6,977	CGTTCGATCTCTCTCTCTCTCT	ATG	22809	1101																																																																																																																																																																																																																																																																																																																																																																											
16	-3,800	2,021	10	-4,495	AGAGCTGGGCTCGACGAGAC	TTG	22490	972																																																																																																																																																																																																																																																																																																																																																																											
17	-5,508	1,222	14	-6,955	GGACGAGCGCTTCGCTTCAGG	ATG	22468	960																																																																																																																																																																																																																																																																																																																																																																											
18	-3,857	1,994	5	-5,857	GGAGTCATTGAAGCTCGGAG	ATG	22400	792																																																																																																																																																																																																																																																																																																																																																																											
19	-4,975	1,471	14	-6,771	CGGCGCGCTCATCGGACGCGCT	ATG	22448	780																																																																																																																																																																																																																																																																																																																																																																											
20	-5,951	1,014	17	-7,951	CATCGGAGCGCTTCATCTCTCG	GTG	22449	741																																																																																																																																																																																																																																																																																																																																																																											
21	-3,899	1,975	16	-5,495	CTCTCTGTAGAGCGCGCGAG	GTG	22392	694																																																																																																																																																																																																																																																																																																																																																																											
22	-3,654	2,089	10	-4,349	CGGAGCGCGCTCGGCGCTCTCC	ATG	22320	612																																																																																																																																																																																																																																																																																																																																																																											
23	-5,309	1,315	14	-6,455	GTTCATCGGAGCGCTTAGCGC	GTG	22302	594																																																																																																																																																																																																																																																																																																																																																																											
24	-3,345	2,234	11	-4,102	GGGCTCTCTCGGAGCGCGCG	GTG	22158	450																																																																																																																																																																																																																																																																																																																																																																											
25	-6,988	0,829	11	-7,745	CTCTGACCGCTCTCTCTCGCT	GTG	22125	417																																																																																																																																																																																																																																																																																																																																																																											
26	-4,259	1,787	13	-5,345	CACATAGCAGCGCTTCGAGCTG	GTG	22056	340																																																																																																																																																																																																																																																																																																																																																																											
27	-3,497	2,049	14	-5,459	TGACCTCTGCTCATTCAGAC	TTG	22041	333																																																																																																																																																																																																																																																																																																																																																																											
28	-6,034	0,975	13	-7,080	GGTCTCATTCAGCGCTTCCTCC	ATG	22035	327																																																																																																																																																																																																																																																																																																																																																																											
29	-4,259	1,787	7	-5,822	CACGCTCTCGAGCGAGAGGAC	ATG	21996	288																																																																																																																																																																																																																																																																																																																																																																											
30	-5,205	1,364	11	-5,962	CAGAGAGCGCATCTCGCGCATC	ATG	21939	231																																																																																																																																																																																																																																																																																																																																																																											
31	-3,240	2,289	10	-4,594	GCTCTTCAGAGCGCTTCAGAC	GTG	21891	189																																																																																																																																																																																																																																																																																																																																																																											
32	-4,705	1,557	14	-6,052	CGAGCGCTGTGCGAGCTTATC	ATG	21837	129																																																																																																																																																																																																																																																																																																																																																																											
33	-5,296	1,321	8	-6,518	GAGGAGCTCTCGGCTCGGC	GTG	21789	81																																																																																																																																																																																																																																																																																																																																																																											
34	-4,291	1,791	9	-5,066	CGAGCTAAGCGCGGAGACAC	ATG	21765	57																																																																																																																																																																																																																																																																																																																																																																											
35	-4,414	1,793	5	-6,416	CATTAGCGCGGCTTCAGAGATC	GTG	21735	27																																																																																																																																																																																																																																																																																																																																																																											
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Avec le start de glimmer l'orf est la plus longue et fait 1677pb.																																																																																																																																																																																																																																																																																																																																																																																		
Est-ce que le start est conservé chez les homologues voir Starterator ?	Gene: Carostasia_30 Start: 23385, Stop: 21709, Start Num: 12 Candidate Starts for Carostasia_30: (Start: 12 @23385 has 143 MA's), (Start: 17 @23352 has 5 MA's), (Start: 22 @23322 has 1 MA's), (Start: 23 @23316 has 7 MA's), (27, 23259), (44, 23115), (52, 23058), (57, 22998), (61, 22965), (69, 22905), (71, 22893), (73, 22860), (76, 22839), (77, 22824), (80, 22809), (100, 22680), (102, 22668), (125, 22500), (131, 22458), (132, 22449), (138, 22392), (150, 22320), (155, 22302), (176, 22158), (183, 22125), (197, 22056), (200, 22041), (202, 22035), (205, 21996), (211, 21939), (217, 21891), (224, 21837), (232, 21789), (235, 21765), (240, 21735),																																																																																																																																																																																																																																																																																																																																																																																		

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?

Distribution of the top 100 Blast Hits on 100 subject sequences

[Download](#) [GenPept](#) [Graphics](#)

RecA-like DNA recombinase [Microbacterium phage Mandalorian]
Sequence ID: [QX013623.1](#) Length: 558 Number of Matches: 1

Range 1: 1 to 558 [GenPept](#) [Graphics](#) [Next Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1138 bits(2943)	0.0	Compositional matrix adjust.	557/558(99%)	557/558(99%)	0/558(0%)
Query 1	MDINPLDLIHRVNHSGVAGHVWMPSTSDIGDKKERFREGAML SARKPSFPDLRDSVDW	60			
Sbjct 1	MDINPLDLIHRVNHSGVAGHVWMPSTSDIGDKKERFREGAML SARKPSFPDLRDSVDW	60			
Query 61	YWTPAVSRGDTTRIKKGTEGTNYPQRAVWDCDDSYDRKLLALRPSYMMETSPGHIQA	120			
Sbjct 61	YWTPAVSRGDTTRIKKGTEGTNYPQRAVWDCDDSYDRKLLALRPSYMMETSPGHIQA	120			
Query 121	IWLKKEALPVSEFHRDGFIGMLTQAIGADKSGVDISQLLRVPGTMHHKREPFLGRVLSS	180			
Sbjct 121	IWLKKEALPVSEFHRDGFIGMLTQAIGADKSGVDISQLLRVPGTMHHKREPFLGRVLSS	180			
Query 181	GRVWTRGMLLSRVAKELGFSAGTAAELGKDDTYGDRSQLLWKFARNAEELGLDQSLAFKM	240			
Sbjct 181	GRVWTRGMLLSRVAKELGFSAGTAAELGKDDTYGDRSQLLWKFARNAEELGLDQSLAFKM	240			
Query 241	IKATEWNKWKDDPKLKKDDIERAYGHVPEKPRKEPKPTQDDSDDDDEEVEIANEMSTLD	300			
Sbjct 241	IKATEWNKWKDDPKLKKDDIERAYGHVPEKPRKEPKPTQDDSDDDDEEVEIANEMSTLD	300			

DECISION: Le start le plus probable est celui proposé par glimmer (23385) en GTG car il a le meilleur score RBS, correspond bien au début du potentiel codant, est retrouvé chez des homologues mais provoque un chev de 24pb (bcp mais <30)

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <div><div>Mouse-over to show define and scores. Click to show alignments</div><div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>200</div></div></div><div><div>1c111</div><div><div>0</div><div>100</div><div>200</div><div>300</div><div>400</div><div>500</div></div></div><div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></</div></div></div>

```
>Nucci_30, function unknown, 558
  Length = 558

Score = 1134 bits (2934), Expect = 0.0
Identities = 558/558 (100%), Positives = 558/558 (100%)

Query: 1 MDINPLDLIHRVWKHSGVAGHVWMPISDIGNDKERFREGAML SARKPSF PDLRDSVDW 60
MDINPLDLIHRVWKHSGVAGHVWMPISDIGNDKERFREGAML SARKPSF PDLRDSVDW
Sbjct: 1 MDINPLDLIHRVWKHSGVAGHVWMPISDIGNDKERFREGAML SARKPSF PDLRDSVDW 60

Query: 61 YWTPAVSRGDTTRIKKGTGTNYPQRAVWVDCDDSYDRKLL EALRPSYHNETSPGHIQA 120
YWTPAVSRGDTTRIKKGTGTNYPQRAVWVDCDDSYDRKLL EALRPSYHNETSPGHIQA
Sbjct: 61 YWTPAVSRGDTTRIKKGTGTNYPQRAVWVDCDDSYDRKLL EALRPSYHNETSPGHIQA 120

>Mandalorian_30, RecA-like DNA recombinase, 558
  Length = 558

Score = 1132 bits (2929), Expect = 0.0
Identities = 557/558 (99%), Positives = 557/558 (99%)

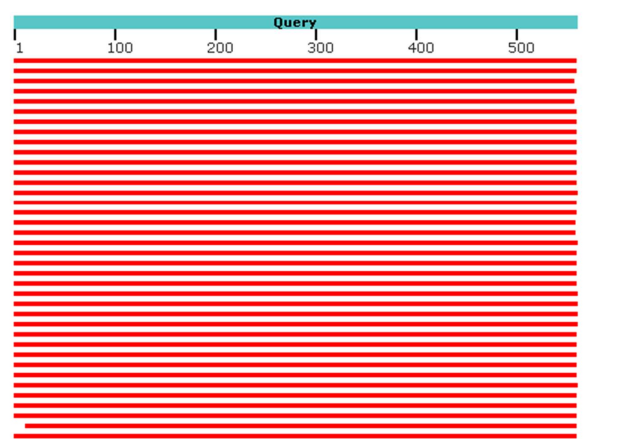
Query: 1 MDINPLDLIHRVWKHSGVAGHVWMPISDIGNDKERFREGAML SARKPSF PDLRDSVDW 60
MDINPLDLIHRVWKHSGVAGHVWMPISDIGNDKERFREGAML SARKPSF PDLRDSVDW
Sbjct: 1 MDINPLDLIHRVWKHSGVAGHVWMPISDIGNDKERFREGAML SARKPSF PDLRDSVDW 60

Query: 61 YWTPAVSRGDTTRIKKGTGTNYPQRAVWVDCDDSYDRKLL EALRPSYHNETSPGHIQA 120
YWTPAVSRGDTTRIKKGTGTNYPQRAVWVDCDDSYDRKLL EALRPSYHNETSPGHIQA
Sbjct: 61 YWTPAVSRGDTTRIKKGTGTNYPQRAVWVDCDDSYDRKLL EALRPSYHNETSPGHIQA 120

Query: 121 IWLK EALPVSEFHRDGF IGHL TQAIGADKSGVDISQLL RVPGTHMHKREPFLGRVLS 180
IWLK EALPVSEFHRDGF IGHL TQAIGADKSGVDISQLL RVPGTHMHKREPFLGRVLS
Sbjct: 121 IWLK EALPVSEFHRDGF IGHL TQAIGADKSGVDISQLL RVPGTHMHKREPFLGRVLS 180

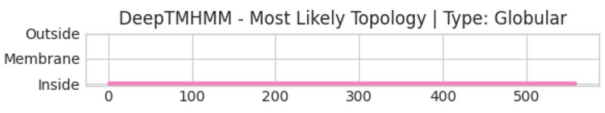
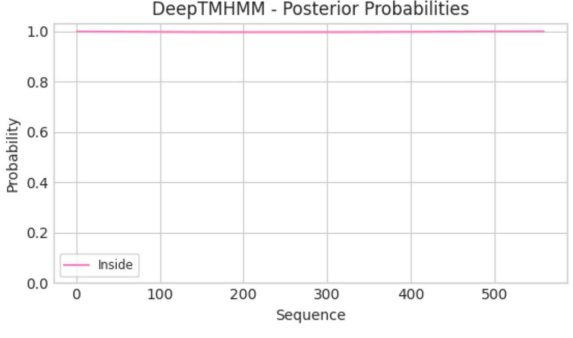
*** nr :
(q#: s#) : [alignment] ; e-value :
```

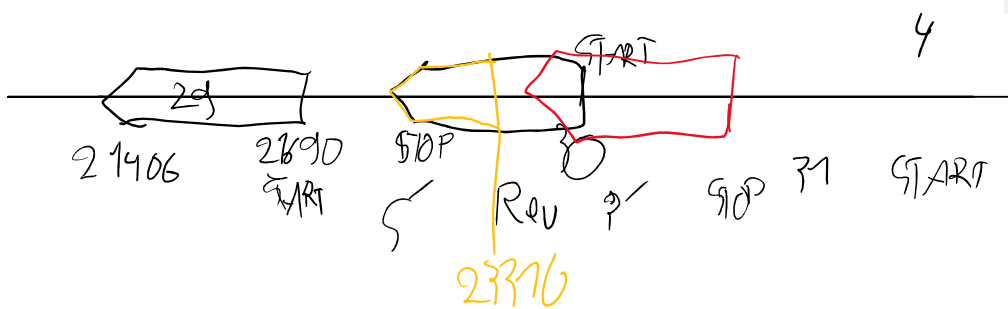
Distribution of the top 100 Blast Hits on 100 subject sequences



select all 100 sequences selected		GenPept	Graphics	Distance tree of results		Multiple alignment		MSA Viewer
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein SEFA_MUCCL_30 (Microbacterium ethaei Nucci)	Microbacterium ethaei Nucci	1139	1139	99%	0.0	100.00%	558	Q045042.1
<input checked="" type="checkbox"/> RecA-like DNA recombinase (Microbacterium ethaei Mandalorian)	Microbacterium ethaei Mandalorian	1138	1138	99%	0.0	99.82%	558	Q013623.1
<input checked="" type="checkbox"/> DNA topoisomerase (Microbacterium ethaei Quartz)	Microbacterium ethaei Quartz	1127	1127	99%	0.0	99.10%	558	U055020.1
<input checked="" type="checkbox"/> RecA-like DNA recombinase (Microbacterium ethaei Juarez)	Microbacterium ethaei Juarez	1039	1039	99%	0.0	94.44%	558	U044781.1
<input checked="" type="checkbox"/> RecA-like DNA recombinase (Microbacterium ethaei Yuxi)	Microbacterium ethaei Yuxi	933	933	99%	0.0	85.46%	559	Q073931.1
<input checked="" type="checkbox"/> DNA recombinase (Microbacterium ethaei WilliamShore)	Microbacterium ethaei WilliamShore	844	844	99%	0.0	73.98%	555	Q039899.1
<input checked="" type="checkbox"/> DNA recombinase (Microbacterium ethaei Sassa)	Microbacterium ethaei Sassa	813	813	99%	0.0	69.11%	559	A048604.1
<input checked="" type="checkbox"/> RecA-like DNA recombinase (Microbacterium ethaei Finlay)	Microbacterium ethaei Finlay	811	811	99%	0.0	69.29%	560	Q071526.1
<input checked="" type="checkbox"/> DNA recombinase (Microbacterium ethaei ColiCosta)	Microbacterium ethaei ColiCosta	810	810	99%	0.0	68.93%	560	A050286.1
<input checked="" type="checkbox"/> RecA-like DNA recombinase (Microbacterium ethaei Sankal)	Microbacterium ethaei Sankal	810	810	99%	0.0	69.93%	560	U020889.1
<input checked="" type="checkbox"/> DNA topoisomerase (Microbacterium ethaei ChikPa)	Microbacterium ethaei ChikPa	810	810	99%	0.0	69.93%	560	U030906.1
<input checked="" type="checkbox"/> DNA recombinase (Microbacterium ethaei Elae)	Microbacterium ethaei Elae	809	809	99%	0.0	68.93%	560	U020820.1

[illegible]

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p>Non</p>  
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui mais doute : RepA-like helicase, DNA primase/helicase</i>
DECISION:	<i>oui DNA primase/helicase car les 2 domaines, primase et helicase sont présents</i>



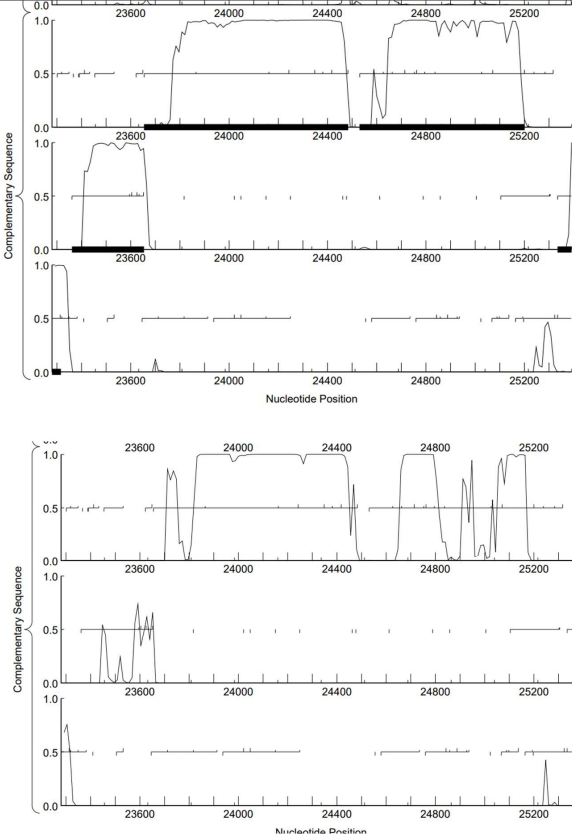
Student Gene Annotation Worksheet

Carosmatique-gene-31

Basic Phage Information	
Nom du Phage	carostasia
Gène #	31
Coordonnées du Stop	23361
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Chev 1
Coordonnées du Start retenu	23654
Fonction prédite	Nuclease

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both</i>
Y-a-t-il des éléments supportant un potentiel codant ?	

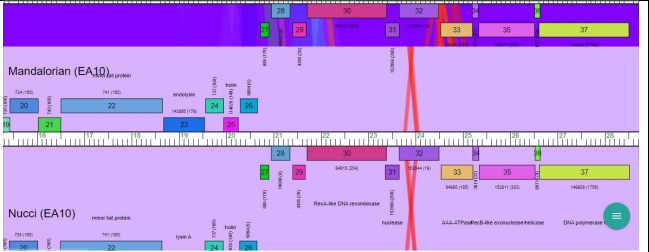
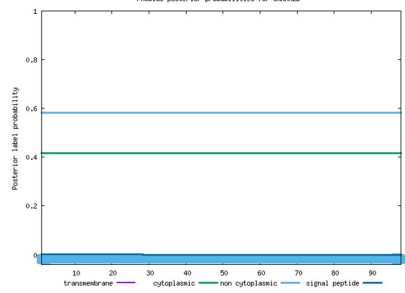
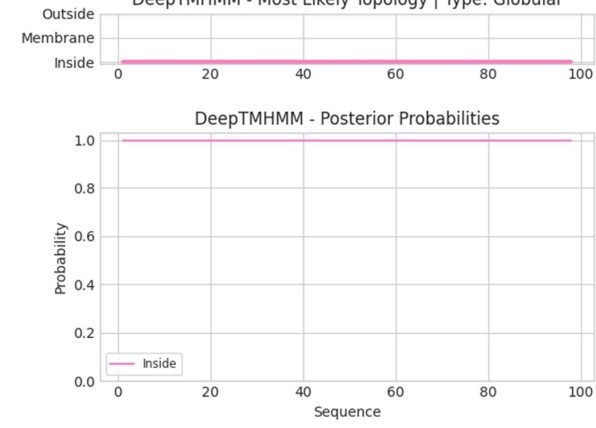

																																								
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_32, nuclease, 97</td><td>200</td><td>1e-51</td></tr><tr><td>Carostasia_Draft_31, function unknown, 97</td><td>200</td><td>1e-51</td></tr><tr><td>Mandalorian_31, nuclease, 97</td><td>199</td><td>2e-51</td></tr><tr><td>Juanyo_32, holliday junction resolvase, 97</td><td>189</td><td>2e-48</td></tr><tr><td>Nucci_31, holliday junction resolvase, 91</td><td>187</td><td>6e-48</td></tr><tr><td>YuuY_33, nuclease, 97</td><td>186</td><td>1e-47</td></tr><tr><td>Juicer_35, nuclease, 97</td><td>176</td><td>2e-44</td></tr><tr><td>Jemerald_35, nuclease, 97</td><td>176</td><td>2e-44</td></tr><tr><td>Zenitsu_33, VRR-Nuc domain protein, 95</td><td>174</td><td>5e-44</td></tr><tr><td>Shamu_Draft_34, function unknown, 95</td><td>174</td><td>5e-44</td></tr><tr><td>Saratos_33, nuclease, 95</td><td>174</td><td>5e-44</td></tr><tr><td>Sansa_32, nuclease, 95</td><td>174</td><td>5e-44</td></tr></table>		Score (bits)	E Value	Quartz_32, nuclease, 97	200	1e-51	Carostasia_Draft_31, function unknown, 97	200	1e-51	Mandalorian_31, nuclease, 97	199	2e-51	Juanyo_32, holliday junction resolvase, 97	189	2e-48	Nucci_31, holliday junction resolvase, 91	187	6e-48	YuuY_33, nuclease, 97	186	1e-47	Juicer_35, nuclease, 97	176	2e-44	Jemerald_35, nuclease, 97	176	2e-44	Zenitsu_33, VRR-Nuc domain protein, 95	174	5e-44	Shamu_Draft_34, function unknown, 95	174	5e-44	Saratos_33, nuclease, 95	174	5e-44	Sansa_32, nuclease, 95	174	5e-44
	Score (bits)	E Value																																						
Quartz_32, nuclease, 97	200	1e-51																																						
Carostasia_Draft_31, function unknown, 97	200	1e-51																																						
Mandalorian_31, nuclease, 97	199	2e-51																																						
Juanyo_32, holliday junction resolvase, 97	189	2e-48																																						
Nucci_31, holliday junction resolvase, 91	187	6e-48																																						
YuuY_33, nuclease, 97	186	1e-47																																						
Juicer_35, nuclease, 97	176	2e-44																																						
Jemerald_35, nuclease, 97	176	2e-44																																						
Zenitsu_33, VRR-Nuc domain protein, 95	174	5e-44																																						
Shamu_Draft_34, function unknown, 95	174	5e-44																																						
Saratos_33, nuclease, 95	174	5e-44																																						
Sansa_32, nuclease, 95	174	5e-44																																						
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Non</i>																																							
DECISION:	<i>Oui</i>																																							

Décision #2 : Quel est le start du gène ?

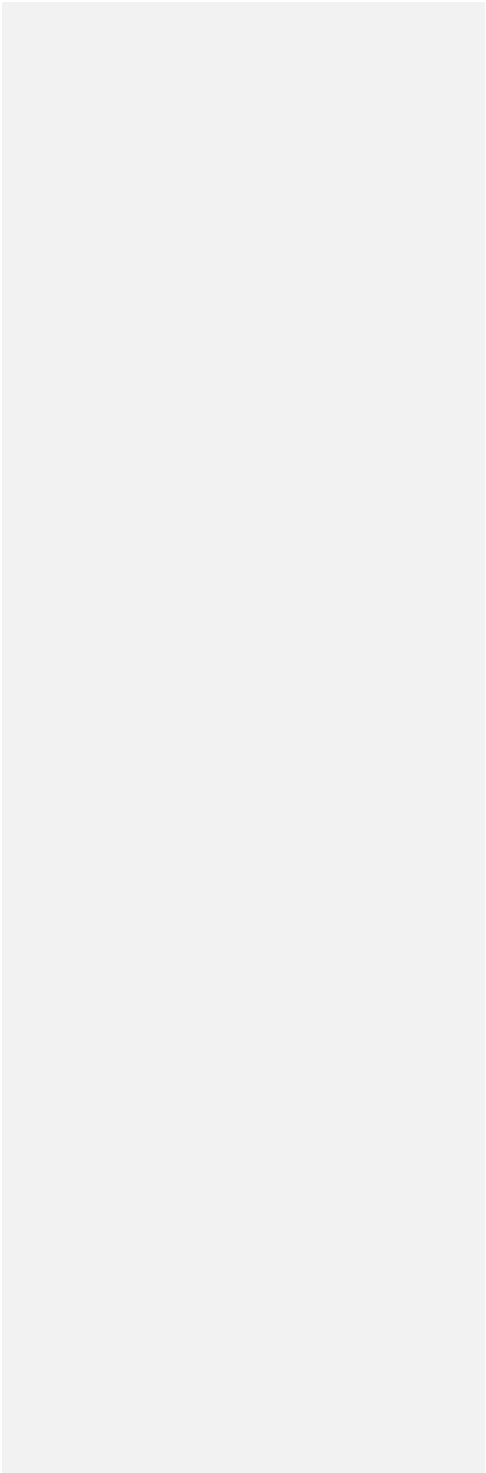
Collection des éléments de réponse	Rationnelle																																																																								
Quel start Glimmer et GeneMark suggèrent-ils ?	23654																																																																								
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Stats: 6 ORF Start: 23597 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix K8mer6 Explore Selected: 1 ORF Stop: 23361 5' End 66.7 66.7 50.0 18 ORF Length: 237 3' End 50.0 65.7 65.7 42 Spacing Weight Matrix Katin Medium Document</div> <table><thead><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>% Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr></thead><tbody><tr><td>1</td><td>-5,693</td><td>1,163</td><td>9</td><td>-6,407</td><td>GTGGACCGCTCAAGGCTGCTG</td><td>ATG</td><td>23654</td><td>294</td></tr><tr><td>2</td><td>-2,736</td><td>2,619</td><td>7</td><td>-4,259</td><td>GCTGATGGATGAAGCGGAGCTC</td><td>GTG</td><td>23636</td><td>276</td></tr><tr><td>3</td><td>-2,736</td><td>2,619</td><td>14</td><td>-4,532</td><td>TGAAGCGGAGGCTGTCGCGGCG</td><td>ATG</td><td>23627</td><td>267</td></tr><tr><td>4</td><td>-5,004</td><td>1,450</td><td>13</td><td>-6,050</td><td>CATGCTGAAGCGGCTCAACGGCG</td><td>ATG</td><td>23606</td><td>246</td></tr><tr><td>5</td><td>-5,872</td><td>1,061</td><td>12</td><td>-6,708</td><td>CGGGCTCAAGCGGATGCTCGGG</td><td>GTG</td><td>23597</td><td>237</td></tr><tr><td>6</td><td>-3,095</td><td>2,351</td><td>14</td><td>-4,441</td><td>AGTAGAGTGATATCAACCTTC</td><td>TTG</td><td>23369</td><td>9</td></tr></tbody></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	% Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-5,693	1,163	9	-6,407	GTGGACCGCTCAAGGCTGCTG	ATG	23654	294	2	-2,736	2,619	7	-4,259	GCTGATGGATGAAGCGGAGCTC	GTG	23636	276	3	-2,736	2,619	14	-4,532	TGAAGCGGAGGCTGTCGCGGCG	ATG	23627	267	4	-5,004	1,450	13	-6,050	CATGCTGAAGCGGCTCAACGGCG	ATG	23606	246	5	-5,872	1,061	12	-6,708	CGGGCTCAAGCGGATGCTCGGG	GTG	23597	237	6	-3,095	2,351	14	-4,441	AGTAGAGTGATATCAACCTTC	TTG	23369	9
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																	
#	Score	% Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																	
1	-5,693	1,163	9	-6,407	GTGGACCGCTCAAGGCTGCTG	ATG	23654	294																																																																	
2	-2,736	2,619	7	-4,259	GCTGATGGATGAAGCGGAGCTC	GTG	23636	276																																																																	
3	-2,736	2,619	14	-4,532	TGAAGCGGAGGCTGTCGCGGCG	ATG	23627	267																																																																	
4	-5,004	1,450	13	-6,050	CATGCTGAAGCGGCTCAACGGCG	ATG	23606	246																																																																	
5	-5,872	1,061	12	-6,708	CGGGCTCAAGCGGATGCTCGGG	GTG	23597	237																																																																	
6	-3,095	2,351	14	-4,441	AGTAGAGTGATATCAACCTTC	TTG	23369	9																																																																	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Score mauvais : -6.407 mais plus longue ORF avec un ATG																																																																								
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																									
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Score = 200 bits (508), Expect = 1e-51 Identities = 97/97 (100%), Positives = 97/97 (100%)</p> <p>Query: 1 MDEAEVVRMLKRLNAMPGVYALRTHGGSFQQKGTPDIIGCAHGHHFAIEAKKTAKEKPS 60 MDEAEVVRMLKRLNAMPGVYALRTHGGSFQQKGTPDIIGCAHGHHFAIEAKKTAKEKPS Sbjct: 1 MDEAEVVRMLKRLNAMPGVYALRTHGGSFQQKGTPDIIGCAHGHHFAIEAKKTAKEKPS 60</p>																																																																								
DECISION:	<p>Le start conservé est 23654 car il n'est pas en contraction avec les principes d'annotation et il est conservé chez d'autres homologues.</p> <p>De plus, il produit le plus long ORF avec un ATG</p>																																																																								

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																				
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <div><div>Sequences producing significant alignments:</div><div><table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_32, nuclease, 97</td><td>200</td><td>1e-51</td></tr><tr><td>Carostasia_Draft_31, function unknown, 97</td><td>200</td><td>1e-51</td></tr><tr><td>Mandalorian_31, nuclease, 97</td><td>199</td><td>2e-51</td></tr><tr><td>Juanyo_32, holliday junction resolvase, 97</td><td>189</td><td>2e-48</td></tr><tr><td>Nucci_31, holliday junction resolvase, 91</td><td>187</td><td>6e-48</td></tr><tr><td>Yuuy_33, nuclease, 97</td><td>186</td><td>1e-47</td></tr><tr><td>Juicer_35, nuclease, 97</td><td>176</td><td>2e-44</td></tr></table></div><p>*** nr : (q#: s#) : [alignment] ; e-value :</p><table><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per Ident</th><th>Ass Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/></td><td>nuclease [Microbacterium phage Quartz]</td><td>Microbacterium phage Quartz</td><td>205</td><td>205</td><td>98%</td><td>2e-66</td><td>100.00%</td><td>97</td><td>U005325.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>nuclease [Microbacterium phage Mandalorian]</td><td>Microbacterium phage Mandalorian</td><td>204</td><td>204</td><td>98%</td><td>4e-66</td><td>98.97%</td><td>97</td><td>GU013624.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>holliday junction resolvase [Microbacterium phage Juanyo]</td><td>Microbacterium phage Juanyo</td><td>196</td><td>196</td><td>98%</td><td>9e-63</td><td>92.78%</td><td>97</td><td>U0944782.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>nuclease [Microbacterium phage Yuuy]</td><td>Microbacterium phage Yuuy</td><td>193</td><td>193</td><td>98%</td><td>1e-61</td><td>91.75%</td><td>97</td><td>GP959362.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>holliday junction resolvase [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>191</td><td>191</td><td>92%</td><td>5e-61</td><td>98.90%</td><td>91</td><td>GU020242.1</td></tr></table></div>		Score (bits)	E Value	Quartz_32, nuclease, 97	200	1e-51	Carostasia_Draft_31, function unknown, 97	200	1e-51	Mandalorian_31, nuclease, 97	199	2e-51	Juanyo_32, holliday junction resolvase, 97	189	2e-48	Nucci_31, holliday junction resolvase, 91	187	6e-48	Yuuy_33, nuclease, 97	186	1e-47	Juicer_35, nuclease, 97	176	2e-44		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Ass Len	Accession	<input checked="" type="checkbox"/>	nuclease [Microbacterium phage Quartz]	Microbacterium phage Quartz	205	205	98%	2e-66	100.00%	97	U005325.1	<input checked="" type="checkbox"/>	nuclease [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	204	204	98%	4e-66	98.97%	97	GU013624.1	<input checked="" type="checkbox"/>	holliday junction resolvase [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	196	196	98%	9e-63	92.78%	97	U0944782.1	<input checked="" type="checkbox"/>	nuclease [Microbacterium phage Yuuy]	Microbacterium phage Yuuy	193	193	98%	1e-61	91.75%	97	GP959362.1	<input checked="" type="checkbox"/>	holliday junction resolvase [Microbacterium phage Nucci]	Microbacterium phage Nucci	191	191	92%	5e-61	98.90%	91	GU020242.1
	Score (bits)	E Value																																																																																			
Quartz_32, nuclease, 97	200	1e-51																																																																																			
Carostasia_Draft_31, function unknown, 97	200	1e-51																																																																																			
Mandalorian_31, nuclease, 97	199	2e-51																																																																																			
Juanyo_32, holliday junction resolvase, 97	189	2e-48																																																																																			
Nucci_31, holliday junction resolvase, 91	187	6e-48																																																																																			
Yuuy_33, nuclease, 97	186	1e-47																																																																																			
Juicer_35, nuclease, 97	176	2e-44																																																																																			
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Ass Len	Accession																																																																												
<input checked="" type="checkbox"/>	nuclease [Microbacterium phage Quartz]	Microbacterium phage Quartz	205	205	98%	2e-66	100.00%	97	U005325.1																																																																												
<input checked="" type="checkbox"/>	nuclease [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	204	204	98%	4e-66	98.97%	97	GU013624.1																																																																												
<input checked="" type="checkbox"/>	holliday junction resolvase [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	196	196	98%	9e-63	92.78%	97	U0944782.1																																																																												
<input checked="" type="checkbox"/>	nuclease [Microbacterium phage Yuuy]	Microbacterium phage Yuuy	193	193	98%	1e-61	91.75%	97	GP959362.1																																																																												
<input checked="" type="checkbox"/>	holliday junction resolvase [Microbacterium phage Nucci]	Microbacterium phage Nucci	191	191	92%	5e-61	98.90%	91	GU020242.1																																																																												
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un criblé HHPred avec une proba >= 90% et une couverture acceptable ?	<table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/></td><td>1</td><td>4QBO_A</td><td>Nuclease; nuclease, HYDROLASE; 1.3A (Streptococcus phage P9) SCOP: c.52.1.35, 1.1.1.1</td><td>99.85</td><td>2.8e-19</td><td>99.57</td><td>13.1</td><td>88</td><td>92</td></tr><tr><td><input type="checkbox"/></td><td>2</td><td>4QBN_B</td><td>Nuclease; Nuclease, HYDROLASE; HET: SO4; 1.85A (Salmonella phage SETP3) SCOP: c.52.1.35</td><td>99.85</td><td>3.3e-19</td><td>99.21</td><td>13.3</td><td>90</td><td>93</td></tr><tr><td><input type="checkbox"/></td><td>3</td><td>cd22365</td><td>VRR-NUC-like; Virus-type replication repair nuclease. This model characterizes a set of nucleases that resemble Holliday</td><td>99.85</td><td>4.2e-19</td><td>98.26</td><td>13.2</td><td>90</td><td>126</td></tr><tr><td><input type="checkbox"/></td><td>4</td><td>Q9T1Q4</td><td>VP44_BPAPS Putative nuclease p44 OS=Acyrtosiphon pisum secondary endosymbiont phage 1 OX=67571 GN=44 PE=3 SV=1</td><td>99.8</td><td>1.8e-17</td><td>91.97</td><td>12.7</td><td>88</td><td>93</td></tr></table> <div><div>1</div><div>98</div><div><div>4QBO_A</div><div>4QBN_B</div><div>cd22365</div><div>Q9T1Q4</div><div>4QBN_B</div><div>cd22365</div><div>VRR-NUC</div><div>VRR-NUC</div></div></div>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/>	1	4QBO_A	Nuclease; nuclease, HYDROLASE; 1.3A (Streptococcus phage P9) SCOP: c.52.1.35, 1.1.1.1	99.85	2.8e-19	99.57	13.1	88	92	<input type="checkbox"/>	2	4QBN_B	Nuclease; Nuclease, HYDROLASE; HET: SO4; 1.85A (Salmonella phage SETP3) SCOP: c.52.1.35	99.85	3.3e-19	99.21	13.3	90	93	<input type="checkbox"/>	3	cd22365	VRR-NUC-like; Virus-type replication repair nuclease. This model characterizes a set of nucleases that resemble Holliday	99.85	4.2e-19	98.26	13.2	90	126	<input type="checkbox"/>	4	Q9T1Q4	VP44_BPAPS Putative nuclease p44 OS=Acyrtosiphon pisum secondary endosymbiont phage 1 OX=67571 GN=44 PE=3 SV=1	99.8	1.8e-17	91.97	12.7	88	93																																			
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																																													
<input type="checkbox"/>	1	4QBO_A	Nuclease; nuclease, HYDROLASE; 1.3A (Streptococcus phage P9) SCOP: c.52.1.35, 1.1.1.1	99.85	2.8e-19	99.57	13.1	88	92																																																																												
<input type="checkbox"/>	2	4QBN_B	Nuclease; Nuclease, HYDROLASE; HET: SO4; 1.85A (Salmonella phage SETP3) SCOP: c.52.1.35	99.85	3.3e-19	99.21	13.3	90	93																																																																												
<input type="checkbox"/>	3	cd22365	VRR-NUC-like; Virus-type replication repair nuclease. This model characterizes a set of nucleases that resemble Holliday	99.85	4.2e-19	98.26	13.2	90	126																																																																												
<input type="checkbox"/>	4	Q9T1Q4	VP44_BPAPS Putative nuclease p44 OS=Acyrtosiphon pisum secondary endosymbiont phage 1 OX=67571 GN=44 PE=3 SV=1	99.8	1.8e-17	91.97	12.7	88	93																																																																												
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Oui, la comparaison est faite avec le Phage mandalorian.																																																																																				

	 <p>En haut Carostasia, en bas mandalorian. Les gènes adjacents correspondent à des protéines de métabolisme de l'ADN.</p>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p>NON</p> <p>ID UNINUED FT TOPO_DOM 1 98 NON CYTOPLASMIC. //</p>  <p>The probability data used in the plot is found here and the gseplot script is here.</p> <p>DeepTMHMM - Most Likely Topology Type: Globular</p>  <p>DeepTMHMM - Posterior Probabilities</p> 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<p>OUI</p>

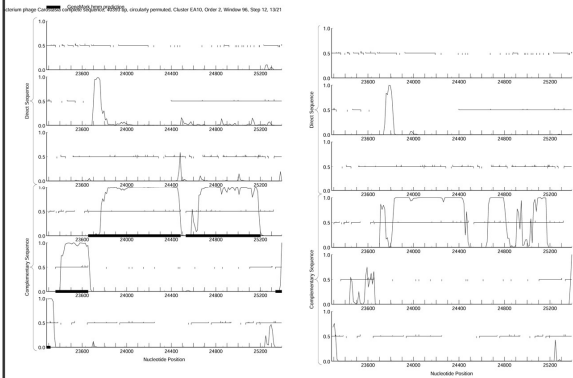
DECISION:	<i>Nuclease</i>
------------------	-----------------

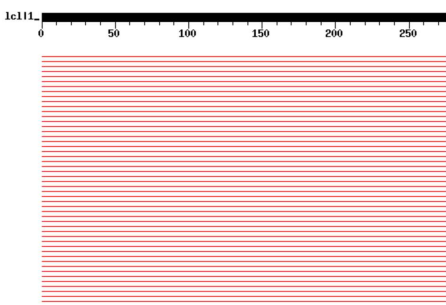


Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	32
Coordonnées du Stop	23654
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	gap 45
Coordonnées du Start retenu	24484
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes both
Y-a-t-il des éléments supportant un potentiel codant ?	 <p>The figure displays genomic tracks for Carostasia phage. The top track shows the Direct Sequence coverage, and the bottom track shows the Complementary Sequence coverage. The x-axis represents the Nucleotide Position, ranging from 23800 to 25200. The y-axis represents the coverage, ranging from 0.0 to 1.0. The tracks show a high level of coverage across the entire region, indicating a strong potential for a coding sequence. The predicted protein structure is shown in the bottom track, with a high level of confidence in the predicted structure.</p>

	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <p>161596 Glamour_33, function unknown, 271..S= 321 E=1e-87</p> <p>Color Key for Alignment Scores</p> <p><40 40-50 50-60 60-200 >200</p>  <p>Sequences producing significant alignments:</p> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_32, function unknown, 276</td><td>549</td><td>e-156</td></tr><tr><td>Mandalorian_32, function unknown, 280</td><td>537</td><td>e-152</td></tr><tr><td>Nucci_32, function unknown, 282</td><td>536</td><td>e-152</td></tr><tr><td>Quartz_33, function unknown, 276</td><td>524</td><td>e-149</td></tr><tr><td>YuuY_34, function unknown, 283</td><td>491</td><td>e-138</td></tr><tr><td>Juanyo_33, function unknown, 285</td><td>484</td><td>e-137</td></tr><tr><td>Zayuliv_33, function unknown, 283</td><td>348</td><td>1e-95</td></tr><tr><td>QuadZero_33, function unknown, 286</td><td>347</td><td>2e-95</td></tr></table>		Score (bits)	E Value	Carostasia_Draft_32, function unknown, 276	549	e-156	Mandalorian_32, function unknown, 280	537	e-152	Nucci_32, function unknown, 282	536	e-152	Quartz_33, function unknown, 276	524	e-149	YuuY_34, function unknown, 283	491	e-138	Juanyo_33, function unknown, 285	484	e-137	Zayuliv_33, function unknown, 283	348	1e-95	QuadZero_33, function unknown, 286	347	2e-95
	Score (bits)	E Value																										
Carostasia_Draft_32, function unknown, 276	549	e-156																										
Mandalorian_32, function unknown, 280	537	e-152																										
Nucci_32, function unknown, 282	536	e-152																										
Quartz_33, function unknown, 276	524	e-149																										
YuuY_34, function unknown, 283	491	e-138																										
Juanyo_33, function unknown, 285	484	e-137																										
Zayuliv_33, function unknown, 283	348	1e-95																										
QuadZero_33, function unknown, 286	347	2e-95																										
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>Longueur OK, gap de 53 avec le gène d'après, même sens que les gènes à côté.</p>																											
DECISION:	Oui																											

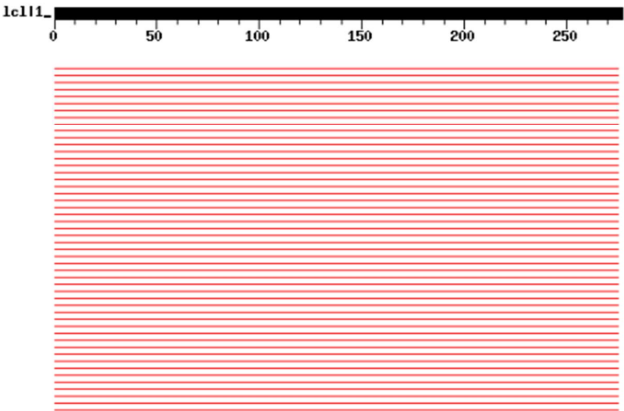
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																	
Quel start Glimmer et GeneMark suggèrent-ils ?	24484 pour les 2																																																																																	
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,951</td><td>2,887</td><td>10</td><td>-2,645</td><td>CAACAGAAAGAGGAGGCCAAC</td><td>ATG</td><td>24484</td><td>831</td></tr><tr><td>2</td><td>-4,013</td><td>1,921</td><td>8</td><td>-5,235</td><td>CTCCGGCATCAACAGAGCGT</td><td>ATG</td><td>24418</td><td>765</td></tr><tr><td>3</td><td>-4,416</td><td>1,733</td><td>11</td><td>-5,173</td><td>GCAAGCCTTCAAGATCGTGGC</td><td>GTG</td><td>24382</td><td>729</td></tr><tr><td>4</td><td>-2,915</td><td>2,435</td><td>13</td><td>-3,961</td><td>GGAGGCCAGGAGGGACCGAC</td><td>ATG</td><td>24349</td><td>696</td></tr><tr><td>5</td><td>-2,482</td><td>2,638</td><td>17</td><td>-4,482</td><td>GTTCGGAAGCTCCGTGACCTC</td><td>ATG</td><td>24244</td><td>591</td></tr><tr><td>6</td><td>-4,875</td><td>1,518</td><td>13</td><td>-5,921</td><td>CCCCAACAGGCCATCGGCAAG</td><td>GTG</td><td>24163</td><td>510</td></tr><tr><td>7</td><td>-2,757</td><td>2,509</td><td>10</td><td>-3,452</td><td>CGGCTCAAGAGGACGAGCTC</td><td>GTG</td><td>23866</td><td>213</td></tr></table> <p>Meilleur score RBS -2.645 correspond au start prédit par les 2 et est un ATG</p>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,951	2,887	10	-2,645	CAACAGAAAGAGGAGGCCAAC	ATG	24484	831	2	-4,013	1,921	8	-5,235	CTCCGGCATCAACAGAGCGT	ATG	24418	765	3	-4,416	1,733	11	-5,173	GCAAGCCTTCAAGATCGTGGC	GTG	24382	729	4	-2,915	2,435	13	-3,961	GGAGGCCAGGAGGGACCGAC	ATG	24349	696	5	-2,482	2,638	17	-4,482	GTTCGGAAGCTCCGTGACCTC	ATG	24244	591	6	-4,875	1,518	13	-5,921	CCCCAACAGGCCATCGGCAAG	GTG	24163	510	7	-2,757	2,509	10	-3,452	CGGCTCAAGAGGACGAGCTC	GTG	23866	213
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																										
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																										
1	-1,951	2,887	10	-2,645	CAACAGAAAGAGGAGGCCAAC	ATG	24484	831																																																																										
2	-4,013	1,921	8	-5,235	CTCCGGCATCAACAGAGCGT	ATG	24418	765																																																																										
3	-4,416	1,733	11	-5,173	GCAAGCCTTCAAGATCGTGGC	GTG	24382	729																																																																										
4	-2,915	2,435	13	-3,961	GGAGGCCAGGAGGGACCGAC	ATG	24349	696																																																																										
5	-2,482	2,638	17	-4,482	GTTCGGAAGCTCCGTGACCTC	ATG	24244	591																																																																										
6	-4,875	1,518	13	-5,921	CCCCAACAGGCCATCGGCAAG	GTG	24163	510																																																																										
7	-2,757	2,509	10	-3,452	CGGCTCAAGAGGACGAGCTC	GTG	23866	213																																																																										
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>L'orf la plus longue correspond avec le start prédit et est de 831pb, on a un gap de 53.</p>																																																																																	
Est-ce que le start est conservé chez les homologues voir Starterator ?	Gene: Carostasia_32 Start: 24484, Stop: 23654, Start Num: 1 Candidate Starts for Carostasia_32: (Start: 1 @24484 has 16 MA's), (2, 24418), (4, 24382), (6, 24349), (9, 24244), (12, 24163), (14, 23866),																																																																																	

Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show define and scores. Click to show alignments

Color Key for Alignment Scores

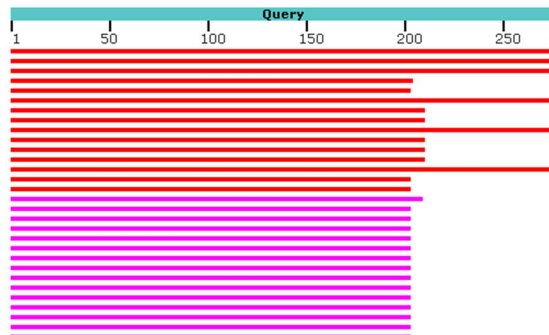


Sequences producing significant alignments:		Score	E
		(bits)	Value
Carostasia_Draft_32, function unknown, 276		549	e-156
Mandalorian_32, function unknown, 280		537	e-152
Nucci_32, function unknown, 282		536	e-152
Quartz_33, function unknown, 276		524	e-149
YuuY_34, function unknown, 283		491	e-138
Juanyo_33, function unknown, 285		484	e-137
Zayuliv_33, function unknown, 283		348	1e-95
QuadZero_33, function unknown, 286		347	2e-95
Hasitha_33, function unknown, 286		347	2e-95
Fulton_Draft_34, function unknown, 286		347	2e-95
Librie_33, function unknown, 287		346	4e-95
CaptainRex_33, function unknown, 287		346	4e-95
LilTerminator_Draft_34, function unknown, 287		345	5e-95
GreenIvy_Draft_34, function unknown, 287		345	7e-95
Zepp_34, function unknown, 280		343	3e-94
>Mandalorian_32, function unknown, 280			
Length = 280			
Score = 537 bits (1383), Expect = e-152			
Identities = 275/280 (98%), Positives = 275/280 (98%), Gaps = 4/280 (1%)			
Query: 1	MAKKIRLDFSKVEERSGINTKRMPGLHAFKIVGVEDKEAQDGDHYVYFYL VPEVEKYKT	60	
	MAKKIRLDFSKVEERSGINTKRMPGLHAFKIVGVEDKEAQDGDHYVYFYL VPEVEKYKT		
Sbjct: 1	MAKKIRLDFSKVEERSGINTKRMPGLHAFKIVGVEDKEAQDGDHYVYFYL VPEVEKYKT	60	
Query: 61	RRFPYYCKLQONQFWKL RDL MVAAGVL DASKAGKAVNIDPNKAIGKVVAGEVEDE TGQY	120	
	RRFPYYCKLQONQFWKL RDL MVAAGVL DASKAGKAVNIDPNKAIGKVVAGEVEDE TGQY		
Sbjct: 61	RRFPYYCKLQONQFWKL RDL MVAAGVL DASKAGKAVNIDPNKAIGKVVAGEVEDE TGQY	120	
Query: 121	EGRSTLANIYGLDILDE DSTGVEDDDEEYDDEADE AEE EYEDDVDEEEADEDELREEIE	180	
	EGRSTLANIYGLDILDE DSTGVEDDDEEYDDEADE AEE EYEDDVDEEEADEDELREEIE		
Sbjct: 121	EGRSTLANIYGLDILDE DSTGVEDDDEEYDDEADE AEE EYEDDVDEEEADEDELREEIE	180	

*** nr :
(q#: s#) : [alignment] ; e-value :

No putative conserved domains have been detected

Distribution of the top 100 Blast Hits on 100 subject sequences



	Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hypothetical protein SEA_MANDALORIAN_32 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	521	521	99%	0.0	99.21%	280	QO013625.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_NUC1_32 [Microbacterium phage Nuc1]	Microbacterium phage Nuc1	520	520	99%	0.0	97.52%	282	QO013625.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_CHEST1_36 [Microbacterium phage Chest1]	Microbacterium phage Chest1	609	609	99%	0.0	98.77%	276	U050262.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_JAN10_33 [Microbacterium phage Jan10]	Microbacterium phage Jan10	317	317	73%	6e-106	90.26%	286	U0504393.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_T037_34 [Microbacterium phage T037]	Microbacterium phage T037	301	301	73%	2e-98	83.61%	283	GF295362.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_CHEST1_36 [Microbacterium phage Chest1]	Microbacterium phage Chest1	218	218	99%	4e-68	95.04%	273	U0504002.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_SANT1_33 [Microbacterium phage Sant1]	Microbacterium phage Sant1	213	213	75%	2e-64	96.19%	272	A0508908.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_GLAMOUR_33 [Microbacterium phage Glamour]	Microbacterium phage Glamour	210	210	75%	3e-63	98.26%	271	QO003692.1

hypothetical protein SEA_MANDALORIAN_32 [Microbacterium phage Mandalorian]

Sequence ID: [QO013625.1](#) Length: 280 Number of Matches: 1

Range 1: 1 to 280 [GenPept](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
521 bits(1342)	0.0	Compositional matrix adjust.	275/280(98%)	275/280(98%)	4/280(1%)
Query 1	MAKKIRLDFSKVEERSGINTKRMEGLHAFKIVGVEDKEAQDGTDMYVFYLPVEVEKYKT	60			
Sbjct 1	MAKKIRLDFSKVEERSGINTKRMEGLHAFKIVGVEDKEAQDGTDMYVFYLPVEVEKYKT	60			
Query 61	RRFPYYCKLQQNQFWKLRDLWVAAGVLDASKAKGKAVNIDPNKATGKVVAGEVEDETGQY	120			
Sbjct 61	RRFPYYCKLQQNQFWKLRDLWVAAGVLDASKAKGKAVNIDPNKATGKVVAGEVEDETGQY	120			
Query 121	EGRSTLANIYGLDILDEDSGVEDDDEEYDDEADEAEVEEYDDWDEEEEADEDLREEIE	180			
Sbjct 121	EGRSTLANIYGLDILDEDSGVEDDDEEYDDEADEAEVEEYDDWDEEEEADEDLREEIE	180			
Query 181	ALTLAALKRRAKGLIGIDTGWKDELVLEVEEELGEGDDEG--DEDDLDEDLDOEELE	238			
Sbjct 181	ALTLAALKRRAKGLIGIDTGWKDELVLEVEEELGEGDDEGDEDDLDEDLDOEELE	240			
Query 239	DEEFEDDEEEEE--EPAPRRRAARKPAAPARARTYKRR	276			
Sbjct 241	DEEFEDDEEEEEPEPAPRRRAARKPAAPARARTYKRR	280			

Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.

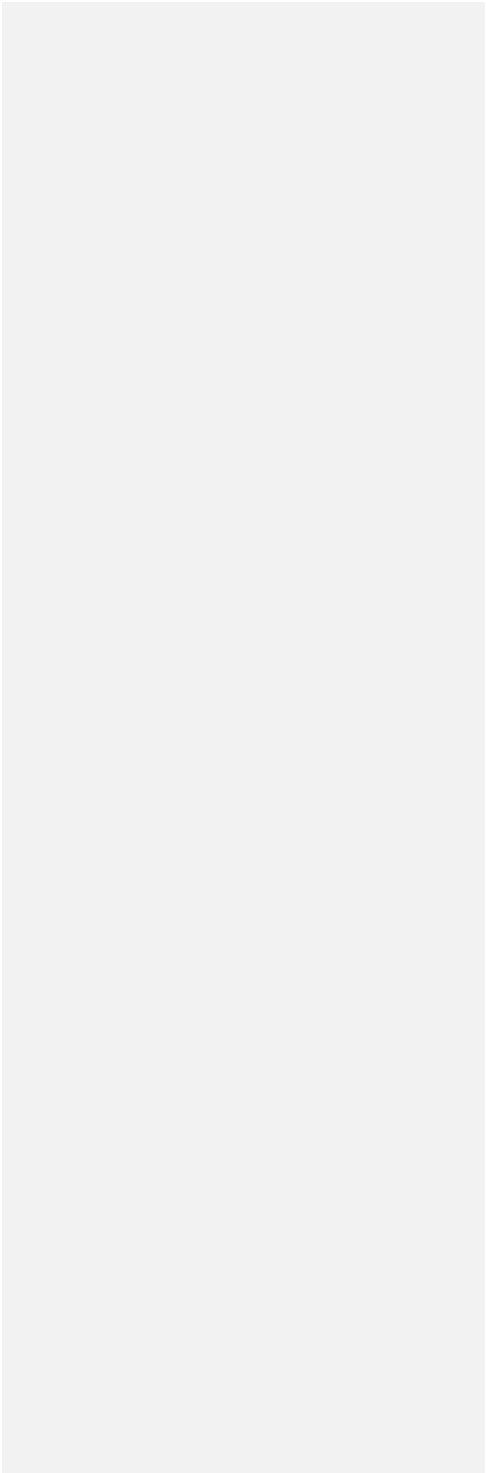
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Number of Hits: 122
Query MSA diversity (Neff): 6.10791
Detected sequence features: •Coiled coil segment(s)

Visualization



DECISION:	<i>NFK</i>
------------------	------------



Student Gene Annotation Worksheet

Carosmatique-gene-33

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	33
Coordonnées du Stop	24530
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Chev 4
Coordonnées du Start retenu	25201
Fonction prédite	AAA-ATPase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>e-Value</th></tr><tr><td>Quartz_34, AAA-ATPase, 223</td><td>439</td><td>e-123</td></tr><tr><td>Carostasia_Draft_33, function unknown, 223</td><td>439</td><td>e-123</td></tr><tr><td>Nucci_33, AAA-ATPase, 223</td><td>437</td><td>e-123</td></tr><tr><td>Mandalorian_33, AAA-ATPase, 223</td><td>437</td><td>e-123</td></tr><tr><td>Juanyo_34, AAA-ATPase, 223</td><td>421</td><td>e-118</td></tr><tr><td>YuuY_35, AAA-ATPase, 223</td><td>419</td><td>e-117</td></tr><tr><td>Morrigan_36, AAA-ATPase, 223</td><td>416</td><td>e-116</td></tr><tr><td>...</td><td>...</td><td>...</td></tr></table>		Score (bits)	e-Value	Quartz_34, AAA-ATPase, 223	439	e-123	Carostasia_Draft_33, function unknown, 223	439	e-123	Nucci_33, AAA-ATPase, 223	437	e-123	Mandalorian_33, AAA-ATPase, 223	437	e-123	Juanyo_34, AAA-ATPase, 223	421	e-118	YuuY_35, AAA-ATPase, 223	419	e-117	Morrigan_36, AAA-ATPase, 223	416	e-116
	Score (bits)	e-Value																										
Quartz_34, AAA-ATPase, 223	439	e-123																										
Carostasia_Draft_33, function unknown, 223	439	e-123																										
Nucci_33, AAA-ATPase, 223	437	e-123																										
Mandalorian_33, AAA-ATPase, 223	437	e-123																										
Juanyo_34, AAA-ATPase, 223	421	e-118																										
YuuY_35, AAA-ATPase, 223	419	e-117																										
Morrigan_36, AAA-ATPase, 223	416	e-116																										
...																										
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non																											
DECISION:	Oui																											

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																											
Quel start Glimmer et GeneMark suggèrent-ils ?	25201																																																																																																																																																																											
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Starts: 17 Selected: 1</div> <div>ORF Start : 25201 ORF Stop : 24530 ORF Length: 672</div> <div>Cdn1 Cdn2 Cdn3 Length 5' End 45.5 27.3 63.6 33 3' End 41.4 48.3 93.1 87</div> <div>SD Scoring Matrix Kibler6</div> <div>Explore</div> <div>Spacing Weight Matrix Karlin Medium</div> <div>Document</div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,983</td><td>1,467</td><td>14</td><td>-6,330</td><td>AAATGAGAGCACTCATCTTGA</td><td>ATG</td><td>25318</td><td>789</td></tr><tr><td>2</td><td>-5,348</td><td>1,297</td><td>17</td><td>-7,348</td><td>GTGTTGACGCCCCCGGCAGGC</td><td>GTG</td><td>25285</td><td>756</td></tr><tr><td>3</td><td>-4,416</td><td>1,733</td><td>7</td><td>-5,939</td><td>AGGTGGGCCCCGTCAGCATTCG</td><td>GTG</td><td>25240</td><td>711</td></tr><tr><td>4</td><td>-2,647</td><td>2,561</td><td>15</td><td>-4,249</td><td>GCGAAGGGAGGCGAGGCAGC</td><td>ATG</td><td>25201</td><td>672</td></tr><tr><td>5</td><td>-4,063</td><td>1,898</td><td>18</td><td>-6,364</td><td>GCGTGGTATCAACCCGAGGCG</td><td>ATG</td><td>25072</td><td>543</td></tr><tr><td>6</td><td>-4,819</td><td>1,544</td><td>5</td><td>-6,819</td><td>CACGGCAATATCAACAGAGCG</td><td>GTG</td><td>25027</td><td>498</td></tr><tr><td>7</td><td>-5,951</td><td>1,014</td><td>5</td><td>-7,951</td><td>GCTGGCCCACTCATCCGACC</td><td>TTG</td><td>24865</td><td>336</td></tr><tr><td>8</td><td>-2,377</td><td>2,687</td><td>10</td><td>-3,071</td><td>TGCAGCAACAGGATGTTATC</td><td>GTG</td><td>24838</td><td>305</td></tr><tr><td>9</td><td>-2,377</td><td>2,687</td><td>13</td><td>-3,422</td><td>CAGCAACAGGATGTTATCGTG</td><td>TTG</td><td>24835</td><td>306</td></tr><tr><td>10</td><td>-3,079</td><td>2,358</td><td>7</td><td>-4,602</td><td>CATCTCCCGAGGAGGACTGG</td><td>GTG</td><td>24796</td><td>247</td></tr><tr><td>11</td><td>-3,079</td><td>2,358</td><td>13</td><td>-4,125</td><td>CGAGCAGGAGGAGGAGGCTC</td><td>ATG</td><td>24763</td><td>234</td></tr><tr><td>12</td><td>-4,650</td><td>1,623</td><td>13</td><td>-5,696</td><td>GGAGCGAGGGTCAAGCAGCG</td><td>GTG</td><td>24754</td><td>225</td></tr><tr><td>13</td><td>-6,171</td><td>0,911</td><td>14</td><td>-7,518</td><td>TGCAGCCAGGCGGCTCTGAG</td><td>ATG</td><td>24715</td><td>186</td></tr><tr><td>14</td><td>-5,150</td><td>1,389</td><td>10</td><td>-5,844</td><td>CGGCCAGCTCAAGCAGAGCG</td><td>GTG</td><td>24664</td><td>135</td></tr><tr><td>15</td><td>-6,213</td><td>0,892</td><td>8</td><td>-7,435</td><td>GCTGGGCGGCTCAGCTCGATT</td><td>GTG</td><td>24628</td><td>99</td></tr><tr><td>16</td><td>-3,410</td><td>2,204</td><td>15</td><td>-5,012</td><td>CTATACTGTACTCGGCTTAC</td><td>TTG</td><td>24583</td><td>54</td></tr><tr><td>17</td><td>-7,321</td><td>0,373</td><td>11</td><td>-8,077</td><td>GAAGCAGGCCACCATCGGCGCG</td><td>TTG</td><td>24559</td><td>30</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,983	1,467	14	-6,330	AAATGAGAGCACTCATCTTGA	ATG	25318	789	2	-5,348	1,297	17	-7,348	GTGTTGACGCCCCCGGCAGGC	GTG	25285	756	3	-4,416	1,733	7	-5,939	AGGTGGGCCCCGTCAGCATTCG	GTG	25240	711	4	-2,647	2,561	15	-4,249	GCGAAGGGAGGCGAGGCAGC	ATG	25201	672	5	-4,063	1,898	18	-6,364	GCGTGGTATCAACCCGAGGCG	ATG	25072	543	6	-4,819	1,544	5	-6,819	CACGGCAATATCAACAGAGCG	GTG	25027	498	7	-5,951	1,014	5	-7,951	GCTGGCCCACTCATCCGACC	TTG	24865	336	8	-2,377	2,687	10	-3,071	TGCAGCAACAGGATGTTATC	GTG	24838	305	9	-2,377	2,687	13	-3,422	CAGCAACAGGATGTTATCGTG	TTG	24835	306	10	-3,079	2,358	7	-4,602	CATCTCCCGAGGAGGACTGG	GTG	24796	247	11	-3,079	2,358	13	-4,125	CGAGCAGGAGGAGGAGGCTC	ATG	24763	234	12	-4,650	1,623	13	-5,696	GGAGCGAGGGTCAAGCAGCG	GTG	24754	225	13	-6,171	0,911	14	-7,518	TGCAGCCAGGCGGCTCTGAG	ATG	24715	186	14	-5,150	1,389	10	-5,844	CGGCCAGCTCAAGCAGAGCG	GTG	24664	135	15	-6,213	0,892	8	-7,435	GCTGGGCGGCTCAGCTCGATT	GTG	24628	99	16	-3,410	2,204	15	-5,012	CTATACTGTACTCGGCTTAC	TTG	24583	54	17	-7,321	0,373	11	-8,077	GAAGCAGGCCACCATCGGCGCG	TTG	24559	30
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																				
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																				
1	-4,983	1,467	14	-6,330	AAATGAGAGCACTCATCTTGA	ATG	25318	789																																																																																																																																																																				
2	-5,348	1,297	17	-7,348	GTGTTGACGCCCCCGGCAGGC	GTG	25285	756																																																																																																																																																																				
3	-4,416	1,733	7	-5,939	AGGTGGGCCCCGTCAGCATTCG	GTG	25240	711																																																																																																																																																																				
4	-2,647	2,561	15	-4,249	GCGAAGGGAGGCGAGGCAGC	ATG	25201	672																																																																																																																																																																				
5	-4,063	1,898	18	-6,364	GCGTGGTATCAACCCGAGGCG	ATG	25072	543																																																																																																																																																																				
6	-4,819	1,544	5	-6,819	CACGGCAATATCAACAGAGCG	GTG	25027	498																																																																																																																																																																				
7	-5,951	1,014	5	-7,951	GCTGGCCCACTCATCCGACC	TTG	24865	336																																																																																																																																																																				
8	-2,377	2,687	10	-3,071	TGCAGCAACAGGATGTTATC	GTG	24838	305																																																																																																																																																																				
9	-2,377	2,687	13	-3,422	CAGCAACAGGATGTTATCGTG	TTG	24835	306																																																																																																																																																																				
10	-3,079	2,358	7	-4,602	CATCTCCCGAGGAGGACTGG	GTG	24796	247																																																																																																																																																																				
11	-3,079	2,358	13	-4,125	CGAGCAGGAGGAGGAGGCTC	ATG	24763	234																																																																																																																																																																				
12	-4,650	1,623	13	-5,696	GGAGCGAGGGTCAAGCAGCG	GTG	24754	225																																																																																																																																																																				
13	-6,171	0,911	14	-7,518	TGCAGCCAGGCGGCTCTGAG	ATG	24715	186																																																																																																																																																																				
14	-5,150	1,389	10	-5,844	CGGCCAGCTCAAGCAGAGCG	GTG	24664	135																																																																																																																																																																				
15	-6,213	0,892	8	-7,435	GCTGGGCGGCTCAGCTCGATT	GTG	24628	99																																																																																																																																																																				
16	-3,410	2,204	15	-5,012	CTATACTGTACTCGGCTTAC	TTG	24583	54																																																																																																																																																																				
17	-7,321	0,373	11	-8,077	GAAGCAGGCCACCATCGGCGCG	TTG	24559	30																																																																																																																																																																				
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>start prédit : -4.249</p> <p>Le start ne correspond pas au plus long ORF mais les plus longs conduisent à un chevauchement excessif : Pour le start 25240 chev de 43, 25286 : chev89, 25318 : chev121</p>																																																																																																																																																																											
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																																																																																																												
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Score = 439 bits (1128), Expect = e-123 Identities = 223/223 (100%), Positives = 223/223 (100%)</p> <p>Query: 1 MSVITIIYGRPKVGKTTLLAKDAPKGTAVFSADRLRGINTQGMILEDLSTRNINKTVN 60 MSVITIIYGRPKVGKTTLLAKDAPKGTAVFSADRLRGINTQGMILEDLSTRNINKTVN Sbjct: 1 MSVITIIYGRPKVGKTTLLAKDAPKGTAVFSADRLRGINTQGMILEDLSTRNINKTVN 60</p>																																																																																																																																																																											

DECISION:	Le start conservé est 225201 car il n'est pas en contraction avec les principes d'annotation et Il est conservé chez d'autres homologues.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Rationnelle

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :

(q#: s#) : [alignment] ; e-value :

Sequences producing significant alignments:	Score (bits)	E Value
Quartz_34, AAA-ATPase, 223	439	e-123
Carostasia_Draft_33, function unknown, 223	439	e-123
Nucci_33, AAA-ATPase, 223	437	e-123
Mandalorian_33, AAA-ATPase, 223	437	e-123
Juanyo_34, AAA-ATPase, 223	421	e-118
YuuY_35, AAA-ATPase, 223	419	e-117

*** nr :

(q#: s#) : [alignment] ; e-value :

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
AAA-ATPase [Microbacterium phage Quartz]	Microbacterium phage Quartz	454	454	99%	4e-161	100.00%	223	U0559253.1
AAA-ATPase [Microbacterium phage Nucci]	Microbacterium phage Nucci	454	454	99%	1e-160	99.99%	223	Q0902426.1
AAA-ATPase [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	407	407	99%	2e-142	95.07%	223	U5164784.1
AAA-ATPase [Microbacterium phage Mercedes]	Microbacterium phage Mercedes	407	407	99%	3e-142	88.34%	223	Q530331.1

Collection des éléments de réponse

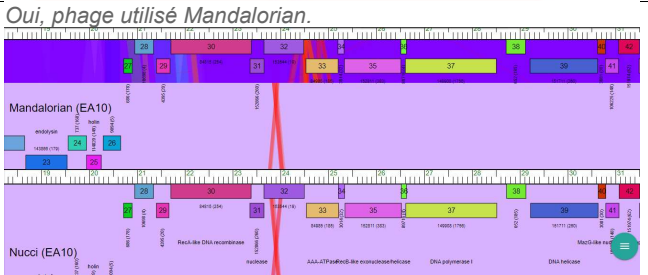
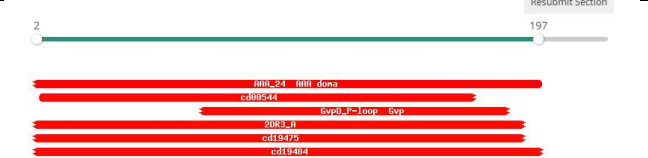
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Rationnelle

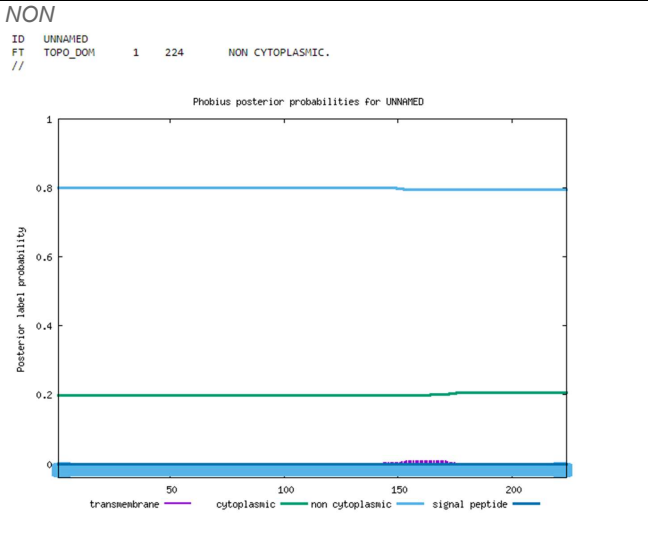
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	1	PF13479.10 ; AAA_24; AAA domain	99.71	8.4e-15	110.17	19.5	177	194
<input type="checkbox"/>	2	cd09044 CobU; Adenosylcobinamide kinase / adenosylcobinamide phosphate guanylyltransferase (CobU).	99.16	7.2e-10	82.03	9.5	144	176
<input type="checkbox"/>	3	PF07088.15 ; GvpD_P-loop ; GvpD gas vesicle protein, P-loop domain	99.16	9.9e-10	82.68	10.3	81	189
<input type="checkbox"/>	4	2DR3_A UPF0273 protein PH0284; RecA superfamily ATPase, Hexamer: Structural Genomics, NPPSFA, National Project on Protein Struc	99.11	4.3e-9	83.13	12.2	143	247

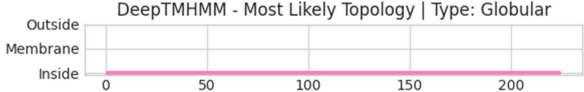
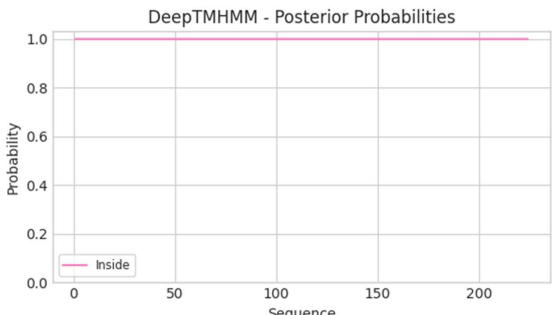
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

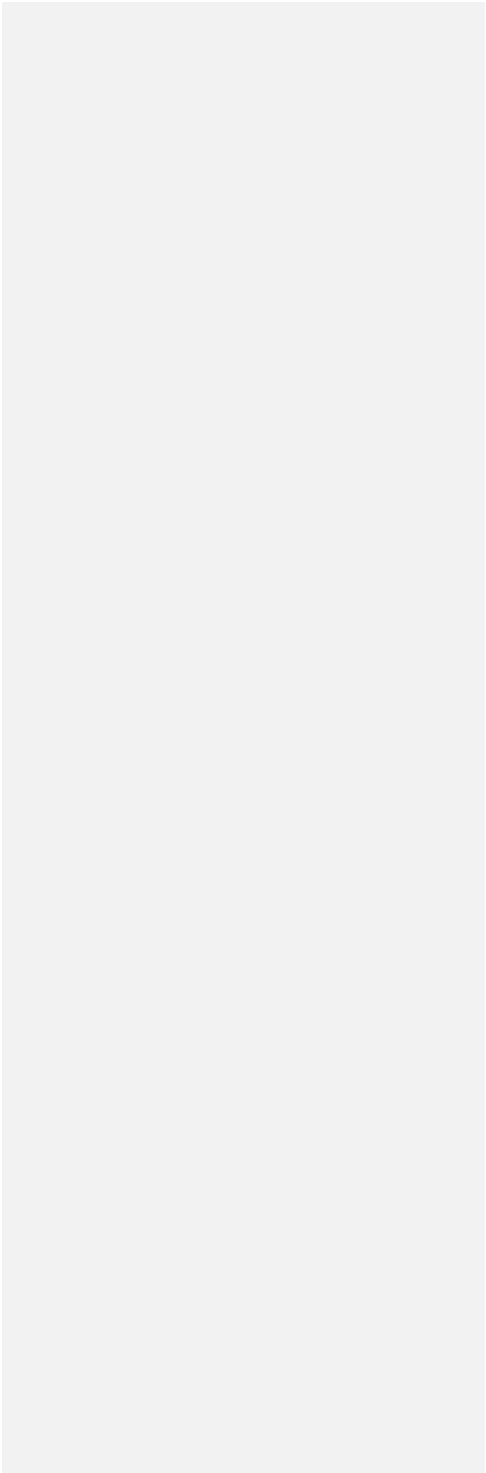
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



Les gènes voisins correspondent à des protéines de métabolisme de l'ADN



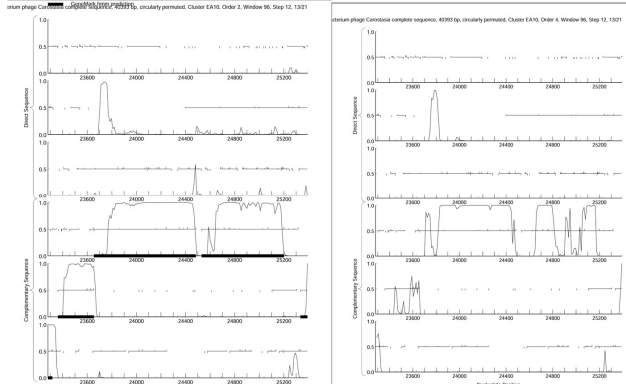
	<div><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Oui</i>
DECISION:	<i>AAA-ATPase</i>



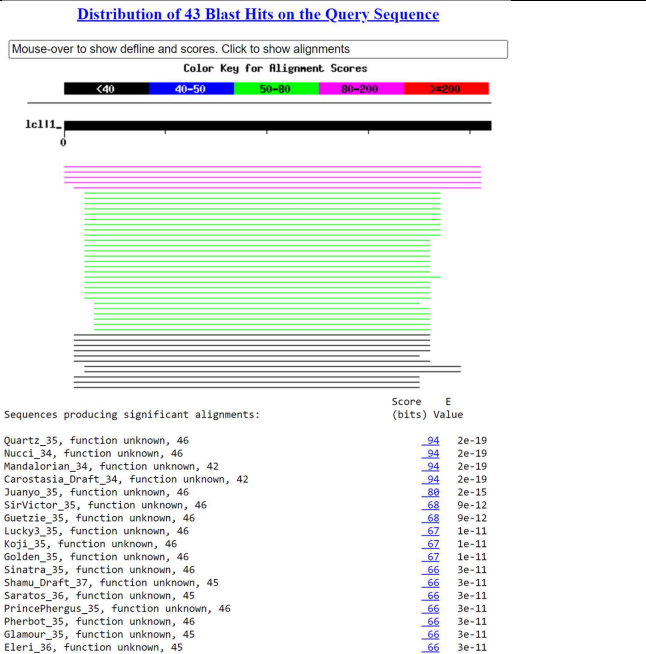
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	34
Coordonnées du Stop	25198
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	gap 8
Coordonnées du Start retenu	25326
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

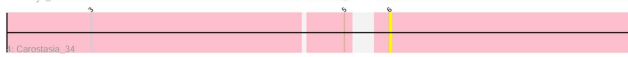

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Glimmer only</i>
Y-a-t-il des éléments supportant un potentiel codant ?	 <p><i>très faible potentiel codant mais présent pour les 2.</i></p>

Est-ce que le candidat est
retrouvé chez d'autres
génomomes annotés ?



	<div><div>>Nucci_34, function unknown, 46 Length = 46 Score = 93.6 bits (231), Expect = 2e-19 Identities = 42/42 (100%), Positives = 42/42 (100%) Query: 1 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 42 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA Sbjct: 5 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 46</div><div>>Mandalorian_34, function unknown, 42 Length = 42 Score = 93.6 bits (231), Expect = 2e-19 Identities = 42/42 (100%), Positives = 42/42 (100%) Query: 1 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 42 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA Sbjct: 1 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 42</div><div>>Carostasia_Draft_34, function unknown, 42 Length = 42 Score = 93.6 bits (231), Expect = 2e-19 Identities = 42/42 (100%), Positives = 42/42 (100%) Query: 1 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 42 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA Sbjct: 1 MFECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 42</div><div>>Juanyo_35, function unknown, 46 Length = 46 Score = 80.1 bits (196), Expect = 2e-15 Identities = 35/41 (85%), Positives = 36/41 (87%) Query: 2 FECSYCGAAPGQACVVTLAPEVGPLSIRIINHRAREEGKAA 42 F CSYCGA PG CVVTLAPEVGPLSIRIINHRAREEGG+ A Sbjct: 6 FACS YCGADPGHFCVVTLAPEVGPLSIRIINHRAREEGQVA 46</div><div><i>Alignement trouvé mais peu</i></div></div>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>La longueur est acceptable chev de 3 avec le gène d'avant et gap de 8 avec celui d'après</i>
DECISION:	<i>oui</i>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																													
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Glimmer : 25326</i> <i>GeneMark : NA</i>																																													
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-5,184</td><td>1,373</td><td>10</td><td>-5,879</td><td>TACAGGCTGCTCGAGCTTCAT</td><td>GTG</td><td>25470</td><td>273</td></tr><tr><td>2</td><td>-3,620</td><td>2,105</td><td>9</td><td>-4,395</td><td>AGAAGAAAAGCGAAGGGCAAGAA</td><td>ATG</td><td>25338</td><td>141</td></tr><tr><td>3</td><td>-3,652</td><td>2,090</td><td>11</td><td>-4,409</td><td>AGGGCAAGAAATGAGAGCACTC</td><td>ATG</td><td>25326</td><td>129</td></tr></table> <p><i>3 score RBS prévu mais le start prédit est le moins bon des 3 (ATG)</i> <i>Le 2° meilleur score RBS est aussi un ATG(25338) et conduit à un chev de 3pb avec le gène 35</i></p>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-5,184	1,373	10	-5,879	TACAGGCTGCTCGAGCTTCAT	GTG	25470	273	2	-3,620	2,105	9	-4,395	AGAAGAAAAGCGAAGGGCAAGAA	ATG	25338	141	3	-3,652	2,090	11	-4,409	AGGGCAAGAAATGAGAGCACTC	ATG	25326	129
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																						
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																						
1	-5,184	1,373	10	-5,879	TACAGGCTGCTCGAGCTTCAT	GTG	25470	273																																						
2	-3,620	2,105	9	-4,395	AGAAGAAAAGCGAAGGGCAAGAA	ATG	25338	141																																						
3	-3,652	2,090	11	-4,409	AGGGCAAGAAATGAGAGCACTC	ATG	25326	129																																						
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p><i>Avec le start prédit on a une ORF de 129pb avec un gap de 8</i> <i>Pour le start avec le 2° meilleur score RBS (25338) on a une ORF de 141 et un chev de 3pb avec le stop du gène 35 ce qui est acceptable</i> <i>On ne retient pas le meilleur score RBS car chev largement > 30pb (140pb)</i></p>																																													
Est-ce que le start est conservé chez les homologues voir Starterator ?	 <p>Gene: Carostasia_34 Start: 25326, Stop: 25198, Start Num: 6 Candidate Starts for Carostasia_34: (3, 25470), (Start: 5 @25338 has 3 MA's), (Start: 6 @25326 has 1 MA's),</p>																																													
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Distribution of the top 15 Blast Hits on 15 subject sequences</p>  <p><i>nécessaire, fournissez le meilleur hit BlastP de NCBI,</i></p>																																													

pour le start 25326 :

Download

GenPept

Graphics

hypothetical protein SEA_NUCC1_34 [Microbacterium phage Nucci]

Sequence ID: [QDK02427.1](#) Length: 46 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 5 to 46 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
89.4 bits(220)	4e-22	Compositional matrix adjust.	42/42(100%)	42/42(100%)	0/42(0%)

Query 1 MFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 42

Sbjct 5 MFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 46

Download

GenPept

Graphics

hypothetical protein SEA_MANDALORIAN_34 [Microbacterium phage Mandalorian]

Sequence ID: [QX013627.1](#) Length: 42 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 42 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
89.0 bits(219)	5e-22	Compositional matrix adjust.	42/42(100%)	42/42(100%)	0/42(0%)

Query 1 MFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 42

Sbjct 1 MFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 42

pour le start 25338 :

Download

GenPept

Graphics

hypothetical protein SEA_NUCC1_34 [Microbacterium phage Nucci]

Sequence ID: [QDK02427.1](#) Length: 46 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 46 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
96.7 bits(239)	6e-25	Compositional matrix adjust.	46/46(100%)	46/46(100%)	0/46(0%)

Query 1 MRALMFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 46

Sbjct 1 MRALMFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 46

Download

GenPept

Graphics

hypothetical protein SEA_MANDALORIAN_34 [Microbacterium phage Mandalorian]

Sequence ID: [QX013627.1](#) Length: 42 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 42 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
89.4 bits(220)	5e-22	Compositional matrix adjust.	42/42(100%)	42/42(100%)	0/42(0%)

Query 5 MFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 46

Sbjct 1 MFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 42

Download

GenPept

Graphics

hypothetical protein SEA_JUANYO_35 [Microbacterium phage Juanyo]

Sequence ID: [USH44815.1](#) Length: 46 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 46 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
84.3 bits(207)	5e-20	Compositional matrix adjust.	39/46(85%)	40/46(86%)	0/46(0%)

Query 1 MRALMFECSYCGAAGQACVVTLAPEVGPLSTRNHRAREEGGKAA 46

Sbjct 1 MRALTFACSYCGADPGFCVVTLAPEVGPLSTRNHRAREEGGQVA 46

DECISION:

Après on a donc 2 start possible entre celui prédit par glimmer 25326 pb qui entraîne un gap de 8 ou alors le start qui est trouvé en analysant les RBS, qui est donc le deuxième meilleur score RBS, 25338 et qui donne un chev de 3 pb qui est acceptable. Mais homologue avec même début pour le start 25326
Start retenu = start prédit = 25326

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Rationnelle

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :

(q#: s#) : [alignment] ; e-value :

[Distribution of 44 Blast Hits on the Query Sequence](#)

Mouse-over to show define and scores. Click to show alignments

Color Key for Alignment Scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

Sequences producing significant alignments:

	Score (bits)	E Value
Quartz_35, function unknown, 46	94	2e-19
Nucci_34, function unknown, 46	94	2e-19
Mandalorian_34, function unknown, 42	94	2e-19
Carostasia_Draft_34, function unknown, 42	94	2e-19
Juanyo_35, function unknown, 46	80	2e-15
SirVictor_35, function unknown, 46	68	9e-12
Guetzle_35, function unknown, 46	68	9e-12
Lucky3_35, function unknown, 46	67	1e-11
Koji_35, function unknown, 46	67	1e-11
Golden_35, function unknown, 46	67	1e-11
Sinatra_35, function unknown, 46	66	3e-11
Shamu_36, function unknown, 45	66	3e-11
Saratos_36, function unknown, 45	66	3e-11
PrincePhergus_35, function unknown, 46	66	3e-11
Pherbot_35, function unknown, 46	66	3e-11
Glamour_35, function unknown, 45	66	3e-11
Eleri_36, function unknown, 45	66	3e-11
ColaCorta_36, function unknown, 45	66	3e-11
ChikPic_36, function unknown, 45	66	3e-11
Bustleton_35, function unknown, 46	66	3e-11
Andromedas_36, function unknown, 45	66	3e-11

```
>Quartz_35, function unknown, 46
  Length = 46

Score = 93.6 bits (231), Expect = 2e-19
Identities = 42/42 (100%), Positives = 42/42 (100%)

Query: 1  MFECASYCGAAPGQACVVTLAPEVGPLSIRWNHRAREEGGKAA 42
          MFECASYCGAAPGQACVVTLAPEVGPLSIRWNHRAREEGGKAA
Sbjct: 5  MFECASYCGAAPGQACVVTLAPEVGPLSIRWNHRAREEGGKAA 46
```

```
>Nucci_34, function unknown, 46
  Length = 46

Score = 93.6 bits (231), Expect = 2e-19
Identities = 42/42 (100%), Positives = 42/42 (100%)

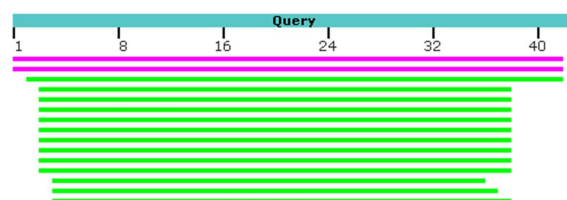
Query: 1  MFECASYCGAAPGQACVVTLAPEVGPLSIRWNHRAREEGGKAA 42
          MFECASYCGAAPGQACVVTLAPEVGPLSIRWNHRAREEGGKAA
Sbjct: 5  MFECASYCGAAPGQACVVTLAPEVGPLSIRWNHRAREEGGKAA 46
```

*** nr :

(q#: s#) : [alignment] ; e-value :

No putative conserved domains have been detected

Distribution of the top 15 Blast Hits on 15 subject sequences



hypothetical protein SEA_NUCCI_34 [Microbacterium phage Nucci]

Sequence ID: [QDK02427.1](#) Length: 46 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(PG\)](#)

Range 1: 5 to 46 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
89.4 bits(220)	4e-22	Compositional matrix adjust.	42/42(100%)	42/42(100%)	0/42(0%)

Query 1 MFECASYGAAPGACVVT LAPEVGPLSLRWIHRAREEGKAA 42
 Sbjct 5 MFECASYGAAPGACVVT LAPEVGPLSLRWIHRAREEGKAA 46

Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.

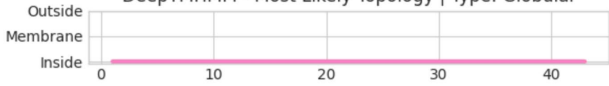
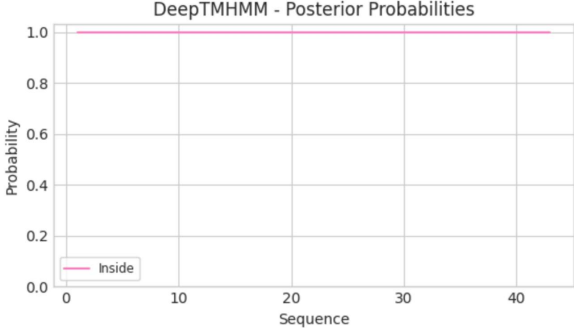
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

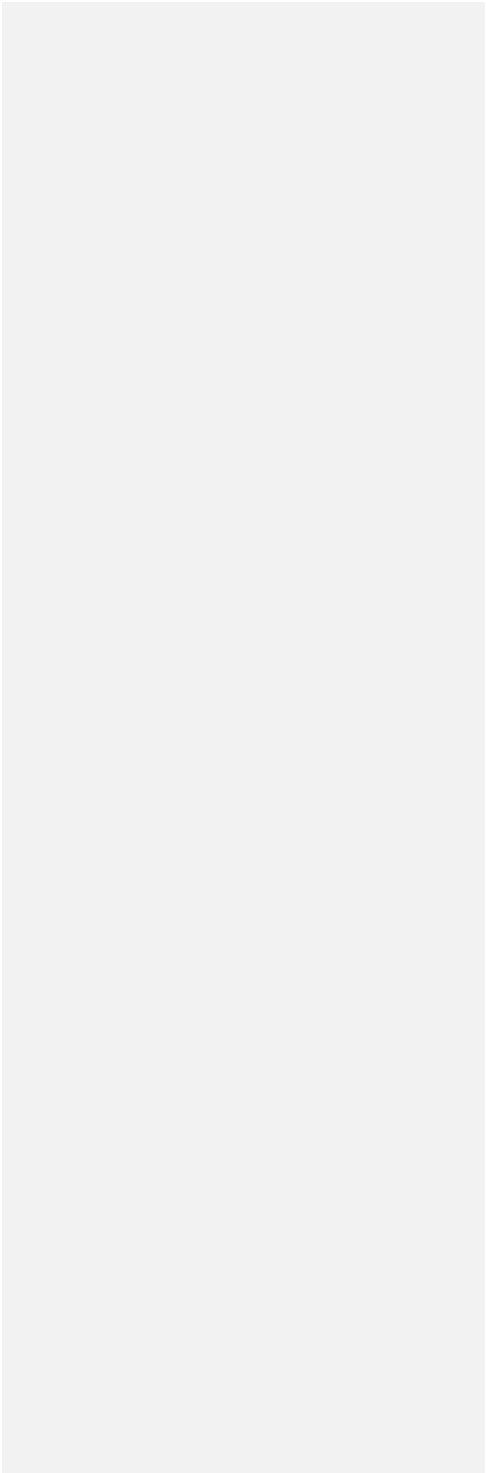
Pas de correspondance pertinente

Number of Hits: **250**
Query MSA diversity (Neff): **4.23951**

Visualization



	<div><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Non</i>
DECISION:	<i>NFK</i>



Student Gene Annotation Worksheet

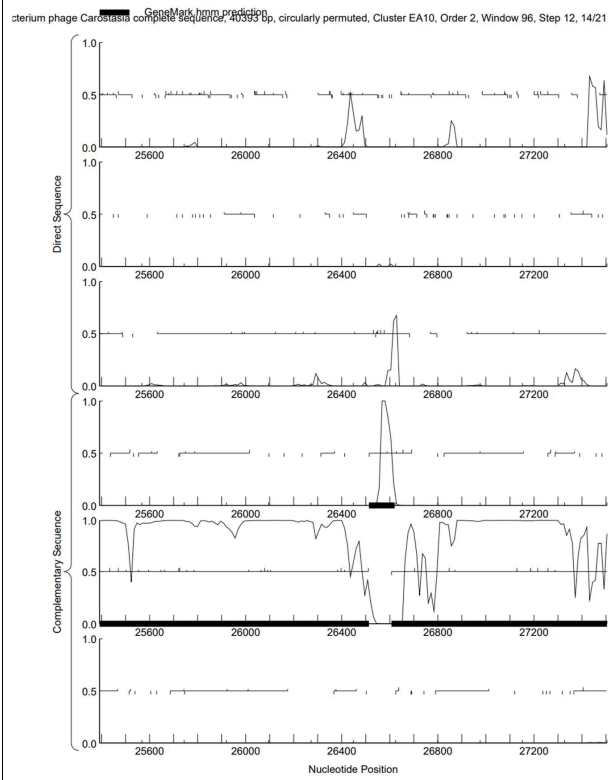
Carosmatique-gene-35

Basic Phage Information	
Nom du Phage	carostasia
Gène #	35
Coordonnées du Stop	25335
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	chev 1
Coordonnées du Start retenu	26513
Fonction prédite	Cas 4 Exonucléase

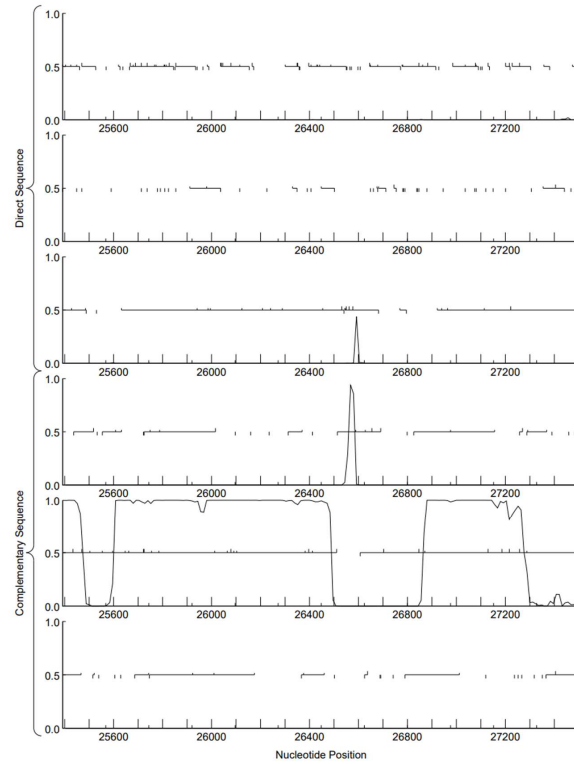
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>Yes both</i>

Y-a-t-il des éléments
supportant un potentiel
codant ?



cterium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 14/21



Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Distribution of 100 Blast Hits on the Query Sequence</p> <div>Mouse-over to show define and scores. Click to show alignments</div> <p>Color Key for Alignment Scores</p> <table><tr><td><40</td><td>40-50</td><td>50-80</td><td>80-200</td><td>>=200</td></tr></table> <p>1c111_</p>	<40	40-50	50-80	80-200	>=200
<40	40-50	50-80	80-200	>=200		
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Non					
DECISION:	oui					

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																															
Quel start Glimmer et GeneMark suggèrent-ils ?	26513																																																																																																																																																																																																															
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Stats: 21 ORF Start : 26015 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler6 Explore Selected: 1 ORF Stop : 25335 5' End 87.9 48.5 42.4 99 Spacing Weight Matrix Karlin Medium Document ORF Length: 681 3' End 58.2 77.6 43.3 402</div> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,910</td><td>1,502</td><td>11</td><td>-5,667</td><td>TCTCTGGGCTGGCGGCTGCTG</td><td>ATG</td><td>26513</td><td>1179</td></tr><tr><td>2</td><td>-5,224</td><td>1,355</td><td>9</td><td>-5,998</td><td>CACCAACGAGAGCGGACAGCTC</td><td>GTG</td><td>26414</td><td>1080</td></tr><tr><td>3</td><td>-5,454</td><td>1,247</td><td>5</td><td>-7,454</td><td>CAGCTGTGTGATCACCAGAGC</td><td>ATG</td><td>26399</td><td>1065</td></tr><tr><td>4</td><td>-3,788</td><td>2,027</td><td>8</td><td>-5,005</td><td>CCAGAGCATGCTCAGCAGCTTC</td><td>GTG</td><td>26384</td><td>1050</td></tr><tr><td>5</td><td>-5,967</td><td>1,007</td><td>14</td><td>-7,314</td><td>CGAAGAGTGTCTACAGCATCTC</td><td>TTG</td><td>26162</td><td>828</td></tr><tr><td>6</td><td>-6,721</td><td>0,654</td><td>17</td><td>-9,721</td><td>CGAGAGCTCACCCCTGTGCGG</td><td>GTG</td><td>26105</td><td>771</td></tr><tr><td>7</td><td>-2,955</td><td>2,416</td><td>10</td><td>-3,650</td><td>CCCTGTGGGCTGGAGCTGAGC</td><td>GTG</td><td>26093</td><td>759</td></tr><tr><td>8</td><td>-2,955</td><td>2,416</td><td>10</td><td>-3,650</td><td>GGAGCTGAGGCTGGAGCGCCCC</td><td>ATG</td><td>26081</td><td>747</td></tr><tr><td>9</td><td>-3,413</td><td>2,202</td><td>8</td><td>-4,634</td><td>GGGCCCCATGTTCCGGGGCAAG</td><td>GTG</td><td>26066</td><td>732</td></tr><tr><td>10</td><td>-4,463</td><td>1,711</td><td>10</td><td>-5,158</td><td>CTGGATCGAGCAGGGGCGAT</td><td>GTG</td><td>26015</td><td>681</td></tr><tr><td>11</td><td>-3,620</td><td>2,105</td><td>16</td><td>-5,416</td><td>CACCAAGGGCAAGATCAGCGGT</td><td>GTG</td><td>25787</td><td>453</td></tr><tr><td>12</td><td>-5,308</td><td>1,315</td><td>11</td><td>-6,065</td><td>GCCCCCGGACAGCTACCCCT</td><td>GTG</td><td>25757</td><td>423</td></tr><tr><td>13</td><td>-5,026</td><td>1,447</td><td>10</td><td>-5,720</td><td>TGAGTGGCTCATCGAGAACGGT</td><td>ATG</td><td>25727</td><td>393</td></tr><tr><td>14</td><td>-3,522</td><td>2,151</td><td>6</td><td>-5,266</td><td>GTGGTCTATCGAGAACGGTATG</td><td>ATG</td><td>25724</td><td>390</td></tr><tr><td>15</td><td>-2,071</td><td>2,831</td><td>13</td><td>-3,116</td><td>AGACACCGAGGAGCGCATCTAC</td><td>GTG</td><td>25664</td><td>330</td></tr><tr><td>16</td><td>-2,071</td><td>2,831</td><td>10</td><td>-2,765</td><td>CATCTAGCTGGAGGAGTTCTCTG</td><td>GTG</td><td>25649</td><td>315</td></tr><tr><td>17</td><td>-6,520</td><td>0,748</td><td>11</td><td>-7,277</td><td>CGAGTTCGGCGCTGACCGGCTG</td><td>GTG</td><td>25598</td><td>264</td></tr><tr><td>18</td><td>-2,939</td><td>2,424</td><td>11</td><td>-3,696</td><td>CGAGGCTCAGCGGAGTGGTTT</td><td>GTG</td><td>25556</td><td>222</td></tr><tr><td>19</td><td>-3,532</td><td>2,147</td><td>17</td><td>-5,532</td><td>CAGTGGGATGACCCCGACTGC</td><td>GTG</td><td>25505</td><td>171</td></tr><tr><td>20</td><td>-5,184</td><td>1,373</td><td>8</td><td>-6,406</td><td>CCCTACAGGCTGCTGAGCTTC</td><td>ATG</td><td>25472</td><td>138</td></tr><tr><td>21</td><td>-4,437</td><td>1,723</td><td>13</td><td>-5,482</td><td>GGACCTCAGGGTTCAGACCTG</td><td>ATG</td><td>25436</td><td>102</td></tr></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,910	1,502	11	-5,667	TCTCTGGGCTGGCGGCTGCTG	ATG	26513	1179	2	-5,224	1,355	9	-5,998	CACCAACGAGAGCGGACAGCTC	GTG	26414	1080	3	-5,454	1,247	5	-7,454	CAGCTGTGTGATCACCAGAGC	ATG	26399	1065	4	-3,788	2,027	8	-5,005	CCAGAGCATGCTCAGCAGCTTC	GTG	26384	1050	5	-5,967	1,007	14	-7,314	CGAAGAGTGTCTACAGCATCTC	TTG	26162	828	6	-6,721	0,654	17	-9,721	CGAGAGCTCACCCCTGTGCGG	GTG	26105	771	7	-2,955	2,416	10	-3,650	CCCTGTGGGCTGGAGCTGAGC	GTG	26093	759	8	-2,955	2,416	10	-3,650	GGAGCTGAGGCTGGAGCGCCCC	ATG	26081	747	9	-3,413	2,202	8	-4,634	GGGCCCCATGTTCCGGGGCAAG	GTG	26066	732	10	-4,463	1,711	10	-5,158	CTGGATCGAGCAGGGGCGAT	GTG	26015	681	11	-3,620	2,105	16	-5,416	CACCAAGGGCAAGATCAGCGGT	GTG	25787	453	12	-5,308	1,315	11	-6,065	GCCCCCGGACAGCTACCCCT	GTG	25757	423	13	-5,026	1,447	10	-5,720	TGAGTGGCTCATCGAGAACGGT	ATG	25727	393	14	-3,522	2,151	6	-5,266	GTGGTCTATCGAGAACGGTATG	ATG	25724	390	15	-2,071	2,831	13	-3,116	AGACACCGAGGAGCGCATCTAC	GTG	25664	330	16	-2,071	2,831	10	-2,765	CATCTAGCTGGAGGAGTTCTCTG	GTG	25649	315	17	-6,520	0,748	11	-7,277	CGAGTTCGGCGCTGACCGGCTG	GTG	25598	264	18	-2,939	2,424	11	-3,696	CGAGGCTCAGCGGAGTGGTTT	GTG	25556	222	19	-3,532	2,147	17	-5,532	CAGTGGGATGACCCCGACTGC	GTG	25505	171	20	-5,184	1,373	8	-6,406	CCCTACAGGCTGCTGAGCTTC	ATG	25472	138	21	-4,437	1,723	13	-5,482	GGACCTCAGGGTTCAGACCTG	ATG	25436	102
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																								
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																								
1	-4,910	1,502	11	-5,667	TCTCTGGGCTGGCGGCTGCTG	ATG	26513	1179																																																																																																																																																																																																								
2	-5,224	1,355	9	-5,998	CACCAACGAGAGCGGACAGCTC	GTG	26414	1080																																																																																																																																																																																																								
3	-5,454	1,247	5	-7,454	CAGCTGTGTGATCACCAGAGC	ATG	26399	1065																																																																																																																																																																																																								
4	-3,788	2,027	8	-5,005	CCAGAGCATGCTCAGCAGCTTC	GTG	26384	1050																																																																																																																																																																																																								
5	-5,967	1,007	14	-7,314	CGAAGAGTGTCTACAGCATCTC	TTG	26162	828																																																																																																																																																																																																								
6	-6,721	0,654	17	-9,721	CGAGAGCTCACCCCTGTGCGG	GTG	26105	771																																																																																																																																																																																																								
7	-2,955	2,416	10	-3,650	CCCTGTGGGCTGGAGCTGAGC	GTG	26093	759																																																																																																																																																																																																								
8	-2,955	2,416	10	-3,650	GGAGCTGAGGCTGGAGCGCCCC	ATG	26081	747																																																																																																																																																																																																								
9	-3,413	2,202	8	-4,634	GGGCCCCATGTTCCGGGGCAAG	GTG	26066	732																																																																																																																																																																																																								
10	-4,463	1,711	10	-5,158	CTGGATCGAGCAGGGGCGAT	GTG	26015	681																																																																																																																																																																																																								
11	-3,620	2,105	16	-5,416	CACCAAGGGCAAGATCAGCGGT	GTG	25787	453																																																																																																																																																																																																								
12	-5,308	1,315	11	-6,065	GCCCCCGGACAGCTACCCCT	GTG	25757	423																																																																																																																																																																																																								
13	-5,026	1,447	10	-5,720	TGAGTGGCTCATCGAGAACGGT	ATG	25727	393																																																																																																																																																																																																								
14	-3,522	2,151	6	-5,266	GTGGTCTATCGAGAACGGTATG	ATG	25724	390																																																																																																																																																																																																								
15	-2,071	2,831	13	-3,116	AGACACCGAGGAGCGCATCTAC	GTG	25664	330																																																																																																																																																																																																								
16	-2,071	2,831	10	-2,765	CATCTAGCTGGAGGAGTTCTCTG	GTG	25649	315																																																																																																																																																																																																								
17	-6,520	0,748	11	-7,277	CGAGTTCGGCGCTGACCGGCTG	GTG	25598	264																																																																																																																																																																																																								
18	-2,939	2,424	11	-3,696	CGAGGCTCAGCGGAGTGGTTT	GTG	25556	222																																																																																																																																																																																																								
19	-3,532	2,147	17	-5,532	CAGTGGGATGACCCCGACTGC	GTG	25505	171																																																																																																																																																																																																								
20	-5,184	1,373	8	-6,406	CCCTACAGGCTGCTGAGCTTC	ATG	25472	138																																																																																																																																																																																																								
21	-4,437	1,723	13	-5,482	GGACCTCAGGGTTCAGACCTG	ATG	25436	102																																																																																																																																																																																																								
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Pas le meilleur score (-5.667) mais plus grand ORF (1179bp)</p>																																																																																																																																																																																																															
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																																																																																																																																																

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?

Cas4 family exonuclease [Microbacterium phage Nucci]

Sequence ID: [QDK02428.1](#) Length: 392 Number of Matches: 1
[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 392 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
781 bits (2017)	0.0	Compositional matrix adjust.	376/392(96%)	384/392(97%)	0/392(0%)
Query 1	MVAKKLASTGKPGSKAMHGPRGTWPTT	EDGKLVTQSMVSSFVECPRETYGII	GLRP	60	
Sbjct 1	MVAKKLASTGKPGSKAMHGPRGTWPTT	EDGKLVTQSMVSSFVECPRETYGII	GLRP	60	
Query 61	RTASKPLTRGTWHALLEARGQCKDWRK	LHRELTLKAETHEFEEVAGLAECYNIV	LSY	120	
Sbjct 61	RI SKPLTRGTWHALLE RG+G DWRK	LH+ELT KAETHEFEEVAGLA+ECYNIV	LSY	120	
Query 121	DWLKETKLHETLTPPAVELTVERPMF	RKGVLRYGRIDIWIIDEGDVMGDHK	THATLPD	180	
Sbjct 121	DWLKETKLHETLTPPAVELTVER M	RKGVLRYGRIDIWIIDEGDVMGDHK	THA LPD	180	
Query 181	WRVRELAFQHYSVLWACETAPEYKAV	RYKGKLP+PKGFIYDVCRTGAIKTPT	LTITGKI	240	
Sbjct 181	WRVRELAFQHYSVLWACETAPEYKA	RYKGKLP+PKGFIYDVCRTGAIKTPT	LTITGKI	240	
Query 241	SRVLKPTGTTYVPFREWLIENGMMTV	IRGKDLAIEDTEERIVVEEFLVELQ	HRDYSTEF	300	
Sbjct 241	SRVLKP+GTTYVPFREWLIENGMMTV	IRGKDLAIED EER+VVEEFLVELQ	HRDYSTEF	300	
Query 301	RRDRLVFSPEQAEQRKSFVTSARRLL	TYKWDPPDCVERNLCSSFCNFKD	LTVDLM	360	
Sbjct 301	RRDRLVFSPEQAEQRKSFVTSARRLL	TYKWDPPDCVERNLCSSFCNFKD	LTVDLM	360	
Query 361	HGTSEIEQRTRYVTTTRDPLDYYPNQ	KKAKGKK	392		
Sbjct 361	HGTSEIEQRTRYVTTTRDPLDYYPNQ	KKAKGKK	392		

Séquences produisant des alignements significatifs

Télécharger Sélectionner des colonnes Montrer 100

☒ Tout sélectionner 100 séquences sélectionnées [GenPept](#) [Graphique](#) [Arbre de résultats de distance](#) [Alignement multiple](#) [Visionneuse MSA](#)

	Description	Nom scientifique	Note maximale	Score total	Couverture de la requête	Valeur E	Par. Identif	Acc. Lect.	Accession
<input checked="" type="checkbox"/>	Exonuclease de la famille Cas4 [Microbacterium phage Nucci]	Phage microbactérien Nucci	781	781	100%	0.0	95.92%	392	QDK02428.1
<input checked="" type="checkbox"/>	Exonuclease de la famille Cas4 [Microbacterium phage Quartz]	Phage microbactérien Quartz	780	780	100%	0.0	95.66%	392	U0550255.1
<input checked="" type="checkbox"/>	Exonuclease de la famille Cas4 [Microbacterium phage Juanao]	Phage microbactérien Juanao	763	763	99%	0.0	93.08%	395	U0544785.1
<input checked="" type="checkbox"/>	Exonuclease de la famille Cas4 [Phage microbactérien YauY]	Phage microbactérien YauY	740	740	99%	0.0	93.59%	398	QEP90365.1
<input checked="" type="checkbox"/>	Exonuclease Cas4 [Inhalo microbactérien Schimmel22]	Phage microbactérien Schimmel22	697	697	99%	0.0	85.35%	393	U0695027.1
<input checked="" type="checkbox"/>	Exonuclease hélicase de type RecB [Phage microbactérien Tivymand]	Phage microbactérien Tivymand	697	697	99%	0.0	85.35%	393	U0695027.1
<input checked="" type="checkbox"/>	Exonuclease hélicase de type RecB [Microbacterium phage HerulesX]	Phage microbactérien HerulesX	696	696	99%	0.0	85.09%	393	U0523821.1
<input checked="" type="checkbox"/>	Exonuclease de la famille Cas4 [Microbacterium phage Finny]	Phage microbactérien Finny	694	694	99%	0.0	86.41%	387	QCF15786.1
<input checked="" type="checkbox"/>	Exonuclease [Microbacterium phage Sarcel]	Phage microbactérien Sarcel	693	693	99%	0.0	86.41%	387	AD488609.1
<input checked="" type="checkbox"/>	Exonuclease [Microbacterium phage Esal]	Microbactérie phage Esal	692	692	99%	0.0	85.93%	396	YP_009628073.1

DECISION:

le start conservé est le 26513 car il n'est pas en contradiction avec les principes d'annotation et Il est conservé chez d'autres homologues. De plus, il induit le plus long ORF avec un ATG

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec	<div>Listez le meilleur hit Blastp pour chaque source :</div> <div>*** PhagesDB :</div> <div>(q#: s#) : [alignment] ; e-value :</div>

une e-value < 10⁻⁴ et une couverture acceptable ?

Sequences producing significant alignments:

	Score	E
	(bits)	Value
Carostasia_Draft_35, function unknown, 392	819	0.0
Nucci_35, Cas4 family exonuclease, 392	790	0.0
Mandalorian_35, RecB-like exonuclease/helicase, 392	790	0.0
Quartz_36, Cas4 family exonuclease, 392	788	0.0
Juanyo_36, Cas4 family exonuclease, 395	770	0.0
Yuuy_36, Cas4 family exonuclease, 398	770	0.0
Sansa_36, exonuclease, 387	710	0.0

*** nr :
(q#: s#) : [alignment] ; e-value :

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▾

Select columns ▾

Show 100 ▾

☒ select all 100 sequences selected

GenPept

Graphics

Distance trees of results

Multiple alignment

MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession
<input checked="" type="checkbox"/> Cas4 family exonuclease [Mycobacterium phage NucoI]	Mycobacterium phage NucoI	781	781	99%	0.0	95.92%	392	GD052428.1
<input checked="" type="checkbox"/> Cas4 family exonuclease [Mycobacterium phage Quat4]	Mycobacterium phage Quat4	780	780	99%	0.0	96.69%	392	U0526525.1
<input checked="" type="checkbox"/> Cas4 family exonuclease [Mycobacterium phage Jantap]	Mycobacterium phage Jantap	763	763	99%	0.0	93.69%	396	U0546106.1
<input checked="" type="checkbox"/> Cas4 family exonuclease [Mycobacterium phage YauY]	Mycobacterium phage YauY	740	740	99%	0.0	93.69%	396	GF595365.1
<input checked="" type="checkbox"/> Cas4 exonuclease [Mycobacterium phage Schimml22]	Mycobacterium phage Schimml22	697	697	98%	0.0	85.36%	393	U0509505.1
<input checked="" type="checkbox"/> RecB-like exonuclease/helicase [Mycobacterium phage Teymard]	Mycobacterium phage Teymard	697	697	98%	0.0	85.36%	393	U0509522.1
<input checked="" type="checkbox"/> RecB-like exonuclease/helicase [Mycobacterium phage HerculesX]	Mycobacterium phage HerculesX	696	696	98%	0.0	85.69%	393	U0531382.1
<input checked="" type="checkbox"/> Cas4 family exonuclease [Mycobacterium phage Eten]	Mycobacterium phage Eten	694	694	99%	0.0	86.41%	387	GF215766.1
<input checked="" type="checkbox"/> exonuclease [Mycobacterium phage Sansa]	Mycobacterium phage Sansa	693	693	99%	0.0	86.41%	387	A0348829.1

Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un cribre HHpred avec une proba >= 90% et une couverture acceptable ?

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	Q05283	VG69_BPMLS Gene 69 protein OS=Mycobacterium phage L5 OX=31757 GN=69 PE=4 SV=1	99.87	1.7e-19	161.34	23.7	159	290
2	Q64262	VG69_BPMD2 Gene 69 protein OS=Mycobacterium phage D29 OX=28369 GN=69 PE=4 SV=1	99.85	7e-19	154.76	23.9	159	269
3	GPPU_A	ATP-dependent DNA helicase (UvrD/REP); DNA, DNA BINDING PROTEIN, DNA BINDING PROTEIN-DNA complex; 3.5A {Mycobacterium sm	99.83	3.9e-18	172.85	25.4	241	724

32

200

Q05283

Q64262

GPPU_A

7147_A

GPPU_A

CAB0050

4EET_B

PDB:2X_1 PD-(07

Lsal KRISPR-ass

G0SP_1

DB-3/20 PDB:3A-

3.8B_0

BMP2000 Protein

cd22318

cd09637

Enc5 Exonuclease

Q0117

Q0104

Alignement avec Pfam 12701.10 PDDEXK_1
Oui, phage utilisé Mandalorian.

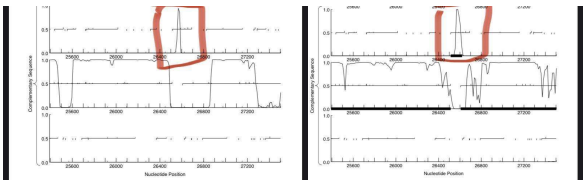
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une

forte conservation de l'ordre des gènes ?	<div><p>Gènes voisins métabolisme ADN</p></div>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div><p>NON</p><p>ID: UNQ1027 TOPLOC_2001 1 393 NON CYTOPLASMIC.</p><p>Probable posterior probabilities for UNQ1027</p><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Oui</i>
DECISION:	<i>Cas4 Exonucléase</i>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	36
Coordonnées du Stop	26513
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	Ove 14
Coordonnées du Start retenu	26620
Fonction prédite	Membrane protein

Décision #1 : Est-ce un gène ?

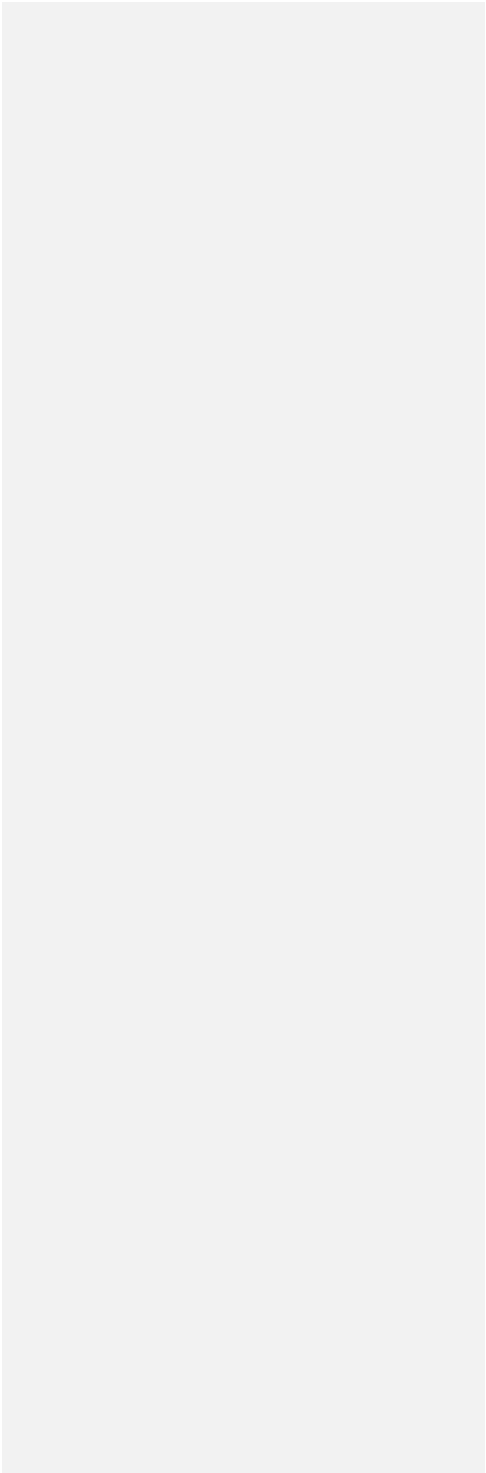
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>si GeneMarkS</p>  <p>GeneMark-host .</p> <p>Dans les deux cas on observe un potentiel codant le long de l'orf</p>
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	Peu d'homologues et ils ne sont pas annoté. Le meilleur hit : Quartz_37, fonction unknown, 35 score 71 et e value 8e-13

Commenté [1]: toute la page si possible

	<div><p>Distribution of 9 Blast Hits on the Query Sequence</p><p>Mouse over to show details and scores. Click to show alignment.</p><p>Color Key for Blastword Scores</p><p>Sequences producing significant alignments:</p><table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_36, function unknown, 38</td><td>81</td><td>8e-16</td></tr><tr><td>Quartz_37, function unknown, 35</td><td>71</td><td>8e-13</td></tr><tr><td>Nucci_36, function unknown, 35</td><td>71</td><td>8e-13</td></tr><tr><td>Mandalorian_36, function unknown, 35</td><td>71</td><td>8e-13</td></tr><tr><td>Zagie_68, function unknown, 37</td><td>34</td><td>0.15</td></tr><tr><td>Rowlf_Draft_66, function unknown, 37</td><td>34</td><td>0.15</td></tr><tr><td>Namago_68, function unknown, 37</td><td>34</td><td>0.15</td></tr><tr><td>Pizzles_67, function unknown, 38</td><td>34</td><td>0.15</td></tr><tr><td>Nike_68, function unknown, 37</td><td>32</td><td>0.56</td></tr></table></div>		Score (bits)	E Value	Carostasia_Draft_36, function unknown, 38	81	8e-16	Quartz_37, function unknown, 35	71	8e-13	Nucci_36, function unknown, 35	71	8e-13	Mandalorian_36, function unknown, 35	71	8e-13	Zagie_68, function unknown, 37	34	0.15	Rowlf_Draft_66, function unknown, 37	34	0.15	Namago_68, function unknown, 37	34	0.15	Pizzles_67, function unknown, 38	34	0.15	Nike_68, function unknown, 37	32	0.56
	Score (bits)	E Value																													
Carostasia_Draft_36, function unknown, 38	81	8e-16																													
Quartz_37, function unknown, 35	71	8e-13																													
Nucci_36, function unknown, 35	71	8e-13																													
Mandalorian_36, function unknown, 35	71	8e-13																													
Zagie_68, function unknown, 37	34	0.15																													
Rowlf_Draft_66, function unknown, 37	34	0.15																													
Namago_68, function unknown, 37	34	0.15																													
Pizzles_67, function unknown, 38	34	0.15																													
Nike_68, function unknown, 37	32	0.56																													
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>. Y a-t-il un chevauchement significatif avec d'autres gènes? Moins de 30pb mais quand même 23 Est-il suffisamment long ? Oui plus que 50pb Les gènes avant et après ce gène sont-ils dans la même direction ? Oui</p>																														
DECISION:	OUI																														


Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																	
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 26629</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 26620</i>																																																																																	
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-5,724</td><td>1,120</td><td>10</td><td>-6,419</td><td>CCATATAAGACAGCTCCAGGG</td><td>TTG</td><td>26728</td><td>216</td></tr><tr><td>2</td><td>-2,144</td><td>2,796</td><td>11</td><td>-2,901</td><td>AAGGAGACAATGGAGCAGGCTG</td><td>ATG</td><td>26692</td><td>180</td></tr><tr><td>3</td><td>-4,069</td><td>1,895</td><td>5</td><td>-6,069</td><td>CGAAGTTCCTCTCAAGGCAG</td><td>ATG</td><td>26656</td><td>144</td></tr><tr><td>4</td><td>-3,654</td><td>2,089</td><td>6</td><td>-5,399</td><td>ACCATCTCCGATGCTGGGGAG</td><td>GTG</td><td>26629</td><td>117</td></tr><tr><td>5</td><td>-2,976</td><td>2,407</td><td>10</td><td>-3,671</td><td>CGATGCTGGGAGGTGAACCCC</td><td>TTG</td><td>26620</td><td>108</td></tr><tr><td>6</td><td>-4,534</td><td>1,677</td><td>10</td><td>-5,229</td><td>GCGAGTTCCTAAGCTGTTTCAGC</td><td>TTG</td><td>26596</td><td>84</td></tr><tr><td>7</td><td>-4,534</td><td>1,677</td><td>16</td><td>-6,330</td><td>TTCTAAGCTGTTTCAGCTTGTC</td><td>GTG</td><td>26590</td><td>78</td></tr></table> <p><i>Meilleur score (-2.901) et meilleure Z Value (2.796) et ORF 180 c'est l'ATG 26692</i></p> <p><i>Score (-5.399) Z value (2.089) et ORF 117 c'est GTG 26629</i></p> <p><i>Score (-3.671) Z value (2.407) et ORF 108 c'est TTG 26620</i></p>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-5,724	1,120	10	-6,419	CCATATAAGACAGCTCCAGGG	TTG	26728	216	2	-2,144	2,796	11	-2,901	AAGGAGACAATGGAGCAGGCTG	ATG	26692	180	3	-4,069	1,895	5	-6,069	CGAAGTTCCTCTCAAGGCAG	ATG	26656	144	4	-3,654	2,089	6	-5,399	ACCATCTCCGATGCTGGGGAG	GTG	26629	117	5	-2,976	2,407	10	-3,671	CGATGCTGGGAGGTGAACCCC	TTG	26620	108	6	-4,534	1,677	10	-5,229	GCGAGTTCCTAAGCTGTTTCAGC	TTG	26596	84	7	-4,534	1,677	16	-6,330	TTCTAAGCTGTTTCAGCTTGTC	GTG	26590	78
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																										
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																										
1	-5,724	1,120	10	-6,419	CCATATAAGACAGCTCCAGGG	TTG	26728	216																																																																										
2	-2,144	2,796	11	-2,901	AAGGAGACAATGGAGCAGGCTG	ATG	26692	180																																																																										
3	-4,069	1,895	5	-6,069	CGAAGTTCCTCTCAAGGCAG	ATG	26656	144																																																																										
4	-3,654	2,089	6	-5,399	ACCATCTCCGATGCTGGGGAG	GTG	26629	117																																																																										
5	-2,976	2,407	10	-3,671	CGATGCTGGGAGGTGAACCCC	TTG	26620	108																																																																										
6	-4,534	1,677	10	-5,229	GCGAGTTCCTAAGCTGTTTCAGC	TTG	26596	84																																																																										
7	-4,534	1,677	16	-6,330	TTCTAAGCTGTTTCAGCTTGTC	GTG	26590	78																																																																										
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p><i>l'ATG 26692 = chevauchement de 86pb</i></p> <p><i>GTG 26629 = chevauchement de 23pb</i></p> <p><i>TTG 26620 = chevauchement de 14pb</i></p> <p><i>Avec le gène d'avant</i></p>																																																																																	
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p><i>Pas de start en vert mais le star le plus annoté c'est le TTG 26620</i></p>																																																																																	



Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p><i>Aucun des START n'est conservé chez les homologues</i></p> <p>>Quartz_37, function unknown, 35 Length = 35</p> <p>Score = 71.2 bits (173), Expect = 8e-13 Identities = 33/35 (94%), Positives = 34/35 (97%)</p> <p>Query: 4 LPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 38 +PSSKLFSLFVVAIGVSFIGGVI TYFGILWAGGC Sbjct: 1 MPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 35</p> <p>>Nucci_36, function unknown, 35 Length = 35</p> <p>Score = 71.2 bits (173), Expect = 8e-13 Identities = 33/35 (94%), Positives = 34/35 (97%)</p> <p>Query: 4 LPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 38 +PSSKLFSLFVVAIGVSFIGGVI TYFGILWAGGC Sbjct: 1 MPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 35</p> <p>>Mandalorian_36, function unknown, 35 Length = 35</p> <p>Score = 71.2 bits (173), Expect = 8e-13 Identities = 33/35 (94%), Positives = 34/35 (97%)</p> <p>Query: 4 LPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 38 +PSSKLFSLFVVAIGVSFIGGVI TYFGILWAGGC Sbjct: 1 MPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 35</p> <p>>Zagie_68, function unknown, 37 Length = 37</p> <p>Score = 33.9 bits (76), Expect = 0.15 Identities = 14/23 (60%), Positives = 18/23 (78%)</p> <p>Query: 14 VVAIGVSFIGGVITTYFGILWAG 36 +V T VSE GV+ TYEG+LW+G</p>
DECISION:	26 620 meilleur compromis ORF, chevauchement, score Z value

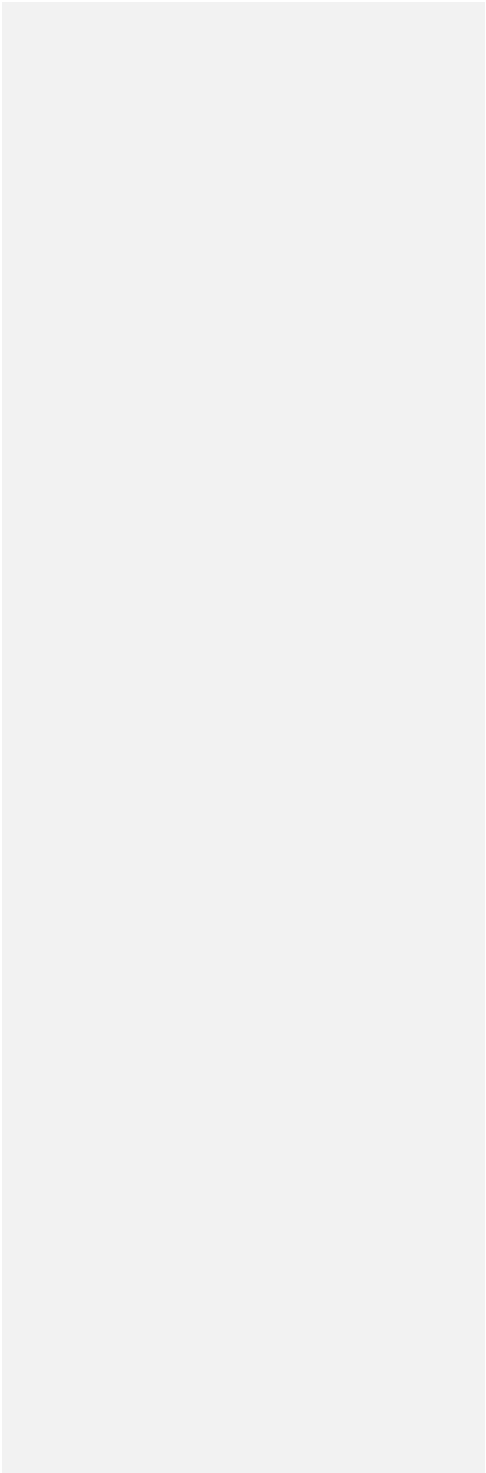
Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																			
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q4: s1) : [alignment] ; e-value : 8e-13</p> <p>Distribution of 9 Blast Hits on the Query Sequence</p> <div><p>Mouse-over to show define and scores. Click to show alignments</p><p>Color Key for Alignment Scores</p><table><tr><td><40</td><td>40-50</td><td>50-80</td><td>80-200</td><td>>=200</td></tr></table></div> <table><thead><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Carostasia_Draft_36, function unknown, 38</td><td>74</td><td>2e-13</td></tr><tr><td>Quartz_37, function unknown, 35</td><td>72</td><td>4e-13</td></tr><tr><td>Nucci_36, function unknown, 35</td><td>72</td><td>4e-13</td></tr><tr><td>Mandalorian_36, function unknown, 35</td><td>72</td><td>4e-13</td></tr><tr><td>Zagie_68, function unknown, 37</td><td>34</td><td>0.14</td></tr><tr><td>Rowlf_Draft_66, function unknown, 37</td><td>34</td><td>0.14</td></tr><tr><td>Namago_68, function unknown, 37</td><td>34</td><td>0.14</td></tr><tr><td>Fizzles_67, function unknown, 38</td><td>34</td><td>0.14</td></tr><tr><td>Nike_68, function unknown, 37</td><td>32</td><td>0.55</td></tr></tbody></table> <p>>Quartz_37, function unknown, 35 Length = 35</p> <p>Score = 72.4 bits (176), Expect = 4e-13 Identities = 34/35 (97%), Positives = 34/35 (97%)</p> <p>Query: 1 MPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 35 MPSSKLFSLFVVAIGVSFIGGVI TYFGILWAGGC Sbjct: 1 MPSSKLFSLFVVAIGVSFIGGVIITYFGILWAGGC 35</p> <p>>Nucci_36, function unknown, 35 Length = 35</p> <p>Score = 72.4 bits (176), Expect = 4e-13 Identities = 34/35 (97%), Positives = 34/35 (97%)</p> <p>Query: 1 MPSSKLFSLFVVAIGVSFIGGVITTYFGILWAGGC 35 MPSSKLFSLFVVAIGVSFIGGVI TYFGILWAGGC Sbjct: 1 MPSSKLFSLFVVAIGVSFIGGVIITYFGILWAGGC 35</p>	<40	40-50	50-80	80-200	>=200	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_36, function unknown, 38	74	2e-13	Quartz_37, function unknown, 35	72	4e-13	Nucci_36, function unknown, 35	72	4e-13	Mandalorian_36, function unknown, 35	72	4e-13	Zagie_68, function unknown, 37	34	0.14	Rowlf_Draft_66, function unknown, 37	34	0.14	Namago_68, function unknown, 37	34	0.14	Fizzles_67, function unknown, 38	34	0.14	Nike_68, function unknown, 37	32	0.55
<40	40-50	50-80	80-200	>=200																																
Sequences producing significant alignments:	Score (bits)	E Value																																		
Carostasia_Draft_36, function unknown, 38	74	2e-13																																		
Quartz_37, function unknown, 35	72	4e-13																																		
Nucci_36, function unknown, 35	72	4e-13																																		
Mandalorian_36, function unknown, 35	72	4e-13																																		
Zagie_68, function unknown, 37	34	0.14																																		
Rowlf_Draft_66, function unknown, 37	34	0.14																																		
Namago_68, function unknown, 37	34	0.14																																		
Fizzles_67, function unknown, 38	34	0.14																																		
Nike_68, function unknown, 37	32	0.55																																		

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

[illegible]

Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Oui</i>
DECISION:	<i>Membrane protein</i>



Student Gene Annotation Worksheet

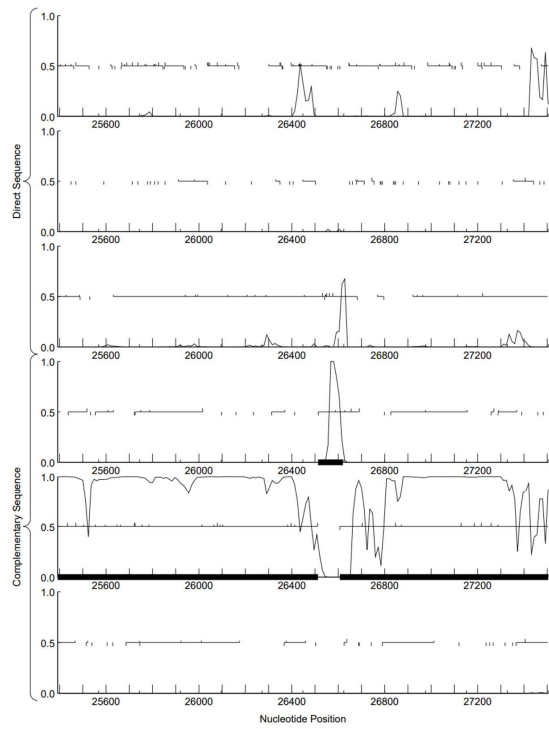
Carosmatique-gene-37

Basic Phage Information	
Nom du Phage	carostasia
Gène #	37
Coordonnées du Stop	26607
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Gap 204
Coordonnées du Start retenu	28508
Fonction prédite	DNA polymérase I

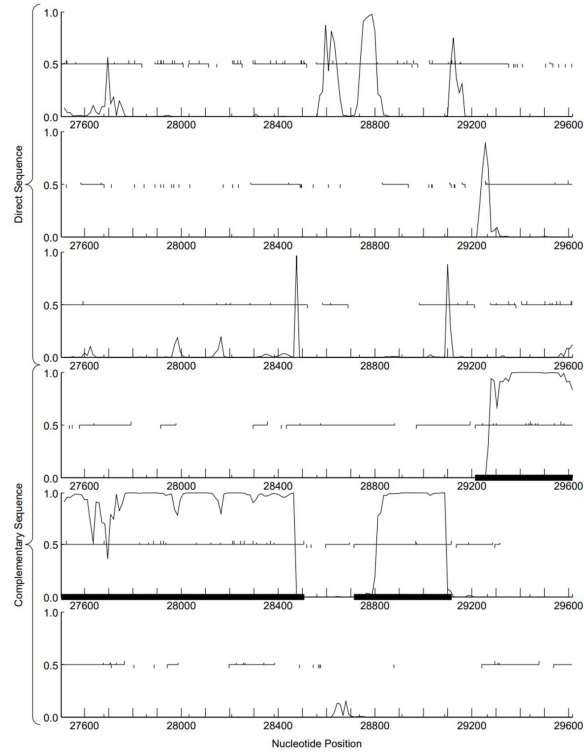
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	

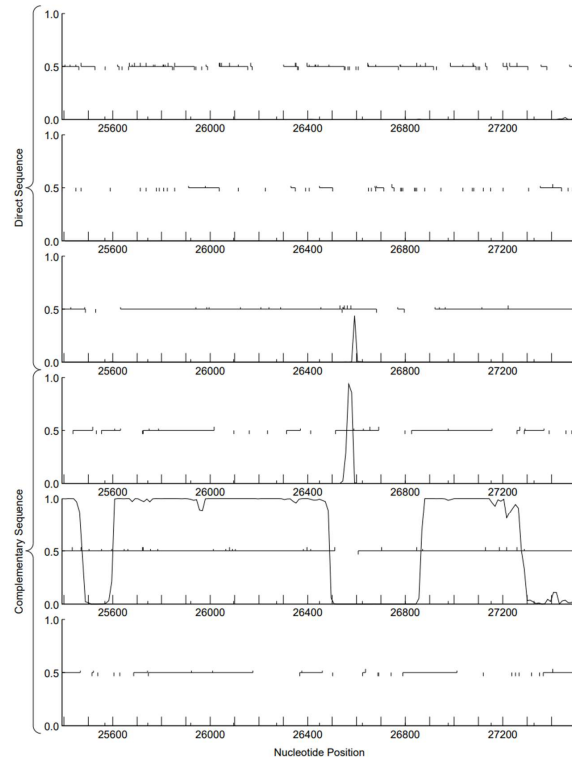
GenMark hmm prediction
cterium phage Carotass complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 14/21



GenMark ES-2 prediction
cterium phage Carotass complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 15/21



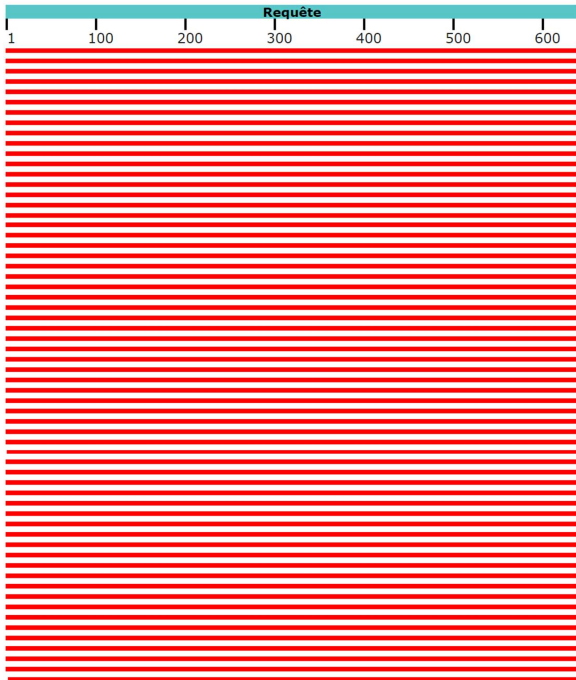
terium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 14/21



Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Oui : chevauchement >30bp</i>
DECISION:	<i>Oui</i>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																																																																																																													
Quel start Glimmer et GeneMark suggèrent-ils ?	28508																																																																																																																																																																																																																																																																																																																																													
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>ORF Length: 1311 J End: 1594 847 522 471</div> <table><thead><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr></thead><tbody><tr><td>1</td><td>-3,119</td><td>2,340</td><td>10</td><td>-3,314</td><td>TCTTTCACACTTAGGACTCCAC</td><td>ATG</td><td>28508</td><td>1502</td></tr><tr><td>2</td><td>-3,420</td><td>2,105</td><td>16</td><td>-5,416</td><td>CGGCAAGGGGCGGACTCTGTC</td><td>GTG</td><td>28395</td><td>1779</td></tr><tr><td>3</td><td>-4,827</td><td>1,640</td><td>13</td><td>-5,873</td><td>CTGTGCTGCTGGGCTTGGGCGAC</td><td>ATG</td><td>28370</td><td>1764</td></tr><tr><td>4</td><td>-3,642</td><td>2,055</td><td>13</td><td>-4,697</td><td>CGGCGACATCGACCGGACGAT</td><td>GTG</td><td>28355</td><td>1749</td></tr><tr><td>5</td><td>-5,339</td><td>1,301</td><td>5</td><td>-7,339</td><td>GTGGCGGCGGACCTTCGGGAG</td><td>TTG</td><td>28322</td><td>1716</td></tr><tr><td>6</td><td>-3,535</td><td>2,145</td><td>13</td><td>-4,581</td><td>CGACCTTGGGAGTTTCGGCAG</td><td>GTG</td><td>28313</td><td>1707</td></tr><tr><td>7</td><td>-2,654</td><td>2,559</td><td>16</td><td>-4,449</td><td>GGGCAAGTGGGACATCGGCGGC</td><td>ATG</td><td>28299</td><td>1692</td></tr><tr><td>8</td><td>-3,967</td><td>1,943</td><td>5</td><td>-5,967</td><td>CTGGCGCTTCGAGCGCGGCTGG</td><td>ATG</td><td>28262</td><td>1656</td></tr><tr><td>9</td><td>-3,594</td><td>2,122</td><td>17</td><td>-5,594</td><td>CGGTGGATGTGGACCTACGGC</td><td>ATG</td><td>28247</td><td>1641</td></tr><tr><td>10</td><td>-3,185</td><td>2,309</td><td>11</td><td>-3,942</td><td>GGTCAACATCTGGTGGGACTTC</td><td>ATG</td><td>28220</td><td>1614</td></tr><tr><td>11</td><td>-3,185</td><td>2,309</td><td>17</td><td>-5,185</td><td>CATCTGGTGGGACTCATGTTTC</td><td>ATG</td><td>28214</td><td>1608</td></tr><tr><td>12</td><td>-2,757</td><td>2,509</td><td>10</td><td>-3,452</td><td>TGTGGGCTCAAGGACTGGTTC</td><td>GTG</td><td>28163</td><td>1557</td></tr><tr><td>13</td><td>-4,174</td><td>1,846</td><td>14</td><td>-5,521</td><td>CCGGGACTGGGAGAGGACAC</td><td>GTG</td><td>28124</td><td>1518</td></tr><tr><td>14</td><td>-4,502</td><td>1,693</td><td>5</td><td>-6,502</td><td>CGACCGGACATCCCGAAGAG</td><td>GTG</td><td>28064</td><td>1458</td></tr><tr><td>15</td><td>-4,669</td><td>1,615</td><td>12</td><td>-5,503</td><td>GGCTGCTTCGGGCGCTGACCTC</td><td>GTG</td><td>27938</td><td>1332</td></tr><tr><td>16</td><td>-4,205</td><td>1,832</td><td>5</td><td>-6,205</td><td>CGGCGCTGACCGCTGTGGGGTTC</td><td>ATG</td><td>27929</td><td>1323</td></tr><tr><td>17</td><td>-3,779</td><td>2,031</td><td>8</td><td>-5,001</td><td>TGTGGGGCTCATGGGGACATC</td><td>ATG</td><td>27917</td><td>1311</td></tr><tr><td>18</td><td>-6,410</td><td>0,796</td><td>13</td><td>-7,464</td><td>GGGCGATCCAGCGGCTCACTGAG</td><td>ATG</td><td>27887</td><td>1281</td></tr><tr><td>19</td><td>-5,046</td><td>1,439</td><td>7</td><td>-6,569</td><td>GATGGAGACACAAAGCTGCC</td><td>GTG</td><td>27866</td><td>1260</td></tr><tr><td>20</td><td>-2,403</td><td>2,675</td><td>7</td><td>-3,526</td><td>GGTCAAGAGACTCAGGACAG</td><td>GTG</td><td>27830</td><td>1224</td></tr><tr><td>21</td><td>-3,779</td><td>2,031</td><td>7</td><td>-5,301</td><td>GCTCTTCGAGTACCGGGGCGCT</td><td>ATG</td><td>27693</td><td>1077</td></tr><tr><td>22</td><td>-3,264</td><td>2,272</td><td>7</td><td>-4,787</td><td>GAGCAAGACTTCCCGGAGGCT</td><td>GTG</td><td>27638</td><td>1032</td></tr><tr><td>23</td><td>-5,145</td><td>1,392</td><td>11</td><td>-5,902</td><td>GACCGGCTTCTGGAACCCATC</td><td>ATG</td><td>27527</td><td>921</td></tr><tr><td>24</td><td>-3,277</td><td>2,266</td><td>5</td><td>-5,277</td><td>TACCTGCTTGGGCGGCGGGG</td><td>TTG</td><td>27332</td><td>726</td></tr><tr><td>25</td><td>-2,791</td><td>2,493</td><td>10</td><td>-3,486</td><td>CTTCAGTCACTGGGACTTCGT</td><td>GTG</td><td>27290</td><td>684</td></tr><tr><td>26</td><td>-4,141</td><td>1,862</td><td>10</td><td>-4,836</td><td>GGGCTGGCCATGAGCGGACCC</td><td>ATG</td><td>27260</td><td>654</td></tr><tr><td>27</td><td>-3,079</td><td>2,358</td><td>16</td><td>-4,875</td><td>TGACGAGGACATCCATACCTAC</td><td>ATG</td><td>27218</td><td>612</td></tr><tr><td>28</td><td>-5,198</td><td>1,367</td><td>9</td><td>-5,972</td><td>GGGCTCATCGAGGCGGGGAG</td><td>ATG</td><td>27188</td><td>582</td></tr><tr><td>29</td><td>-1,951</td><td>2,887</td><td>13</td><td>-2,956</td><td>GATGGCGAAGGAGACCGGAGC</td><td>TTG</td><td>27167</td><td>561</td></tr><tr><td>30</td><td>-5,068</td><td>1,427</td><td>14</td><td>-6,415</td><td>CAACTTGGGTTCTGTCTACGGC</td><td>ATG</td><td>27131</td><td>525</td></tr><tr><td>31</td><td>-5,034</td><td>1,444</td><td>16</td><td>-6,830</td><td>CGGCGAGGCGCATCACTCCCGC</td><td>GTG</td><td>26973</td><td>267</td></tr><tr><td>32</td><td>-4,775</td><td>1,565</td><td>13</td><td>-5,820</td><td>GGATCCAGGGGCAAGGACTTC</td><td>ATG</td><td>26849</td><td>243</td></tr><tr><td>33</td><td>-5,748</td><td>1,110</td><td>8</td><td>-6,969</td><td>CTTGGCGCGAATAAGTGGGAC</td><td>TTG</td><td>26816</td><td>210</td></tr><tr><td>34</td><td>-4,543</td><td>1,673</td><td>10</td><td>-5,238</td><td>CGAATAAGTGGGACTTGGGC</td><td>TTG</td><td>26810</td><td>204</td></tr><tr><td>35</td><td>-1,748</td><td>2,981</td><td>7</td><td>-3,271</td><td>TGGCGGTATCTCAAGGAGACA</td><td>ATG</td><td>26705</td><td>99</td></tr></tbody></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,119	2,340	10	-3,314	TCTTTCACACTTAGGACTCCAC	ATG	28508	1502	2	-3,420	2,105	16	-5,416	CGGCAAGGGGCGGACTCTGTC	GTG	28395	1779	3	-4,827	1,640	13	-5,873	CTGTGCTGCTGGGCTTGGGCGAC	ATG	28370	1764	4	-3,642	2,055	13	-4,697	CGGCGACATCGACCGGACGAT	GTG	28355	1749	5	-5,339	1,301	5	-7,339	GTGGCGGCGGACCTTCGGGAG	TTG	28322	1716	6	-3,535	2,145	13	-4,581	CGACCTTGGGAGTTTCGGCAG	GTG	28313	1707	7	-2,654	2,559	16	-4,449	GGGCAAGTGGGACATCGGCGGC	ATG	28299	1692	8	-3,967	1,943	5	-5,967	CTGGCGCTTCGAGCGCGGCTGG	ATG	28262	1656	9	-3,594	2,122	17	-5,594	CGGTGGATGTGGACCTACGGC	ATG	28247	1641	10	-3,185	2,309	11	-3,942	GGTCAACATCTGGTGGGACTTC	ATG	28220	1614	11	-3,185	2,309	17	-5,185	CATCTGGTGGGACTCATGTTTC	ATG	28214	1608	12	-2,757	2,509	10	-3,452	TGTGGGCTCAAGGACTGGTTC	GTG	28163	1557	13	-4,174	1,846	14	-5,521	CCGGGACTGGGAGAGGACAC	GTG	28124	1518	14	-4,502	1,693	5	-6,502	CGACCGGACATCCCGAAGAG	GTG	28064	1458	15	-4,669	1,615	12	-5,503	GGCTGCTTCGGGCGCTGACCTC	GTG	27938	1332	16	-4,205	1,832	5	-6,205	CGGCGCTGACCGCTGTGGGGTTC	ATG	27929	1323	17	-3,779	2,031	8	-5,001	TGTGGGGCTCATGGGGACATC	ATG	27917	1311	18	-6,410	0,796	13	-7,464	GGGCGATCCAGCGGCTCACTGAG	ATG	27887	1281	19	-5,046	1,439	7	-6,569	GATGGAGACACAAAGCTGCC	GTG	27866	1260	20	-2,403	2,675	7	-3,526	GGTCAAGAGACTCAGGACAG	GTG	27830	1224	21	-3,779	2,031	7	-5,301	GCTCTTCGAGTACCGGGGCGCT	ATG	27693	1077	22	-3,264	2,272	7	-4,787	GAGCAAGACTTCCCGGAGGCT	GTG	27638	1032	23	-5,145	1,392	11	-5,902	GACCGGCTTCTGGAACCCATC	ATG	27527	921	24	-3,277	2,266	5	-5,277	TACCTGCTTGGGCGGCGGGG	TTG	27332	726	25	-2,791	2,493	10	-3,486	CTTCAGTCACTGGGACTTCGT	GTG	27290	684	26	-4,141	1,862	10	-4,836	GGGCTGGCCATGAGCGGACCC	ATG	27260	654	27	-3,079	2,358	16	-4,875	TGACGAGGACATCCATACCTAC	ATG	27218	612	28	-5,198	1,367	9	-5,972	GGGCTCATCGAGGCGGGGAG	ATG	27188	582	29	-1,951	2,887	13	-2,956	GATGGCGAAGGAGACCGGAGC	TTG	27167	561	30	-5,068	1,427	14	-6,415	CAACTTGGGTTCTGTCTACGGC	ATG	27131	525	31	-5,034	1,444	16	-6,830	CGGCGAGGCGCATCACTCCCGC	GTG	26973	267	32	-4,775	1,565	13	-5,820	GGATCCAGGGGCAAGGACTTC	ATG	26849	243	33	-5,748	1,110	8	-6,969	CTTGGCGCGAATAAGTGGGAC	TTG	26816	210	34	-4,543	1,673	10	-5,238	CGAATAAGTGGGACTTGGGC	TTG	26810	204	35	-1,748	2,981	7	-3,271	TGGCGGTATCTCAAGGAGACA	ATG	26705	99
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																																																																																																																																																						
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																																																																																																																																																						
1	-3,119	2,340	10	-3,314	TCTTTCACACTTAGGACTCCAC	ATG	28508	1502																																																																																																																																																																																																																																																																																																																																						
2	-3,420	2,105	16	-5,416	CGGCAAGGGGCGGACTCTGTC	GTG	28395	1779																																																																																																																																																																																																																																																																																																																																						
3	-4,827	1,640	13	-5,873	CTGTGCTGCTGGGCTTGGGCGAC	ATG	28370	1764																																																																																																																																																																																																																																																																																																																																						
4	-3,642	2,055	13	-4,697	CGGCGACATCGACCGGACGAT	GTG	28355	1749																																																																																																																																																																																																																																																																																																																																						
5	-5,339	1,301	5	-7,339	GTGGCGGCGGACCTTCGGGAG	TTG	28322	1716																																																																																																																																																																																																																																																																																																																																						
6	-3,535	2,145	13	-4,581	CGACCTTGGGAGTTTCGGCAG	GTG	28313	1707																																																																																																																																																																																																																																																																																																																																						
7	-2,654	2,559	16	-4,449	GGGCAAGTGGGACATCGGCGGC	ATG	28299	1692																																																																																																																																																																																																																																																																																																																																						
8	-3,967	1,943	5	-5,967	CTGGCGCTTCGAGCGCGGCTGG	ATG	28262	1656																																																																																																																																																																																																																																																																																																																																						
9	-3,594	2,122	17	-5,594	CGGTGGATGTGGACCTACGGC	ATG	28247	1641																																																																																																																																																																																																																																																																																																																																						
10	-3,185	2,309	11	-3,942	GGTCAACATCTGGTGGGACTTC	ATG	28220	1614																																																																																																																																																																																																																																																																																																																																						
11	-3,185	2,309	17	-5,185	CATCTGGTGGGACTCATGTTTC	ATG	28214	1608																																																																																																																																																																																																																																																																																																																																						
12	-2,757	2,509	10	-3,452	TGTGGGCTCAAGGACTGGTTC	GTG	28163	1557																																																																																																																																																																																																																																																																																																																																						
13	-4,174	1,846	14	-5,521	CCGGGACTGGGAGAGGACAC	GTG	28124	1518																																																																																																																																																																																																																																																																																																																																						
14	-4,502	1,693	5	-6,502	CGACCGGACATCCCGAAGAG	GTG	28064	1458																																																																																																																																																																																																																																																																																																																																						
15	-4,669	1,615	12	-5,503	GGCTGCTTCGGGCGCTGACCTC	GTG	27938	1332																																																																																																																																																																																																																																																																																																																																						
16	-4,205	1,832	5	-6,205	CGGCGCTGACCGCTGTGGGGTTC	ATG	27929	1323																																																																																																																																																																																																																																																																																																																																						
17	-3,779	2,031	8	-5,001	TGTGGGGCTCATGGGGACATC	ATG	27917	1311																																																																																																																																																																																																																																																																																																																																						
18	-6,410	0,796	13	-7,464	GGGCGATCCAGCGGCTCACTGAG	ATG	27887	1281																																																																																																																																																																																																																																																																																																																																						
19	-5,046	1,439	7	-6,569	GATGGAGACACAAAGCTGCC	GTG	27866	1260																																																																																																																																																																																																																																																																																																																																						
20	-2,403	2,675	7	-3,526	GGTCAAGAGACTCAGGACAG	GTG	27830	1224																																																																																																																																																																																																																																																																																																																																						
21	-3,779	2,031	7	-5,301	GCTCTTCGAGTACCGGGGCGCT	ATG	27693	1077																																																																																																																																																																																																																																																																																																																																						
22	-3,264	2,272	7	-4,787	GAGCAAGACTTCCCGGAGGCT	GTG	27638	1032																																																																																																																																																																																																																																																																																																																																						
23	-5,145	1,392	11	-5,902	GACCGGCTTCTGGAACCCATC	ATG	27527	921																																																																																																																																																																																																																																																																																																																																						
24	-3,277	2,266	5	-5,277	TACCTGCTTGGGCGGCGGGG	TTG	27332	726																																																																																																																																																																																																																																																																																																																																						
25	-2,791	2,493	10	-3,486	CTTCAGTCACTGGGACTTCGT	GTG	27290	684																																																																																																																																																																																																																																																																																																																																						
26	-4,141	1,862	10	-4,836	GGGCTGGCCATGAGCGGACCC	ATG	27260	654																																																																																																																																																																																																																																																																																																																																						
27	-3,079	2,358	16	-4,875	TGACGAGGACATCCATACCTAC	ATG	27218	612																																																																																																																																																																																																																																																																																																																																						
28	-5,198	1,367	9	-5,972	GGGCTCATCGAGGCGGGGAG	ATG	27188	582																																																																																																																																																																																																																																																																																																																																						
29	-1,951	2,887	13	-2,956	GATGGCGAAGGAGACCGGAGC	TTG	27167	561																																																																																																																																																																																																																																																																																																																																						
30	-5,068	1,427	14	-6,415	CAACTTGGGTTCTGTCTACGGC	ATG	27131	525																																																																																																																																																																																																																																																																																																																																						
31	-5,034	1,444	16	-6,830	CGGCGAGGCGCATCACTCCCGC	GTG	26973	267																																																																																																																																																																																																																																																																																																																																						
32	-4,775	1,565	13	-5,820	GGATCCAGGGGCAAGGACTTC	ATG	26849	243																																																																																																																																																																																																																																																																																																																																						
33	-5,748	1,110	8	-6,969	CTTGGCGCGAATAAGTGGGAC	TTG	26816	210																																																																																																																																																																																																																																																																																																																																						
34	-4,543	1,673	10	-5,238	CGAATAAGTGGGACTTGGGC	TTG	26810	204																																																																																																																																																																																																																																																																																																																																						
35	-1,748	2,981	7	-3,271	TGGCGGTATCTCAAGGAGACA	ATG	26705	99																																																																																																																																																																																																																																																																																																																																						
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Orf le plus long</p> <p>4^e meilleur score (-3,314)</p>																																																																																																																																																																																																																																																																																																																																													
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>28508: sample_31</p> <p>28508: sample_31 + 2</p> <p>28508: sample_31 + 2</p>																																																																																																																																																																																																																																																																																																																																													

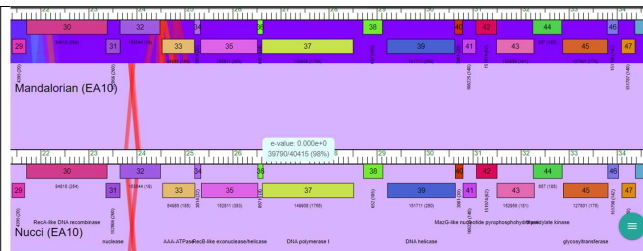
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>ADN polymérase I [Microbacterium phage Nucci]</p> <p>ID de séquence : QDK02430.1 Longueur : 633 Nombre de correspondances : 1</p> <p>Voir 1 titre(s) supplémentaire(s) ▼ Voir toutes les protéines identiques (IPG)</p> <p>Page 1 : 1 à 633 graphiques GenPept Prochain match Match précédent</p> <table><tr><td>Score</td><td>Attendre</td><td>Méthode</td><td>Identités</td><td>Points positifs</td><td>Lacunes</td></tr><tr><td>1314 bits(3401)</td><td>0,0</td><td>Ajustement de la matrice de composition.</td><td>633/633(100%)</td><td>633/633(100%)</td><td>0/633(0%)</td></tr></table> <p>Requête 1 MQLIISDDLAEVQAFLDQSGAPVVVDIETTSLTVGKGRILCVGFGHMDRDDVLVWMPDRL 60 MQLIISDDLAEVQAFLDQSGAPVVVDIETTSLTVGKGRILCVGFGHMDRDDVLVWMPDRL Sbjct 1 MQLIISDDLAEVQAFLDQSGAPVVVDIETTSLTVGKGRILCVGFGHMDRDDVLVWMPDRL 60</p> 	Score	Attendre	Méthode	Identités	Points positifs	Lacunes	1314 bits(3401)	0,0	Ajustement de la matrice de composition.	633/633(100%)	633/633(100%)	0/633(0%)
	Score	Attendre	Méthode	Identités	Points positifs	Lacunes							
1314 bits(3401)	0,0	Ajustement de la matrice de composition.	633/633(100%)	633/633(100%)	0/633(0%)								
DECISION:	<p>le start conservé est le 28508 car il n'est pas en contradiction avec les principes d'annotation. De plus, il induit l'ORF le plus long et est conservé chez les homologues</p>												

Décision #3 : Quelle est la fonction de la protéine putative ?

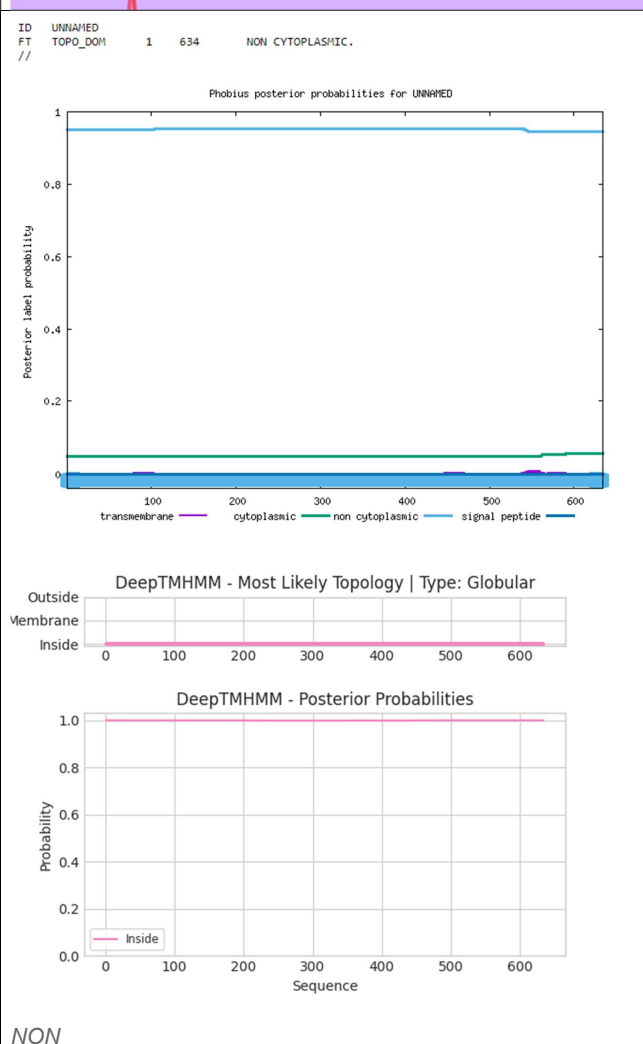
Collection des éléments de réponse	Rationnelle
------------------------------------	-------------

Oui, phage utilisé Mandalorian. Fonctions gènes voisins métabolisme ADN.

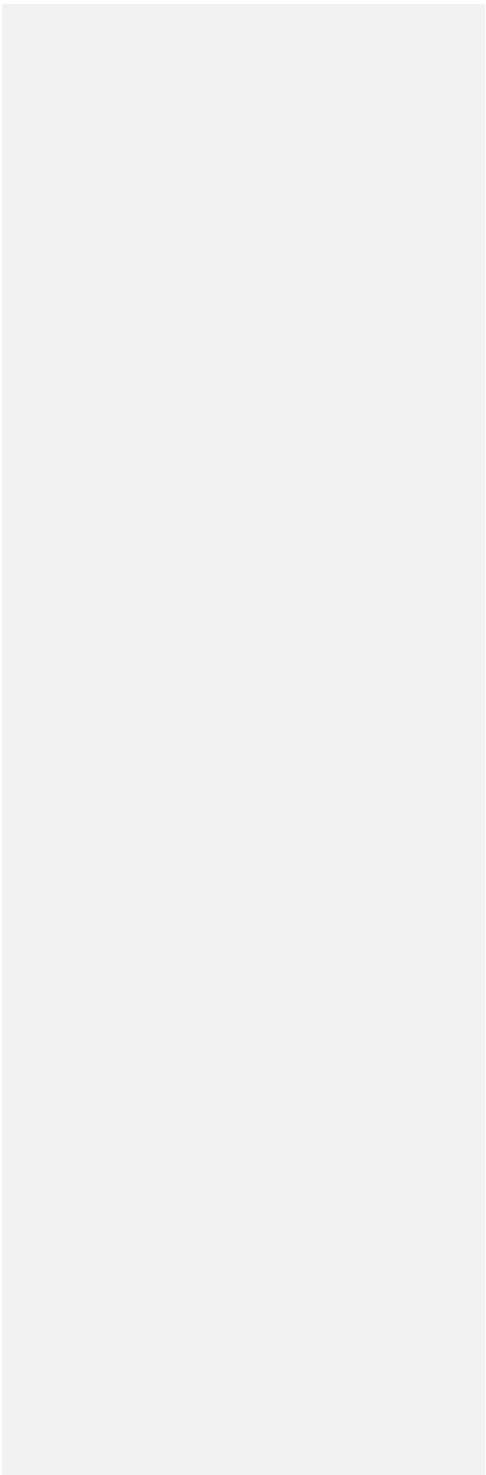
du génome qui montre une forte conservation de l'ordre des gènes ?



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



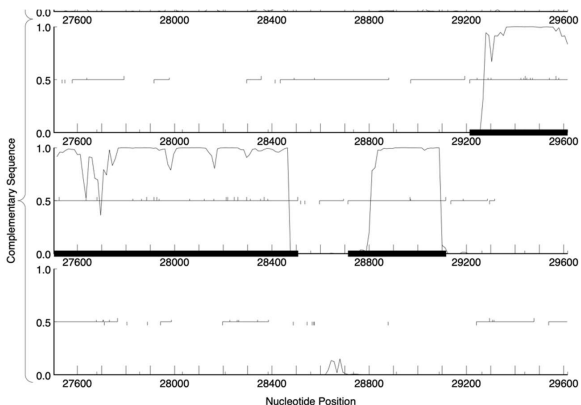
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Oui</i>
DECISION:	<i>DNA polymerase I</i>



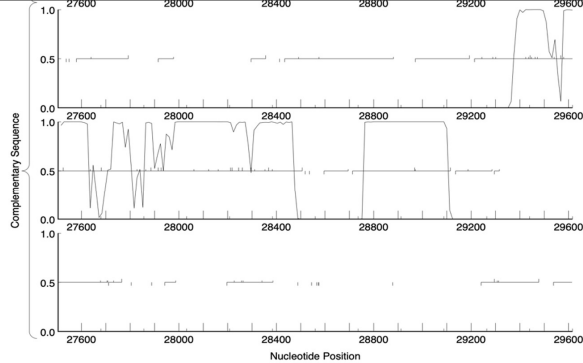
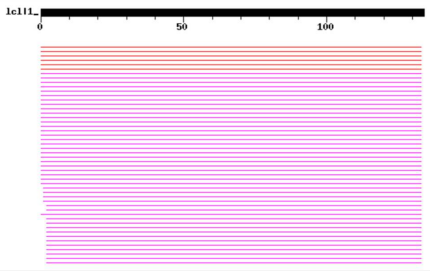
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	38
Coordonnées du Stop	28713
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	gap 95
Coordonnées du Start retenu	29117
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

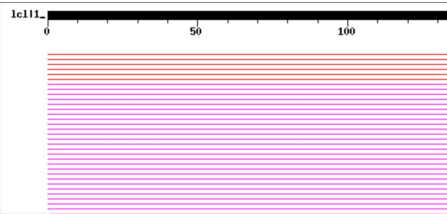
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	

Commenté [2]: page entière si possible

																																																							
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_38, function unknown, 134</td><td>261</td><td>6e-70</td></tr><tr><td>Mandalorian_38, function unknown, 134</td><td>258</td><td>5e-69</td></tr><tr><td>Quartz_39, function unknown, 134</td><td>256</td><td>1e-68</td></tr><tr><td>Nucci_38, function unknown, 130</td><td>247</td><td>6e-66</td></tr><tr><td>Yuyu_39, function unknown, 129</td><td>221</td><td>5e-58</td></tr><tr><td>Juanyo_38, function unknown, 132</td><td>221</td><td>5e-58</td></tr><tr><td>Willandstrong_39, function unknown, 136</td><td>192</td><td>3e-49</td></tr><tr><td>WestPM_Draft_36, function unknown, 138</td><td>186</td><td>1e-47</td></tr><tr><td>Tinyman4_38, function unknown, 138</td><td>181</td><td>4e-46</td></tr><tr><td>HerculesXII_39, function unknown, 138</td><td>181</td><td>4e-46</td></tr><tr><td>Schimmels22_38, function unknown, 138</td><td>179</td><td>3e-45</td></tr><tr><td>Luna18_41, function unknown, 140</td><td>174</td><td>7e-44</td></tr><tr><td>KatChar_41, function unknown, 140</td><td>174</td><td>7e-44</td></tr><tr><td>Chepli_41, function unknown, 140</td><td>174</td><td>7e-44</td></tr><tr><td>Finny_40, function unknown, 134</td><td>166</td><td>2e-41</td></tr><tr><td>Zenitau_39, function unknown, 134</td><td>164</td><td>5e-41</td></tr><tr><td>MCubed_39, function unknown, 134</td><td>164</td><td>5e-41</td></tr></table>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_38, function unknown, 134	261	6e-70	Mandalorian_38, function unknown, 134	258	5e-69	Quartz_39, function unknown, 134	256	1e-68	Nucci_38, function unknown, 130	247	6e-66	Yuyu_39, function unknown, 129	221	5e-58	Juanyo_38, function unknown, 132	221	5e-58	Willandstrong_39, function unknown, 136	192	3e-49	WestPM_Draft_36, function unknown, 138	186	1e-47	Tinyman4_38, function unknown, 138	181	4e-46	HerculesXII_39, function unknown, 138	181	4e-46	Schimmels22_38, function unknown, 138	179	3e-45	Luna18_41, function unknown, 140	174	7e-44	KatChar_41, function unknown, 140	174	7e-44	Chepli_41, function unknown, 140	174	7e-44	Finny_40, function unknown, 134	166	2e-41	Zenitau_39, function unknown, 134	164	5e-41	MCubed_39, function unknown, 134	164	5e-41
Sequences producing significant alignments:	Score (bits)	E Value																																																					
Carostasia_Draft_38, function unknown, 134	261	6e-70																																																					
Mandalorian_38, function unknown, 134	258	5e-69																																																					
Quartz_39, function unknown, 134	256	1e-68																																																					
Nucci_38, function unknown, 130	247	6e-66																																																					
Yuyu_39, function unknown, 129	221	5e-58																																																					
Juanyo_38, function unknown, 132	221	5e-58																																																					
Willandstrong_39, function unknown, 136	192	3e-49																																																					
WestPM_Draft_36, function unknown, 138	186	1e-47																																																					
Tinyman4_38, function unknown, 138	181	4e-46																																																					
HerculesXII_39, function unknown, 138	181	4e-46																																																					
Schimmels22_38, function unknown, 138	179	3e-45																																																					
Luna18_41, function unknown, 140	174	7e-44																																																					
KatChar_41, function unknown, 140	174	7e-44																																																					
Chepli_41, function unknown, 140	174	7e-44																																																					
Finny_40, function unknown, 134	166	2e-41																																																					
Zenitau_39, function unknown, 134	164	5e-41																																																					
MCubed_39, function unknown, 134	164	5e-41																																																					
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Pas de chevauchement, espaces de 46 avec la séquence codante suivante (qui est un ARN), potentiel codant prédit par Gene Mark, ORF =405 pb., 96pb entre le codon START de ce gene et le codon STOP de celui d'avant,</i>																																																						
DECISION:	<i>oui</i>																																																						

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																													
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 29117 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 29117																																													
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	ATG 29117 Z value = 2,981 Final score = -2,443 <div>Starts : 3 ORF Start : 28970 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler6 Explore Selected : 1 ORF Stop : 28713 5' End 56.3 52.1 47.9 144 Spacing Weight Matrix Karlin Medium Document ORF Length: 258 3' End 100.0 0.0 50.0 6</div> <table><tr><td>Sta</td><td>Raw SD</td><td>Genomic</td><td>Spacer</td><td>Final</td><td>Sequence of the Region</td><td>Start</td><td>Start</td><td>ORF</td></tr><tr><td>#</td><td>Score</td><td>Z Value</td><td>Distance</td><td>Score</td><td>Upstream of the Start</td><td>Codon</td><td>Position</td><td>Length</td></tr><tr><td>1</td><td>-1,748</td><td>2,981</td><td>10</td><td>-2,443</td><td>GTAACTCATCAAGAGACATC</td><td>ATG</td><td>29117</td><td>405</td></tr><tr><td>2</td><td>-5,308</td><td>1,315</td><td>16</td><td>-7,104</td><td>CCTCCAGCAGCCCGCTACAAG</td><td>GTC</td><td>28973</td><td>261</td></tr><tr><td>3</td><td>-5,302</td><td>1,315</td><td>5</td><td>-7,302</td><td>CCAGCAGCCCGCTACAAGGTC</td><td>ATG</td><td>28970</td><td>258</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,748	2,981	10	-2,443	GTAACTCATCAAGAGACATC	ATG	29117	405	2	-5,308	1,315	16	-7,104	CCTCCAGCAGCCCGCTACAAG	GTC	28973	261	3	-5,302	1,315	5	-7,302	CCAGCAGCCCGCTACAAGGTC	ATG	28970	258
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																						
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																						
1	-1,748	2,981	10	-2,443	GTAACTCATCAAGAGACATC	ATG	29117	405																																						
2	-5,308	1,315	16	-7,104	CCTCCAGCAGCCCGCTACAAG	GTC	28973	261																																						
3	-5,302	1,315	5	-7,302	CCAGCAGCCCGCTACAAGGTC	ATG	28970	258																																						
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le plus long ORF oui (405 pb) Distance avec le STOP en amont = 2pb Potentiel codant démarre au START 29117																																													
Est-ce que le start est conservé chez les homologues voir Starterator ?	Vous devrez également fournir les informations suivantes à partir de Starterator: le start correspond-il au start consensus prédit par Starterator ? OUI																																													
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Oui le start est conserve avec (au moins) les homologues de la capture d'écran :																																													



Query:	1	MARSTKAAVLEETDEVEEVEQTPKQINEGISRFIDAGGLDVQHARYKVMRAIYAQFVN	60
Query:	1	MARSTKAAVLEETDEVEDE EQTPKQINEGISRFIDAGGLDVQHARYKVMRAIYAQFVN	60
Sbjct:	1	MARSTKAAVLEETDEVEAEQTPKQINEGISRFIDAGGLDVQHARYKVMRAIYAQFVN	60
Query:	61	AIEETGTFDELVDIAINAGDLPSGWELERVAKDEAPAKPATKAPAKTAKKVPKAKA	120
Sbjct:	61	AIEETGTFDELVDIAINAGDLPSGWELERVAKDEAPAKPATKAPAKTAKKPAKAAKAKA	120
Query:	121	KAAAKPAARRKRPT	134
Sbjct:	121	KAAAKPAARRKRPT	134

```
>Quartz_39, function unknown, 134
      Length = 134
```

Score = 256 bits (655), Expect = 1e-68
Identities = 132/134 (98%), Positives = 132/134 (98%)

Query: 1 MARSTKAAVLEETVDEVEVEQTPKQIINEGISRFLDAGLDVQHARYKVMRAIAYQAFVN 60
MARSTKAAVLEETVDEVE EQTPKQIINEGISRFLDAGLDVQHARYKVMRAIAYQAFVN
Sbjct: 1 MARSTKAAVLEETVDEVEGEQTPKQIINEGISRFLDAGLDVQHARYKVMRAIAYQAFVN 60

Query: 61 AIEEGTFDELVD EAIANAGDLP SGWELERVAKDEAPAKPAKATKPAKTAAKKVPAPAKA 120
 AIEEGTFDELVD EAIANAGDLP SGWELERVAKDEAPAKPAKATKPAKTAAKK PAKAPA
 Sbjct: 61 AIEEGTFDELVD EAIANAGDLP SGWELERVAKDEAPAKPAKATKPAKTAAKKAPAKAPA 120

```
Query: 121 KAAAKPAARKRPTR 134
      KAAAKPAARKRPTR
Sbjct: 121 KAAAKPAARKRPTR 134
```

```
>Nucci_38, function unknown, 130
      Length = 130
```

Score = 247 bits (631), Expect = 6e-66
Identities = 130/134 (97%), Positives = 130/134 (97%), Gaps = 4/134 (2%)

Query: 1 MARSTKAIVLEETVDEVEVEQTPKQIINEGISRFLDGLDVQHARYKVMRAIAYQAFVN 60
MARSTKAIVLEETVDEVEVEQTPKQIINEGISRFLDGLDVQHARYKVMRAIAYQAFVN
Sbjct: 1 MARSTKAIVLEETVDEVEVEQTPKQIINEGISRFLDGLDVQHARYKVMRAIAYQAFVN 60

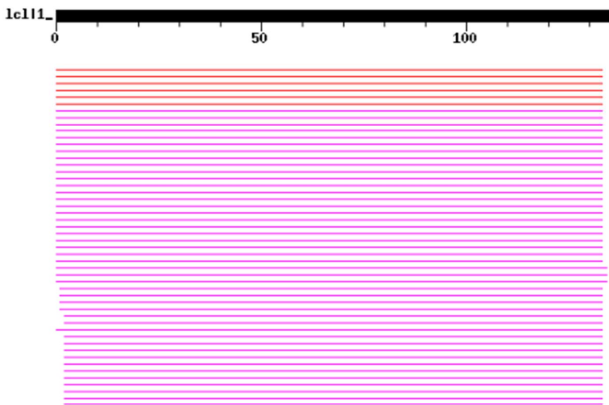
Query: 61 ATERCTENETVDEATANACNI.DSCWET.ERVAKNRADAKDAKATKADAKTAAKKUPAKADA 120

DECISION:

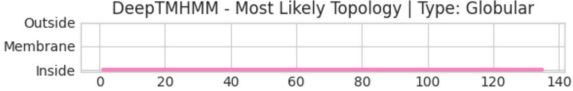
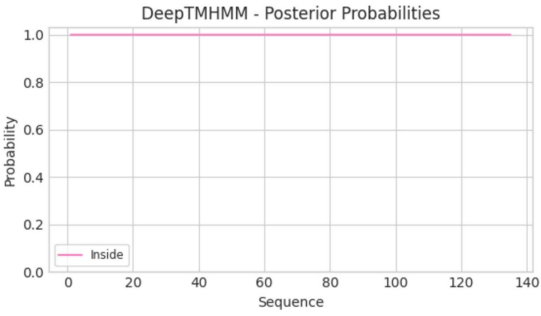
29117

Commenté [3]: montrer l'alignement si possible

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																													
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>Distribution of 100 Blast Hits on the Query Sequence</p> <div><div>Mouse-over to show define and scores. Click to show alignments</div><div><div>Color Key for Alignment Scores</div><div><div><40</div><div>40-50</div><div>50-80</div><div>80-200</div><div>>=200</div></div></div><p>Sequences producing significant alignments:</p><table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Carostasia_Draft_38, function unknown, 134</td><td>261</td><td>6e-70</td></tr><tr><td>Mandalorian_38, function unknown, 134</td><td>258</td><td>5e-69</td></tr><tr><td>Quartz_39, function unknown, 134</td><td>256</td><td>1e-68</td></tr><tr><td>Nucci_38, function unknown, 130</td><td>247</td><td>6e-66</td></tr><tr><td>YuuY_39, function unknown, 129</td><td>221</td><td>5e-58</td></tr><tr><td>Juanyo_38, function unknown, 132</td><td>221</td><td>5e-58</td></tr><tr><td>WilliamStrong_39, function unknown, 136</td><td>192</td><td>3e-49</td></tr><tr><td>WestPM_Draft_36, function unknown, 138</td><td>186</td><td>1e-47</td></tr><tr><td>Tinyman4_38, function unknown, 138</td><td>181</td><td>4e-46</td></tr><tr><td>HerculesXL_39, function unknown, 138</td><td>181</td><td>4e-46</td></tr><tr><td>Schimmels22_38, function unknown, 138</td><td>179</td><td>3e-45</td></tr><tr><td>Luna18_41, function unknown, 140</td><td>174</td><td>7e-44</td></tr><tr><td>KatChan_41, function unknown, 140</td><td>174</td><td>7e-44</td></tr><tr><td>Chepli_41, function unknown, 140</td><td>174</td><td>7e-44</td></tr></tbody></table><p>>Mandalorian_38, function unknown, 134 Length = 134</p><p>Score = 258 bits (658), Expect = 5e-69 Identities = 132/134 (98%), Positives = 132/134 (98%)</p><p>Query: 1 HARSTKAHVLEETVDEVEVEQTPKQIINEGISRFLDGAGLDVQHARYKVMRAIAYQAFVN 60 HARSTKAHVLEETVDEVE EQTPKQIINEGISRFLDGAGLDVQHARYKVMRAIAYQAFVN Sbjct: 1 HARSTKAHVLEETVDEVEAEQTPKQIINEGISRFLDGAGLDVQHARYKVMRAIAYQAFVN 60</p><p>Query: 61 AIEEGTFDELVDIAIANAGDLPNGMELERVAKDEAPAKPAKATKAPAKTAACKVPAPAKA 120 AIEEGTFDELVDIAIANAGDLPNGMELERVAKDEAPAKPAKATKAPAKTAACK PAKAPA Sbjct: 61 AIEEGTFDELVDIAIANAGDLPNGMELERVAKDEAPAKPAKATKAPAKTAACKPAKAPA 120</p><p>Query: 121 KAAAKPAARKRPTR 134 KAAAKPAARKRPTR Sbjct: 121 KAAAKPAARKRPTR 134</p></div>		Score (bits)	E Value	Carostasia_Draft_38, function unknown, 134	261	6e-70	Mandalorian_38, function unknown, 134	258	5e-69	Quartz_39, function unknown, 134	256	1e-68	Nucci_38, function unknown, 130	247	6e-66	YuuY_39, function unknown, 129	221	5e-58	Juanyo_38, function unknown, 132	221	5e-58	WilliamStrong_39, function unknown, 136	192	3e-49	WestPM_Draft_36, function unknown, 138	186	1e-47	Tinyman4_38, function unknown, 138	181	4e-46	HerculesXL_39, function unknown, 138	181	4e-46	Schimmels22_38, function unknown, 138	179	3e-45	Luna18_41, function unknown, 140	174	7e-44	KatChan_41, function unknown, 140	174	7e-44	Chepli_41, function unknown, 140	174	7e-44
	Score (bits)	E Value																																												
Carostasia_Draft_38, function unknown, 134	261	6e-70																																												
Mandalorian_38, function unknown, 134	258	5e-69																																												
Quartz_39, function unknown, 134	256	1e-68																																												
Nucci_38, function unknown, 130	247	6e-66																																												
YuuY_39, function unknown, 129	221	5e-58																																												
Juanyo_38, function unknown, 132	221	5e-58																																												
WilliamStrong_39, function unknown, 136	192	3e-49																																												
WestPM_Draft_36, function unknown, 138	186	1e-47																																												
Tinyman4_38, function unknown, 138	181	4e-46																																												
HerculesXL_39, function unknown, 138	181	4e-46																																												
Schimmels22_38, function unknown, 138	179	3e-45																																												
Luna18_41, function unknown, 140	174	7e-44																																												
KatChan_41, function unknown, 140	174	7e-44																																												
Chepli_41, function unknown, 140	174	7e-44																																												

[illegible]

	<div><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p><p><i>Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebeius par ex).</i></p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Non</i>
DECISION:	<i>NFK</i>

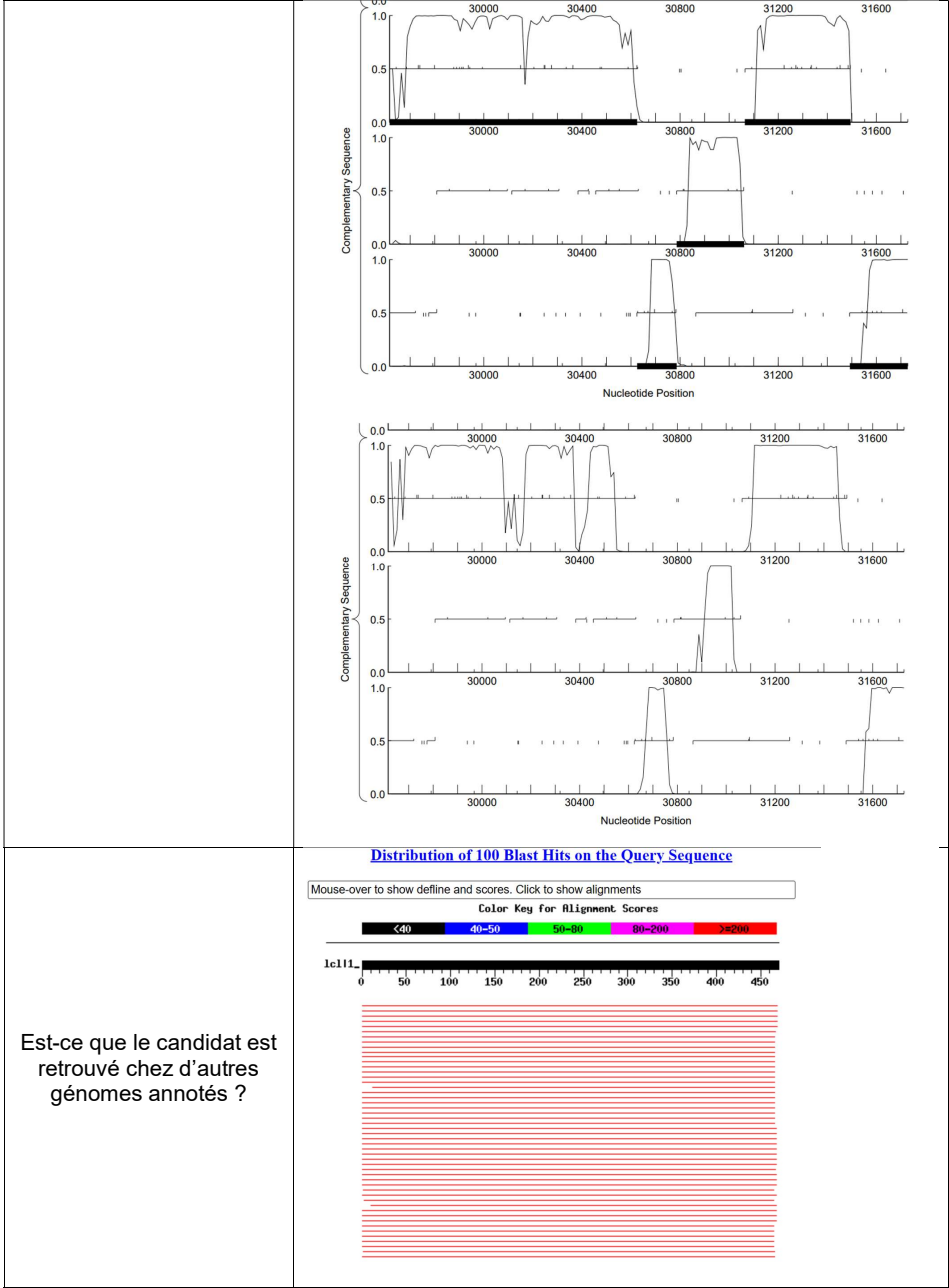
Student Gene Annotation Worksheet

Carosmatique-gene-39

Basic Phage Information	
Nom du Phage	carostasia
Gène #	39
Coordonnées du Stop	29213
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	chev 1
Coordonnées du Start retenu	30625
Fonction prédite	DNA helicase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	



Séquences produisant des alignements significatifs									
Télécharger Sélectionner des colonnes Montrer 100									
tout sélectionner 100 séquences sélectionnées									
GenPest Grambow Arbre de résultats de distance Alignement multiple Visionneuse MS									
	Description	Nom scientifique	Note maximale	Score total	Couverture de la requête	Valeur E	Par. identifier	Acc. Lect.	Accession
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Neigel]	Phage_micobacterien_Neigel	969	969	100%	0.0	100.00%	470	GD0202432.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Montadiner]	Phage_micobacterien_Montadiner	967	967	100%	0.0	99.79%	470	QX013031.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Gault]	Phage_micobacterien_Gault	964	964	100%	0.0	99.36%	470	U8598299.1
<input checked="" type="checkbox"/>	ADN helicase [Phage_micobacterien_YsaU]	Phage_micobacterien_YsaU	819	819	100%	0.0	86.17%	456	QF995068.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Elbel]	Microbactérie shrevei Elbel	818	818	99%	0.0	84.15%	470	YP_009503078
<input checked="" type="checkbox"/>	helicase [shrevei_micobacterien_ChiKPhi]	Phage_micobacterien_ChiKPhi	816	816	99%	0.0	83.84%	470	WN965834.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Saratz]	Phage_micobacterien_Saratz	816	816	99%	0.0	83.84%	470	AXH48812.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Saratz]	Phage_micobacterien_Saratz	816	816	99%	0.0	83.84%	470	LDX078998.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei MCubed]	Phage_micobacterien_MCubed	813	813	99%	0.0	83.73%	470	QD003700.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Juanelo]	Phage_micobacterien_Juanelo	812	812	99%	0.0	86.57%	456	US844788.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Finn]	Phage_micobacterien_Finn	797	797	96%	0.0	83.96%	458	QOF15789.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Zeop]	Phage_micobacterien_Zeop	724	724	99%	0.0	74.42%	475	QJG58128.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Hassila]	Phage_micobacterien_Hassila	722	722	99%	0.0	74.21%	475	QZ095168.1
<input checked="" type="checkbox"/>	ADN helicase [Microbacterium shrevei Zeyuli]	Phage_micobacterien_Zeyuli	721	721	99%	0.0	74.21%	475	QOP65236.1

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse

Quel start Glimmer et GeneMark suggèrent-ils ?

Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?

30625

Starts: 42
Selected: 1

ORF Start : 30337
ORF Stop : 29213
ORF Length: 1125

Cdn1 Cdn2 Cdn3 Length
100.0 0.0 100.0 3
69.1 49.5 71.1 291

SD Scoring Matrix
Spacing Weight Matrix

Kibler6
Karlin Medium

Explore
Document

Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length
1	-3,766	2,037	9	-4,541	TCTGGCAGCGGAGCGGGCGTGT	GTG	30628	1416
2	-3,766	2,037	12	-4,602	GGCAGCGGAGCGGGCGGTGTGTG	ATG	30625	1413
3	-4,228	1,821	8	-5,450	ACGCGAGTGCAGCGGGGTAGCC	TTG	30595	1393
4	-4,228	1,821	14	-5,575	GTGCGAGCGGGTAGCGTTGAAG	GTG	30589	1377
5	-4,705	1,597	18	-7,006	ACCTGGTGCAGCTTCGCTGCC	TTG	30556	1344
6	-3,821	2,011	7	-5,344	GCACCAACAGCCCGAGGTCCGC	GTG	30491	1269
7	-3,821	2,011	13	-4,866	CAAGCCCGAGGTCCCGGTGATC	GTG	30475	1263
8	-5,472	1,239	16	-7,268	TAAAGAGTACCGCTCCCGCGGC	ATG	30364	1152
9	-6,201	0,897	10	-6,896	CGGGGGTACGAGCTGCTGTTTC	GTG	30337	1125
10	-5,184	1,373	10	-5,879	GTGGAGCGGCTGAAGCCCTTC	TTG	30286	1074
11	-5,738	1,114	8	-6,960	GCTGAAGCCCTTCTTGACAAG	ATG	30277	1065
12	-4,502	1,693	11	-5,259	CAAGCTGTGCAGCAACGAGGGC	ATG	30250	1038
13	-4,141	1,862	10	-4,836	GCTGTGCAGCAACGAGGGCATG	ATG	30247	1035
14	-4,796	1,555	14	-6,143	CATCATCAAGAGCGCCAGCAGC	GTG	30205	993
15	-2,669	2,551	17	-4,669	CAAGTGGAAGTACCGTCTCTGT	ATG	30151	939
16	-3,620	2,105	16	-5,416	CGTCAAGGGCTTCCCGGAGCTG	GTG	29995	783
17	-5,331	1,304	8	-6,553	CGCTACATCCAGCGGAACTGC	GTG	29944	732
18	-5,331	1,304	14	-6,678	CATCCAGCGGAACTGCTGAGC	ATG	29938	726
19	-5,918	1,030	9	-6,693	GATGGTGGTCTGCGGACGCG	GTG	29917	705
20	-3,967	1,943	7	-5,490	TGGTCTGGGCGGCGGTGCCG	GTG	29911	699
21	-3,967	1,943	10	-4,661	CCAGCGGTGGCGGTGGCGCGC	GTG	29902	690
22	-6,214	0,891	8	-7,436	GGTGGCGCGGTGGGTACAGC	GTG	29890	678
23	-6,201	0,897	13	-7,247	GGGTACAGCGGTGGCGAGAC	GTG	29878	666
24	-5,308	1,315	10	-6,003	GCAGTACGTGCATGACCCAG	ATG	29800	588
25	-6,070	0,959	14	-7,417	CCCGCTGAGCGGCTGCTCCGG	ATG	29740	528
26	-3,178	2,312	8	-4,400	GAGCGGCTGCTCGGATGCGC	ATG	29734	522
27	-3,942	1,954	16	-5,738	CGAGGAGGGGACTCGTTCAAG	GTG	29696	474
28	-4,174	1,846	7	-5,697	ADGGCTGGGCACTGGGCAGG	GTG	29644	432
29	-3,945	2,234	14	-4,692	GCATCTCTGGGATCAAGCTG	GTG	29581	369
30	-3,697	2,069	13	-4,742	GATCAAGCTGGTGGCGGTAC	ATG	29569	357
31	-3,620	2,105	13	-4,666	GCGGATCAAGGCGATCGGCAC	TTG	29545	333
32	-3,620	2,105	16	-5,416	GATCAAGGCGATCGGCACCTG	GTG	29542	330
33	-3,690	2,073	10	-4,384	CCAGAGGAGCGGATGCCCGC	GTG	29473	261
34	-6,946	0,549	11	-7,703	CCCGGATGCCCGGTGCTGCTG	GTG	29464	252
35	-3,697	2,069	16	-5,493	GCTGCTGGTGCAGCCCGGACC	GTG	29449	237
36	-4,070	1,895	11	-4,827	GGTGCAGCCCGGACCGTGGCG	ATG	29443	231
37	-5,074	1,425	10	-5,768	GCCCCGACCGTGGCGATGGCC	GTG	29437	225
38	-3,964	1,944	10	-4,659	GGCGATGGCGGTGGCATCTCG	GTG	29425	213
39	-4,817	1,545	10	-5,511	AGCCAGCCCGAGCGTCTGGTTC	TTG	29320	108
40	-5,570	1,192	12	-6,406	GTTCTTGTGTGGCGGTGGCACT	GTG	29302	90
41	-3,072	2,362	10	-3,766	CGTGGCACTGTGGATGAGGAC	GTG	29290	78
42	-4,259	1,787	7	-5,822	CCACAACATCTCCAGAGGTC	GTG	29245	33

Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?

Est-ce que le start est conservé chez les

Pas le meilleur score (2e) : -4,602
ORF le plus long avec un ATG

Pro2: dmmpc_30

Pro3: mmpc_30 + 2

homologues voir Starterator ?													
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div><div>DNA helicase [Microbacterium phage Nucci]</div><div>Sequence ID: QDK02432.1 Length: 470 Number of Matches: 1</div><div>Range 1: 1 to 470 GenPept Graphics ▼ Next Match ▲ Previous Match</div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>969 bits(2504)</td><td>0.0</td><td>Compositional matrix adjust.</td><td>470/470(100%)</td><td>470/470(100%)</td><td>0/470(0%)</td></tr></table><div><div>Query 1</div><div>MQLRRESQRVALKVLTEPGRTFAALWAEPRSGKTAVALKWIEHHKQVAVIVGPKIAEATW</div><div>Sbjct 1</div><div>MQLRRESQRVALKVLTEPGRTFAALWAEPRSGKTAVALKWIEHHKQVAVIVGPKIAEATW</div></div><div><div>Query 61</div><div>RTEAAKWLGITVRFPLTAGNEYPSPAMFRGTTLLFVNVDQFNTAPWKRLKPFLDKMAKL</div><div>Sbjct 61</div><div>RTEAAKWLGITVRFPLTAGNEYPSPAMFRGTTLLFVNVDQFNTAPWKRLKPFLDKMAKL</div></div><div><div>Query 121</div><div>CSNEGMMLLDESHIITKTPSSVQGRSTRPLAHKWYRLMTGTPTVNPQIDAVYGQWTF</div><div>Sbjct 121</div><div>CSNEGMMLLDESHIITKTPSSVQGRSTRPLAHKWYRLMTGTPTVNPQIDAVYGQWTF</div></div><div><div>Query 181</div><div>DPAIRNRMETARDFREYFGEWSTVKGPELVPRRQAEINAYIQPNVVTMVGPGQVPVPR</div><div>Sbjct 181</div><div>DPAIRNRMETARDFREYFGEWSTVKGPELVPRRQAEINAYIQPNVVTMVGPGQVPVPR</div></div><div><div>Query 241</div><div>RVRVTVPSDVL EAHKWLLKGVASFQDRQVYHDPKMERDDTRVAIGLNPLTRLLRMRLV</div><div>Sbjct 241</div><div>RVRVTVPSDVL EAHKWLLKGVASFQDRQVYHDPKMERDDTRVAIGLNPLTRLLRMRLV</div></div><div><div>Query 301</div><div>GGWLKDEGDSFTVPAARARLAALGRVLRCKGKTIACHTLWEIKLVRRYMRIGIGH</div><div>Sbjct 301</div><div>GGWLKDEGDSFTVPAARARLAALGRVLRCKGKTIACHTLWEIKLVRRYMRIGIGH</div></div><div><div>Query 361</div><div>LVITGATPKDRVIERFQTPDARVLLVQPRTVAMAVDISVAQDLINWTSDFNVTFFKQA</div><div>Sbjct 361</div><div>LVITGATPKDRVIERFQTPDARVLLVQPRTVAMAVDISVAQDLINWTSDFNVTFFKQA</div></div><div><div>Query 421</div><div>SDRIKLSPASPTVWFLCGRGTVDEDWTTLQDDHNLQKVKVRIKQSF</div><div>Sbjct 421</div><div>SDRIKLSPASPTVWFLCGRGTVDEDWTTLQDDHNLQKVKVRIKQSF</div></div></div>	Score	Expect	Method	Identities	Positives	Gaps	969 bits(2504)	0.0	Compositional matrix adjust.	470/470(100%)	470/470(100%)	0/470(0%)
Score	Expect	Method	Identities	Positives	Gaps								
969 bits(2504)	0.0	Compositional matrix adjust.	470/470(100%)	470/470(100%)	0/470(0%)								
DECISION:	Le start retenu est le 30625 car il n'est pas en contradiction avec les principes d'annotation et Il est conservé chez d'autres homologues.												

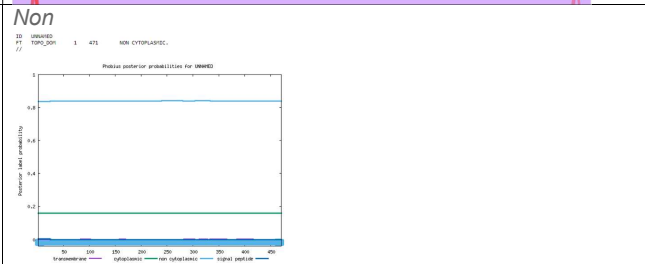
Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																	
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<div>Listez le meilleur hit Blastp pour chaque source :</div> <div>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</div> <div><table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Nucci_39, DNA helicase, 470</td><td>262</td><td>0.0</td></tr><tr><td>Cerostasia_Draft_39, Function unknown, 470</td><td>262</td><td>0.0</td></tr><tr><td>Mandalorian_39, DNA helicase, 470</td><td>261</td><td>0.0</td></tr><tr><td>Quartz_40, DNA helicase, 470</td><td>258</td><td>0.0</td></tr><tr><td>Yuuy_40, DNA helicase, 456</td><td>231</td><td>0.0</td></tr><tr><td>Juanyo_39, DNA helicase, 456</td><td>226</td><td>0.0</td></tr><tr><td>Shamu_41, DNA helicase, 470</td><td>212</td><td>0.0</td></tr><tr><td>Glamour_40, DNA helicase, 470</td><td>212</td><td>0.0</td></tr><tr><td>Eleri_40, DNA helicase, 470</td><td>212</td><td>0.0</td></tr><tr><td>ColaCorta_40, DNA helicase, 470</td><td>212</td><td>0.0</td></tr></table></div> <div>*** nr : (q#: s#) : [alignment] ; e-value :</div>	Sequences producing significant alignments:	Score (bits)	E Value	Nucci_39, DNA helicase, 470	262	0.0	Cerostasia_Draft_39, Function unknown, 470	262	0.0	Mandalorian_39, DNA helicase, 470	261	0.0	Quartz_40, DNA helicase, 470	258	0.0	Yuuy_40, DNA helicase, 456	231	0.0	Juanyo_39, DNA helicase, 456	226	0.0	Shamu_41, DNA helicase, 470	212	0.0	Glamour_40, DNA helicase, 470	212	0.0	Eleri_40, DNA helicase, 470	212	0.0	ColaCorta_40, DNA helicase, 470	212	0.0
Sequences producing significant alignments:	Score (bits)	E Value																																
Nucci_39, DNA helicase, 470	262	0.0																																
Cerostasia_Draft_39, Function unknown, 470	262	0.0																																
Mandalorian_39, DNA helicase, 470	261	0.0																																
Quartz_40, DNA helicase, 470	258	0.0																																
Yuuy_40, DNA helicase, 456	231	0.0																																
Juanyo_39, DNA helicase, 456	226	0.0																																
Shamu_41, DNA helicase, 470	212	0.0																																
Glamour_40, DNA helicase, 470	212	0.0																																
Eleri_40, DNA helicase, 470	212	0.0																																
ColaCorta_40, DNA helicase, 470	212	0.0																																

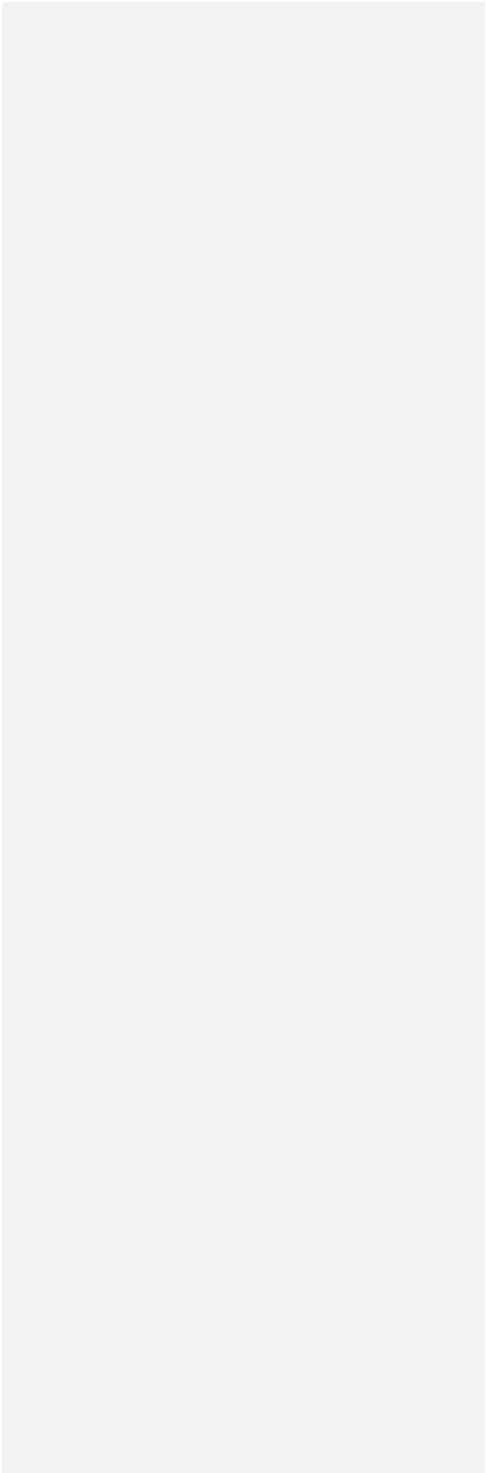
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



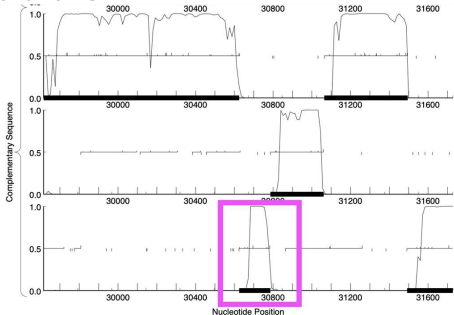
	<div><div>DeepTMHMM - Most Likely Topology Type: Globular</div><div><div>Outside</div><div>Membrane</div><div>Inside</div></div><div><div>0100200300400</div></div></div> <div><div>DeepTMHMM - Posterior Probabilities</div><div><div>1.0</div><div>0.8</div><div>0.6</div><div>0.4</div><div>0.2</div><div>0.0</div></div><div>Probability</div><div><div>0100200300400</div></div><div>Sequence</div><div><div>Inside</div></div></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui</i>
DECISION:	<i>DNA helicase</i>



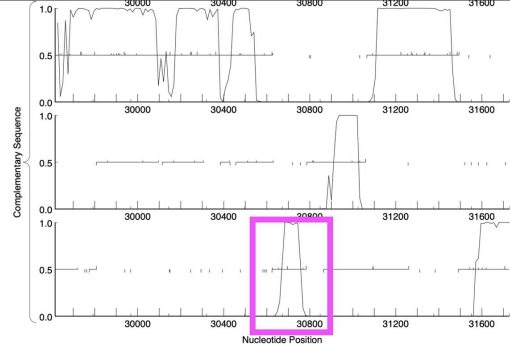
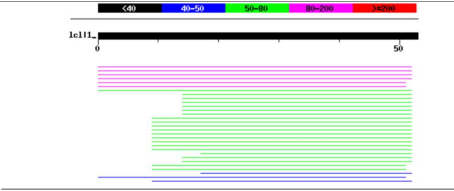
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	40
Coordonnées du Stop	30625
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	ove 1
Coordonnées du Start retenu	30786
Fonction prédite	NFK

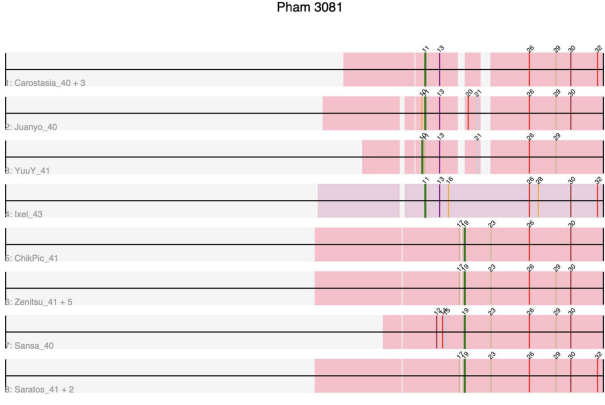
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GeneMarkS</i></p>  <p><i>Gene Mark host :</i></p>

Commenté [4]: page entière si possible

																																																																												
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 <p>Sequences producing significant alignments:</p> <table><tr><th></th><th>Score</th><th>E</th></tr><tr><th></th><th>(bits)</th><th>Value</th></tr><tr><td>Quartz_41, function unknown, 53</td><td>105</td><td>3e-23</td></tr><tr><td>Nucci_40, function unknown, 53</td><td>105</td><td>3e-23</td></tr><tr><td>Mandalorian_40, function unknown, 53</td><td>105</td><td>3e-23</td></tr><tr><td>Carostasia_Draft_40, function unknown, 53</td><td>105</td><td>3e-23</td></tr><tr><td>YuuY_41, function unknown, 54</td><td>96</td><td>4e-20</td></tr><tr><td>Juanyo_40, function unknown, 53</td><td>93</td><td>3e-19</td></tr><tr><td>Ixel_43, function unknown, 59</td><td>76</td><td>3e-14</td></tr><tr><td>Lucky3_40, function unknown, 46</td><td>62</td><td>7e-10</td></tr><tr><td>Golden_40, function unknown, 46</td><td>62</td><td>7e-10</td></tr><tr><td>Sinatra_41, function unknown, 46</td><td>61</td><td>1e-09</td></tr><tr><td>PrincePhergus_40, function unknown, 46</td><td>61</td><td>1e-09</td></tr><tr><td>Pherbot_40, function unknown, 47</td><td>61</td><td>1e-09</td></tr><tr><td>Bustleton_40, function unknown, 47</td><td>61</td><td>1e-09</td></tr><tr><td>Theresita_39, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Shamu_Draft_42, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Saratos_41, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Sansa_40, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Glamour_41, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Fimny_42, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Elteri_41, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>ColaCorta_41, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>ChikPic_41, function unknown, 46</td><td>60</td><td>2e-09</td></tr><tr><td>Andromedas_41, function unknown, 46</td><td>60</td><td>2e-09</td></tr></table> <p>On remarque que les e-values ne sont pas aussi basses que pour les gènes identifiés précédemment. De plus, on voit que il y a peu de hit qui couvrent en entier la séquence du gène potentiel.</p>		Score	E		(bits)	Value	Quartz_41, function unknown, 53	105	3e-23	Nucci_40, function unknown, 53	105	3e-23	Mandalorian_40, function unknown, 53	105	3e-23	Carostasia_Draft_40, function unknown, 53	105	3e-23	YuuY_41, function unknown, 54	96	4e-20	Juanyo_40, function unknown, 53	93	3e-19	Ixel_43, function unknown, 59	76	3e-14	Lucky3_40, function unknown, 46	62	7e-10	Golden_40, function unknown, 46	62	7e-10	Sinatra_41, function unknown, 46	61	1e-09	PrincePhergus_40, function unknown, 46	61	1e-09	Pherbot_40, function unknown, 47	61	1e-09	Bustleton_40, function unknown, 47	61	1e-09	Theresita_39, function unknown, 46	60	2e-09	Shamu_Draft_42, function unknown, 46	60	2e-09	Saratos_41, function unknown, 46	60	2e-09	Sansa_40, function unknown, 46	60	2e-09	Glamour_41, function unknown, 46	60	2e-09	Fimny_42, function unknown, 46	60	2e-09	Elteri_41, function unknown, 46	60	2e-09	ColaCorta_41, function unknown, 46	60	2e-09	ChikPic_41, function unknown, 46	60	2e-09	Andromedas_41, function unknown, 46	60	2e-09
	Score	E																																																																										
	(bits)	Value																																																																										
Quartz_41, function unknown, 53	105	3e-23																																																																										
Nucci_40, function unknown, 53	105	3e-23																																																																										
Mandalorian_40, function unknown, 53	105	3e-23																																																																										
Carostasia_Draft_40, function unknown, 53	105	3e-23																																																																										
YuuY_41, function unknown, 54	96	4e-20																																																																										
Juanyo_40, function unknown, 53	93	3e-19																																																																										
Ixel_43, function unknown, 59	76	3e-14																																																																										
Lucky3_40, function unknown, 46	62	7e-10																																																																										
Golden_40, function unknown, 46	62	7e-10																																																																										
Sinatra_41, function unknown, 46	61	1e-09																																																																										
PrincePhergus_40, function unknown, 46	61	1e-09																																																																										
Pherbot_40, function unknown, 47	61	1e-09																																																																										
Bustleton_40, function unknown, 47	61	1e-09																																																																										
Theresita_39, function unknown, 46	60	2e-09																																																																										
Shamu_Draft_42, function unknown, 46	60	2e-09																																																																										
Saratos_41, function unknown, 46	60	2e-09																																																																										
Sansa_40, function unknown, 46	60	2e-09																																																																										
Glamour_41, function unknown, 46	60	2e-09																																																																										
Fimny_42, function unknown, 46	60	2e-09																																																																										
Elteri_41, function unknown, 46	60	2e-09																																																																										
ColaCorta_41, function unknown, 46	60	2e-09																																																																										
ChikPic_41, function unknown, 46	60	2e-09																																																																										
Andromedas_41, function unknown, 46	60	2e-09																																																																										
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>Y a-t-il un chevauchement significatif avec d'autres gènes? Non, seulement de 1 base avec le gène d'avant et celui d'après</p> <p>Est-il suffisamment long ? 162 pb donc ca va c'est assez long</p> <p>Les gènes avant et après ce gène sont-ils dans la même direction ? oui</p>																																																																											
DECISION:	OUI																																																																											

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 30786 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 30786
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	ATG 30786 = (le meilleur score \square -2,523 ; la meilleure Z value \square 2,981 et l'ORF la plus longue \square 162)
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Oui l'ORF la plus longue (162) <i>On voit avec GeneMARK (et aussi gene MARK host) que le potentiel codant semble débuté à l'ATG 30786</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<i>Le gene 40 fait parti des : Genes that do not have the "Most Annotated" start</i> <i>Mais start vert = le start 30786 (? je crois ?)</i> 
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<i>Parmis les 6 meilleurs Hit au BLAST, 5 débute au même ATG.</i>

Commenté [5]: screen

Commenté [6]: screen

	<p>Query: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV Sbjct: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53</p> <p>>Nucci_40, function unknown, 53 Length = 53</p> <p>Score = 105 bits (263), Expect = 3e-23 Identities = 53/53 (100%), Positives = 53/53 (100%)</p> <p>Query: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV Sbjct: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53</p> <p>>Mandalorian_40, function unknown, 53 Length = 53</p> <p>Score = 105 bits (263), Expect = 3e-23 Identities = 53/53 (100%), Positives = 53/53 (100%)</p> <p>Query: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV Sbjct: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53</p> <p>>Carostasia_Draft_40, function unknown, 53 Length = 53</p> <p>Score = 105 bits (263), Expect = 3e-23 Identities = 53/53 (100%), Positives = 53/53 (100%)</p> <p>Query: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV Sbjct: 1 MDPDKVLATVSPALRASLERQLIQINRRNYMAALARLYQVDIRTVLIWHATGRV 53</p>
DECISION:	<p>30786 ORF la plus longue, chevauchement convenable, meilleur score retrouvé chez les hits au blast, ...</p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																					
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <p>Distribution of 30 Blast Hits on the Query Sequence</p> <div> <div>Mouse-over to show define and scores. Click to show alignments</div> <div> <p>Color Key for Alignment Scores</p> <table> <tr> <td><40</td> <td>40-50</td> <td>50-80</td> <td>80-200</td> <td>>=200</td> </tr> </table> </div> <table> <thead> <tr> <th>Sequences producing significant alignments:</th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr><td>Quartz_41, function unknown, 53</td><td>105</td><td>3e-23</td></tr> <tr><td>Nucci_40, function unknown, 53</td><td>105</td><td>3e-23</td></tr> <tr><td>Mandalorian_40, function unknown, 53</td><td>105</td><td>3e-23</td></tr> <tr><td>Carostasia_Draft_40, function unknown, 53</td><td>105</td><td>3e-23</td></tr> <tr><td>YuuY_41, function unknown, 54</td><td>96</td><td>4e-20</td></tr> <tr><td>Juanyo_40, function unknown, 53</td><td>93</td><td>3e-19</td></tr> <tr><td>Ixel_43, function unknown, 59</td><td>76</td><td>3e-14</td></tr> <tr><td>Lucky3_40, function unknown, 46</td><td>62</td><td>7e-10</td></tr> <tr><td>Golden_40, function unknown, 46</td><td>62</td><td>7e-10</td></tr> <tr><td>Sinatra_41, function unknown, 46</td><td>61</td><td>1e-09</td></tr> <tr><td>PrincePhergus_40, function unknown, 46</td><td>61</td><td>1e-09</td></tr> <tr><td>Pherbot_40, function unknown, 47</td><td>61</td><td>1e-09</td></tr> <tr><td>Bustleton_40, function unknown, 47</td><td>61</td><td>1e-09</td></tr> <tr><td>Theresita_39, function unknown, 46</td><td>60</td><td>2e-09</td></tr> <tr><td>Shamu_42, function unknown, 46</td><td>60</td><td>2e-09</td></tr> </tbody> </table> <p>>Quartz_41, function unknown, 53 Length = 53</p> <p>Score = 105 bits (263), Expect = 3e-23 Identities = 53/53 (100%), Positives = 53/53 (100%)</p> <p>Query: 1 MDPDKVLATVSPALRASLERQLIQINRNYMAALARLYQVDIRTVLIWHATGRV 53 MDPKVLATVSPALRASLERQLIQINRNYMAALARLYQVDIRTVLIWHATGRV Sbjct: 1 MDPDKVLATVSPALRASLERQLIQINRNYMAALARLYQVDIRTVLIWHATGRV 53</p> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p> </div>	<40	40-50	50-80	80-200	>=200	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_41, function unknown, 53	105	3e-23	Nucci_40, function unknown, 53	105	3e-23	Mandalorian_40, function unknown, 53	105	3e-23	Carostasia_Draft_40, function unknown, 53	105	3e-23	YuuY_41, function unknown, 54	96	4e-20	Juanyo_40, function unknown, 53	93	3e-19	Ixel_43, function unknown, 59	76	3e-14	Lucky3_40, function unknown, 46	62	7e-10	Golden_40, function unknown, 46	62	7e-10	Sinatra_41, function unknown, 46	61	1e-09	PrincePhergus_40, function unknown, 46	61	1e-09	Pherbot_40, function unknown, 47	61	1e-09	Bustleton_40, function unknown, 47	61	1e-09	Theresita_39, function unknown, 46	60	2e-09	Shamu_42, function unknown, 46	60	2e-09
<40	40-50	50-80	80-200	>=200																																																		
Sequences producing significant alignments:	Score (bits)	E Value																																																				
Quartz_41, function unknown, 53	105	3e-23																																																				
Nucci_40, function unknown, 53	105	3e-23																																																				
Mandalorian_40, function unknown, 53	105	3e-23																																																				
Carostasia_Draft_40, function unknown, 53	105	3e-23																																																				
YuuY_41, function unknown, 54	96	4e-20																																																				
Juanyo_40, function unknown, 53	93	3e-19																																																				
Ixel_43, function unknown, 59	76	3e-14																																																				
Lucky3_40, function unknown, 46	62	7e-10																																																				
Golden_40, function unknown, 46	62	7e-10																																																				
Sinatra_41, function unknown, 46	61	1e-09																																																				
PrincePhergus_40, function unknown, 46	61	1e-09																																																				
Pherbot_40, function unknown, 47	61	1e-09																																																				
Bustleton_40, function unknown, 47	61	1e-09																																																				
Theresita_39, function unknown, 46	60	2e-09																																																				
Shamu_42, function unknown, 46	60	2e-09																																																				

No putative conserved domains have been detected

Distribution of the top 14 Blast Hits on 14 subject sequences

	Description	Scientific Name	Max Score	Total Score	Query Cov	E value	Per Ident	Acc Len	Accession
<input checked="" type="checkbox"/>	hypothetical protein SEA_NUCCL_40 [Microbacterium phage Nucci]	Microbacterium phage Nucci	108	108	98%	3e-29	100.00%	53	Q2K02433.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_YUUV_41 [Microbacterium phage YuuV]	Microbacterium phage YuuV	97.8	97.8	96%	4e-25	90.38%	54	Q2E96369.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_AJUNYD_40 [Microbacterium phage Ajunyo]	Microbacterium phage Ajunyo	94.7	94.7	96%	6e-24	89.46%	53	U2H4299.1
<input checked="" type="checkbox"/>	hypothetical protein HSL_43 [Microbacterium phage hsl]	Microbacterium phage hsl	78.2	78.2	98%	3e-17	71.19%	59	Q2H82743.1
<input checked="" type="checkbox"/>	hypothetical protein FDUCL_40 [Microbacterium phage Golden]	Microbacterium phage Golden	63.2	63.2	81%	2e-11	68.18%	46	YP_009624181.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_PHERROT_41 [Microbacterium phage Pherot]	Microbacterium phage Pherot	61.6	61.6	81%	7e-11	68.18%	47	Q8Z73091.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_PINKEPHERGUS_41 [Microbacterium phage PincePhery]	Microbacterium phage PincePhery	61.6	61.6	81%	7e-11	68.18%	46	Q8Z73091.1

Download

[GenPept](#)

[Graphics](#)

hypothetical protein SEA_NUCCL_40 [Microbacterium phage Nucci]

Sequence ID: [Q2K02433.1](#) Length: 53 Number of Matches: 1

[See 2 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 53

[GenPept](#)

[Graphics](#)

▼ Next Match

▲ Previous Match

	Score	Expect	Method	Identities	Positives	Gaps
	108 bits(269)	3e-29	Compositional matrix adjust.	53/53(100%)	53/53(100%)	0/53(0%)

Query

1

MDPDKVLATVSPALRASLERQLIQINRNYMAALARLYQVDIRTVLIWHATGRV

53

Sbjct

1

MDPDKVLATVSPALRASLERQLIQINRNYMAALARLYQVDIRTVLIWHATGRV

53

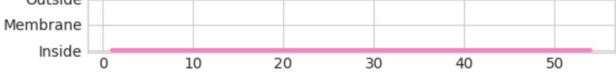
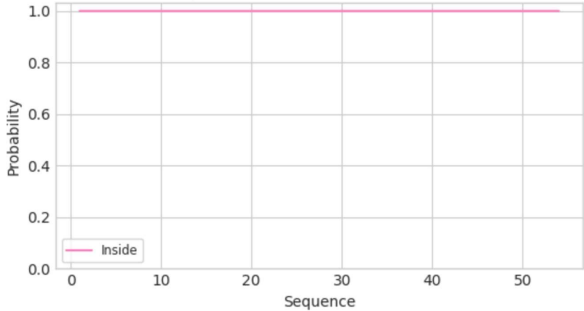
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

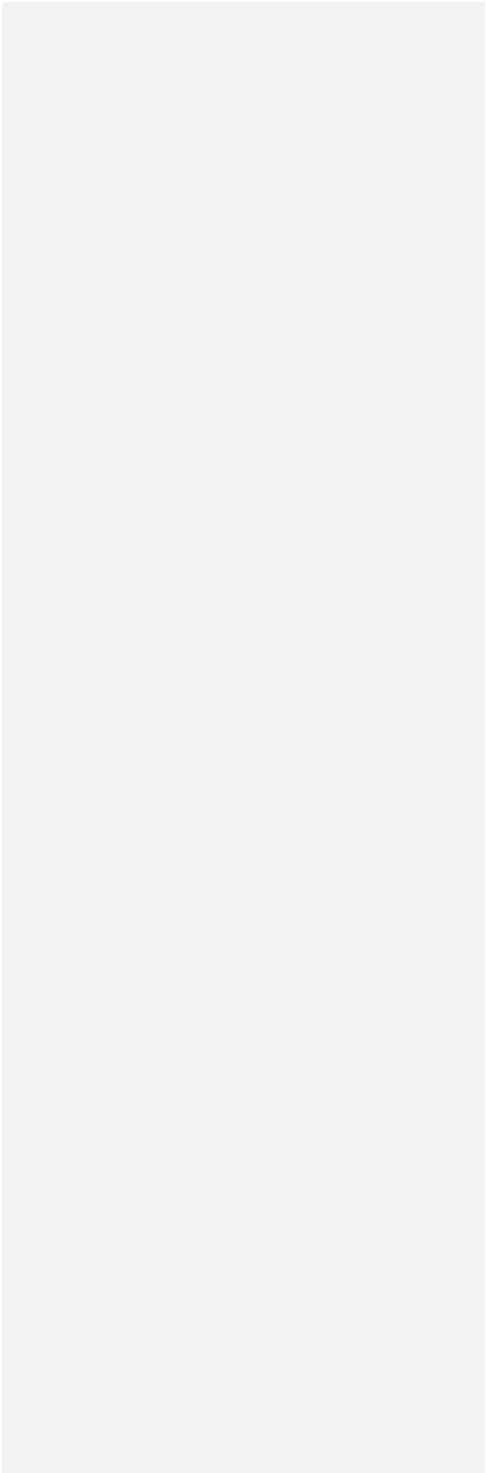
Number of Hits: 250

Query MSA diversity (Neff): 2.11124

Detected sequence features: **•**Called coil segment(s)

Visualization

	<div><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p><p><i>Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebeius par ex).</i></p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Non
DECISION:	NFK



Student Gene Annotation Worksheet

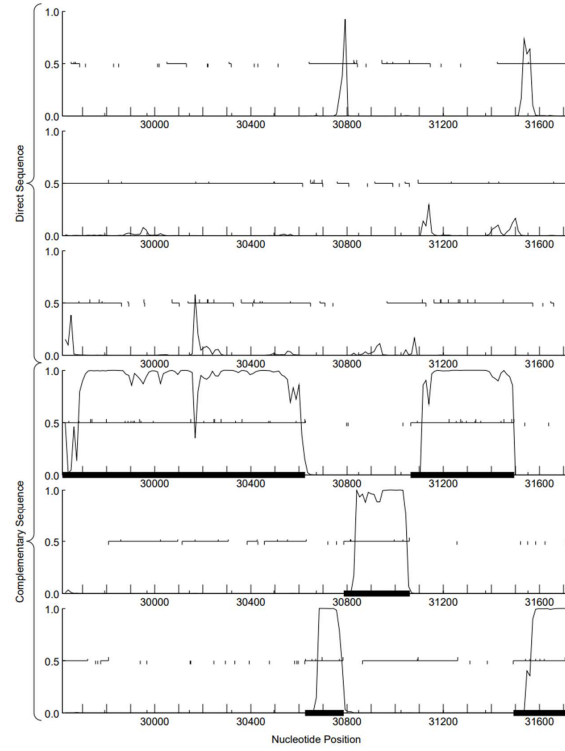
Carosmatique-gene-41

Basic Phage Information	
Nom du Phage	carostasia
Gène #	41
Coordonnées du Stop	30786
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Gap 2
Coordonnées du Start retenu	31061
Fonction prédite	NKF

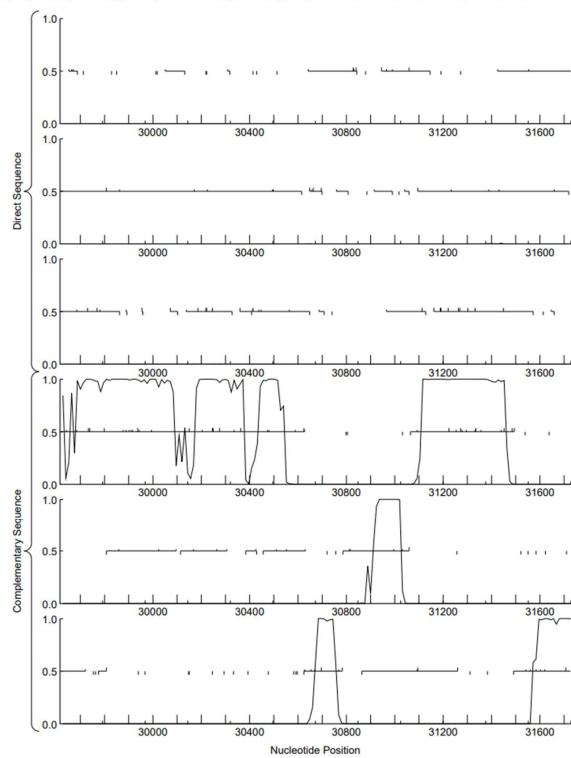
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>YES BOTH</i>
Y-a-t-il des éléments supportant un potentiel codant ?	

GenMark.hmm prediction
cterium phage Carotavirus complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 2, Window 96, Step 12, 16/21



terium phage Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 16/21



Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?

Sequences producing significant alignments:	Score (bits)	E Value
Nucci_41, function unknown, 91	187	6e-48
Mandalorian_41, function unknown, 91	187	6e-48
Carostasia_Draft_41, function unknown, 91	187	6e-48
Quartz_42, function unknown, 91	186	2e-47
Yuuy_42, function unknown, 92	103	1e-22
Juanyo_41, function unknown, 92	98	6e-21
Zada_45, function unknown, 92	84	2e-16
TatarkaPM_44, function unknown, 92	84	2e-16
Stanktossa_44, function unknown, 92	84	2e-16
SoilGremlin_Draft_45, function unknown, 92	84	2e-16
Shee_44, function unknown, 92	84	2e-16
Riyhil_44, function unknown, 92	84	2e-16
Rapheph_44, function unknown, 92	84	2e-16
Phrienda_44, function unknown, 92	84	2e-16
PhredPlintston_44, function unknown, 92	84	2e-16
Phireproof_44, function unknown, 92	84	2e-16
Phiderman_Draft_46, function unknown, 92	84	2e-16

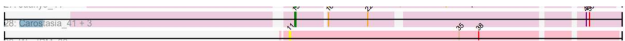
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

Non

DECISION:

oui

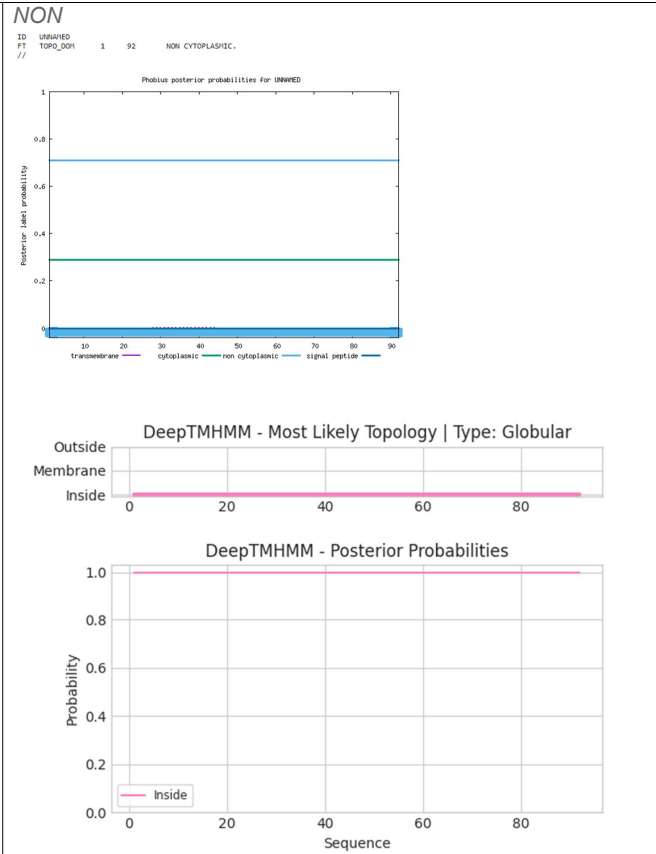
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																															
Quel start Glimmer et GeneMark suggèrent-ils ?	31061																																																															
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Stats: 5 Selected: 1</div><div>ORF Start : 30998 ORF Stop : 30786 ORF Length : 213</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 66.7 33.3 66.7 27 3' End 84.6 53.8 84.6 39</div><div>SD Scoring Matrix Kibler6 Spacing Weight Matrix Karlin Medium</div><div>Explore Document</div></div> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,867</td><td>1,990</td><td>9</td><td>-4,641</td><td>AGGATCTGGCCAGGCGTAATC</td><td>ATG</td><td>31061</td><td>276</td></tr><tr><td>2</td><td>-4,463</td><td>1,711</td><td>16</td><td>-6,259</td><td>CGGGAGCAGCCACTTCGGTTC</td><td>GTG</td><td>31034</td><td>249</td></tr><tr><td>3</td><td>-4,141</td><td>1,862</td><td>16</td><td>-5,937</td><td>CAGCAAGAAGTCCGAGACCTG</td><td>GTG</td><td>30998</td><td>213</td></tr><tr><td>4</td><td>-6,206</td><td>0,855</td><td>8</td><td>-7,428</td><td>ACGAAACACATCTCGTTCCGC</td><td>GTG</td><td>30818</td><td>33</td></tr><tr><td>5</td><td>-6,206</td><td>0,855</td><td>11</td><td>-6,963</td><td>AAACACATCTCGTTCCGGCTG</td><td>GTG</td><td>30815</td><td>30</td></tr></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,867	1,990	9	-4,641	AGGATCTGGCCAGGCGTAATC	ATG	31061	276	2	-4,463	1,711	16	-6,259	CGGGAGCAGCCACTTCGGTTC	GTG	31034	249	3	-4,141	1,862	16	-5,937	CAGCAAGAAGTCCGAGACCTG	GTG	30998	213	4	-6,206	0,855	8	-7,428	ACGAAACACATCTCGTTCCGC	GTG	30818	33	5	-6,206	0,855	11	-6,963	AAACACATCTCGTTCCGGCTG	GTG	30815	30
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																								
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																								
1	-3,867	1,990	9	-4,641	AGGATCTGGCCAGGCGTAATC	ATG	31061	276																																																								
2	-4,463	1,711	16	-6,259	CGGGAGCAGCCACTTCGGTTC	GTG	31034	249																																																								
3	-4,141	1,862	16	-5,937	CAGCAAGAAGTCCGAGACCTG	GTG	30998	213																																																								
4	-6,206	0,855	8	-7,428	ACGAAACACATCTCGTTCCGC	GTG	30818	33																																																								
5	-6,206	0,855	11	-6,963	AAACACATCTCGTTCCGGCTG	GTG	30815	30																																																								
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Plus long ORF avec un ATG et meilleur score																																																															
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div><div>Télécharger GenPept Graphique</div><div>Suivant Précédent Descriptions</div><div>protéine hypothétique SEA_NUCCI_41 [Microbacterium phage Nucci] ID de séquence : Q0K02434.1 Longueur : 91 Nombre de correspondances : 1 Voir 1 titre(s) supplémentaire(s) Voir toutes les protéines identiques (RPG)</div><div><div>Plage 1 : 1 à 91 graphiques GenPept</div><div><div>Score</div><div>190 bits(483)</div><div>Se-60</div><div>Ajustement de la matrice de composition</div><div>91/91(100%)</div><div>91/91(100%)</div><div>0/91(0%)</div></div><div><div>Requête 1</div><div>MFGEQPLPVVYVPSKSETLVVTEDLFGHTNGVQGLADKLAERARLERLLPDPPTG 68</div><div>MFGEQPLPVVYVPSKSETLVVTEDLFGHTNGVQGLADKLAERARLERLLPDPPTG 68</div><div>Subjet 1</div><div>MFGEQPLPVVYVPSKSETLVVTEDLFGHTNGVQGLADKLAERARLERLLPDPPTG 68</div></div><div><div>Requête 61</div><div>YCHEFQGSDDYIRNTTSFRVATLKLKEI 91</div><div>YCHEFQGSDDYIRNTTSFRVATLKLKEI 91</div><div>Subjet 61</div><div>YCHEFQGSDDYIRNTTSFRVATLKLKEI 91</div></div></div><div><div>Informations connexes</div><div>Protéines identiques - Protéines identiques à Q0K02434.1</div></div></div>																																																															
	<div><div>Télécharger GenPept Graphique</div><div>Suivant Précédent Descriptions</div><div>protéine hypothétique SEA_QUARTZ_42 [Microbacterium phage Quartz] ID de séquence : UWK9261.1 Longueur : 91 Nombre de correspondances : 1</div><div><div>Plage 1 : 1 à 91 graphiques GenPept</div><div><div>Score</div><div>188 bits(478)</div><div>Se-60</div><div>Ajustement de la matrice de composition</div><div>90/91(99%)</div><div>90/91(98%)</div><div>0/91(0%)</div></div><div><div>Requête 1</div><div>MFGEQPLPVVYVPSKSETLVVTEDLFGHTNGVQGLADKLAERARLERLLPDPPTG 68</div><div>MFGEQPLPVVYVPSKSETLVVTEDLFGHTNGVQGLADKLAERARLERLLPDPPTG 68</div><div>Subjet 1</div><div>MFGEQPLPVVYVPSKSETLVVTEDLFGHTNGVQGLADKLAERARLERLLPDPPTG 68</div></div><div><div>Requête 61</div><div>YCHEFQGSDDYIRNTTSFRVATLKLKEI 91</div><div>YCHEFQGSDDYIRNTTSFRVATLKLKEI 91</div><div>Subjet 61</div><div>YCHEFQGSDDYIRNTTSFRVATLKLKEI 91</div></div></div></div>																																																															
DECISION:	Le start conservé est le 31061 car il n'est pas en contradiction avec les principes d'annotation et est conservé chez les homologues. De plus, il conduit au plus long ORF avec un ATG et meilleur score																																																															

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																							
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Nucci_41, function unknown, 91</td><td>187</td><td>6e-48</td></tr><tr><td>Mandalorian_41, function unknown, 91</td><td>187</td><td>6e-48</td></tr><tr><td>Carostasia_Draft_41, function unknown, 91</td><td>187</td><td>6e-48</td></tr><tr><td>Quartz_42, function unknown, 91</td><td>186</td><td>2e-47</td></tr></table> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p> <table><tr><th></th><th>GenProt</th><th>Graphics</th><th>Distance tree of results</th><th>Multiple alignment</th><th>MSA Viewer</th></tr><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_NUCCI_41 [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>190</td><td>190</td><td>98%</td><td>1e-60</td><td>100.00%</td><td>91</td><td>GD252434.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_QUARTZ_42 [Microbacterium phage Quartz]</td><td>Microbacterium phage Quartz</td><td>188</td><td>188</td><td>98%</td><td>5e-60</td><td>98.90%</td><td>91</td><td>LVX59261.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_YUZZY_42 [Microbacterium phage Yuzz]</td><td>Microbacterium phage Yuzz</td><td>105</td><td>105</td><td>94%</td><td>5e-27</td><td>89.00%</td><td>92</td><td>QEP56378.1</td></tr><tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein SEA_JUANITO_41 [Microbacterium phage Juanito]</td><td>Microbacterium phage Juanito</td><td>99.0</td><td>99.0</td><td>94%</td><td>2e-24</td><td>85.56%</td><td>92</td><td>US644792.1</td></tr></table>	Sequences producing significant alignments:	Score (bits)	E Value	Nucci_41, function unknown, 91	187	6e-48	Mandalorian_41, function unknown, 91	187	6e-48	Carostasia_Draft_41, function unknown, 91	187	6e-48	Quartz_42, function unknown, 91	186	2e-47		GenProt	Graphics	Distance tree of results	Multiple alignment	MSA Viewer		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/>	hypothetical protein SEA_NUCCI_41 [Microbacterium phage Nucci]	Microbacterium phage Nucci	190	190	98%	1e-60	100.00%	91	GD252434.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_QUARTZ_42 [Microbacterium phage Quartz]	Microbacterium phage Quartz	188	188	98%	5e-60	98.90%	91	LVX59261.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_YUZZY_42 [Microbacterium phage Yuzz]	Microbacterium phage Yuzz	105	105	94%	5e-27	89.00%	92	QEP56378.1	<input checked="" type="checkbox"/>	hypothetical protein SEA_JUANITO_41 [Microbacterium phage Juanito]	Microbacterium phage Juanito	99.0	99.0	94%	2e-24	85.56%	92	US644792.1
Sequences producing significant alignments:	Score (bits)	E Value																																																																						
Nucci_41, function unknown, 91	187	6e-48																																																																						
Mandalorian_41, function unknown, 91	187	6e-48																																																																						
Carostasia_Draft_41, function unknown, 91	187	6e-48																																																																						
Quartz_42, function unknown, 91	186	2e-47																																																																						
	GenProt	Graphics	Distance tree of results	Multiple alignment	MSA Viewer																																																																			
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																																															
<input checked="" type="checkbox"/>	hypothetical protein SEA_NUCCI_41 [Microbacterium phage Nucci]	Microbacterium phage Nucci	190	190	98%	1e-60	100.00%	91	GD252434.1																																																															
<input checked="" type="checkbox"/>	hypothetical protein SEA_QUARTZ_42 [Microbacterium phage Quartz]	Microbacterium phage Quartz	188	188	98%	5e-60	98.90%	91	LVX59261.1																																																															
<input checked="" type="checkbox"/>	hypothetical protein SEA_YUZZY_42 [Microbacterium phage Yuzz]	Microbacterium phage Yuzz	105	105	94%	5e-27	89.00%	92	QEP56378.1																																																															
<input checked="" type="checkbox"/>	hypothetical protein SEA_JUANITO_41 [Microbacterium phage Juanito]	Microbacterium phage Juanito	99.0	99.0	94%	2e-24	85.56%	92	US644792.1																																																															
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un cribble HHpred avec une proba >= 90% et une couverture acceptable ?	<table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/> 1</td><td>PF11776.12</td><td>RcnB; Nickel/cobalt transporter regulator</td><td>79.26</td><td>3.8</td><td>24.49</td><td>1.7</td><td>12</td><td>51</td></tr><tr><td><input type="checkbox"/> 2</td><td>2L15_A</td><td>Uncharacterized protein yohN; Pfam PF11776, DUF3315, JCSG, Structural Genomics, PSI-Biology, Protein Structure Initiative</td><td>74.03</td><td>5.4</td><td>27.17</td><td>1.5</td><td>11</td><td>83</td></tr><tr><td><input type="checkbox"/> 3</td><td>PF19777.3</td><td>DUF6263; Family of unknown function (DUF6263)</td><td>73.07</td><td>49</td><td>23.82</td><td>6</td><td>47</td><td>192</td></tr><tr><td><input type="checkbox"/> 4</td><td>2UWJ_E</td><td>TYPE III EXPORT PROTEIN PSCE; VIRULENCE, CHAPERONES, COILED COIL, NEEDLE FORMATION, TYPE III SECRETION, BACTERIAL PATHOG</td><td>68.16</td><td>19</td><td>24.05</td><td>3</td><td>42</td><td>70</td></tr></table> <p>NON</p> <p>Visualization</p>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/> 1	PF11776.12	RcnB; Nickel/cobalt transporter regulator	79.26	3.8	24.49	1.7	12	51	<input type="checkbox"/> 2	2L15_A	Uncharacterized protein yohN; Pfam PF11776, DUF3315, JCSG, Structural Genomics, PSI-Biology, Protein Structure Initiative	74.03	5.4	27.17	1.5	11	83	<input type="checkbox"/> 3	PF19777.3	DUF6263; Family of unknown function (DUF6263)	73.07	49	23.82	6	47	192	<input type="checkbox"/> 4	2UWJ_E	TYPE III EXPORT PROTEIN PSCE; VIRULENCE, CHAPERONES, COILED COIL, NEEDLE FORMATION, TYPE III SECRETION, BACTERIAL PATHOG	68.16	19	24.05	3	42	70																										
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																																
<input type="checkbox"/> 1	PF11776.12	RcnB; Nickel/cobalt transporter regulator	79.26	3.8	24.49	1.7	12	51																																																																
<input type="checkbox"/> 2	2L15_A	Uncharacterized protein yohN; Pfam PF11776, DUF3315, JCSG, Structural Genomics, PSI-Biology, Protein Structure Initiative	74.03	5.4	27.17	1.5	11	83																																																																
<input type="checkbox"/> 3	PF19777.3	DUF6263; Family of unknown function (DUF6263)	73.07	49	23.82	6	47	192																																																																
<input type="checkbox"/> 4	2UWJ_E	TYPE III EXPORT PROTEIN PSCE; VIRULENCE, CHAPERONES, COILED COIL, NEEDLE FORMATION, TYPE III SECRETION, BACTERIAL PATHOG	68.16	19	24.05	3	42	70																																																																
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	Fonction inconnue																																																																							

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue

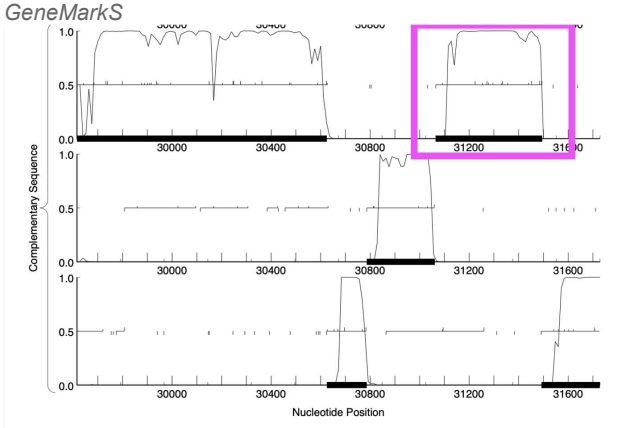
DECISION:

NKF

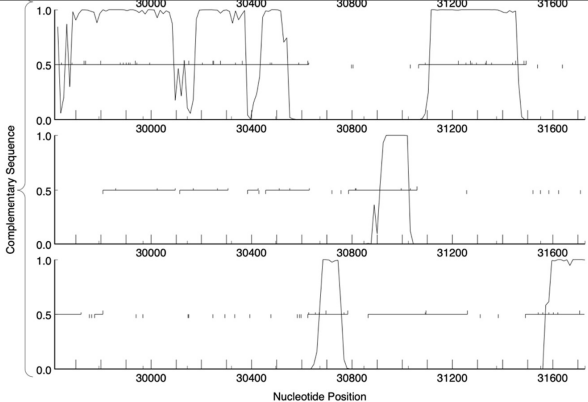
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	42
Coordonnées du Stop	31064
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	ove 4
Coordonnées du Start retenu	31495
Fonction prédite	MazG-like nucleotide pyrophosphohydrolase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS</p>  <p>GeneMark-host</p>

Commenté [7]: page entière si possible

	 <p>Potentiel codant tout le long de l'ORF</p>																																				
	<p>De nombreux hits et plusieurs sont annotés : Le meilleur hit : Quartz_43, MazG-like nucleotide pyrophosphohydrolase, 143 SCORE 285 e-value 3e-77</p> <div><div></div><div></div></div> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_43, MazG-like nucleotide pyrophosphohydrolase, 143</td><td>285</td><td>3e-77</td></tr><tr><td>Carostasia_Draft_42, function unknown, 143</td><td>285</td><td>3e-77</td></tr><tr><td>Nucci_42, MazG-like nucleotide pyrophosphohydrolase, 143</td><td>283</td><td>8e-77</td></tr><tr><td>Mandalorian_42, MazG-like nucleotide pyrophosphohydrolase, 143</td><td>283</td><td>8e-77</td></tr><tr><td>Juanyo_42, MazG-like nucleotide pyrophosphohydrolase, 143</td><td>267</td><td>6e-72</td></tr><tr><td>Yuuf_43, MazG-like nucleotide pyrophosphohydrolase, 143</td><td>266</td><td>1e-71</td></tr><tr><td>Zenitsu_43, MazG-like nucleotide pyrophosphohydrolase, 147</td><td>260</td><td>9e-70</td></tr><tr><td>MCubed_43, nucleotide pyrophosphohydrolase, 147</td><td>260</td><td>9e-70</td></tr><tr><td>Finny_44, MazG-like nucleotide pyrophosphohydrolase, 147</td><td>260</td><td>9e-70</td></tr><tr><td>Shamu_Draft_44, function unknown, 147</td><td>257</td><td>6e-69</td></tr><tr><td>Sansa_42, nucleotide pyrophosphohydrolase, 147</td><td>257</td><td>6e-69</td></tr></table>	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_43, MazG-like nucleotide pyrophosphohydrolase, 143	285	3e-77	Carostasia_Draft_42, function unknown, 143	285	3e-77	Nucci_42, MazG-like nucleotide pyrophosphohydrolase, 143	283	8e-77	Mandalorian_42, MazG-like nucleotide pyrophosphohydrolase, 143	283	8e-77	Juanyo_42, MazG-like nucleotide pyrophosphohydrolase, 143	267	6e-72	Yuuf_43, MazG-like nucleotide pyrophosphohydrolase, 143	266	1e-71	Zenitsu_43, MazG-like nucleotide pyrophosphohydrolase, 147	260	9e-70	MCubed_43, nucleotide pyrophosphohydrolase, 147	260	9e-70	Finny_44, MazG-like nucleotide pyrophosphohydrolase, 147	260	9e-70	Shamu_Draft_44, function unknown, 147	257	6e-69	Sansa_42, nucleotide pyrophosphohydrolase, 147	257	6e-69
Sequences producing significant alignments:	Score (bits)	E Value																																			
Quartz_43, MazG-like nucleotide pyrophosphohydrolase, 143	285	3e-77																																			
Carostasia_Draft_42, function unknown, 143	285	3e-77																																			
Nucci_42, MazG-like nucleotide pyrophosphohydrolase, 143	283	8e-77																																			
Mandalorian_42, MazG-like nucleotide pyrophosphohydrolase, 143	283	8e-77																																			
Juanyo_42, MazG-like nucleotide pyrophosphohydrolase, 143	267	6e-72																																			
Yuuf_43, MazG-like nucleotide pyrophosphohydrolase, 143	266	1e-71																																			
Zenitsu_43, MazG-like nucleotide pyrophosphohydrolase, 147	260	9e-70																																			
MCubed_43, nucleotide pyrophosphohydrolase, 147	260	9e-70																																			
Finny_44, MazG-like nucleotide pyrophosphohydrolase, 147	260	9e-70																																			
Shamu_Draft_44, function unknown, 147	257	6e-69																																			
Sansa_42, nucleotide pyrophosphohydrolase, 147	257	6e-69																																			
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>Y a-t-il un chevauchement significatif avec d'autres gènes? Pas de chevauchement significatif Est-il suffisamment long ? 432pb donc oui Les gènes avant et après ce gène sont-ils dans la même direction ? oui</p>																																				
DECISION:	OUI																																				

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 31495 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 31495
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	Le meilleur score (-2,765) et Z value (2,090): GTG 31357 mais ORF = 294 L'ORF la plus longue = 432 et score = -4,427 et Z value = 2,831 : ATG 31495
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Avec le GTG 31357 : on a +134 gap (donc trop élevé) Avec le ATG 31495 : on a -4 chevauchement Le démarrage du potentiel codant est vers 31 495
Est-ce que le start est conservé chez les homologues voir Starterator ?	Le Start consensus = 31 495
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Très conservé chez les différents homologues
DECISION:	ATG 31495 -bon score -bonne zvalue - très conservé chez homologues -chevauchement inférieur à 30pb -ORF>50pb et la plus longue

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction	*** PhagesDB : (q#Quartz_43, MazG-like nucleotide pyrophosphohydrolase, 143

annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

: s#285) : [100%] ; e-value : 3e-77

>Quartz_43, MazG-like nucleotide pyrophosphohydrolase, 143
Length = 143

Score = 285 bits (729), Expect = 3e-77
Identities = 143/143 (100%), Positives = 143/143 (100%)

Query: 1 MNKMQQQVLEFQAMELPVGHRARAVPEDQKAVRHELIREEEFLVALQTDDMVEQADA 60
Sbjct: 1 MNKMQQQVLEFQAMELPVGHRARAVPEDQKAVRHELIREEEFLVALQTDDMVEQADA 60

Query: 61 LIDILYVAFGGLVVMGIDAEVLFDEVHRSNMSKFGEDGKPIIAGPNPDGIFEGRVKKG 120
Sbjct: 61 LIDILYVAFGGLVVMGIDAEVLFDEVHRSNMSKFGEDGKPIIAGPNPDGIFEGRVKKG 120

Query: 121 NYFRPNIRALLAHGVADQDLAKA 143
Sbjct: 121 NYFRPNIRALLAHGVADQDLAKA 143

*** nr :
(q#MazG-like nucleotide pyrophosphohydrolase
[Microbacterium phage Quartz]: s#288) : [100%] ; e-value : 6e-98

MazG-like nucleotide pyrophosphohydrolase [Microbacterium phage Quartz]
Sequence ID: [UVK59262.1](#) Length: 143 Number of Matches: 1

Range 1: 1 to 143 [GenPept](#) [Graphics](#) [Next Match](#) [Previous](#)

Score	Expect	Method	Identities	Positives	Gaps
288 bits(737)	6e-98	Compositional matrix adjust.	143/143(100%)	143/143(100%)	0/143(0%)

Query 1 MNKMQQQVLEFQAMELPVGHRARAVPEDQKAVRHELIREEEFLVALQTDDMVEQADA 60
Sbjct 1 MNKMQQQVLEFQAMELPVGHRARAVPEDQKAVRHELIREEEFLVALQTDDMVEQADA 60

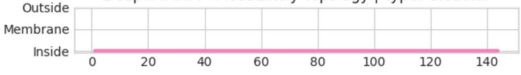
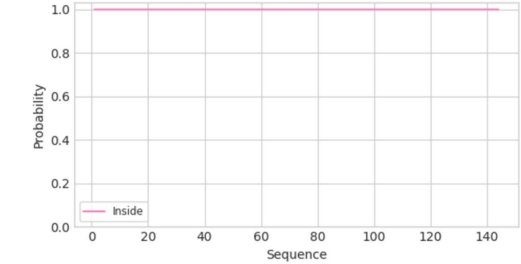
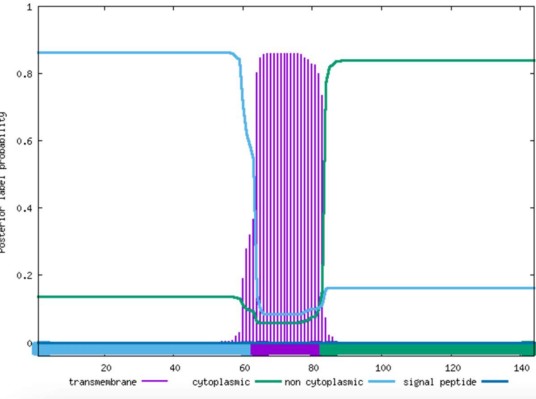
Query 61 LIDILYVAFGGLVVMGIDAEVLFDEVHRSNMSKFGEDGKPIIAGPNPDGIFEGRVKKG 120
Sbjct 61 LIDILYVAFGGLVVMGIDAEVLFDEVHRSNMSKFGEDGKPIIAGPNPDGIFEGRVKKG 120

Query 121 NYFRPNIRALLAHGVADQDLAKA 143
Sbjct 121 NYFRPNIRALLAHGVADQDLAKA 143

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

HHPred : putative NTP pyrophosphohydrolase; Structural Genomics, Joint Center for Structural Genomics, JCSG, Protein Structure In

Hit [3NL9_A](#)
Database PDB
Probability 99,97

	<div><div>DeepTMHMM - Most Likely Topology Type: Globular</div><div></div><div>DeepTMHMM - Posterior Probabilities</div><div></div><div>Phoebeius : ID UNNAMED FT TOPO_DOM 1 63 NON CYTOPLASMIC. FT TRANSMEM 64 83 FT TOPO_DOM 84 144 CYTOPLASMIC. //</div><div><div>Phoebeius posterior probabilities for UNNAMED</div><div></div></div></div> <tr><td>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</td><td>Yes</td></tr> <tr><td>DECISION:</td><td>MazG-like nucleotide pyrophosphohydrolase selon les blasp</td></tr>	Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Yes	DECISION:	MazG-like nucleotide pyrophosphohydrolase selon les blasp
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Yes				
DECISION:	MazG-like nucleotide pyrophosphohydrolase selon les blasp				

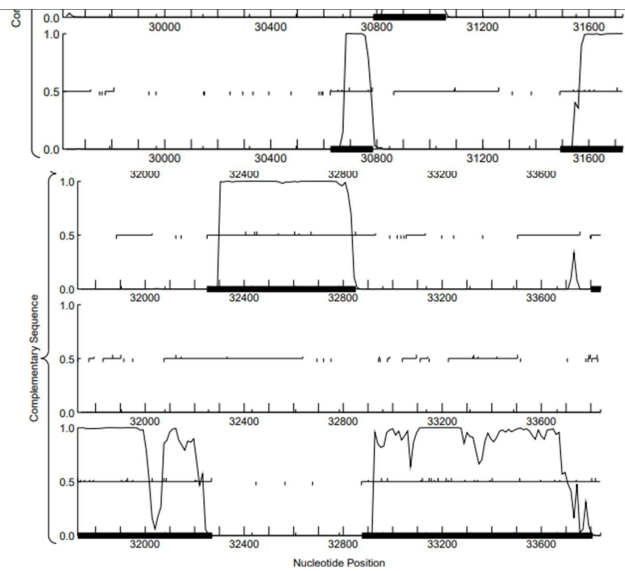
Student Gene Annotation Worksheet

Carbone Melwin

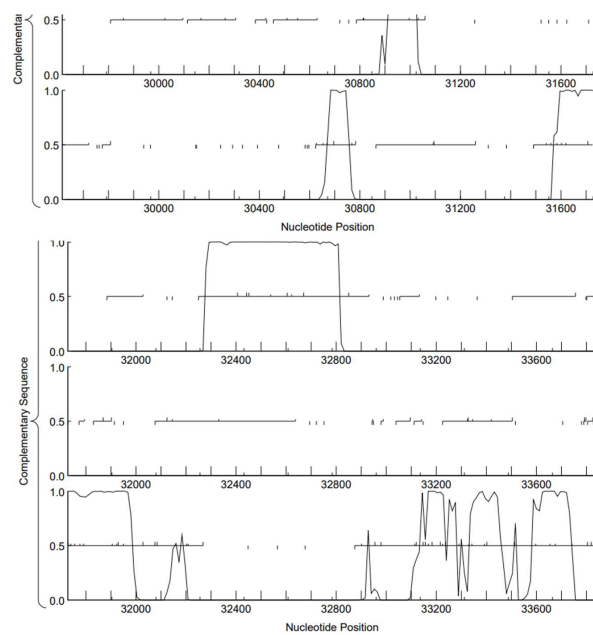
Basic Phage Information	
Nom du Phage	Carostasia
Gène #	43
Coordonnées du Stop	31492
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	chev 23
Coordonnées du Start retenu	32271
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	Potentiel codant pour les 2, avec une baisse entre 32000 et 32400 GenemarkS :



Genemark :



Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	Sequences producing significant alignments:		Score (bits)	E Value
	Quartz_44, function unknown, 259		521	e-148
	Nucci_43, function unknown, 259		521	e-148
	Mandalorian_43, function unknown, 259		521	e-148
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions			
DECISION:	OUI			

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																															
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 32271 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 32271																																																																																																																																																																																																															
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Stats: 21 Selected: 1</div> <div>ORF Start : 32271 ORF Stop : 31492 ORF Length: 780</div> <div>Cdn1 Cdn2 Cdn3 Length 5' End 20.0 75.0 95.0 60 3' End 47.6 61.9 95.2 63</div> <div>SD Scoring Matrix Kibler6 Spacing Weight Matrix Karlin Medium</div> <div>Explore Document</div> <div>32270</div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,791</td><td>2,493</td><td>14</td><td>-4,138</td><td>TCGCTGCTGGAGCTGCTCGCT</td><td>ATG</td><td>32271</td><td>780</td></tr><tr><td>2</td><td>-4,876</td><td>1,518</td><td>17</td><td>-6,876</td><td>CCACAGCGACCACTGTTCAAA</td><td>GTG</td><td>32211</td><td>720</td></tr><tr><td>3</td><td>-4,515</td><td>1,686</td><td>6</td><td>-6,260</td><td>CGACCACTGTTCAAAGTGGG</td><td>GTG</td><td>32205</td><td>714</td></tr><tr><td>4</td><td>-1,748</td><td>2,981</td><td>10</td><td>-2,443</td><td>GCAAAACGAGGAGGCAACC</td><td>ATG</td><td>32088</td><td>597</td></tr><tr><td>5</td><td>-4,352</td><td>1,744</td><td>16</td><td>-6,187</td><td>GAAAGAGCGAACCATGCTGG</td><td>ATG</td><td>32079</td><td>588</td></tr><tr><td>6</td><td>-4,595</td><td>1,649</td><td>12</td><td>-5,431</td><td>CGAGTCCCGCGTTCCACGCC</td><td>ATG</td><td>32031</td><td>540</td></tr><tr><td>7</td><td>-2,770</td><td>2,503</td><td>17</td><td>-4,770</td><td>CGGACGGGGCATCTGTTGCC</td><td>GTG</td><td>31953</td><td>462</td></tr><tr><td>8</td><td>-5,472</td><td>1,239</td><td>13</td><td>-6,518</td><td>CGTGATCGACCGCGAGAG</td><td>ATG</td><td>31932</td><td>441</td></tr><tr><td>9</td><td>-4,463</td><td>1,711</td><td>10</td><td>-5,158</td><td>CGACGACCGCGAGAGTGGC</td><td>GTG</td><td>31926</td><td>435</td></tr><tr><td>10</td><td>-2,856</td><td>2,463</td><td>10</td><td>-3,551</td><td>GCGCTGGGCGAGGTAACGAC</td><td>GTG</td><td>31908</td><td>417</td></tr><tr><td>11</td><td>-5,974</td><td>1,004</td><td>10</td><td>-6,668</td><td>CAAGCACATACCGACAAGCT</td><td>GTG</td><td>31794</td><td>303</td></tr><tr><td>12</td><td>-6,640</td><td>0,692</td><td>11</td><td>-7,397</td><td>CAAGCGTGTGCGCTTCAATCC</td><td>GTG</td><td>31779</td><td>288</td></tr><tr><td>13</td><td>-3,654</td><td>2,089</td><td>9</td><td>-4,429</td><td>CGTCTGAAGGCTGGGACACC</td><td>GTG</td><td>31758</td><td>267</td></tr><tr><td>14</td><td>-4,827</td><td>1,540</td><td>13</td><td>-5,873</td><td>GGACACCGTGGGCGGCACTG</td><td>GTG</td><td>31743</td><td>252</td></tr><tr><td>15</td><td>-3,697</td><td>2,069</td><td>13</td><td>-4,742</td><td>CGCGCACTGGTGGGCAAGTC</td><td>ATG</td><td>31731</td><td>240</td></tr><tr><td>16</td><td>-4,463</td><td>1,711</td><td>7</td><td>-5,586</td><td>CATGACGCTTACGACGAGCTG</td><td>ATG</td><td>31710</td><td>219</td></tr><tr><td>17</td><td>-5,207</td><td>1,363</td><td>12</td><td>-6,042</td><td>GTTGCTGAAGGATCCAGTGA</td><td>GTG</td><td>31623</td><td>132</td></tr><tr><td>18</td><td>-6,499</td><td>0,758</td><td>8</td><td>-7,721</td><td>GGTAGTGTCAACCGCTTCTCG</td><td>GTG</td><td>31605</td><td>114</td></tr><tr><td>19</td><td>-3,158</td><td>2,322</td><td>16</td><td>-4,954</td><td>CTGGTGGAGAGCGCGAGGAC</td><td>GTG</td><td>31587</td><td>96</td></tr><tr><td>20</td><td>-3,368</td><td>2,223</td><td>8</td><td>-4,590</td><td>GACCTGGTCAACTGGGAGAA</td><td>GTG</td><td>31563</td><td>72</td></tr><tr><td>21</td><td>-3,475</td><td>2,173</td><td>16</td><td>-5,271</td><td>CGAAGTGGACCACTGACCGAA</td><td>GTG</td><td>31545</td><td>54</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,791	2,493	14	-4,138	TCGCTGCTGGAGCTGCTCGCT	ATG	32271	780	2	-4,876	1,518	17	-6,876	CCACAGCGACCACTGTTCAAA	GTG	32211	720	3	-4,515	1,686	6	-6,260	CGACCACTGTTCAAAGTGGG	GTG	32205	714	4	-1,748	2,981	10	-2,443	GCAAAACGAGGAGGCAACC	ATG	32088	597	5	-4,352	1,744	16	-6,187	GAAAGAGCGAACCATGCTGG	ATG	32079	588	6	-4,595	1,649	12	-5,431	CGAGTCCCGCGTTCCACGCC	ATG	32031	540	7	-2,770	2,503	17	-4,770	CGGACGGGGCATCTGTTGCC	GTG	31953	462	8	-5,472	1,239	13	-6,518	CGTGATCGACCGCGAGAG	ATG	31932	441	9	-4,463	1,711	10	-5,158	CGACGACCGCGAGAGTGGC	GTG	31926	435	10	-2,856	2,463	10	-3,551	GCGCTGGGCGAGGTAACGAC	GTG	31908	417	11	-5,974	1,004	10	-6,668	CAAGCACATACCGACAAGCT	GTG	31794	303	12	-6,640	0,692	11	-7,397	CAAGCGTGTGCGCTTCAATCC	GTG	31779	288	13	-3,654	2,089	9	-4,429	CGTCTGAAGGCTGGGACACC	GTG	31758	267	14	-4,827	1,540	13	-5,873	GGACACCGTGGGCGGCACTG	GTG	31743	252	15	-3,697	2,069	13	-4,742	CGCGCACTGGTGGGCAAGTC	ATG	31731	240	16	-4,463	1,711	7	-5,586	CATGACGCTTACGACGAGCTG	ATG	31710	219	17	-5,207	1,363	12	-6,042	GTTGCTGAAGGATCCAGTGA	GTG	31623	132	18	-6,499	0,758	8	-7,721	GGTAGTGTCAACCGCTTCTCG	GTG	31605	114	19	-3,158	2,322	16	-4,954	CTGGTGGAGAGCGCGAGGAC	GTG	31587	96	20	-3,368	2,223	8	-4,590	GACCTGGTCAACTGGGAGAA	GTG	31563	72	21	-3,475	2,173	16	-5,271	CGAAGTGGACCACTGACCGAA	GTG	31545	54
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																								
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																								
1	-2,791	2,493	14	-4,138	TCGCTGCTGGAGCTGCTCGCT	ATG	32271	780																																																																																																																																																																																																								
2	-4,876	1,518	17	-6,876	CCACAGCGACCACTGTTCAAA	GTG	32211	720																																																																																																																																																																																																								
3	-4,515	1,686	6	-6,260	CGACCACTGTTCAAAGTGGG	GTG	32205	714																																																																																																																																																																																																								
4	-1,748	2,981	10	-2,443	GCAAAACGAGGAGGCAACC	ATG	32088	597																																																																																																																																																																																																								
5	-4,352	1,744	16	-6,187	GAAAGAGCGAACCATGCTGG	ATG	32079	588																																																																																																																																																																																																								
6	-4,595	1,649	12	-5,431	CGAGTCCCGCGTTCCACGCC	ATG	32031	540																																																																																																																																																																																																								
7	-2,770	2,503	17	-4,770	CGGACGGGGCATCTGTTGCC	GTG	31953	462																																																																																																																																																																																																								
8	-5,472	1,239	13	-6,518	CGTGATCGACCGCGAGAG	ATG	31932	441																																																																																																																																																																																																								
9	-4,463	1,711	10	-5,158	CGACGACCGCGAGAGTGGC	GTG	31926	435																																																																																																																																																																																																								
10	-2,856	2,463	10	-3,551	GCGCTGGGCGAGGTAACGAC	GTG	31908	417																																																																																																																																																																																																								
11	-5,974	1,004	10	-6,668	CAAGCACATACCGACAAGCT	GTG	31794	303																																																																																																																																																																																																								
12	-6,640	0,692	11	-7,397	CAAGCGTGTGCGCTTCAATCC	GTG	31779	288																																																																																																																																																																																																								
13	-3,654	2,089	9	-4,429	CGTCTGAAGGCTGGGACACC	GTG	31758	267																																																																																																																																																																																																								
14	-4,827	1,540	13	-5,873	GGACACCGTGGGCGGCACTG	GTG	31743	252																																																																																																																																																																																																								
15	-3,697	2,069	13	-4,742	CGCGCACTGGTGGGCAAGTC	ATG	31731	240																																																																																																																																																																																																								
16	-4,463	1,711	7	-5,586	CATGACGCTTACGACGAGCTG	ATG	31710	219																																																																																																																																																																																																								
17	-5,207	1,363	12	-6,042	GTTGCTGAAGGATCCAGTGA	GTG	31623	132																																																																																																																																																																																																								
18	-6,499	0,758	8	-7,721	GGTAGTGTCAACCGCTTCTCG	GTG	31605	114																																																																																																																																																																																																								
19	-3,158	2,322	16	-4,954	CTGGTGGAGAGCGCGAGGAC	GTG	31587	96																																																																																																																																																																																																								
20	-3,368	2,223	8	-4,590	GACCTGGTCAACTGGGAGAA	GTG	31563	72																																																																																																																																																																																																								
21	-3,475	2,173	16	-5,271	CGAAGTGGACCACTGACCGAA	GTG	31545	54																																																																																																																																																																																																								
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start conduit à l'ORF le plus long. C'est un ATG avec un bon score final, et le start correspond au départ du pouvoir codant.																																																																																																																																																																																																															
Est-ce que le start est conservé chez les homologues voir Starterator ?	• Track 25 : Carostasia_43, Quartz_44 Le start proposé est le même, 32271																																																																																																																																																																																																															
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?																																																																																																																																																																																																																

	<p>>Quartz_44, function unknown, 259 Length = 259</p> <p>Score = 521 bits (1343), Expect = e-148 Identities = 259/259 (100%), Positives = 259/259 (100%)</p> <p>Query: 1 MPTSTSAKASQHPHSDHLFKVGVLLQRSKRYPYPKATRKEAARELDRYWDEGLILQTTKEA 60 MPTSTSAKASQHPHSDHLFKVGVLLQRSKRYPYPKATRKEAARELDRYWDEGLILQTTKEA 60 Sbjct: 1 MPTSTSAKASQHPHSDHLFKVGVLLQRSKRYPYPKATRKEAARELDRYWDEGLILQTTKEA 60</p> <p>Query: 61 TMPGMKNSGFCCTFESPGSHAMCSGRHCDCHKEAKAFSGRGILFAVIDDREKMRVGKVT 120 TMPGMKNSGFCCTFESPGSHAMCSGRHCDCHKEAKAFSGRGILFAVIDDREKMRVGKVT 120 Sbjct: 61 TMPGMKNSGFCCTFESPGSHAMCSGRHCDCHKEAKAFSGRGILFAVIDDREKMRVGKVT 120</p> <p>Query: 121 TVDVIADGIEAALDAGQETSEQIARFLTTSFAFKHIPDKRVRFNSVLKAGDTVGAHLVRKL 180 TVDVIADGIEAALDAGQETSEQIARFLTTSFAFKHIPDKRVRFNSVLKAGDTVGAHLVRKL 180 Sbjct: 121 TVDVIADGIEAALDAGQETSEQIARFLTTSFAFKHIPDKRVRFNSVLKAGDTVGAHLVRKL 180</p> <p>Query: 181 MNAYDELMETADKSRSPKDKKEALTNEARGFAEAIQVVLNPFVSVEDAEDVTLWNWDEVHDL 240 MNAYDELMETADKSRSPKDKKEALTNEARGFAEAIQVVLNPFVSVEDAEDVTLWNWDEVHDL 240 Sbjct: 181 MNAYDELMETADKSRSPKDKKEALTNEARGFAEAIQVVLNPFVSVEDAEDVTLWNWDEVHDL 240</p> <p>Query: 241 TEVFEKERRVVRADKKNR 259 TEVFEKERRVVRADKKNR Sbjct: 241 TEVFEKERRVVRADKKNR 259</p>
DECISION:	<p>32271 retenu. Correspond aux critères et aux principes fondamentaux.</p> <p>32271</p> <ul style="list-style-type: none"> - bon score -bonne zvalue - assez conservé chez homologues -chevauchement inférieur à 30pb -ORF>50pb et la plus longue

Commenté [8]: si possible mettre le screen pour un autre homologue

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																								
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_44, function unknown, 259</td><td>521</td><td>e-148</td></tr><tr><td>Nucci_43, function unknown, 259</td><td>521</td><td>e-148</td></tr><tr><td>Mandelorian_43, function unknown, 259</td><td>521</td><td>e-148</td></tr><tr><td>Carostasia_Draft_43, Function unknown, 259</td><td>521</td><td>e-148</td></tr><tr><td>YuuV_44, function unknown, 259</td><td>503</td><td>e-142</td></tr><tr><td>Juanyo_43, function unknown, 259</td><td>493</td><td>e-139</td></tr><tr><td>Zepp_44, function unknown, 260</td><td>410</td><td>e-114</td></tr></table> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p>	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_44, function unknown, 259	521	e-148	Nucci_43, function unknown, 259	521	e-148	Mandelorian_43, function unknown, 259	521	e-148	Carostasia_Draft_43, Function unknown, 259	521	e-148	YuuV_44, function unknown, 259	503	e-142	Juanyo_43, function unknown, 259	493	e-139	Zepp_44, function unknown, 260	410	e-114
Sequences producing significant alignments:	Score (bits)	E Value																							
Quartz_44, function unknown, 259	521	e-148																							
Nucci_43, function unknown, 259	521	e-148																							
Mandelorian_43, function unknown, 259	521	e-148																							
Carostasia_Draft_43, Function unknown, 259	521	e-148																							
YuuV_44, function unknown, 259	503	e-142																							
Juanyo_43, function unknown, 259	493	e-139																							
Zepp_44, function unknown, 260	410	e-114																							

Sequences producing significant alignments

Download

Select columns

Show

100

select all

63 sequences selected

GenPept

Graphics

Distance tree of results

Multiple alignment

MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Hypothetical protein SEA_NUCC1_43 [Microbacterium ethaps Nucci]	Microbacterium ethaps Nucci	539	539	99%	0.0	100.00%	259	Q0002436.1
<input checked="" type="checkbox"/>	Hypothetical protein SEA_YUUY_44 [Microbacterium ethaps Yuuy]	Microbacterium ethaps Yuuy	519	519	99%	0.0	95.75%	259	QEP35372.1
<input checked="" type="checkbox"/>	Hypothetical protein SEA_YUUYO_43 [Microbacterium ethaps YuuyO]	Microbacterium ethaps YuuyO	510	510	99%	0.0	94.21%	259	US0164792.1
<input checked="" type="checkbox"/>	Hypothetical protein SEA_ZZFP_44 [Microbacterium ethaps Zzep]	Microbacterium ethaps Zzep	424	424	99%	8e-140	77.69%	260	Q0068132.1
<input checked="" type="checkbox"/>	Hypothetical protein SEA_HSD170_44 [Microbacterium ethaps Healt]	Microbacterium ethaps Healt	422	422	99%	4e-147	77.31%	260	Q2209112.1

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Non

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	2JBW_A	Cytochrome C; C; HEME, IRON, TRANSPORT, CYTOCHROME, METAL-BINDING, ELECTRON TRANSFER, ELECTRON TRANSPORT, RUBRIVIVAX G	69.78	95	24.3	8.6	93	129
<input type="checkbox"/>	7URH_K	Ferritin; Ferroxidase, OXIDOREDUCTASE; HET: MPD; 1.468A (Caenorhabditis elegans)	65.36	110	23.44	6.9	80	168
<input type="checkbox"/>	PF16148.9	DUF4855; Domain of unknown function (DUF4855)	65.15	110	30.85	8.4	63	360

Visualization

Resubmit Section

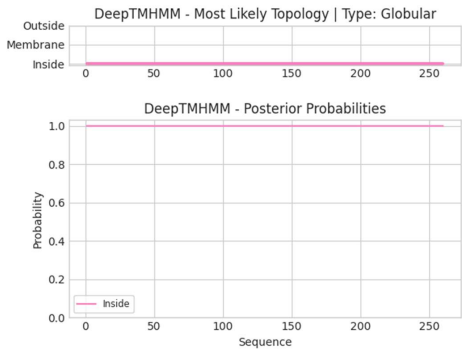
166

259

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

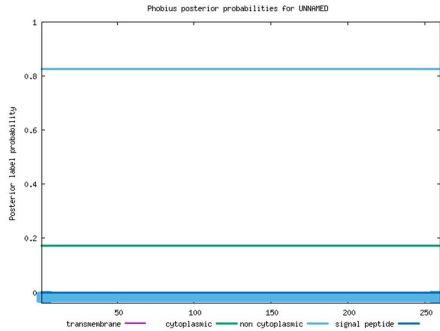
Fonction inconnue

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



NON

ID UNPAID
FT TOPO_DOM 1 268 NON CYTOPLASMIC.
//



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

NFK

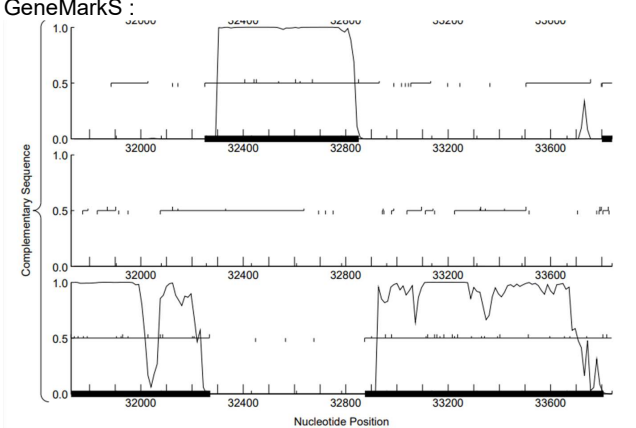
DECISION:

NFK

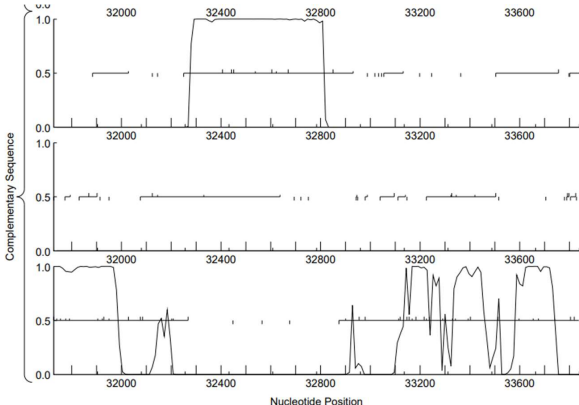
Student Gene Annotation Worksheet *Carbone Melwin*

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	44
Coordonnées du Stop	32249
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	Gap 23
Coordonnées du Start retenu	32851
Fonction prédite	thymidylate kinase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>Fort potentiel codant constant. GeneMarkS :</p> 

Commenté [9]: page entière si possible

	<div>GeneMark : </div>												
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Nucci_44, thymidylate kinase, 200</td><td>414</td><td>e-116</td></tr><tr><td>Mandalorian_44, thymidylate kinase, 200</td><td>414</td><td>e-116</td></tr><tr><td>Carostasia_Draft_44, function unknown, 200</td><td>414</td><td>e-116</td></tr></table>		Score (bits)	E Value	Nucci_44, thymidylate kinase, 200	414	e-116	Mandalorian_44, thymidylate kinase, 200	414	e-116	Carostasia_Draft_44, function unknown, 200	414	e-116
	Score (bits)	E Value											
Nucci_44, thymidylate kinase, 200	414	e-116											
Mandalorian_44, thymidylate kinase, 200	414	e-116											
Carostasia_Draft_44, function unknown, 200	414	e-116											
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradiction												
DECISION:	OUI												

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																			
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 32851</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 32851</i>																																																																																																			
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Stats : 9 Selected : 1</div><div>ORF Start : 32851 ORF Stop : 32249 ORF Length: 603</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 81.5 40.7 59.3 81 3' End 100.0 0.0 0.0 3</div><div>SD Scoring Matrix Kibler6</div><div>Explore</div><div>Spacing Weight Matrix Karlin Medium</div><div>Document</div></div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,302</td><td>2,722</td><td>7</td><td>-3,825</td><td>GGGTAGGTGTAACTGGAGGCA</td><td>GTG</td><td>32932</td><td>684</td></tr><tr><td>2</td><td>-1,748</td><td>2,981</td><td>9</td><td>-2,623</td><td>CCACACAACTCAAGGAGAGCAC</td><td>ATG</td><td>32851</td><td>603</td></tr><tr><td>3</td><td>-5,049</td><td>1,436</td><td>11</td><td>-5,806</td><td>CGGATGCACTGGCTGACGCAC</td><td>ATG</td><td>32671</td><td>423</td></tr><tr><td>4</td><td>-2,669</td><td>2,551</td><td>14</td><td>-4,016</td><td>GTTCAGTGGAAAGACGACGCGC</td><td>GTG</td><td>32623</td><td>375</td></tr><tr><td>5</td><td>-4,817</td><td>1,545</td><td>10</td><td>-5,511</td><td>CGCGTGGCGACGGCTCTTGGCC</td><td>ATG</td><td>32605</td><td>357</td></tr><tr><td>6</td><td>-5,833</td><td>1,070</td><td>13</td><td>-6,879</td><td>CGAGGGCCAGCTTGCAGCCAAAG</td><td>GTG</td><td>32539</td><td>291</td></tr><tr><td>7</td><td>-3,912</td><td>1,969</td><td>16</td><td>-5,707</td><td>CTTCAGGAGTGGCTGGCGCGCC</td><td>ATG</td><td>32452</td><td>204</td></tr><tr><td>8</td><td>-4,502</td><td>1,693</td><td>8</td><td>-5,724</td><td>CTGGCTGGCGCCATGAACACAG</td><td>ATG</td><td>32443</td><td>195</td></tr><tr><td>9</td><td>-4,796</td><td>1,555</td><td>14</td><td>-6,143</td><td>CTGTTCAGAACCATCTCGGTC</td><td>ATG</td><td>32407</td><td>159</td></tr></table> <p>On choisit le 2^e start, un ATG. ORF de 603pb (le plus long possible = 684pb) mais grand chevauchement avec le gène précédent. Meilleur Final score et Z value.</p>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,302	2,722	7	-3,825	GGGTAGGTGTAACTGGAGGCA	GTG	32932	684	2	-1,748	2,981	9	-2,623	CCACACAACTCAAGGAGAGCAC	ATG	32851	603	3	-5,049	1,436	11	-5,806	CGGATGCACTGGCTGACGCAC	ATG	32671	423	4	-2,669	2,551	14	-4,016	GTTCAGTGGAAAGACGACGCGC	GTG	32623	375	5	-4,817	1,545	10	-5,511	CGCGTGGCGACGGCTCTTGGCC	ATG	32605	357	6	-5,833	1,070	13	-6,879	CGAGGGCCAGCTTGCAGCCAAAG	GTG	32539	291	7	-3,912	1,969	16	-5,707	CTTCAGGAGTGGCTGGCGCGCC	ATG	32452	204	8	-4,502	1,693	8	-5,724	CTGGCTGGCGCCATGAACACAG	ATG	32443	195	9	-4,796	1,555	14	-6,143	CTGTTCAGAACCATCTCGGTC	ATG	32407	159
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																												
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																												
1	-2,302	2,722	7	-3,825	GGGTAGGTGTAACTGGAGGCA	GTG	32932	684																																																																																												
2	-1,748	2,981	9	-2,623	CCACACAACTCAAGGAGAGCAC	ATG	32851	603																																																																																												
3	-5,049	1,436	11	-5,806	CGGATGCACTGGCTGACGCAC	ATG	32671	423																																																																																												
4	-2,669	2,551	14	-4,016	GTTCAGTGGAAAGACGACGCGC	GTG	32623	375																																																																																												
5	-4,817	1,545	10	-5,511	CGCGTGGCGACGGCTCTTGGCC	ATG	32605	357																																																																																												
6	-5,833	1,070	13	-6,879	CGAGGGCCAGCTTGCAGCCAAAG	GTG	32539	291																																																																																												
7	-3,912	1,969	16	-5,707	CTTCAGGAGTGGCTGGCGCGCC	ATG	32452	204																																																																																												
8	-4,502	1,693	8	-5,724	CTGGCTGGCGCCATGAACACAG	ATG	32443	195																																																																																												
9	-4,796	1,555	14	-6,143	CTGTTCAGAACCATCTCGGTC	ATG	32407	159																																																																																												
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start choisit (32851) ne conduit pas à l'ORF le plus long, mais permet un chevauchement suffisamment court pour ne pas entrer en conflit avec les principes fondamentaux de l'annotation (22pb, alors qu'on aurait 103pb de chevauchement avec l'ORF le plus long). C'est un ATG, et correspond au départ du pouvoir codant.																																																																																																			
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Track 22 : <u>Carostasia_44</u></p> <p>• <u>Track 23 : Carostasia_44</u></p> <p>Gene: Carostasia_44 Start: 32851, Stop: 32249, Start Num: 5 Candidate Starts for Carostasia_44: (3, 32932), (Start: 5 @32851 has 155 MA's), (24, 32671), (27, 32623), (29, 32605), (40, 32539), (55, 32452), (56, 32443), (61, 32407),</p>																																																																																																			

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>>Nucci_44, thymidylate kinase, 200 Length = 200</p> <p>Score = 414 bits (1064), Expect = e-116 Identities = 200/200 (100%), Positives = 200/200 (100%)</p> <p>Query: 1 MLIAFEGPDETGKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEFPDAVITFDRIDWLTH 60 MLIAFEGPDETGKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEFPDAVITFDRIDWLTH 60 Sbjct: 1 MLIAFEGPDETGKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEFPDAVITFDRIDWLTH 60</p> <p>Query: 61 MVYRLSLPGFEWNDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV 120 MVYRLSLPGFEWNDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV 120 Sbjct: 61 MVYRLSLPGFEWNDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV 120</p> <p>Query: 121 TQTYHHFTDWLAAMNQMRDPSLFKTI SVMVHHNQETGEFTQRLVVSAPGFPWGTQYEK 180 TQTYHHFTDWLAAMNQMRDPSLFKTI SVMVHHNQETGEFTQRLVVSAPGFPWGTQYEK 180 Sbjct: 121 TQTYHHFTDWLAAMNQMRDPSLFKTI SVMVHHNQETGEFTQRLVVSAPGFPWGTQYEK 180</p> <p>Query: 181 LVHDDLESLELLRYADQHIG 200 LVHDDLESLELLRYADQHIG Sbjct: 181 LVHDDLESLELLRYADQHIG 200</p>
DECISION:	Start à 32851, respecte les principes fondamentaux

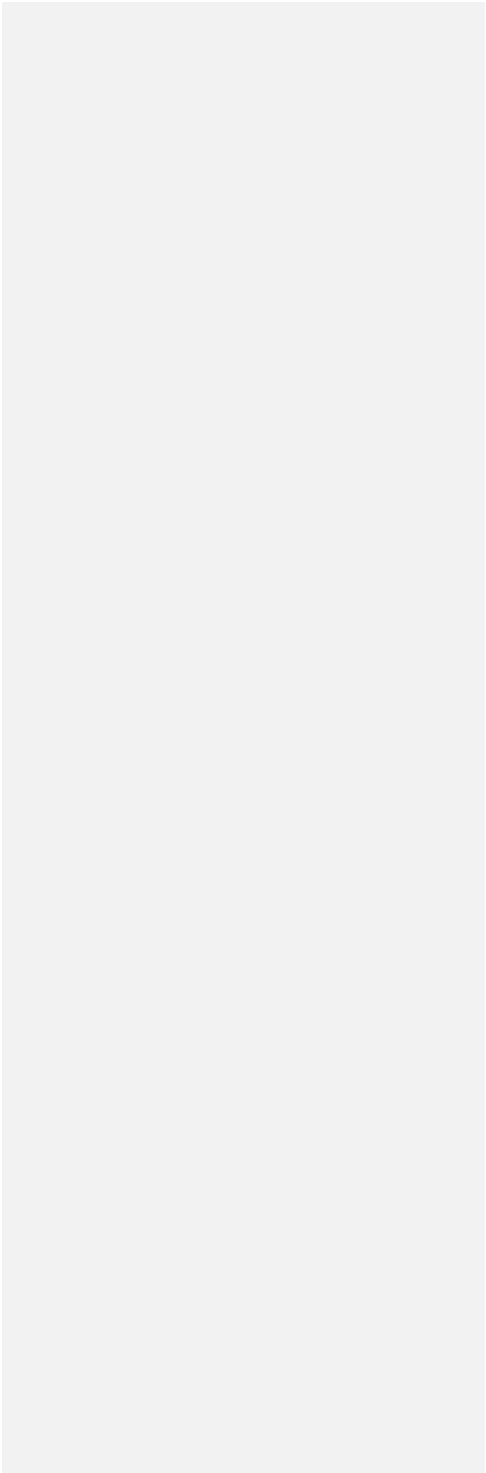
Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>*** <i>PhagesDB</i> :</p> <p>(q# Nucci_44, thymidylate kinase, 200 : s#414) : [100%] ; e-value : e-116</p> <p>>Nucci_44, thymidylate kinase, 200 Length = 200</p> <p>Score = 414 bits (1064), Expect = e-116 Identities = 200/200 (100%), Positives = 200/200 (100%)</p> <p>Query: 1 MLIAFEGPDETGKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEFPDAVITFDRIDWLTH 60 MLIAFEGPDETGKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEFPDAVITFDRIDWLTH 60 Sbjct: 1 MLIAFEGPDETGKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEFPDAVITFDRIDWLTH 60</p> <p>Query: 61 MVYRLSLPGFEWNDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV 120 MVYRLSLPGFEWNDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV 120 Sbjct: 61 MVYRLSLPGFEWNDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV 120</p> <p>Query: 121 TQTYHHFTDWLAAMNQMRDPSLFKTI SVMVHHNQETGEFTQRLVVSAPGFPWGTQYEK 180 TQTYHHFTDWLAAMNQMRDPSLFKTI SVMVHHNQETGEFTQRLVVSAPGFPWGTQYEK 180 Sbjct: 121 TQTYHHFTDWLAAMNQMRDPSLFKTI SVMVHHNQETGEFTQRLVVSAPGFPWGTQYEK 180</p> <p>Query: 181 LVHDDLESLELLRYADQHIG 200 LVHDDLESLELLRYADQHIG Sbjct: 181 LVHDDLESLELLRYADQHIG 200</p> <p>*** <i>nr</i> :</p> <p>(q#thymidylate kinase [Microbacterium phage Nucci] : s#418) : [100%] ; e-value : 2e-147</p>

	<div><div>thymidylate kinase [Microbacterium phage Nucci]</div><div>Sequence ID: QDK02437.1 Length: 200 Number of Matches: 1</div><div>See 1 more title(s) See all Identical Proteins(IPG)</div><div><div>Range 1: 1 to 200 GenPept Graphics</div><div><div>▼ Next Match ▲ Previous</div><table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>418 bits(1074)</td><td>2e-147</td><td>Compositional matrix adjust.</td><td>200/200(100%)</td><td>200/200(100%)</td><td>0/200(0%)</td></tr><tr><td>Query 1</td><td>MLIAFEGPDET</td><td>GKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEF</td><td>PDVITTFDRIDWLTH</td><td>60</td><td></td></tr><tr><td>Sbjct 1</td><td>MLIAFEGPDET</td><td>GKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEF</td><td>PDVITTFDRIDWLTH</td><td>60</td><td></td></tr><tr><td>Query 61</td><td>MVYRLSLPGFEW</td><td>NDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV</td><td>120</td><td></td><td></td></tr><tr><td>Sbjct 61</td><td>MVYRLSLPGFEW</td><td>NDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV</td><td>120</td><td></td><td></td></tr><tr><td>Query 121</td><td>TQTYHHFTDWLAAMNQMRD</td><td>FSFKTISVMEVHHNQETGEFTQRLVVFSA</td><td>PGFPWGTQYEK</td><td>180</td><td></td></tr><tr><td>Sbjct 121</td><td>TQTYHHFTDWLAAMNQMRD</td><td>FSFKTISVMEVHHNQETGEFTQRLVVFSA</td><td>PGFPWGTQYEK</td><td>180</td><td></td></tr><tr><td>Query 181</td><td>LVHDESLELLRLRYADQHIG</td><td>200</td><td></td><td></td><td></td></tr><tr><td>Sbjct 181</td><td>LVHDESLELLRLRYADQHIG</td><td>200</td><td></td><td></td><td></td></tr></table></div></div></div>	Score	Expect	Method	Identities	Positives	Gaps	418 bits(1074)	2e-147	Compositional matrix adjust.	200/200(100%)	200/200(100%)	0/200(0%)	Query 1	MLIAFEGPDET	GKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEF	PDVITTFDRIDWLTH	60		Sbjct 1	MLIAFEGPDET	GKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEF	PDVITTFDRIDWLTH	60		Query 61	MVYRLSLPGFEW	NDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV	120			Sbjct 61	MVYRLSLPGFEW	NDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV	120			Query 121	TQTYHHFTDWLAAMNQMRD	FSFKTISVMEVHHNQETGEFTQRLVVFSA	PGFPWGTQYEK	180		Sbjct 121	TQTYHHFTDWLAAMNQMRD	FSFKTISVMEVHHNQETGEFTQRLVVFSA	PGFPWGTQYEK	180		Query 181	LVHDESLELLRLRYADQHIG	200				Sbjct 181	LVHDESLELLRLRYADQHIG	200				
Score	Expect	Method	Identities	Positives	Gaps																																																									
418 bits(1074)	2e-147	Compositional matrix adjust.	200/200(100%)	200/200(100%)	0/200(0%)																																																									
Query 1	MLIAFEGPDET	GKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEF	PDVITTFDRIDWLTH	60																																																										
Sbjct 1	MLIAFEGPDET	GKSTSAQNLAWSAGTIYNANKDNYAQAQADVAEF	PDVITTFDRIDWLTH	60																																																										
Query 61	MVYRLSLPGFEW	NDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV	120																																																											
Sbjct 61	MVYRLSLPGFEW	NDDRVRTVFAMPDTHLVKTHRLTGEGQLAAKVEAVGEGYQSGNLDTV	120																																																											
Query 121	TQTYHHFTDWLAAMNQMRD	FSFKTISVMEVHHNQETGEFTQRLVVFSA	PGFPWGTQYEK	180																																																										
Sbjct 121	TQTYHHFTDWLAAMNQMRD	FSFKTISVMEVHHNQETGEFTQRLVVFSA	PGFPWGTQYEK	180																																																										
Query 181	LVHDESLELLRLRYADQHIG	200																																																												
Sbjct 181	LVHDESLELLRLRYADQHIG	200																																																												
<div>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?</div>	<div><div>HHPred : Thymidylate kinase; Structural Genomics, PSI-Biology, Midwest Center for Structural Genomics, MCSG, Mtb Proteins Conferr</div><div>Hit 4MQB_A</div><div>Database PDB</div><div>Probability 99,44</div><div><div><div>1106</div><div></div></div><div><div>4MQB_R</div><div>1GTV_R</div><div>4EDH_R</div><div>cd01672</div><div>2PLR_B</div><div>5ZBB_R</div><div>3V9P_R</div><div>1NKS_C</div></div></div><table><tr><td><input type="checkbox"/></td><td>1</td><td>4MQB_A</td><td>Thymidylate kinase; Structural Genomics, PSI-Biology, Midwest Center for Structural Genomics, MCSG, Mtb Proteins Conferr</td><td>99.44</td><td>7.1e-12</td><td>91.36</td><td>12.3</td><td>104</td><td>208</td></tr><tr><td><input type="checkbox"/></td><td>2</td><td>1GTV_A</td><td>THYMIDYLATE KINASE; TRANSFERASE, TRANSFERASE (ATP:TMP PHOSPHOTRANSFERASE), KINASE; HET: TMP, ACT, TYD, SO4; 1.55A (MYCOB</td><td>99.43</td><td>1.6e-11</td><td>89.5</td><td>13.2</td><td>106</td><td>214</td></tr><tr><td><input type="checkbox"/></td><td>3</td><td>4EDH_A</td><td>Thymidylate kinase; structural genomics, PSI-Biology, protein structure initiative, midwest center for structural genomi</td><td>99.41</td><td>3.9e-11</td><td>87.95</td><td>14.5</td><td>124</td><td>213</td></tr><tr><td><input type="checkbox"/></td><td>4</td><td>cd01672</td><td>TMPK; Thymidine monophosphate kinase (TMPK), also known as thymidylate kinase, catalyzes the phosphorylation of thymidin</td><td>99.41</td><td>4.6e-11</td><td>85.79</td><td>14.5</td><td>129</td><td>208</td></tr><tr><td><input type="checkbox"/></td><td>5</td><td>2PLR_B</td><td>Probable thymidylate kinase; TMP-binding, ATP-binding, Thymidylate Kinase, Structural Genomics, NPPSFA, National Project</td><td>99.4</td><td>3.1e-11</td><td>89.28</td><td>13.4</td><td>145</td><td>213</td></tr></table></div>	<input type="checkbox"/>	1	4MQB_A	Thymidylate kinase; Structural Genomics, PSI-Biology, Midwest Center for Structural Genomics, MCSG, Mtb Proteins Conferr	99.44	7.1e-12	91.36	12.3	104	208	<input type="checkbox"/>	2	1GTV_A	THYMIDYLATE KINASE; TRANSFERASE, TRANSFERASE (ATP:TMP PHOSPHOTRANSFERASE), KINASE; HET: TMP, ACT, TYD, SO4; 1.55A (MYCOB	99.43	1.6e-11	89.5	13.2	106	214	<input type="checkbox"/>	3	4EDH_A	Thymidylate kinase; structural genomics, PSI-Biology, protein structure initiative, midwest center for structural genomi	99.41	3.9e-11	87.95	14.5	124	213	<input type="checkbox"/>	4	cd01672	TMPK; Thymidine monophosphate kinase (TMPK), also known as thymidylate kinase, catalyzes the phosphorylation of thymidin	99.41	4.6e-11	85.79	14.5	129	208	<input type="checkbox"/>	5	2PLR_B	Probable thymidylate kinase; TMP-binding, ATP-binding, Thymidylate Kinase, Structural Genomics, NPPSFA, National Project	99.4	3.1e-11	89.28	13.4	145	213											
<input type="checkbox"/>	1	4MQB_A	Thymidylate kinase; Structural Genomics, PSI-Biology, Midwest Center for Structural Genomics, MCSG, Mtb Proteins Conferr	99.44	7.1e-12	91.36	12.3	104	208																																																					
<input type="checkbox"/>	2	1GTV_A	THYMIDYLATE KINASE; TRANSFERASE, TRANSFERASE (ATP:TMP PHOSPHOTRANSFERASE), KINASE; HET: TMP, ACT, TYD, SO4; 1.55A (MYCOB	99.43	1.6e-11	89.5	13.2	106	214																																																					
<input type="checkbox"/>	3	4EDH_A	Thymidylate kinase; structural genomics, PSI-Biology, protein structure initiative, midwest center for structural genomi	99.41	3.9e-11	87.95	14.5	124	213																																																					
<input type="checkbox"/>	4	cd01672	TMPK; Thymidine monophosphate kinase (TMPK), also known as thymidylate kinase, catalyzes the phosphorylation of thymidin	99.41	4.6e-11	85.79	14.5	129	208																																																					
<input type="checkbox"/>	5	2PLR_B	Probable thymidylate kinase; TMP-binding, ATP-binding, Thymidylate Kinase, Structural Genomics, NPPSFA, National Project	99.4	3.1e-11	89.28	13.4	145	213																																																					
<div>Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?</div>	<div>Aucune synténie observée</div>																																																													

	<div>Carostasia_Draft (EA10)</div> <div><p>Sequence alignment of Carostasia_Draft (EA10) and Juanyo (EA10). The alignment shows a high degree of similarity between the two sequences, with a score of 5366/6300 (85%) and an e-value of 0.000e+00. The alignment is displayed as a purple bar with colored segments representing different domains or features. The x-axis represents the sequence position from 122 to 152.</p></div> <div><p>Sequence alignment of Carostasia_Draft (EA10) and Mandalorian (EA10). The alignment shows a high degree of similarity between the two sequences, with a score of 5366/6300 (85%) and an e-value of 0.000e+00. The alignment is displayed as a purple bar with colored segments representing different domains or features. The x-axis represents the sequence position from 122 to 152.</p></div>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div>No</div> <div>DeepTMHMM : DeepTMHMM - Most Likely Topology Type: Globular</div> <div><p>DeepTMHMM - Most Likely Topology Type: Globular. The plot shows the probability of the protein being transmembrane (TM) across the sequence. The y-axis represents the probability (0.0 to 1.0) and the x-axis represents the sequence position (0 to 200). The plot shows a single peak at position 122, indicating a high probability of TM at this position.</p></div> <div>DeepTMHMM - Posterior Probabilities</div> <div><p>DeepTMHMM - Posterior Probabilities. The plot shows the probability of the protein being transmembrane (TM) across the sequence. The y-axis represents the probability (0.0 to 1.0) and the x-axis represents the sequence position (0 to 200). The plot shows a single peak at position 122, indicating a high probability of TM at this position.</p></div> <div>Phoebeius :</div>

	<div><div><div>ID UNNAMED</div><div>FT TOPO_DOM</div><div>//</div></div><div><div>1</div><div>201</div><div>NON CYTOPLASMIC.</div></div></div> <div><p>Phobius posterior probabilities for UNNAMED</p><table border="1"><caption>Phobius Posterior Probabilities Data</caption><thead><tr><th>Label</th><th>Probability</th></tr></thead><tbody><tr><td>transmembrane</td><td>~0.01</td></tr><tr><td>cytoplasmic</td><td>~0.01</td></tr><tr><td>non cytoplasmic</td><td>~0.18</td></tr><tr><td>signal peptide</td><td>~0.82</td></tr></tbody></table></div>	Label	Probability	transmembrane	~0.01	cytoplasmic	~0.01	non cytoplasmic	~0.18	signal peptide	~0.82
Label	Probability										
transmembrane	~0.01										
cytoplasmic	~0.01										
non cytoplasmic	~0.18										
signal peptide	~0.82										
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Yes										
DECISION:	<i>thymidylate kinase</i>										

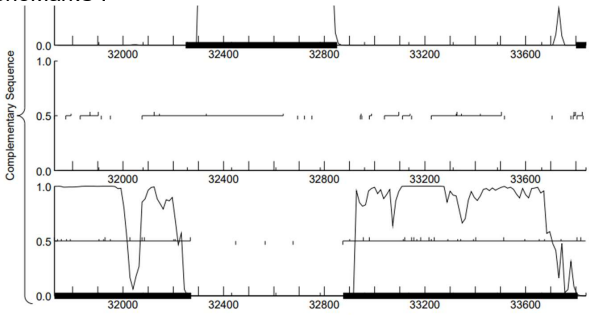


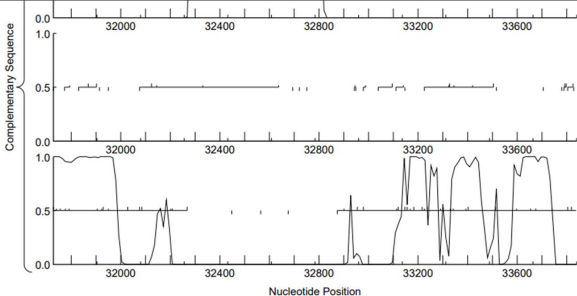
Student Gene Annotation Worksheet

Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	45
Coordonnées du Stop	32875
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	Chev 8
Coordonnées du Start retenu	33807
Fonction prédite	Glycosyltransferase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS :</p>  <p>GeneMark :</p>

	 <p>Pouvoir codant peu constant</p>									
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_45, function unknown, 310</td><td>629</td><td>e-180</td></tr><tr><td>Quartz_46, glycosyltransferase, 310</td><td>627</td><td>e-179</td></tr></table>		Score (bits)	E Value	Carostasia_Draft_45, function unknown, 310	629	e-180	Quartz_46, glycosyltransferase, 310	627	e-179
	Score (bits)	E Value								
Carostasia_Draft_45, function unknown, 310	629	e-180								
Quartz_46, glycosyltransferase, 310	627	e-179								
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions									
DECISION:	OUI									

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																																														
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 33807 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 33807																																																																																																																																																																																																																																																																														
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Stats: 28 ORF Start : 33807 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler6 Explore Selected: 1 ORF Stop : 32875 5' End 33.3 57.6 87.9 198 Spacing Weight Matrix Karlin Medium Document ORF Length: 933 3' End 62.2 77.8 66.7 135</div> <table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,642</td><td>2,095</td><td>17</td><td>-5,642</td><td>ACCATGGAGCGCTTCTCGCTG</td><td>GTG</td><td>34137</td><td>1263</td></tr><tr><td>2</td><td>-6,108</td><td>0,941</td><td>15</td><td>-7,710</td><td>GGCTCAGCCACTTCCACATCC</td><td>GTG</td><td>33939</td><td>1065</td></tr><tr><td>3</td><td>-5,296</td><td>1,321</td><td>12</td><td>-6,131</td><td>ATGCTGCCACCGAGCTTCTCT</td><td>ATG</td><td>33912</td><td>1038</td></tr><tr><td>4</td><td>-6,050</td><td>0,968</td><td>10</td><td>-6,745</td><td>GTGCGGCACTCCGTTACCGAGT</td><td>TTG</td><td>33831</td><td>957</td></tr><tr><td>5</td><td>-4,784</td><td>1,561</td><td>14</td><td>-6,131</td><td>TCCGTCACCGAGTTTGAATGGG</td><td>ATG</td><td>33822</td><td>948</td></tr><tr><td>6</td><td>-3,791</td><td>2,493</td><td>9</td><td>-3,566</td><td>GAATGGATGGCTGGAGGACC</td><td>ATG</td><td>33807</td><td>933</td></tr><tr><td>7</td><td>-3,420</td><td>2,199</td><td>17</td><td>-5,420</td><td>CAGCTGGATATGCCGATCTAC</td><td>GTG</td><td>33744</td><td>870</td></tr><tr><td>8</td><td>-5,650</td><td>1,155</td><td>13</td><td>-6,495</td><td>GCTCAGACGGCTCCGCCAGC</td><td>GTG</td><td>33678</td><td>804</td></tr><tr><td>9</td><td>-4,553</td><td>1,669</td><td>11</td><td>-5,310</td><td>CGTCAGACCGCGGTCACATC</td><td>GTG</td><td>33657</td><td>783</td></tr><tr><td>10</td><td>-4,663</td><td>1,617</td><td>11</td><td>-5,420</td><td>GCACTACCGCTGGGCGACACT</td><td>GTG</td><td>33600</td><td>726</td></tr><tr><td>11</td><td>-4,328</td><td>1,774</td><td>18</td><td>-6,429</td><td>GAGAGGCTACGAGCGCATGTA</td><td>ATG</td><td>33516</td><td>642</td></tr><tr><td>12</td><td>-1,907</td><td>2,907</td><td>7</td><td>-3,430</td><td>GGTCAGTGGCATCCAGGAGCCG</td><td>ATG</td><td>33405</td><td>531</td></tr><tr><td>13</td><td>-1,907</td><td>2,907</td><td>16</td><td>-3,703</td><td>CATCCAGGAGCGGATGCTGCTG</td><td>GTG</td><td>33396</td><td>522</td></tr><tr><td>14</td><td>-5,021</td><td>1,450</td><td>15</td><td>-6,423</td><td>CGCTCAGCGGCTCTTCTCCAGT</td><td>GTG</td><td>33345</td><td>471</td></tr><tr><td>15</td><td>-5,570</td><td>1,192</td><td>10</td><td>-6,265</td><td>CTTCTCCAGTGTGGCCGAGGCT</td><td>GTG</td><td>33333</td><td>459</td></tr><tr><td>16</td><td>-2,736</td><td>2,519</td><td>6</td><td>-4,481</td><td>CAACGCGCTCTTCAGCGGAGAC</td><td>GTG</td><td>33294</td><td>420</td></tr><tr><td>17</td><td>-5,808</td><td>1,081</td><td>10</td><td>-6,503</td><td>GGGCACTTCCCGGCTAGCGTC</td><td>ATG</td><td>33240</td><td>366</td></tr><tr><td>18</td><td>-5,321</td><td>1,309</td><td>11</td><td>-6,078</td><td>GGCTAGGCTCATGTTCTACGAC</td><td>GTG</td><td>33228</td><td>354</td></tr><tr><td>19</td><td>-5,369</td><td>1,287</td><td>10</td><td>-6,063</td><td>CATGTTCTACGAGCTGACGAGG</td><td>ATG</td><td>33215</td><td>345</td></tr><tr><td>20</td><td>-5,501</td><td>1,225</td><td>10</td><td>-6,195</td><td>CGAGCTGCTTAAGCGCTTCCAC</td><td>ATG</td><td>33186</td><td>312</td></tr><tr><td>21</td><td>-5,021</td><td>1,450</td><td>9</td><td>-5,796</td><td>CTTCCACATGCCAGGCGAGGAC</td><td>GTG</td><td>33171</td><td>297</td></tr><tr><td>22</td><td>-3,079</td><td>2,358</td><td>16</td><td>-4,875</td><td>CGGCGAGGAGCTGGGCTTCCGC</td><td>ATG</td><td>33155</td><td>285</td></tr><tr><td>23</td><td>-3,642</td><td>2,095</td><td>7</td><td>-5,164</td><td>CGTGGGCTTGGCATGAGGTC</td><td>ATG</td><td>33150</td><td>276</td></tr><tr><td>24</td><td>-4,000</td><td>1,928</td><td>8</td><td>-5,222</td><td>CGGTGGGCGAGCGTGGTTCAGG</td><td>ATG</td><td>33123</td><td>249</td></tr><tr><td>25</td><td>-4,000</td><td>1,928</td><td>14</td><td>-5,347</td><td>GCAGCAGTGGTTACGATGCCCT</td><td>GTG</td><td>33117</td><td>243</td></tr><tr><td>26</td><td>-6,008</td><td>0,988</td><td>8</td><td>-7,230</td><td>CCACCGGTTCTCCGTTAGTCC</td><td>ATG</td><td>32982</td><td>108</td></tr><tr><td>27</td><td>-4,025</td><td>1,916</td><td>9</td><td>-4,800</td><td>GAAGACAAGGCGAGGCGGGTC</td><td>ATG</td><td>32958</td><td>94</td></tr><tr><td>28</td><td>-6,304</td><td>0,849</td><td>13</td><td>-7,350</td><td>CAAGGCCACCGACTCCGAGCCG</td><td>GTG</td><td>32904</td><td>30</td></tr></table>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,642	2,095	17	-5,642	ACCATGGAGCGCTTCTCGCTG	GTG	34137	1263	2	-6,108	0,941	15	-7,710	GGCTCAGCCACTTCCACATCC	GTG	33939	1065	3	-5,296	1,321	12	-6,131	ATGCTGCCACCGAGCTTCTCT	ATG	33912	1038	4	-6,050	0,968	10	-6,745	GTGCGGCACTCCGTTACCGAGT	TTG	33831	957	5	-4,784	1,561	14	-6,131	TCCGTCACCGAGTTTGAATGGG	ATG	33822	948	6	-3,791	2,493	9	-3,566	GAATGGATGGCTGGAGGACC	ATG	33807	933	7	-3,420	2,199	17	-5,420	CAGCTGGATATGCCGATCTAC	GTG	33744	870	8	-5,650	1,155	13	-6,495	GCTCAGACGGCTCCGCCAGC	GTG	33678	804	9	-4,553	1,669	11	-5,310	CGTCAGACCGCGGTCACATC	GTG	33657	783	10	-4,663	1,617	11	-5,420	GCACTACCGCTGGGCGACACT	GTG	33600	726	11	-4,328	1,774	18	-6,429	GAGAGGCTACGAGCGCATGTA	ATG	33516	642	12	-1,907	2,907	7	-3,430	GGTCAGTGGCATCCAGGAGCCG	ATG	33405	531	13	-1,907	2,907	16	-3,703	CATCCAGGAGCGGATGCTGCTG	GTG	33396	522	14	-5,021	1,450	15	-6,423	CGCTCAGCGGCTCTTCTCCAGT	GTG	33345	471	15	-5,570	1,192	10	-6,265	CTTCTCCAGTGTGGCCGAGGCT	GTG	33333	459	16	-2,736	2,519	6	-4,481	CAACGCGCTCTTCAGCGGAGAC	GTG	33294	420	17	-5,808	1,081	10	-6,503	GGGCACTTCCCGGCTAGCGTC	ATG	33240	366	18	-5,321	1,309	11	-6,078	GGCTAGGCTCATGTTCTACGAC	GTG	33228	354	19	-5,369	1,287	10	-6,063	CATGTTCTACGAGCTGACGAGG	ATG	33215	345	20	-5,501	1,225	10	-6,195	CGAGCTGCTTAAGCGCTTCCAC	ATG	33186	312	21	-5,021	1,450	9	-5,796	CTTCCACATGCCAGGCGAGGAC	GTG	33171	297	22	-3,079	2,358	16	-4,875	CGGCGAGGAGCTGGGCTTCCGC	ATG	33155	285	23	-3,642	2,095	7	-5,164	CGTGGGCTTGGCATGAGGTC	ATG	33150	276	24	-4,000	1,928	8	-5,222	CGGTGGGCGAGCGTGGTTCAGG	ATG	33123	249	25	-4,000	1,928	14	-5,347	GCAGCAGTGGTTACGATGCCCT	GTG	33117	243	26	-6,008	0,988	8	-7,230	CCACCGGTTCTCCGTTAGTCC	ATG	32982	108	27	-4,025	1,916	9	-4,800	GAAGACAAGGCGAGGCGGGTC	ATG	32958	94	28	-6,304	0,849	13	-7,350	CAAGGCCACCGACTCCGAGCCG	GTG	32904	30
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																																																																																							
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																																																																																							
1	-3,642	2,095	17	-5,642	ACCATGGAGCGCTTCTCGCTG	GTG	34137	1263																																																																																																																																																																																																																																																																							
2	-6,108	0,941	15	-7,710	GGCTCAGCCACTTCCACATCC	GTG	33939	1065																																																																																																																																																																																																																																																																							
3	-5,296	1,321	12	-6,131	ATGCTGCCACCGAGCTTCTCT	ATG	33912	1038																																																																																																																																																																																																																																																																							
4	-6,050	0,968	10	-6,745	GTGCGGCACTCCGTTACCGAGT	TTG	33831	957																																																																																																																																																																																																																																																																							
5	-4,784	1,561	14	-6,131	TCCGTCACCGAGTTTGAATGGG	ATG	33822	948																																																																																																																																																																																																																																																																							
6	-3,791	2,493	9	-3,566	GAATGGATGGCTGGAGGACC	ATG	33807	933																																																																																																																																																																																																																																																																							
7	-3,420	2,199	17	-5,420	CAGCTGGATATGCCGATCTAC	GTG	33744	870																																																																																																																																																																																																																																																																							
8	-5,650	1,155	13	-6,495	GCTCAGACGGCTCCGCCAGC	GTG	33678	804																																																																																																																																																																																																																																																																							
9	-4,553	1,669	11	-5,310	CGTCAGACCGCGGTCACATC	GTG	33657	783																																																																																																																																																																																																																																																																							
10	-4,663	1,617	11	-5,420	GCACTACCGCTGGGCGACACT	GTG	33600	726																																																																																																																																																																																																																																																																							
11	-4,328	1,774	18	-6,429	GAGAGGCTACGAGCGCATGTA	ATG	33516	642																																																																																																																																																																																																																																																																							
12	-1,907	2,907	7	-3,430	GGTCAGTGGCATCCAGGAGCCG	ATG	33405	531																																																																																																																																																																																																																																																																							
13	-1,907	2,907	16	-3,703	CATCCAGGAGCGGATGCTGCTG	GTG	33396	522																																																																																																																																																																																																																																																																							
14	-5,021	1,450	15	-6,423	CGCTCAGCGGCTCTTCTCCAGT	GTG	33345	471																																																																																																																																																																																																																																																																							
15	-5,570	1,192	10	-6,265	CTTCTCCAGTGTGGCCGAGGCT	GTG	33333	459																																																																																																																																																																																																																																																																							
16	-2,736	2,519	6	-4,481	CAACGCGCTCTTCAGCGGAGAC	GTG	33294	420																																																																																																																																																																																																																																																																							
17	-5,808	1,081	10	-6,503	GGGCACTTCCCGGCTAGCGTC	ATG	33240	366																																																																																																																																																																																																																																																																							
18	-5,321	1,309	11	-6,078	GGCTAGGCTCATGTTCTACGAC	GTG	33228	354																																																																																																																																																																																																																																																																							
19	-5,369	1,287	10	-6,063	CATGTTCTACGAGCTGACGAGG	ATG	33215	345																																																																																																																																																																																																																																																																							
20	-5,501	1,225	10	-6,195	CGAGCTGCTTAAGCGCTTCCAC	ATG	33186	312																																																																																																																																																																																																																																																																							
21	-5,021	1,450	9	-5,796	CTTCCACATGCCAGGCGAGGAC	GTG	33171	297																																																																																																																																																																																																																																																																							
22	-3,079	2,358	16	-4,875	CGGCGAGGAGCTGGGCTTCCGC	ATG	33155	285																																																																																																																																																																																																																																																																							
23	-3,642	2,095	7	-5,164	CGTGGGCTTGGCATGAGGTC	ATG	33150	276																																																																																																																																																																																																																																																																							
24	-4,000	1,928	8	-5,222	CGGTGGGCGAGCGTGGTTCAGG	ATG	33123	249																																																																																																																																																																																																																																																																							
25	-4,000	1,928	14	-5,347	GCAGCAGTGGTTACGATGCCCT	GTG	33117	243																																																																																																																																																																																																																																																																							
26	-6,008	0,988	8	-7,230	CCACCGGTTCTCCGTTAGTCC	ATG	32982	108																																																																																																																																																																																																																																																																							
27	-4,025	1,916	9	-4,800	GAAGACAAGGCGAGGCGGGTC	ATG	32958	94																																																																																																																																																																																																																																																																							
28	-6,304	0,849	13	-7,350	CAAGGCCACCGACTCCGAGCCG	GTG	32904	30																																																																																																																																																																																																																																																																							
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start choisi ne donne pas l'ORF le plus long mais son final score est meilleur et correspond à un plus haut départ de potentiel codant. C'est un ATG. Les starts 34137,33939, 33912 et 33831 impliquent des chevauchements avec le gène précédent. Le start 33822 s'éloigne du départ du pouvoir codant.																																																																																																																																																																																																																																																																														
Est-ce que le start est conservé chez les homologues voir Starterator ?	• Track 33 : Carostasia_45, Nucci_45																																																																																																																																																																																																																																																																														

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>>Quartz_46, glycosyltransferase, 310 Length = 310</p> <p>Score = 627 bits (1617), Expect = e-179 Identities = 309/310 (99%), Positives = 310/310 (100%)</p> <p>Query: 1 MSNEHVKLLRRAKSSWIWPIYVPSYTRGGTAPFLNLLQTAPASVQRRVHIVRREERKLY 60 MSNEHVKLLRRAKSSWIWPIYVPSYTRGGTAPFLNLLQTAPASVQRRVHIVRREERKLY Sbjct: 1 MSNEHVKLLRRAKSSWIWPIYVPSYTRGGTAPFLNLLQTAPASVQRRVHIVRREERKLY 60</p> <p>Query: 61 QKHYPWATTVLVRRPGLGPARTQALEDALERGYERIVMFDDIRHLSLLERIAREGKPDH 120 QKHYPWATTVLVRRPGLGPARTQALEDALERGYERIVMFDDIRHLSLLERIAREGKPDH Sbjct: 61 QKHYPWATTVLVRRPGLGPARTQALEDALERGYERIVMFDDIRHLSLLERIAREGKPDH 120</p> <p>Query: 121 TRRYSTGVSGIQEPHLLVRS LAVACRLADGVFSSVADAVYGAARNALFSGDVTDSIGATL 180 TRRYSTGVSGIQEPHLLVRS LAVACRLADGVFSSVADAVYGAARNALFSGDVTDSIGATL Sbjct: 121 TRRYSTGVSGIQEPHLLVRS LAVACRLADGVFSSVADAVYGAARNALFSGDVTDSIGATL 180</p> <p>Query: 181 NKGTFFPASVMFYDVTIRMENFELPKRFHMHGEDVAFANDVMTRGQONFTMPVVAFDQDQNI 240 NKGTFFPASVMFYDVTIRMENFELPKRFHMHGEDVAFANDVMTRGQONFTMPVVAFDQDQNI Sbjct: 181 NKGTFFPASVMFYDVTIRMENFELPKRFHMHGEDVAFANDVMTRGQONFTMPVVAFDQDQNI 240</p> <p>Query: 241 ETIIPLDPTDPEARRVDIENAAIDYDPIHPFLRESMKNKAGGVMRVGVNMRQWYKATDSE 300 ETIIPLDPTDPEARRVDIENAAIDYDPIHPFLRESMKNKAGGVMRVGVNMRQWYKATDSE Sbjct: 241 ETIIPLDPTDPEARRVDIENAAIDYDPIHPFLRESMKNKAGGVMRVGVNMRQWYKATDSE 300</p> <p>Query: 301 PVELPLADLL 310 PVELPLADLL Sbjct: 301 PVELPLADLL 310</p>
DECISION:	33807, départ du pouvoir codant, start conservé chez d'autres phages.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																				
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <table><thead><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Carostasia_Draft_45, function unknown, 310</td><td>629</td><td>e-180</td></tr><tr><td>Quartz_46, glycosyltransferase, 310</td><td>627</td><td>e-179</td></tr><tr><td>Nucci_45, glycosyltransferase, 310</td><td>627</td><td>e-179</td></tr><tr><td>Mandalorian_45, glycosyltransferase, 310</td><td>627</td><td>e-179</td></tr><tr><td>YuuY_46, glycosyltransferase, 310</td><td>612</td><td>e-175</td></tr><tr><td>Cheeto1_48, glycosyltransferase, 309</td><td>385</td><td>e-107</td></tr><tr><td>GaeCeo_48, glycosyltransferase, 309</td><td>384</td><td>e-106</td></tr><tr><td>ChickenKing_46, glycosyltransferase, 309</td><td>384</td><td>e-106</td></tr><tr><td>Schubert_45, glycosyltransferase, 317</td><td>381</td><td>e-105</td></tr></tbody></table> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p> <table><thead><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per Ident</th><th>Acc. Len</th><th>Accession</th></tr></thead><tbody><tr><td><input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Nucl]</td><td>Mombacterium phage Nucl</td><td>638</td><td>638</td><td>99%</td><td>0.0</td><td>99.68%</td><td>310</td><td>GD053438.1</td></tr><tr><td><input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Yau1]</td><td>Mombacterium phage Yau1</td><td>621</td><td>621</td><td>99%</td><td>0.0</td><td>96.13%</td><td>310</td><td>QF593314.1</td></tr><tr><td><input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Cheeto1]</td><td>Mombacterium phage Cheeto1</td><td>396</td><td>396</td><td>99%</td><td>6e-133</td><td>60.71%</td><td>309</td><td>UW0334611.1</td></tr><tr><td><input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage ChickenKing]</td><td>Mombacterium phage ChickenKing</td><td>392</td><td>392</td><td>99%</td><td>1e-133</td><td>60.06%</td><td>309</td><td>QF5934716.1</td></tr><tr><td><input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Etna]</td><td>Mombacterium phage Etna</td><td>388</td><td>388</td><td>99%</td><td>6e-132</td><td>60.19%</td><td>310</td><td>AW556313.1</td></tr></tbody></table>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_45, function unknown, 310	629	e-180	Quartz_46, glycosyltransferase, 310	627	e-179	Nucci_45, glycosyltransferase, 310	627	e-179	Mandalorian_45, glycosyltransferase, 310	627	e-179	YuuY_46, glycosyltransferase, 310	612	e-175	Cheeto1_48, glycosyltransferase, 309	385	e-107	GaeCeo_48, glycosyltransferase, 309	384	e-106	ChickenKing_46, glycosyltransferase, 309	384	e-106	Schubert_45, glycosyltransferase, 317	381	e-105	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Nucl]	Mombacterium phage Nucl	638	638	99%	0.0	99.68%	310	GD053438.1	<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Yau1]	Mombacterium phage Yau1	621	621	99%	0.0	96.13%	310	QF593314.1	<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Cheeto1]	Mombacterium phage Cheeto1	396	396	99%	6e-133	60.71%	309	UW0334611.1	<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage ChickenKing]	Mombacterium phage ChickenKing	392	392	99%	1e-133	60.06%	309	QF5934716.1	<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Etna]	Mombacterium phage Etna	388	388	99%	6e-132	60.19%	310	AW556313.1
Sequences producing significant alignments:	Score (bits)	E Value																																																																																			
Carostasia_Draft_45, function unknown, 310	629	e-180																																																																																			
Quartz_46, glycosyltransferase, 310	627	e-179																																																																																			
Nucci_45, glycosyltransferase, 310	627	e-179																																																																																			
Mandalorian_45, glycosyltransferase, 310	627	e-179																																																																																			
YuuY_46, glycosyltransferase, 310	612	e-175																																																																																			
Cheeto1_48, glycosyltransferase, 309	385	e-107																																																																																			
GaeCeo_48, glycosyltransferase, 309	384	e-106																																																																																			
ChickenKing_46, glycosyltransferase, 309	384	e-106																																																																																			
Schubert_45, glycosyltransferase, 317	381	e-105																																																																																			
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession																																																																													
<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Nucl]	Mombacterium phage Nucl	638	638	99%	0.0	99.68%	310	GD053438.1																																																																													
<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Yau1]	Mombacterium phage Yau1	621	621	99%	0.0	96.13%	310	QF593314.1																																																																													
<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Cheeto1]	Mombacterium phage Cheeto1	396	396	99%	6e-133	60.71%	309	UW0334611.1																																																																													
<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage ChickenKing]	Mombacterium phage ChickenKing	392	392	99%	1e-133	60.06%	309	QF5934716.1																																																																													
<input checked="" type="checkbox"/> glycosyltransferase [Mombacterium phage Etna]	Mombacterium phage Etna	388	388	99%	6e-132	60.19%	310	AW556313.1																																																																													

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	Q06718	GSTG_BPT6 Beta-glucosyl-HMC-alpha-glucosyl-transferase O5-Enterobacteria phage T6 OX=10666 PE=4 SV=1	99.95	2.6e-26	208.41	21.2	252	280
2	PF20691.1	; TAGT ; TET-Associated Glycosyltransferase	99.48	6.2e-12	108.33	16.6	189	212
3	3BCV_B	Putative glycosyltransferase protein; Glycosyltransferase, Protein Structure Initiative II, PSI-II, 12059a, NYSGXRC, Str	99.19	1.6e-9	92.83	14.6	193	240
4	cd04184	GT2_RfbC_Mn_like: Myxococcus xanthus RfbC-like proteins are required for O-antigen biosynthesis. The rfbC gene encodes a	99.15	7.6e-9	86.63	16.5	189	202
5	6P61_D	Glycosyltransferase: SSGCID, Structural Genomics, Seattle Structural Genomics Center for Infectious Disease, TRANSFERASE	99.06	7.6e-8	82.56	19	213	228

OUI

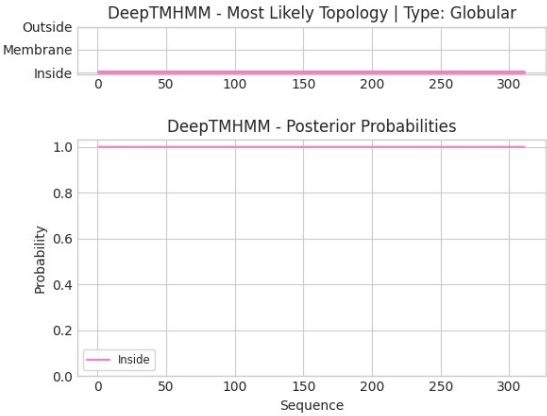
Visualization

Resubmit Section

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Oui, modifications des molécules.

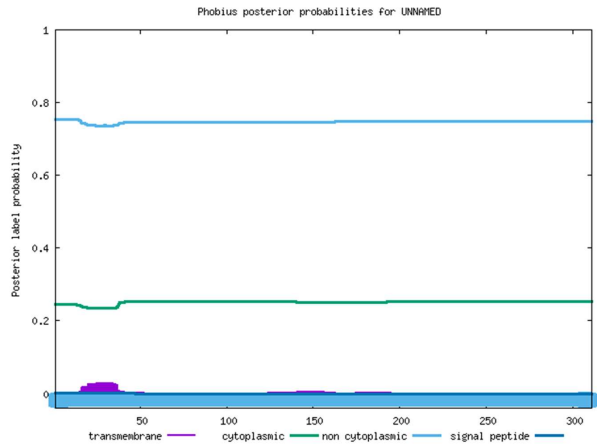
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



NON

Prediction of UNNAMED

ID UNNAMED
FT TOPO_DOM 1 311 NON CYTOPLASMIC.
//



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

oui

DECISION:

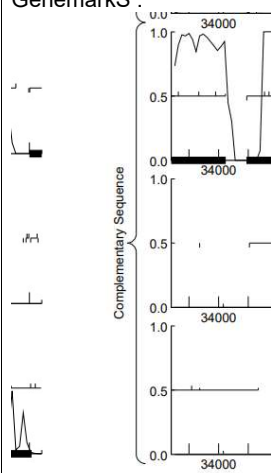
Glycosyltransferase

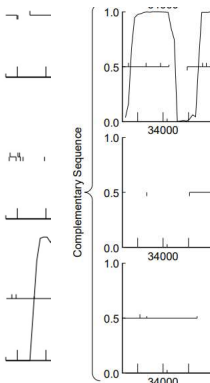
Student Gene Annotation Worksheet

Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	46
Coordonnées du Stop	33800
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	Gap de 69
Coordonnées du Start retenu	34024
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenemarkS :</p>  <p>GeneMark :</p>

													
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_47, function unknown, 74</td><td>162</td><td>2e-40</td></tr><tr><td>Mandalorian_46, function unknown, 74</td><td>162</td><td>2e-40</td></tr><tr><td>Carostasia_Draft_46, function unknown, 74</td><td>162</td><td>2e-40</td></tr></table>		Score (bits)	E Value	Quartz_47, function unknown, 74	162	2e-40	Mandalorian_46, function unknown, 74	162	2e-40	Carostasia_Draft_46, function unknown, 74	162	2e-40
	Score (bits)	E Value											
Quartz_47, function unknown, 74	162	2e-40											
Mandalorian_46, function unknown, 74	162	2e-40											
Carostasia_Draft_46, function unknown, 74	162	2e-40											
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions.												
DECISION:	OUI												

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 34024 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 34024																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Stats : 4 Selected : 1</div><div>ORF Start : 34093 ORF Stop : 33800 ORF Length : 294</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 90.9 81.8 27.3 33 3' End 68.6 82.9 51.4 105</div><div>SD Scoring Matrix Kibler6 Spacing Weight Matrix Karlin Medium</div></div> <table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,460</td><td>2,648</td><td>13</td><td>-3,506</td><td>CCACTATCAGGGAGTCCACCCC</td><td>GTG</td><td>34024</td><td>225</td></tr><tr><td>2</td><td>-5,046</td><td>1,438</td><td>16</td><td>-6,842</td><td>CATCAAGCTGTTCGGCGGTCCC</td><td>ATG</td><td>33991</td><td>192</td></tr><tr><td>3</td><td>-6,008</td><td>0,988</td><td>5</td><td>-8,008</td><td>CAGCCACTTCACATCCGTGAG</td><td>ATG</td><td>33994</td><td>135</td></tr><tr><td>4</td><td>-2,972</td><td>2,409</td><td>8</td><td>-4,194</td><td>CCTGGTCCGCACACGGGACGGC</td><td>ATG</td><td>33865</td><td>66</td></tr></table>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,460	2,648	13	-3,506	CCACTATCAGGGAGTCCACCCC	GTG	34024	225	2	-5,046	1,438	16	-6,842	CATCAAGCTGTTCGGCGGTCCC	ATG	33991	192	3	-6,008	0,988	5	-8,008	CAGCCACTTCACATCCGTGAG	ATG	33994	135	4	-2,972	2,409	8	-4,194	CCTGGTCCGCACACGGGACGGC	ATG	33865	66
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																															
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																															
1	-2,460	2,648	13	-3,506	CCACTATCAGGGAGTCCACCCC	GTG	34024	225																																															
2	-5,046	1,438	16	-6,842	CATCAAGCTGTTCGGCGGTCCC	ATG	33991	192																																															
3	-6,008	0,988	5	-8,008	CAGCCACTTCACATCCGTGAG	ATG	33994	135																																															
4	-2,972	2,409	8	-4,194	CCTGGTCCGCACACGGGACGGC	ATG	33865	66																																															
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ?	Le start choisit induit l'ORF le plus long mais implique un chevauchement de 7pb.																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?	• Track 12 : Mandalorian_46, Carostasia_46, Quartz_47 • Track 13 : Mucosa_46																																																						
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	>Quartz_47, function unknown, 74 Length = 74 Score = 162 bits (411), Expect = 2e-40 Identities = 74/74 (100%), Positives = 74/74 (100%) Query: 1 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH Sbjct: 1 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60 Query: 61 SRTEFEWDGWSHE 74 SRTEFEWDGWSHE Sbjct: 61 SRTEFEWDGWSHE 74																																																						
DECISION:	34024																																																						

Commenté [10]: justification

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
------------------------------------	-------------

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

*** PhagesDB :
(q# Quartz_47, function unknown, 74 : s#162) : [100%] ; e-value : 2e-40

>Quartz_47, function unknown, 74
Length = 74

Score = 162 bits (411), Expect = 2e-40
Identities = 74/74 (100%), Positives = 74/74 (100%)

Query: 1 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60
MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60
Sbjct: 1 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60

Query: 61 SRTEFEWDGWSHE 74
SRTEFEWDGWSHE
Sbjct: 61 SRTEFEWDGWSHE 74

*** nr :
(q# hypothetical protein SEA_MANDALORIAN_46
[Microbacterium phage Mandalorian]: s#156) : [100%] ; e-value : 1e-47

hypothetical protein SEA_MANDALORIAN_46 [Microbacterium phage Mandalorian]
Sequence ID: [OX013638.1](#) Length: 74 Number of Matches: 1
[See 1 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 74 [GenPept](#) [Graphics](#) [Next Match](#) [Pre](#)

Score	Expect	Method	Identities	Positives	Gaps
156 bits(394)	1e-47	Compositional matrix adjust.	74/74(100%)	74/74(100%)	0/74(0%)

Query 1 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60
MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60
Sbjct 1 MSQHIKLFGGPMHGKEIALEDGVSHFHIREMLPPSFVYGETNNPADLVRTRDGMYSRVGH 60

Query 61 SRTEFEWDGWSHE 74
SRTEFEWDGWSHE
Sbjct 61 SRTEFEWDGWSHE 74

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un criblé HHPred avec une proba >= 90% et une couverture acceptable ?

Pas de correspondance pertinente

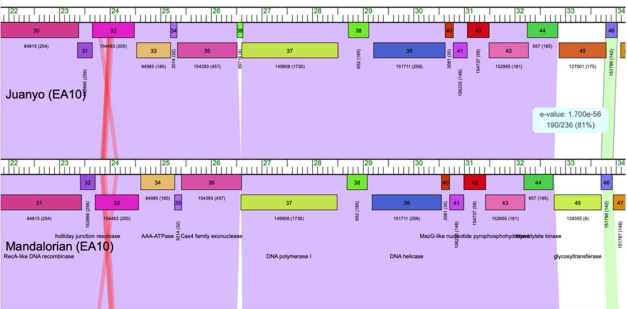
<input type="checkbox"/>	1	4G25_F	Protein prgH; FHA domain, CELL INVASION; 1.858A (Salmonella enterica subsp. enterica serovar Typhimurium)	63.87	37	22.44	3.7	28	111
<input type="checkbox"/>	2	5GM9_A	Glycoside hydrolase family 45 protein; substrate binding, cellulase, inhibitor, HYDROLASE; HET: GLC, BGC; 1.36A (Thielav	51.45	59	26.2	3.6	28	213
<input type="checkbox"/>	3	5H4U_A	Endo-beta-1,4-glucanase; cold-active, Cellulase, Antarctic springtail, Cryptosporidius antarcticus, HYDROLASE; 2.6A (Cryptop	41.36	110	25.2	3.6	28	231
<input type="checkbox"/>	4	6MVJ_A	Endoglucanase V; Endoglucanase V Glycoside hydrolase 45 family A Cellobiose complex, HYDROLASE; HET: BGC; 1.809A (Neuro	41.1	100	25.54	3.5	28	293
<input type="checkbox"/>	5	PFO2015.20	; Glyco_hydro_45; Glycosyl hydrolase family 45	40.12	110	24.58	3.5	28	210

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Aucune synténie observée

Région conservée entre divers phages EA10 e-value = 1,7 e-56

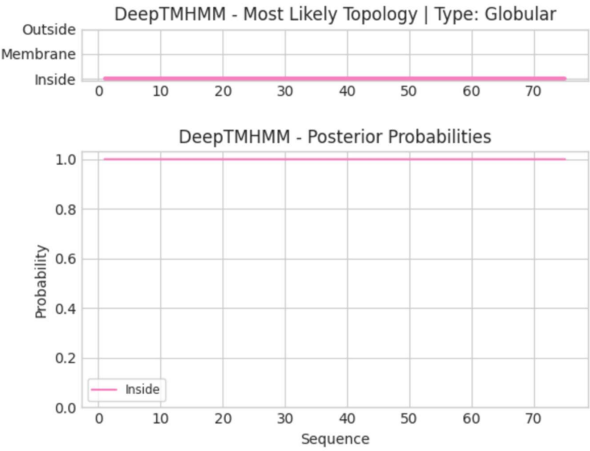
Carostasia_Draft (EA10)



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

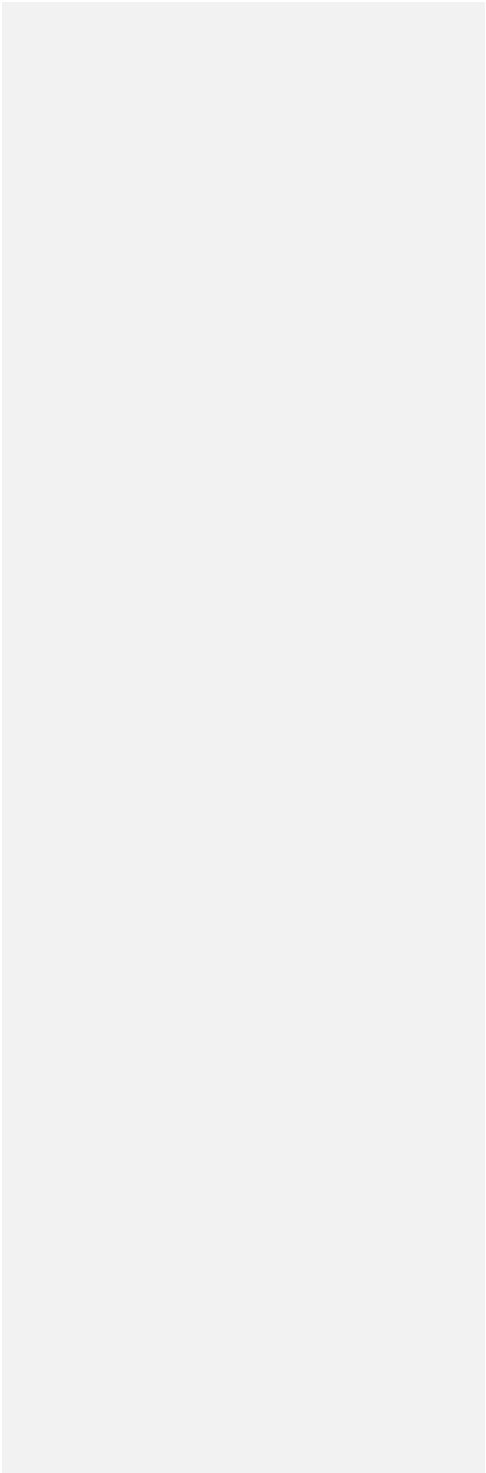
No

DeepTMHMM :



Phoebeius :

	<div><div><div>ID</div><div>UNNAMED</div></div><div><div>FT</div><div>TOPO_DOM</div></div><div><div>//</div><div></div></div></div> <div><div>1</div><div>75</div></div> <div><div>NON CYTOPLASMIC.</div></div>															
	<div><div>Phobius posterior probabilities for UNNAMED</div><table><caption>Phobius Posterior Probabilities Data</caption><tr><th>Residue Range</th><th>Label</th><th>Posterior Probability</th></tr><tr><td>0 - 70</td><td>transmembrane</td><td>~0.01</td></tr><tr><td>0 - 70</td><td>cytoplasmic</td><td>~0.39</td></tr><tr><td>0 - 70</td><td>non cytoplasmic</td><td>~0.62</td></tr><tr><td>0 - 70</td><td>signal peptide</td><td>~0.62</td></tr></table></div>	Residue Range	Label	Posterior Probability	0 - 70	transmembrane	~0.01	0 - 70	cytoplasmic	~0.39	0 - 70	non cytoplasmic	~0.62	0 - 70	signal peptide	~0.62
Residue Range	Label	Posterior Probability														
0 - 70	transmembrane	~0.01														
0 - 70	cytoplasmic	~0.39														
0 - 70	non cytoplasmic	~0.62														
0 - 70	signal peptide	~0.62														
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?																
DECISION:	NKF															

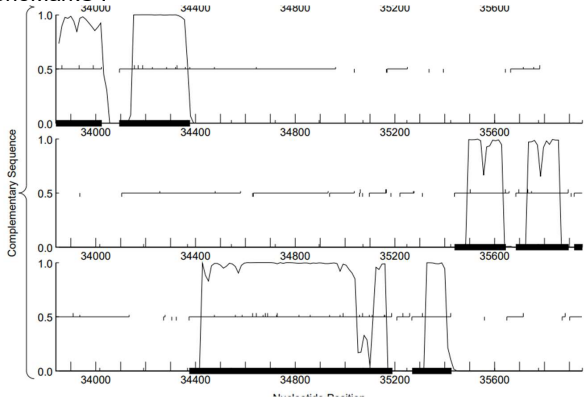


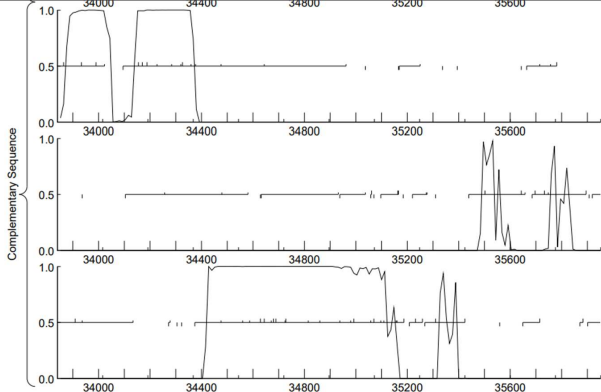
Student Gene Annotation Worksheet

Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	47
Coordonnées du Stop	34094
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	Chev 4
Coordonnées du Start retenu	34378
Fonction prédite	<i>Membrane protein</i>

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS :</p>  <p>GeneMark :</p>

													
	Fort potentiel codant stable												
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_48, function unknown, 94</td><td>199</td><td>2e-51</td></tr><tr><td>Mandalorian_47, function unknown, 94</td><td>199</td><td>2e-51</td></tr><tr><td>Carostasia Draft 47, function unknown, 94</td><td>199</td><td>2e-51</td></tr></table>		Score (bits)	E Value	Quartz_48, function unknown, 94	199	2e-51	Mandalorian_47, function unknown, 94	199	2e-51	Carostasia Draft 47, function unknown, 94	199	2e-51
	Score (bits)	E Value											
Quartz_48, function unknown, 94	199	2e-51											
Mandalorian_47, function unknown, 94	199	2e-51											
Carostasia Draft 47, function unknown, 94	199	2e-51											
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions												
DECISION:	OUI												

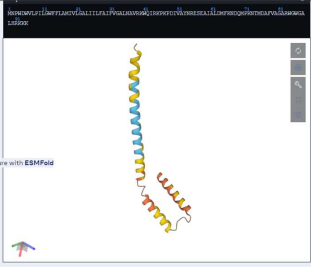
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																							
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 34378 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 34378																																																																																																																																							
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>Starts: 13 Selected: 1</div> <div>ORF Start : 34477 ORF Stop : 34094 ORF Length: 384</div> <div>Cdn1 Cdn2 Cdn3 Length 5' End 77.3 50.0 54.5 66 3' End 70.6 49.1 65.6 489</div> <div>SD Scoring Matrix: Kibler6 Spacing Weight Matrix: Karlin Medium</div> <div>Explore Document</div> <table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,857</td><td>1,526</td><td>6</td><td>-6,602</td><td>ACTGGGTTGGCGGTGAGCTGGC</td><td>TTG</td><td>35029</td><td>936</td></tr><tr><td>2</td><td>-5,046</td><td>1,438</td><td>12</td><td>-5,892</td><td>CGGTGCGGAAGCTGTTCCAGGC</td><td>GTG</td><td>34963</td><td>870</td></tr><tr><td>3</td><td>-4,234</td><td>1,818</td><td>12</td><td>-5,070</td><td>GCTACCGCGCGGACTACAGCAT</td><td>GTG</td><td>34645</td><td>552</td></tr><tr><td>4</td><td>-5,918</td><td>1,030</td><td>12</td><td>-6,754</td><td>AGGAGCGGCTGGCCCGTGTGT</td><td>GTG</td><td>34477</td><td>884</td></tr><tr><td>5</td><td>-3,311</td><td>2,250</td><td>12</td><td>-4,147</td><td>CGGAGACTGGACTTGGAGCT</td><td>GTG</td><td>34378</td><td>285</td></tr><tr><td>6</td><td>-3,857</td><td>1,594</td><td>8</td><td>-5,075</td><td>AGCTGTGAACCGCTGGGACTGG</td><td>GTG</td><td>34360</td><td>267</td></tr><tr><td>7</td><td>-4,325</td><td>1,775</td><td>14</td><td>-5,672</td><td>CTTGGCTGGTTCTTCTGGCC</td><td>ATG</td><td>34327</td><td>234</td></tr><tr><td>8</td><td>-5,918</td><td>1,030</td><td>10</td><td>-6,613</td><td>CTGGTTCTTCTGGCCATGATC</td><td>GTG</td><td>34321</td><td>228</td></tr><tr><td>9</td><td>-6,609</td><td>0,706</td><td>13</td><td>-7,655</td><td>GATCATACTTGGCCATCTTC</td><td>GTG</td><td>34285</td><td>192</td></tr><tr><td>10</td><td>-4,070</td><td>1,895</td><td>7</td><td>-5,593</td><td>CGGAGCGCCAGCGGACATC</td><td>GTG</td><td>34228</td><td>135</td></tr><tr><td>11</td><td>-6,462</td><td>0,775</td><td>15</td><td>-9,064</td><td>CTGGAGCCATCGCTCTGGAC</td><td>ATG</td><td>34189</td><td>96</td></tr><tr><td>12</td><td>-3,990</td><td>1,932</td><td>11</td><td>-4,747</td><td>GGACATGTTCAAGAACATCAG</td><td>ATG</td><td>34171</td><td>78</td></tr><tr><td>13</td><td>-4,502</td><td>1,693</td><td>8</td><td>-5,724</td><td>CGATCAGATGCCCAAGAACCC</td><td>ATG</td><td>34156</td><td>63</td></tr></table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,857	1,526	6	-6,602	ACTGGGTTGGCGGTGAGCTGGC	TTG	35029	936	2	-5,046	1,438	12	-5,892	CGGTGCGGAAGCTGTTCCAGGC	GTG	34963	870	3	-4,234	1,818	12	-5,070	GCTACCGCGCGGACTACAGCAT	GTG	34645	552	4	-5,918	1,030	12	-6,754	AGGAGCGGCTGGCCCGTGTGT	GTG	34477	884	5	-3,311	2,250	12	-4,147	CGGAGACTGGACTTGGAGCT	GTG	34378	285	6	-3,857	1,594	8	-5,075	AGCTGTGAACCGCTGGGACTGG	GTG	34360	267	7	-4,325	1,775	14	-5,672	CTTGGCTGGTTCTTCTGGCC	ATG	34327	234	8	-5,918	1,030	10	-6,613	CTGGTTCTTCTGGCCATGATC	GTG	34321	228	9	-6,609	0,706	13	-7,655	GATCATACTTGGCCATCTTC	GTG	34285	192	10	-4,070	1,895	7	-5,593	CGGAGCGCCAGCGGACATC	GTG	34228	135	11	-6,462	0,775	15	-9,064	CTGGAGCCATCGCTCTGGAC	ATG	34189	96	12	-3,990	1,932	11	-4,747	GGACATGTTCAAGAACATCAG	ATG	34171	78	13	-4,502	1,693	8	-5,724	CGATCAGATGCCCAAGAACCC	ATG	34156	63
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																
1	-4,857	1,526	6	-6,602	ACTGGGTTGGCGGTGAGCTGGC	TTG	35029	936																																																																																																																																
2	-5,046	1,438	12	-5,892	CGGTGCGGAAGCTGTTCCAGGC	GTG	34963	870																																																																																																																																
3	-4,234	1,818	12	-5,070	GCTACCGCGCGGACTACAGCAT	GTG	34645	552																																																																																																																																
4	-5,918	1,030	12	-6,754	AGGAGCGGCTGGCCCGTGTGT	GTG	34477	884																																																																																																																																
5	-3,311	2,250	12	-4,147	CGGAGACTGGACTTGGAGCT	GTG	34378	285																																																																																																																																
6	-3,857	1,594	8	-5,075	AGCTGTGAACCGCTGGGACTGG	GTG	34360	267																																																																																																																																
7	-4,325	1,775	14	-5,672	CTTGGCTGGTTCTTCTGGCC	ATG	34327	234																																																																																																																																
8	-5,918	1,030	10	-6,613	CTGGTTCTTCTGGCCATGATC	GTG	34321	228																																																																																																																																
9	-6,609	0,706	13	-7,655	GATCATACTTGGCCATCTTC	GTG	34285	192																																																																																																																																
10	-4,070	1,895	7	-5,593	CGGAGCGCCAGCGGACATC	GTG	34228	135																																																																																																																																
11	-6,462	0,775	15	-9,064	CTGGAGCCATCGCTCTGGAC	ATG	34189	96																																																																																																																																
12	-3,990	1,932	11	-4,747	GGACATGTTCAAGAACATCAG	ATG	34171	78																																																																																																																																
13	-4,502	1,693	8	-5,724	CGATCAGATGCCCAAGAACCC	ATG	34156	63																																																																																																																																
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le pouvoir codant démarre au start à 34378. L'ORF le plus long impliquerait un chevauchement très grand avec un autre gène.																																																																																																																																							
Est-ce que le start est conservé chez les homologues voir Starterator ?	<ul style="list-style-type: none">Track 28 : Carostasia_47, Quartz_48Start 19:Found in 5 of 148 (3.4%) of genes in phamManual Annotations of this start: 4 of 127Called 100.0% of time when presentPhage (with cluster) where this start called: Carostasia_47 (EA10), Mandalorian_47 (EA10), Nucci_47 (EA10), Quartz_48 (EA10), Yuyu_48 (EA10),																																																																																																																																							

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>>Quartz_48, function unknown, 94 Length = 94</p> <p>Score = 199 bits (507), Expect = 2e-51 Identities = 94/94 (100%), Positives = 94/94 (100%)</p> <p>Query: 1 MNPWDWVLPILGWFFLAMIVLGALIIILFAIFVGALNAVRKWQIRKPKPDIVAYNRESEAI 60 MNPWDWVLPILGWFFLAMIVLGALIIILFAIFVGALNAVRKWQIRKPKPDIVAYNRESEAI Sbjct: 1 MNPWDWVLPILGWFFLAMIVLGALIIILFAIFVGALNAVRKWQIRKPKPDIVAYNRESEAI 60</p> <p>Query: 61 ALDMFKNDQMPKNTMDAFVAGARWGWALHRKKK 94 ALDMFKNDQMPKNTMDAFVAGARWGWALHRKKK Sbjct: 61 ALDMFKNDQMPKNTMDAFVAGARWGWALHRKKK 94</p>
DECISION:	34378. D'après le départ du pouvoir codant, et pour éviter le chevauchement de gènes.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																											
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>(q#: s#) : [alignment] ; e-value :</p> <table><thead><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Quartz_48, function unknown, 94</td><td><u>199</u></td><td>2e-51</td></tr><tr><td>Mandalorian_47, function unknown, 94</td><td><u>199</u></td><td>2e-51</td></tr><tr><td>Carostasia_Draft_47, function unknown, 94</td><td><u>199</u></td><td>2e-51</td></tr><tr><td>Nucci_47, function unknown, 94</td><td><u>198</u></td><td>5e-51</td></tr><tr><td>YuuY_48, function unknown, 93</td><td><u>191</u></td><td>4e-49</td></tr><tr><td>SirVictor_47, function unknown, 98</td><td><u>92</u></td><td>6e-19</td></tr></tbody></table> <p>*** nr :</p> <p>(q#: s#) : [alignment] ; e-value :</p> <table><thead><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Pos. Ident</th><th>Acc. Len</th><th>Accession</th></tr></thead><tbody><tr><td><input checked="" type="checkbox"/> membrane protein (Microbacterium phage Mandalorian)</td><td>Microbacterium phage Mandalorian</td><td>189</td><td>189</td><td>98%</td><td>3e-60</td><td>100.00%</td><td>94</td><td>QX013639.1</td></tr><tr><td><input checked="" type="checkbox"/> membrane protein (Microbacterium phage Nucci)</td><td>Microbacterium phage Nucci</td><td>188</td><td>188</td><td>98%</td><td>6e-60</td><td>98.94%</td><td>94</td><td>QX024440.1</td></tr><tr><td><input checked="" type="checkbox"/> membrane protein (Microbacterium phage YuuY)</td><td>Microbacterium phage YuuY</td><td>182</td><td>182</td><td>97%</td><td>2e-57</td><td>94.62%</td><td>93</td><td>QF756376.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein T3J42_w48 (Microbacterium phage Golden)</td><td>Microbacterium phage Golden</td><td>89.0</td><td>89.0</td><td>96%</td><td>2e-20</td><td>45.92%</td><td>90</td><td>YP_009524187.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SCA_SIRVICTOR_47 (Microbacterium phage SirVictor)</td><td>Microbacterium phage SirVictor</td><td>89.0</td><td>89.0</td><td>96%</td><td>3e-20</td><td>45.92%</td><td>90</td><td>U0885019.1</td></tr></tbody></table>	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_48, function unknown, 94	<u>199</u>	2e-51	Mandalorian_47, function unknown, 94	<u>199</u>	2e-51	Carostasia_Draft_47, function unknown, 94	<u>199</u>	2e-51	Nucci_47, function unknown, 94	<u>198</u>	5e-51	YuuY_48, function unknown, 93	<u>191</u>	4e-49	SirVictor_47, function unknown, 98	<u>92</u>	6e-19	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Pos. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Mandalorian)	Microbacterium phage Mandalorian	189	189	98%	3e-60	100.00%	94	QX013639.1	<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Nucci)	Microbacterium phage Nucci	188	188	98%	6e-60	98.94%	94	QX024440.1	<input checked="" type="checkbox"/> membrane protein (Microbacterium phage YuuY)	Microbacterium phage YuuY	182	182	97%	2e-57	94.62%	93	QF756376.1	<input checked="" type="checkbox"/> hypothetical protein T3J42_w48 (Microbacterium phage Golden)	Microbacterium phage Golden	89.0	89.0	96%	2e-20	45.92%	90	YP_009524187.1	<input checked="" type="checkbox"/> hypothetical protein SCA_SIRVICTOR_47 (Microbacterium phage SirVictor)	Microbacterium phage SirVictor	89.0	89.0	96%	3e-20	45.92%	90	U0885019.1
Sequences producing significant alignments:	Score (bits)	E Value																																																																										
Quartz_48, function unknown, 94	<u>199</u>	2e-51																																																																										
Mandalorian_47, function unknown, 94	<u>199</u>	2e-51																																																																										
Carostasia_Draft_47, function unknown, 94	<u>199</u>	2e-51																																																																										
Nucci_47, function unknown, 94	<u>198</u>	5e-51																																																																										
YuuY_48, function unknown, 93	<u>191</u>	4e-49																																																																										
SirVictor_47, function unknown, 98	<u>92</u>	6e-19																																																																										
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Pos. Ident	Acc. Len	Accession																																																																				
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Mandalorian)	Microbacterium phage Mandalorian	189	189	98%	3e-60	100.00%	94	QX013639.1																																																																				
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage Nucci)	Microbacterium phage Nucci	188	188	98%	6e-60	98.94%	94	QX024440.1																																																																				
<input checked="" type="checkbox"/> membrane protein (Microbacterium phage YuuY)	Microbacterium phage YuuY	182	182	97%	2e-57	94.62%	93	QF756376.1																																																																				
<input checked="" type="checkbox"/> hypothetical protein T3J42_w48 (Microbacterium phage Golden)	Microbacterium phage Golden	89.0	89.0	96%	2e-20	45.92%	90	YP_009524187.1																																																																				
<input checked="" type="checkbox"/> hypothetical protein SCA_SIRVICTOR_47 (Microbacterium phage SirVictor)	Microbacterium phage SirVictor	89.0	89.0	96%	3e-20	45.92%	90	U0885019.1																																																																				
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHpred avec une proba >= 90% et une couverture acceptable ?	<table><thead><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr></thead><tbody><tr><td><input type="checkbox"/> 1</td><td>PF20087.3</td><td>; DUF6479; Family of unknown function (DUF6479)</td><td>87.11</td><td>5.7</td><td>30.27</td><td>5.6</td><td>52</td><td>121</td></tr><tr><td><input type="checkbox"/> 2</td><td>8I4T_K</td><td>Envelopment polyprotein; SFTSV, virion, icosahedral reconstruction, VIRUS, VIRAL PROTEIN; HET: NAG; 5.2A (Severe fever w</td><td>86.86</td><td>4.1</td><td>35.1</td><td>5.3</td><td>44</td><td>513</td></tr><tr><td><input type="checkbox"/> 3</td><td>Q77376</td><td>VPU_HV1AN Protein Vpu OS=Human immunodeficiency virus type 1 group O (isolate ANT70) OX=327105 GN=vpu PE=3 SV=1</td><td>85.68</td><td>5.8</td><td>28.89</td><td>4.8</td><td>55</td><td>85</td></tr><tr><td><input type="checkbox"/> 4</td><td>PF13131.10</td><td>; DUF3951; Protein of unknown function (DUF3951)</td><td>84.31</td><td>7.8</td><td>25.71</td><td>4.5</td><td>33</td><td>52</td></tr></tbody></table> <p>Non.</p>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/> 1	PF20087.3	; DUF6479; Family of unknown function (DUF6479)	87.11	5.7	30.27	5.6	52	121	<input type="checkbox"/> 2	8I4T_K	Envelopment polyprotein; SFTSV, virion, icosahedral reconstruction, VIRUS, VIRAL PROTEIN; HET: NAG; 5.2A (Severe fever w	86.86	4.1	35.1	5.3	44	513	<input type="checkbox"/> 3	Q77376	VPU_HV1AN Protein Vpu OS=Human immunodeficiency virus type 1 group O (isolate ANT70) OX=327105 GN=vpu PE=3 SV=1	85.68	5.8	28.89	4.8	55	85	<input type="checkbox"/> 4	PF13131.10	; DUF3951; Protein of unknown function (DUF3951)	84.31	7.8	25.71	4.5	33	52																														
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																																				
<input type="checkbox"/> 1	PF20087.3	; DUF6479; Family of unknown function (DUF6479)	87.11	5.7	30.27	5.6	52	121																																																																				
<input type="checkbox"/> 2	8I4T_K	Envelopment polyprotein; SFTSV, virion, icosahedral reconstruction, VIRUS, VIRAL PROTEIN; HET: NAG; 5.2A (Severe fever w	86.86	4.1	35.1	5.3	44	513																																																																				
<input type="checkbox"/> 3	Q77376	VPU_HV1AN Protein Vpu OS=Human immunodeficiency virus type 1 group O (isolate ANT70) OX=327105 GN=vpu PE=3 SV=1	85.68	5.8	28.89	4.8	55	85																																																																				
<input type="checkbox"/> 4	PF13131.10	; DUF3951; Protein of unknown function (DUF3951)	84.31	7.8	25.71	4.5	33	52																																																																				

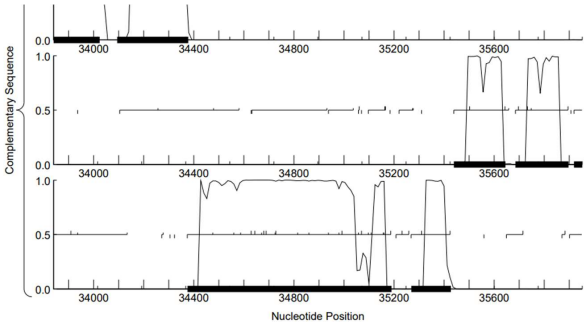
	 <p>Structure predite avec esmFold</p>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Fonction inconnue
DECISION:	Membrane protein

Student Gene Annotation Worksheet

Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	48
Coordonnées du Stop	34375
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	Gap 78
Coordonnées du Start retenu	35190
Fonction prédite	thymidylate synthase

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS :</p>  <p>GeneMark :</p>

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <div> <div>Score</div> <div>E</div> <div>(bits) Value</div> </div> <div> <div>Carostasia_Draft_48, function unknown, 271</div> <div>553</div> <div>e-157</div> </div> <div> <div>Quartz_49, thymidylate synthase, 271</div> <div>549</div> <div>e-156</div> </div> <div> <div>Mandalorian_48, thymidylate synthase, 271</div> <div>548</div> <div>e-156</div> </div>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions
DECISION:	OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																								
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 35190 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 35190																																																																																																																																																																																																																								
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Starts: 22 Selected: 1</div><div>ORF Start : 35100 ORF Stop : 34375 ORF Length: 726</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 44.4 88.9 33.3 27 3' End 45.5 72.7 45.5 66</div><div>SD Scoring Matrix Kibler6 Spacing Weight Matrix Karlin Medium</div><div>Explore Document</div></div> <table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,748</td><td>2,981</td><td>10</td><td>-2,443</td><td>ATAGCAGTCCAGGAGGACAC</td><td>ATG</td><td>35190</td><td>816</td></tr><tr><td>2</td><td>-5,600</td><td>1,178</td><td>9</td><td>-6,375</td><td>GAAGCAGACACTGTGATGAT</td><td>GTG</td><td>35163</td><td>789</td></tr><tr><td>3</td><td>-4,580</td><td>1,656</td><td>7</td><td>-6,103</td><td>GCACACACTGTGATGATGTG</td><td>GTG</td><td>35160</td><td>786</td></tr><tr><td>4</td><td>-3,638</td><td>2,097</td><td>9</td><td>-4,413</td><td>CGACGAGCGGACCGGGAACGA</td><td>GTG</td><td>35112</td><td>738</td></tr><tr><td>5</td><td>-4,807</td><td>1,550</td><td>15</td><td>-6,409</td><td>CGGGAACCGAGTGTCTACTCG</td><td>GTG</td><td>35100</td><td>726</td></tr><tr><td>6</td><td>-4,069</td><td>1,895</td><td>16</td><td>-5,865</td><td>CTGGAGGCAACACCGAGTCC</td><td>ATG</td><td>35073</td><td>699</td></tr><tr><td>7</td><td>-4,928</td><td>1,493</td><td>11</td><td>-5,685</td><td>CACGAGTCCATGATCGGCAAT</td><td>GTG</td><td>35061</td><td>687</td></tr><tr><td>8</td><td>-4,502</td><td>1,693</td><td>17</td><td>-6,502</td><td>CGGCAAGCAAGCTCGCGCAT</td><td>ATG</td><td>34995</td><td>621</td></tr><tr><td>9</td><td>-2,835</td><td>2,473</td><td>10</td><td>-3,530</td><td>GCTCGCGGACATGGGAAGGCCG</td><td>GTG</td><td>34983</td><td>609</td></tr><tr><td>10</td><td>-3,857</td><td>1,994</td><td>11</td><td>-4,614</td><td>GTGCGCTGGCTGGGAGGCGCG</td><td>GTG</td><td>34941</td><td>567</td></tr><tr><td>11</td><td>-7,210</td><td>0,425</td><td>13</td><td>-8,255</td><td>CGTCACACGCTCTACGACCGC</td><td>GTG</td><td>34863</td><td>489</td></tr><tr><td>12</td><td>-2,791</td><td>2,493</td><td>16</td><td>-4,587</td><td>AGGGCTGGAGACCGCGGACGCC</td><td>GTG</td><td>34818</td><td>444</td></tr><tr><td>13</td><td>-6,259</td><td>0,870</td><td>13</td><td>-7,305</td><td>CTGCACACAGCGGCTCACTTC</td><td>ATG</td><td>34731</td><td>357</td></tr><tr><td>14</td><td>-6,259</td><td>0,870</td><td>16</td><td>-8,055</td><td>CACCAAGCGGCTCACTTCATG</td><td>GTG</td><td>34728</td><td>354</td></tr><tr><td>15</td><td>-6,832</td><td>0,602</td><td>11</td><td>-7,588</td><td>TGGCTCGACATCATGCGCCAG</td><td>ATG</td><td>34692</td><td>318</td></tr><tr><td>16</td><td>-5,205</td><td>1,364</td><td>10</td><td>-5,899</td><td>CATCATGCGCCATGCGGACG</td><td>ATG</td><td>34683</td><td>309</td></tr><tr><td>17</td><td>-3,642</td><td>2,095</td><td>7</td><td>-5,164</td><td>CGAGATCGGCGAGTGGAGGCT</td><td>GTG</td><td>34674</td><td>300</td></tr><tr><td>18</td><td>-4,234</td><td>1,818</td><td>10</td><td>-4,925</td><td>CGGCTACCGCGGCGACTACAG</td><td>ATG</td><td>34647</td><td>273</td></tr><tr><td>19</td><td>-3,509</td><td>2,157</td><td>11</td><td>-4,266</td><td>CTACAGCATGTGGGACGACCTG</td><td>ATG</td><td>34632</td><td>258</td></tr><tr><td>20</td><td>-3,435</td><td>2,192</td><td>11</td><td>-4,192</td><td>CTCAGACACCGGGGTGGGAC</td><td>GTG</td><td>34590</td><td>216</td></tr><tr><td>21</td><td>-3,818</td><td>2,012</td><td>18</td><td>-6,119</td><td>TGTTGGGACATCACCTTCAG</td><td>GTG</td><td>34563</td><td>189</td></tr><tr><td>22</td><td>-5,918</td><td>1,030</td><td>10</td><td>-6,613</td><td>CGAGGAGCGGCTGGCCGTGCT</td><td>GTG</td><td>34475</td><td>105</td></tr></table> <p>On retient le premier, avec un Z score le plus proche de 0, départ avec un ATG, plus grand ORF</p>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,748	2,981	10	-2,443	ATAGCAGTCCAGGAGGACAC	ATG	35190	816	2	-5,600	1,178	9	-6,375	GAAGCAGACACTGTGATGAT	GTG	35163	789	3	-4,580	1,656	7	-6,103	GCACACACTGTGATGATGTG	GTG	35160	786	4	-3,638	2,097	9	-4,413	CGACGAGCGGACCGGGAACGA	GTG	35112	738	5	-4,807	1,550	15	-6,409	CGGGAACCGAGTGTCTACTCG	GTG	35100	726	6	-4,069	1,895	16	-5,865	CTGGAGGCAACACCGAGTCC	ATG	35073	699	7	-4,928	1,493	11	-5,685	CACGAGTCCATGATCGGCAAT	GTG	35061	687	8	-4,502	1,693	17	-6,502	CGGCAAGCAAGCTCGCGCAT	ATG	34995	621	9	-2,835	2,473	10	-3,530	GCTCGCGGACATGGGAAGGCCG	GTG	34983	609	10	-3,857	1,994	11	-4,614	GTGCGCTGGCTGGGAGGCGCG	GTG	34941	567	11	-7,210	0,425	13	-8,255	CGTCACACGCTCTACGACCGC	GTG	34863	489	12	-2,791	2,493	16	-4,587	AGGGCTGGAGACCGCGGACGCC	GTG	34818	444	13	-6,259	0,870	13	-7,305	CTGCACACAGCGGCTCACTTC	ATG	34731	357	14	-6,259	0,870	16	-8,055	CACCAAGCGGCTCACTTCATG	GTG	34728	354	15	-6,832	0,602	11	-7,588	TGGCTCGACATCATGCGCCAG	ATG	34692	318	16	-5,205	1,364	10	-5,899	CATCATGCGCCATGCGGACG	ATG	34683	309	17	-3,642	2,095	7	-5,164	CGAGATCGGCGAGTGGAGGCT	GTG	34674	300	18	-4,234	1,818	10	-4,925	CGGCTACCGCGGCGACTACAG	ATG	34647	273	19	-3,509	2,157	11	-4,266	CTACAGCATGTGGGACGACCTG	ATG	34632	258	20	-3,435	2,192	11	-4,192	CTCAGACACCGGGGTGGGAC	GTG	34590	216	21	-3,818	2,012	18	-6,119	TGTTGGGACATCACCTTCAG	GTG	34563	189	22	-5,918	1,030	10	-6,613	CGAGGAGCGGCTGGCCGTGCT	GTG	34475	105
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																																																																																																																																	
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																																																																																																																																	
1	-1,748	2,981	10	-2,443	ATAGCAGTCCAGGAGGACAC	ATG	35190	816																																																																																																																																																																																																																	
2	-5,600	1,178	9	-6,375	GAAGCAGACACTGTGATGAT	GTG	35163	789																																																																																																																																																																																																																	
3	-4,580	1,656	7	-6,103	GCACACACTGTGATGATGTG	GTG	35160	786																																																																																																																																																																																																																	
4	-3,638	2,097	9	-4,413	CGACGAGCGGACCGGGAACGA	GTG	35112	738																																																																																																																																																																																																																	
5	-4,807	1,550	15	-6,409	CGGGAACCGAGTGTCTACTCG	GTG	35100	726																																																																																																																																																																																																																	
6	-4,069	1,895	16	-5,865	CTGGAGGCAACACCGAGTCC	ATG	35073	699																																																																																																																																																																																																																	
7	-4,928	1,493	11	-5,685	CACGAGTCCATGATCGGCAAT	GTG	35061	687																																																																																																																																																																																																																	
8	-4,502	1,693	17	-6,502	CGGCAAGCAAGCTCGCGCAT	ATG	34995	621																																																																																																																																																																																																																	
9	-2,835	2,473	10	-3,530	GCTCGCGGACATGGGAAGGCCG	GTG	34983	609																																																																																																																																																																																																																	
10	-3,857	1,994	11	-4,614	GTGCGCTGGCTGGGAGGCGCG	GTG	34941	567																																																																																																																																																																																																																	
11	-7,210	0,425	13	-8,255	CGTCACACGCTCTACGACCGC	GTG	34863	489																																																																																																																																																																																																																	
12	-2,791	2,493	16	-4,587	AGGGCTGGAGACCGCGGACGCC	GTG	34818	444																																																																																																																																																																																																																	
13	-6,259	0,870	13	-7,305	CTGCACACAGCGGCTCACTTC	ATG	34731	357																																																																																																																																																																																																																	
14	-6,259	0,870	16	-8,055	CACCAAGCGGCTCACTTCATG	GTG	34728	354																																																																																																																																																																																																																	
15	-6,832	0,602	11	-7,588	TGGCTCGACATCATGCGCCAG	ATG	34692	318																																																																																																																																																																																																																	
16	-5,205	1,364	10	-5,899	CATCATGCGCCATGCGGACG	ATG	34683	309																																																																																																																																																																																																																	
17	-3,642	2,095	7	-5,164	CGAGATCGGCGAGTGGAGGCT	GTG	34674	300																																																																																																																																																																																																																	
18	-4,234	1,818	10	-4,925	CGGCTACCGCGGCGACTACAG	ATG	34647	273																																																																																																																																																																																																																	
19	-3,509	2,157	11	-4,266	CTACAGCATGTGGGACGACCTG	ATG	34632	258																																																																																																																																																																																																																	
20	-3,435	2,192	11	-4,192	CTCAGACACCGGGGTGGGAC	GTG	34590	216																																																																																																																																																																																																																	
21	-3,818	2,012	18	-6,119	TGTTGGGACATCACCTTCAG	GTG	34563	189																																																																																																																																																																																																																	
22	-5,918	1,030	10	-6,613	CGAGGAGCGGCTGGCCGTGCT	GTG	34475	105																																																																																																																																																																																																																	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	ORF le plus long avec 816pb.Gap de 79pb avec le gène suivant.																																																																																																																																																																																																																								
Est-ce que le start est conservé chez les homologues voir Starterator ?	• Track 27 : Carostasia_48, Mandalorian_48 Track 28 : Mandalorian_48, Carostasia_48																																																																																																																																																																																																																								

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>>Quartz_49, thymidylate synthase, 271 Length = 271</p> <p>Score = 549 bits (1415), Expect = e-156 Identities = 269/271 (99%), Positives = 270/271 (99%)</p> <p>Query: 1 MLKHNTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPdWVRRELAWFNsGS 60 MLK NTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPdWVRRELAWFNsGS Sbjct: 1 MLKQNTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPdWVRRELAWFNsGS 60</p> <p>Query: 61 NKladMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE 120 NKladMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE Sbjct: 61 NKladMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE 120</p> <p>Query: 121 TRHAVAIISDRDIHQlASyNGRNDFICTNALNFMVDAGRLHIIAQMRsMDAVFGYRADY 180 TRHAVAIISDRDIHQlASyNGRNDFICTNALNFMVDAGRLHIIAQMRsMDAVFGYRADY Sbjct: 121 TRHAVAIISDRDIHQlASyNGRNDFICTNALNFMVDAGRLHIIAQMRsMDAVFGYRADY 180</p> <p>Query: 181 SMWDDLmKRLVIDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL 240 SMWDDLmKRL+IDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL Sbjct: 181 SMWDDLmKRLVIDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL 240</p> <p>Query: 241 DQAERYGDTSAQRDPHSTEPAGTGETLDFEL 271 DQAERYGDTSAQRDPHSTEPAGTGETLDFEL Sbjct: 241 DQAERYGDTSAQRDPHSTEPAGTGETLDFEL 271</p>
DECISION:	35190. Meilleur score, départ du potentiel codant.

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>*** <i>PhagesDB</i> : (<i>q#uartz_49, thymidylate synthase, 271 : s#549</i>) : [99%] ; e-value : e-156</p> <p>>Quartz_49, thymidylate synthase, 271 Length = 271</p> <p>Score = 549 bits (1415), Expect = e-156 Identities = 269/271 (99%), Positives = 270/271 (99%)</p> <p>Query: 1 MLKHNTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPdWVRRELAWFNsGS 60 MLK NTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPdWVRRELAWFNsGS Sbjct: 1 MLKQNTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPdWVRRELAWFNsGS 60</p> <p>Query: 61 NKladMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE 120 NKladMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE Sbjct: 61 NKladMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE 120</p> <p>Query: 121 TRHAVAIISDRDIHQlASyNGRNDFICTNALNFMVDAGRLHIIAQMRsMDAVFGYRADY 180 TRHAVAIISDRDIHQlASyNGRNDFICTNALNFMVDAGRLHIIAQMRsMDAVFGYRADY Sbjct: 121 TRHAVAIISDRDIHQlASyNGRNDFICTNALNFMVDAGRLHIIAQMRsMDAVFGYRADY 180</p> <p>Query: 181 SMWDDLmKRLVIDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL 240 SMWDDLmKRL+IDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL Sbjct: 181 SMWDDLmKRLVIDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL 240</p> <p>Query: 241 DQAERYGDTSAQRDPHSTEPAGTGETLDFEL 271 DQAERYGDTSAQRDPHSTEPAGTGETLDFEL Sbjct: 241 DQAERYGDTSAQRDPHSTEPAGTGETLDFEL 271</p>

*** nr :

(q#thymidylate synthase [Microbacterium phage Quartz] : s#557) : [99%] ; e-value : 0.0

thymidylate synthase [Microbacterium phage Quartz]

Sequence ID: [UVK59268.1](#) Length: 271 Number of Matches: 1

Range 1: 1 to 271 [GenPept](#) [Graphics](#) [Next Match](#) [Previous](#)

Score	Expect	Method	Identities	Positives	Gaps
557 bits(1435)	0.0	Compositional matrix adjust.	269/271(99%)	270/271(99%)	0/271(0%)
Query 1	MLKHNTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPDWVRRELAWFN	60			
Sbjct 1	MLK NTVDDVVLALASAPIDERTGNRVLYSVNWKATTESMIGNVNPDWVRRELAWFN	60			
Query 61	NKLADMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE	120			
Sbjct 61	NKLADMEGPVPKLFQACAGWDGAVNSAYGHILFSRDELLPRVTTYDRVLITLLKEGLE	120			
Query 121	TRHAVAIISDRDIHQLASYNGRNDFICTNALNFMVDADGRLHIIAQMRSDAVFGYRADY	180			
Sbjct 121	TRHAVAIISDRDIHQLASYNGRNDFICTNALNFMVDADGRLHIIAQMRSDAVFGYRADY	180			
Query 181	SMWDDLKMKRLVIDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL	240			
Sbjct 181	SMWDDLKMKRLVIDLNNRGRDVRRGDITFQVANLHVYPRHFQLLDETAQDTEERLARRVWL	240			
Query 241	DQAERYGDTSAQRDPHSTEPAGTGETLDFEL 271				
Sbjct 241	DQAERYGDTSAQRDPHSTEPAGTGETLDFEL 271				

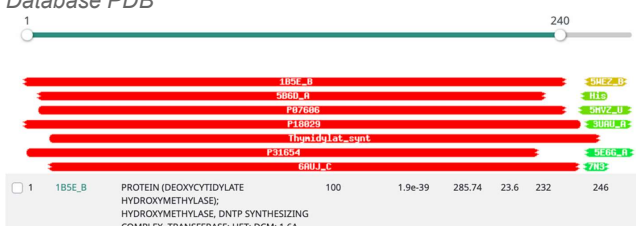
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

HHPred : PROTEIN (DEOXYCYTIDYLATE HYDROXYMETHYLASE); HYDROXYMETHYLASE, DNTP SYNTHESIZING COMPLEX, TRANSFERASE; HET: DCM; 1.6A

Hit 1B5E_B

Probability 100%

Database PDB



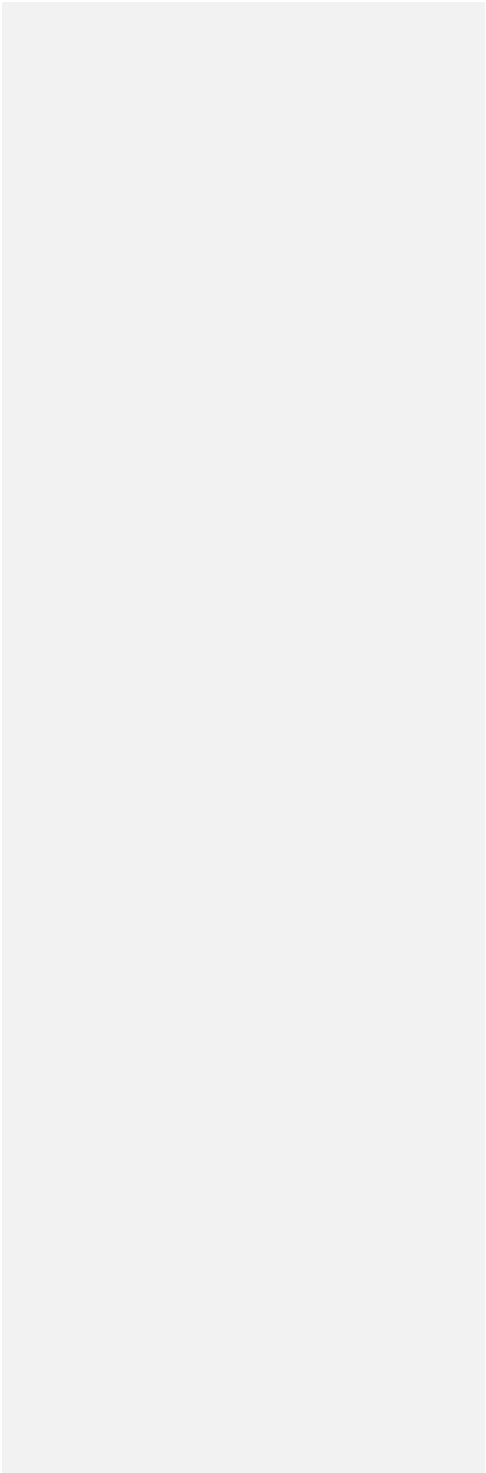
<input type="checkbox"/> 1	1B5E_B	PROTEIN (DEOXYCYTIDYLATE HYDROXYMETHYLASE); HYDROXYMETHYLASE, DNTP SYNTHESIZING COMPLEX, TRANSFERASE; HET: DCM; 1.6A (En	100	1.9e-39	285.74	23.6	232	246
<input type="checkbox"/> 2	5B6D_A	CMP 5-hydroxymethylase; CMP hydroxymethylase, TRANSFERASE; HET: CSP; 1.65A (Streptomyces rimofaciens)	100	6.5e-39	293.48	23.3	209	325
<input type="checkbox"/> 3	P07606	TYSY_BPHT Thymidylate synthase OS=Bacillus phage phi3T OX=10736 GN=thyP3 PE=3 SV=1	100	1.7e-38	284.9	21.6	222	279
<input type="checkbox"/> 4	P18029	DCHM_BPT2 Deoxycytidylate 5-hydroxymethyltransferase OS=Enterobacteria phage T2 OX=10664 GN=42 PE=3 SV=1	100	3.7e-38	277.71	23.3	231	246
<input type="checkbox"/> 5	PF00303.23	; Thymidylat_synt; Thymidylate synthase	100	6.2e-38	278.94	20.1	231	263

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Aucune synténie observée

	<div>Carostasia_Draft (EA10)</div> <div><p>Diagram of protein domain architecture for Carostasia_Draft (EA10). The protein is 275 amino acids long. Domains include: B47.1 (20-35), B47.1 (35-50), B47.1 (50-65), B47.1 (65-80), B47.1 (80-95), B47.1 (95-110), B47.1 (110-125), B47.1 (125-140), B47.1 (140-155), B47.1 (155-170), B47.1 (170-185), B47.1 (185-200), B47.1 (200-215), B47.1 (215-230), B47.1 (230-245), B47.1 (245-260), B47.1 (260-275). A red line indicates a transmembrane domain between residues 20 and 35.</p></div> <div><p>Diagram of protein domain architecture for Juanyo (EA10). The protein is 275 amino acids long. Domains include: B47.1 (20-35), B47.1 (35-50), B47.1 (50-65), B47.1 (65-80), B47.1 (80-95), B47.1 (95-110), B47.1 (110-125), B47.1 (125-140), B47.1 (140-155), B47.1 (155-170), B47.1 (170-185), B47.1 (185-200), B47.1 (200-215), B47.1 (215-230), B47.1 (230-245), B47.1 (245-260), B47.1 (260-275). A red line indicates a transmembrane domain between residues 20 and 35.</p></div> <div><p>Diagram of protein domain architecture for Mandalorian (EA10). The protein is 275 amino acids long. Domains include: B47.1 (20-35), B47.1 (35-50), B47.1 (50-65), B47.1 (65-80), B47.1 (80-95), B47.1 (95-110), B47.1 (110-125), B47.1 (125-140), B47.1 (140-155), B47.1 (155-170), B47.1 (170-185), B47.1 (185-200), B47.1 (200-215), B47.1 (215-230), B47.1 (230-245), B47.1 (245-260), B47.1 (260-275). A red line indicates a transmembrane domain between residues 20 and 35.</p></div>
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<div>No</div> <div>DeepTMHMM : <p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>Outside Membrane Inside</p><p>0 50 100 150 200 250</p></div> <div><p>DeepTMHMM - Posterior Probabilities</p><p>Probability</p><p>1.0 0.8 0.6 0.4 0.2 0.0</p><p>0 50 100 150 200 250</p><p>Sequence</p><p>Inside</p></div> <div>Phoebeius :</div>

	<div><div><div>ID</div><div>FT</div><div>//</div></div><div>UNNAMED TOPO_DOM</div><div>1272</div><div>NON CYTOPLASMIC.</div></div> <div><p>Phobius posterior probabilities for UNNAMED</p><table><caption>Approximate data from Phobius plot</caption><tr><th>Residue Range</th><th>transmembrane</th><th>cytoplasmic</th><th>non cytoplasmic</th><th>signal peptide</th></tr><tr><td>0 - 250</td><td>~0.00</td><td>~0.15</td><td>~0.85</td><td>~0.00</td></tr></table></div>	Residue Range	transmembrane	cytoplasmic	non cytoplasmic	signal peptide	0 - 250	~0.00	~0.15	~0.85	~0.00
Residue Range	transmembrane	cytoplasmic	non cytoplasmic	signal peptide							
0 - 250	~0.00	~0.15	~0.85	~0.00							
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Yes.										
DECISION:	<i>thymidylate synthase</i>										

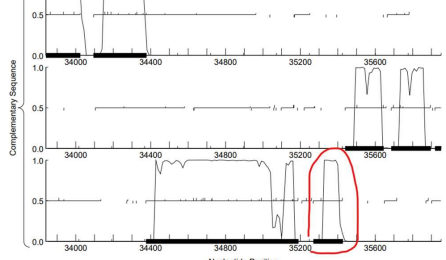


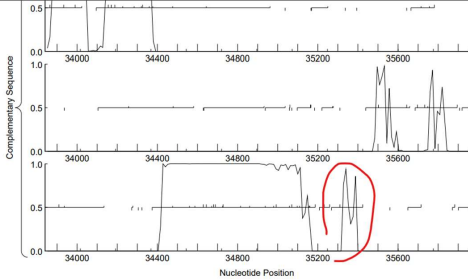
Student Gene Annotation Worksheet

Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	49
Coordonnées du Stop	35269
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	gap 11
Coordonnées du Start retenu	35427
Fonction prédite	NFK (hypotetical protein)

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS :</p>  <p>GeneMark :</p>

																										
	Le pouvoir codant semble peu constant.																									
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><thead><tr><th></th><th>Score</th><th>E</th><th>(bits)</th><th>Value</th></tr></thead><tbody><tr><td>Quartz_50, function unknown, 52</td><td>108</td><td>6e-24</td><td></td><td></td></tr><tr><td>Nucci_49, function unknown, 52</td><td>108</td><td>6e-24</td><td></td><td></td></tr><tr><td>Mandalorian_49, function unknown, 52</td><td>108</td><td>6e-24</td><td></td><td></td></tr><tr><td>Carostasia_Draft_49, function unknown, 52</td><td>108</td><td>6e-24</td><td></td><td></td></tr></tbody></table>		Score	E	(bits)	Value	Quartz_50, function unknown, 52	108	6e-24			Nucci_49, function unknown, 52	108	6e-24			Mandalorian_49, function unknown, 52	108	6e-24			Carostasia_Draft_49, function unknown, 52	108	6e-24		
	Score	E	(bits)	Value																						
Quartz_50, function unknown, 52	108	6e-24																								
Nucci_49, function unknown, 52	108	6e-24																								
Mandalorian_49, function unknown, 52	108	6e-24																								
Carostasia_Draft_49, function unknown, 52	108	6e-24																								
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions.																									
DECISION:	OUI																									

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																				
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 35427 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 35427																																				
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Starts : 2 Selected : 1</div><div>ORF Start : 35427 ORF Stop : 35269 ORF Length: 159</div><div>Cdn1 Cdn2 Cdn3 Length 5' End 76.3 50.0 65.8 114 3' End 71.8 28.2 59.0 117</div><div>SD Scoring Matrix Kibler6</div><div>Explore</div><div>Spacing Weight Matrix Karlin Medium</div><div>Document</div></div> <table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-1,499</td><td>9,126</td><td>11</td><td>-2,196</td><td>TGCTCAAGTAAGGAACCCCATC</td><td>ATG</td><td>35427</td><td>159</td></tr><tr><td>2</td><td>-3,599</td><td>2,115</td><td>11</td><td>-4,366</td><td>GGCTCGGAGCGGGGCGACAC</td><td>ATG</td><td>35313</td><td>45</td></tr></table> <p>On choisit le premier start, meilleur score, ORF plus long</p>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-1,499	9,126	11	-2,196	TGCTCAAGTAAGGAACCCCATC	ATG	35427	159	2	-3,599	2,115	11	-4,366	GGCTCGGAGCGGGGCGACAC	ATG	35313	45
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																													
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																													
1	-1,499	9,126	11	-2,196	TGCTCAAGTAAGGAACCCCATC	ATG	35427	159																													
2	-3,599	2,115	11	-4,366	GGCTCGGAGCGGGGCGACAC	ATG	35313	45																													
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	On considère l'ORF le plus long, 159pb. C'est un ATG. Il a un Final Score plus proche de 0.																																				
Est-ce que le start est conservé chez les homologues voir Starterator ?	• Track 2 : Carostasia_49, Nucci_49, Quartz_50, Mandalorian_49																																				
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	>Quartz_50, function unknown, 52 Length = 52 Score = 108 bits (269), Expect = 6e-24 Identities = 52/52 (100%), Positives = 52/52 (100%) Query: 1 MARKSAHALRAEIHQANIQLRILEKKLASAEACERGDHMKPNKAQPCILCGA 52 MARKSAHALRAEIHQANIQLRILEKKLASAEACERGDHMKPNKAQPCILCGA Sbjct: 1 MARKSAHALRAEIHQANIQLRILEKKLASAEACERGDHMKPNKAQPCILCGA 52																																				
DECISION:	35427																																				

Commenté [11]: justification

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
------------------------------------	-------------

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :
(q#: s#) : [alignment] ; e-value :
Quartz_50, function unknown, 52 108 6e-24
Nucci_49, function unknown, 52 108 6e-24
Mandalorian_49, function unknown, 52 108 6e-24
Carostasia_Draft_49, function unknown, 52 108 6e-24
Yuuy_50, function unknown, 52 105 4e-23
Juanyo_50, function unknown, 52 101 6e-22
*** nr :
(q#: s#) : [alignment] ; e-value :

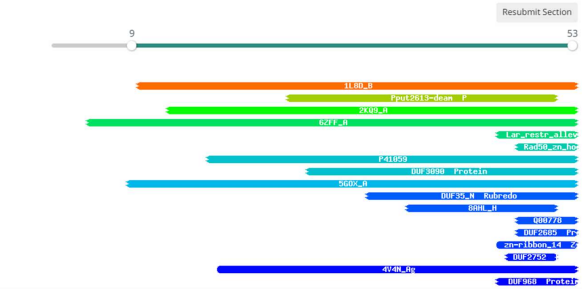
	Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hypothetical protein SEA_NUCC_49 (Microbacterium phage Nucci)	Microbacterium phage Nucci	105	105	98%	4e-28	100.00%	52	GD032442.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_YUUY_50 (Microbacterium phage Yuuy)	Microbacterium phage Yuuy	102	102	98%	4e-27	96.15%	52	GF295478.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_JUANYO_50 (Microbacterium phage Juanyo)	Microbacterium phage Juanyo	99.8	99.8	98%	7e-26	92.31%	52	GU344758.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_CAROSTASIA_49 (Microbacterium phage Carostasia)	Microbacterium phage Carostasia	45.8	45.8	99%	5e-04	60.42%	119	GU002088.1
<input checked="" type="checkbox"/>	hypothetical protein SEA_SANSI_51 (Microbacterium phage Sansi)	Microbacterium phage Sansi	45.8	45.8	99%	5e-04	60.42%	119	AB048924.1

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un criblé HHPred avec une proba >= 90% et une couverture acceptable ?

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	
<input type="checkbox"/>	1	1LRD_B	DNA double-strand break repair rad50 ATPase; zinc finger, rad50, DNA repair, Recombination, hook motif, REPLICATION; HET	90.9	2.7	24.11	4.8	35	112
<input type="checkbox"/>	2	PF14427.10	; Pput2613-deam ; Pput_2613-like deaminase	79.6	1.2	28.58	-0.4	28	117
<input type="checkbox"/>	3	2KQ9_A	DnaK suppressor protein; Zinc binding protein, Structural Genomics, PSI-2, Protein Structure Initiative, Northeast Struc	70.4	38	19.96	4.7	36	112
<input type="checkbox"/>	4	6ZFF_A	DNA double-strand break repair Rad50 ATPase; double strand break repair; DNA damage response, scaffold, closed coiled coil	64.43	61	20.13	5.4	40	140

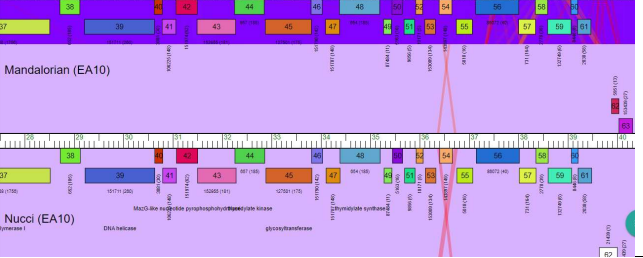
Oui

Visualization

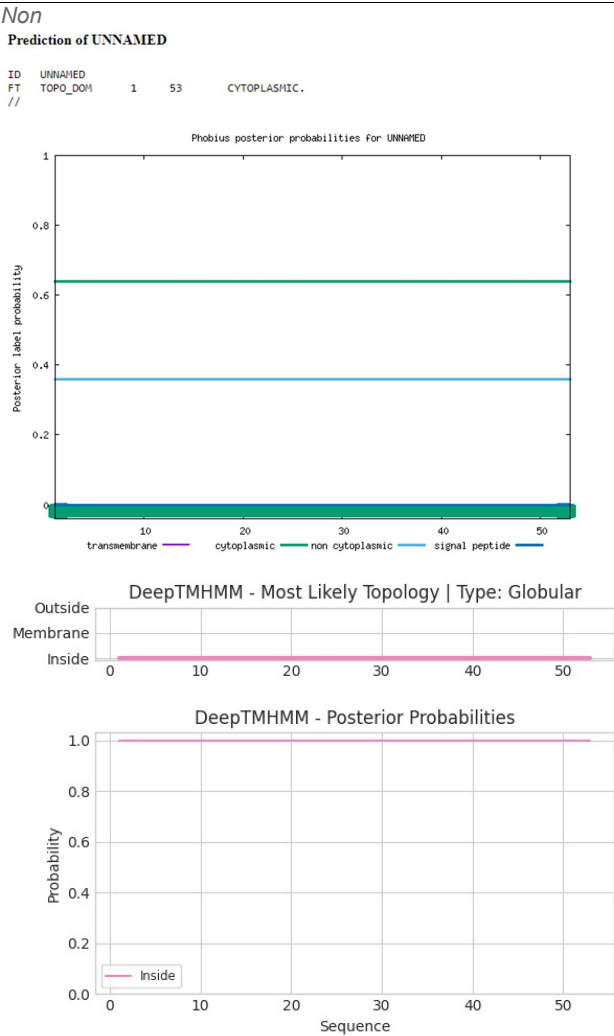


Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Fonction inconnue.



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue ;

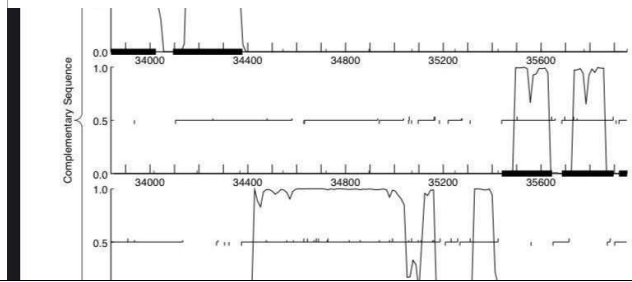
DECISION:

NFK (hypotetical protein)

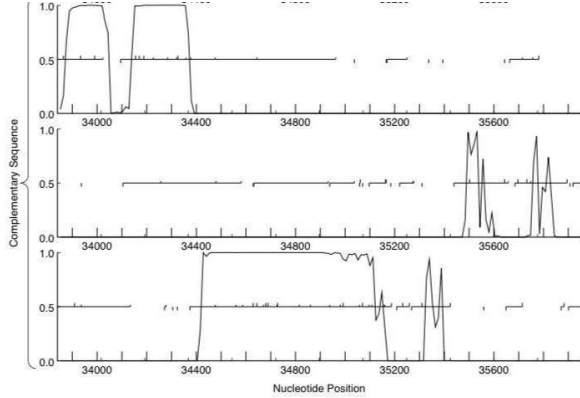
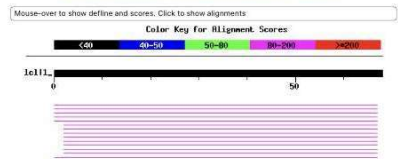
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	50
Coordonnées du Stop	35439
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	gap 39
Coordonnées du Start retenu	35645
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenemarkS pouvoir codant élevé dans la region de l'ORF avec la sequence complementaire</p> <p>Genemark folio. Pareil</p> 

Commenté [12]: page entière

																																										
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Distribution of 16 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <table><tr><td><40</td><td>40-50</td><td>50-60</td><td>60-200</td><td>>200</td></tr></table> <p>16111</p>  <p>quences producing significant alignments:</p> <table><thead><tr><th></th><th>Score</th><th>E</th></tr><tr><th></th><th>(bits)</th><th>Val</th></tr></thead><tbody><tr><td>rostasia_Draft_50, function unknown, 68</td><td>128</td><td></td></tr><tr><td>idalorian_50, function unknown, 68</td><td>126</td><td></td></tr><tr><td>irtz_51, function unknown, 68</td><td>122</td><td></td></tr><tr><td>ci_50, function unknown, 68</td><td>112</td><td></td></tr><tr><td>Y_51, function unknown, 68</td><td>102</td><td></td></tr><tr><td>atra_50, function unknown, 67</td><td>93</td><td></td></tr><tr><td>robot_49, function unknown, 67</td><td>93</td><td></td></tr><tr><td>stleton_49, function unknown, 67</td><td>93</td><td></td></tr><tr><td>Victor_50, function unknown, 67</td><td>92</td><td></td></tr><tr><td>incePhergus_49, function unknown, 67</td><td>92</td><td></td></tr></tbody></table> <p>Non annotés mais beaucoup d' homologues</p>	<40	40-50	50-60	60-200	>200		Score	E		(bits)	Val	rostasia_Draft_50, function unknown, 68	128		idalorian_50, function unknown, 68	126		irtz_51, function unknown, 68	122		ci_50, function unknown, 68	112		Y_51, function unknown, 68	102		atra_50, function unknown, 67	93		robot_49, function unknown, 67	93		stleton_49, function unknown, 67	93		Victor_50, function unknown, 67	92		incePhergus_49, function unknown, 67	92	
<40	40-50	50-60	60-200	>200																																						
	Score	E																																								
	(bits)	Val																																								
rostasia_Draft_50, function unknown, 68	128																																									
idalorian_50, function unknown, 68	126																																									
irtz_51, function unknown, 68	122																																									
ci_50, function unknown, 68	112																																									
Y_51, function unknown, 68	102																																									
atra_50, function unknown, 67	93																																									
robot_49, function unknown, 67	93																																									
stleton_49, function unknown, 67	93																																									
Victor_50, function unknown, 67	92																																									
incePhergus_49, function unknown, 67	92																																									
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE COBTRADICTION</p> <p>-plus de 50pb</p> <p>-PAS chevauchement</p> <p>-direction similaire aux gènes qui l'entourent</p>																																									
DECISION:	OUI																																									

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 35645</p> <p>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :35645</p>

Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?

Choose ORF start

Start: 3	ORF Start: 35645	Cdn1	Cdn2	Cdn3	Length	SD Scoring Matrix	Kibler6	Explore
Selected: 1	ORF Stop: 35439	5' End	60.0	20.0	40.0	15	Spacing Weight Matrix	Karin Medium
	ORF Length: 207	3' End	100.0	100.0	100.0	3		Document

35654

Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length
1	-7,290	0,388	10	-7,985	GCTGGGCCACTTCAACTAAGG	GTG	35660	222
2	-2,728	2,823	17	-4,728	ACTAAGGGTGTACACTGAAGTT	ATG	35645	207
3	-5,919	1,030	10	-6,613	GCTGGTCACCTGGCCGACATT	ATG	35604	66

Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?

Le start 35660 conduit à une ORF de 222pb, la plus longue PAS DE chevauchement
Le start 35645 conduit à une ORF très légèrement plus courte de 207pb

Est-ce que le start est conservé chez les homologues voir Starterator ?



Start le plus conservé est 35645

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Query: 1 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA</p> <p>Sbjct: 1 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60</p> <p>Query: 61 VSALALVK 68 VSALALVK</p> <p>Sbjct: 61 VSALALVK 68</p> <p>>Mandalorian_50, function unknown, 68 Length = 68</p> <p>Score = 126 bits (316), Expect = 2e-29 Identities = 66/68 (97%), Positives = 67/68 (98%)</p> <p>Query: 1 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA</p> <p>Sbjct: 1 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60</p> <p>Query: 61 VSALALVK 68 SALALVK</p> <p>Sbjct: 61 ASALALVK 68</p> <p>>Quartz_51, function unknown, 68 Length = 68</p> <p>Score = 122 bits (307), Expect = 2e-28 Identities = 65/68 (95%), Positives = 65/68 (95%)</p> <p>Query: 1 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60 MN KKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA</p> <p>Sbjct: 1 MNTTKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60</p> <p>Query: 61 VSALALVK 68 SALALVK</p> <p>Sbjct: 61 ASALALVK 68</p> <p>>Nucci_50, function unknown, 68 Length = 68</p> <p>Score = 112 bits (281), Expect = 2e-25 Identities = 59/68 (86%), Positives = 62/68 (91%)</p> <p>Query: 1 MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKLGHLADVMYTGSDYVRKRAA 60 MN IKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATK+G+LAD MYTG DVR RAA</p> <p>Sbjct: 1 MNTTKKLLIAAINDQDSRVIRRLADAVETKVVTTQAVATKIGYLDAMYTGRKADVRKRAA 60</p> <p>Query: 61 VSALALVK 68 SALALVK</p> <p>Sbjct: 61 ASALALVK 68</p> <p>Le Start 35645 est conservé</p>
DECISION:	<p>35645:</p> <ul style="list-style-type: none"> -bon score -bonne zvalue - tres conservé chez homologues -pas chevauchement -ORF>50pb et la plus longue

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>*** PhagesDB :</p> <p>(q# >Mandalorian_50, function unknown, 68 : s#126) : [97%] ; e-value : 2e-29</p>

*** nr :
(*q# hypothetical protein SEA_MANDALORIAN_50*
[Microbacterium phage Mandalorian]: s#129) : [97%] ; e-
value : 2e-37
hypothetical protein SEA_MANDALORIAN_50 [Microbacterium phage Mandalorian]
Sequence ID: [OXO13642.1](#) Length: 68 Number of Matches: 1

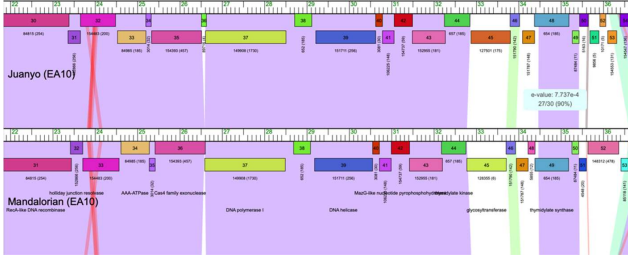
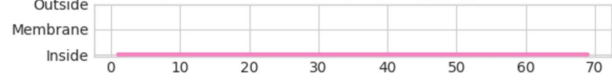
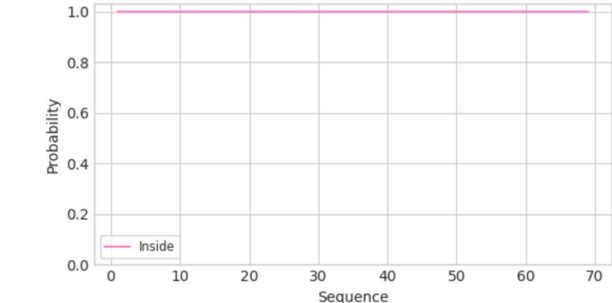
Range 1: 1 to 68 [GenPept](#) [Graphics](#) [▼ Next Match](#) [▶ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
129 bits(325)	2e-37	Compositional matrix adjust.	66/68(97%)	67/68(98%)	0/68(0%)
Query 1	MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTQAVATKGLHADLVMMYTGSDYVRKRAA	60			
Sbjct 1	MNNIKKLLIAAINDQDSRVIRRLADAVETKVVTQAVATK+GLHADLVMMYTGSDYVRKRAA	60			
Query 61	VSALALVK SALALVK	68			
Sbjct 61	ASALALVK	68			

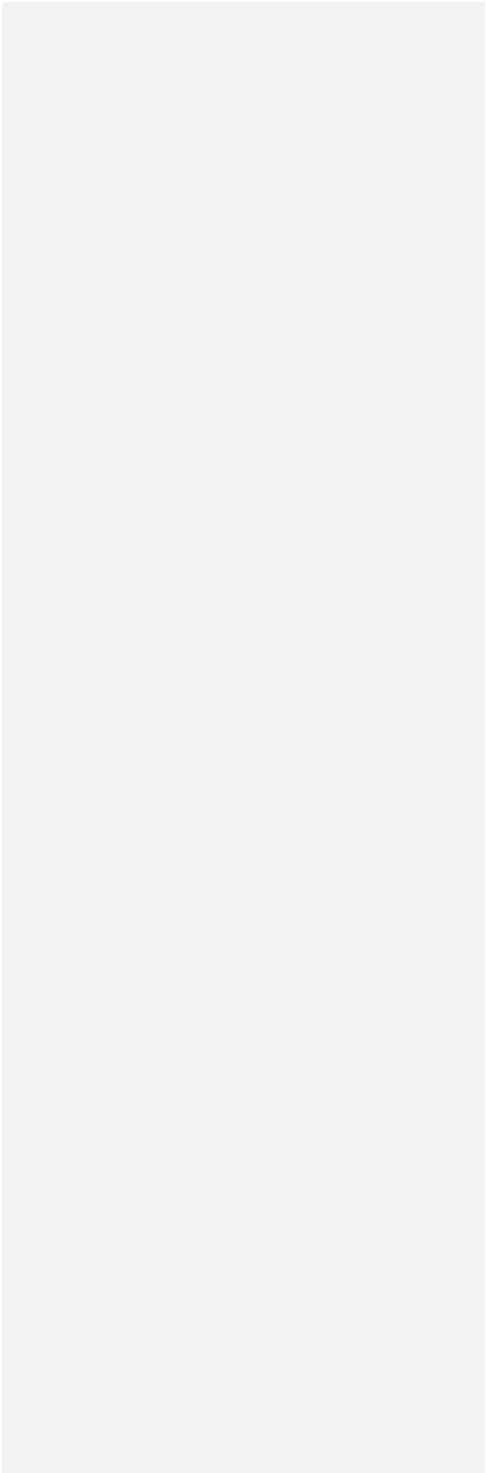
HHPred : DUF577 ; Family of unknown function (DUF577)
Hit [PF04510.16](#)
Probability 80,55
Database Pfam



Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

	<p><i>Pas de correspondance pertinente car l'alignement ne donne pas de fonction</i></p>
<p>Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?</p>	<p><i>Aucune synténie observée</i></p> <p><i>Fin du gène avec région conservée entre plusieurs EA10 avec e-value = 7,737 e-4</i></p> <p>Carostasia_Draft (EA10)</p>  <p><i>Juanyo (EA10)</i></p> <p><i>Mandalorian (EA10)</i></p>
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p><i>No</i></p> <p><i>DeepTMHMM :</i></p> <p>DeepTMHMM - Most Likely Topology Type: Globular</p>  <p>DeepTMHMM - Posterior Probabilities</p>  <p><i>Phoebeius :</i></p>

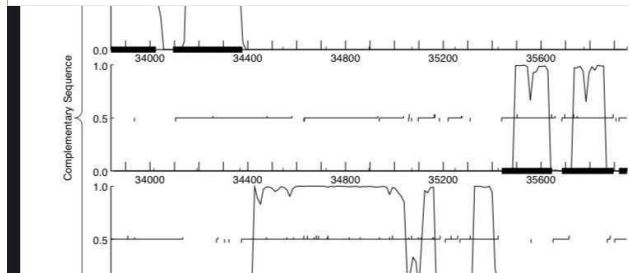
	<div><div>IDUNNAMED FT //</div><div>TOPO_DOM169NON CYTOPLASMIC.</div></div> <div><p>Phobius posterior probabilities for UNNAMED</p><p>Posterior label probability</p><p>transmembrane cytoplasmic non cytoplasmic signal peptide</p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	
DECISION:	NKF



Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	51
Coordonnées du Stop	35685
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Gap 21
Coordonnées du Start retenu	35897
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GenemarkS pouvoir codant élevé dans la region de l'ORF avec la sequence complementaire</i></p> <p><i>Genemark folio. Pareil</i></p> 

Commenté [13]: page entière



Mouse-over to show define and scores. Click to show alignments:

Distribution of 21 Blast Hits on the Query Sequence

Color Key for Alignment Scores

<40 40-50 50-60 60-70 70-80

0 50 100

Carostasia_Draft_51, function unknown, 70

Mandalorian_51, function unknown, 70

Nucci_51, function unknown, 70

Quartz_52, function unknown, 70

Yuyu_52, function unknown, 65

Zenitsu_52, function unknown, 119

Saratos_52, function unknown, 119

MCubed_52, function unknown, 119

Finny_53, function unknown, 119

Eleri_52, function unknown, 119

ColaCorta_52, function unknown, 119

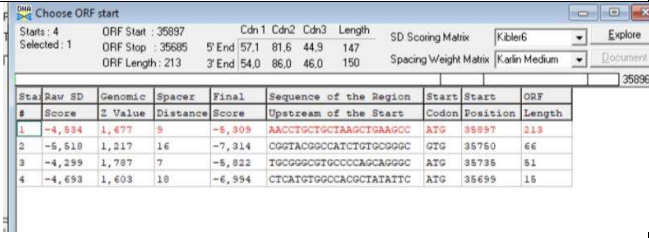
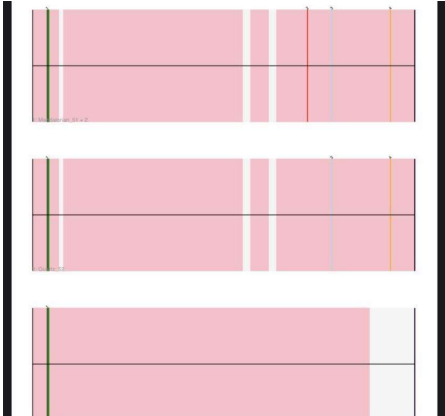
Sequences producing significant alignments:	Score (bits)	E Value
Carostasia_Draft_51, function unknown, 70	145	3e-36
Mandalorian_51, function unknown, 70	137	7e-36
Nucci_51, function unknown, 70	137	9e-36
Quartz_52, function unknown, 70	136	2e-36
Yuyu_52, function unknown, 65	66	3e-16
Zenitsu_52, function unknown, 119	47	2e-11
Saratos_52, function unknown, 119	47	2e-11
MCubed_52, function unknown, 119	47	2e-11
Finny_53, function unknown, 119	47	2e-11
Eleri_52, function unknown, 119	47	2e-11
ColaCorta_52, function unknown, 119	47	2e-11

Non annotés mais beaucoup d' homologues

PAS DE CONTRADICTION
-plus de 50pb
-PAS chevauchement
-direction similaire aux gènes qui l'entourent

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 35897</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 35897</i>
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? <i>Si le plus long pas retenu, quel est l'écart intergène résultant ?</i>	<i>Le start 35897 conduit à une ORF de 213pb, la plus longue</i> <i>PAS DE chevauchement</i> <i>Les autres starts conduisent à des ORFs nettement plus courtes</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	 <i>Start le plus conservé est 35897</i>

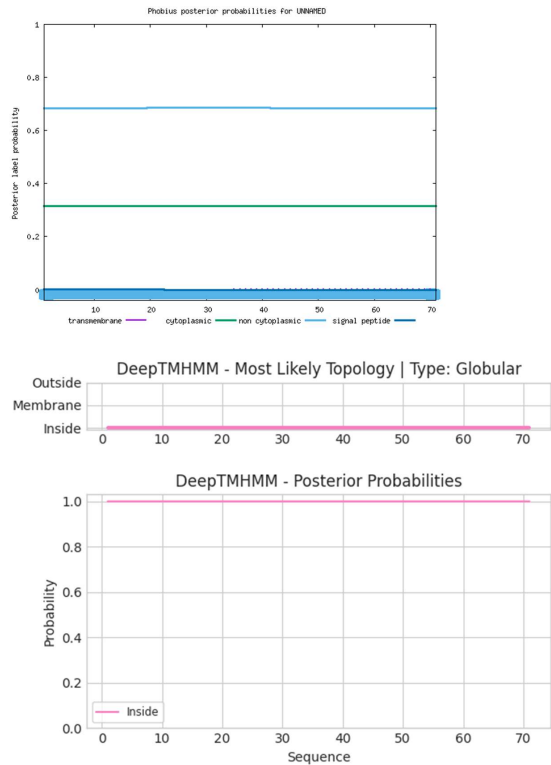
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<pre> Query: 1 MDITCSAAKYTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTAICAGVPQQGMFYTF 60 MDITCSAAKYTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTAICAGVPQQGMFYTF Sbjct: 1 MDITCSAAKYTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTAICAGVPQQGMFYTF 60 Query: 61 CGHAIFMESA 70 CGHAIFMESA Sbjct: 61 CGHAIFMESA 70 >Mandalorian_51, function unknown, 70 Length = 70 Score = 137 bits (346), Expect = 7e-33 Identities = 66/70 (94%), Positives = 67/70 (95%) Query: 1 MDITCSAAKYTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTAICAGVPQQGMFYTF 60 MDITCSA K TQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTA CAGVPQQGMFYTF Sbjct: 1 MDITCSAGKLTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTATCAGVPQQGMFYTF 60 Query: 61 CGHAIFMESA 70 CGHAIFM+SA Sbjct: 61 CGHAIFMESA 70 >Nucci_51, function unknown, 70 Length = 70 Score = 137 bits (345), Expect = 9e-33 Identities = 64/70 (91%), Positives = 67/70 (95%) Query: 1 MDITCSAAKYTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTAICAGVPQQGMFYTF 60 MDITCSA K TQVERAWLATASSFDRLQFQQQRHGV VVDVEAGTA+CAGVPQQGMFYTF Sbjct: 1 MDITCSAGKLTQVERAWLATASSFDRLQFQQQRHGA VVDVEAGTAVCAGVPQQGMFYTF 60 Query: 61 CGHAIFMESA 70 CGHA+FM+SA Sbjct: 61 CGHAIFMESA 70 >Quartz_52, function unknown, 70 Length = 70 Score = 136 bits (342), Expect = 2e-32 Identities = 65/70 (92%), Positives = 66/70 (94%) Query: 1 MDITCSAAKYTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTAICAGVPQQGMFYTF 60 MDITCSA K TQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTA CAG PQGMFYTF Sbjct: 1 MDITCSAGKLTQVERAWLATASSFDRLQFQQQRHGVVVDVEAGTATCAGAPQGMFYTF 60 Query: 61 CGHAIFMESA 70 CGHAIFM+SA Sbjct: 61 CGHAIFMESA 70 </pre> <p>Le Start 35897est conservé</p>
<p>DECISION:</p>	<p>35897: -bon score -bonne zvalue - tres conservé chez homologues -pas chevauchement -ORF>50pb et la plus longue</p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																						
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value : Carostasia_Draft_51, function unknown, 70 145 3e-35 Mandalorian_51, function unknown, 70 137 7e-33 Nucci_51, function unknown, 70 137 9e-33 Quartz_52, function unknown, 70 136 2e-32 YuuY_52, function unknown, 65 66 3e-11 Zenitsu_52, function unknown, 119 48 1e-05</p> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p> <table><thead><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr></thead><tbody><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_51 [Microbacterium phage Mandalorian]</td><td>Microbacterium phage Mandalorian</td><td>139</td><td>139</td><td>98%</td><td>6e-41</td><td>94.29%</td><td>70</td><td>QX013643.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_51 [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>137</td><td>137</td><td>98%</td><td>2e-40</td><td>91.43%</td><td>70</td><td>QX052444.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_52 [Microbacterium phage Quartz]</td><td>Microbacterium phage Quartz</td><td>137</td><td>137</td><td>98%</td><td>3e-40</td><td>92.86%</td><td>70</td><td>LV055271.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_YUY_52 [Microbacterium phage YuuY]</td><td>Microbacterium phage YuuY</td><td>67.0</td><td>67.0</td><td>91%</td><td>1e-12</td><td>52.31%</td><td>65</td><td>QF35380.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein [Microbacterium sp. AK031]</td><td>Microbacterium sp. AK031</td><td>52.0</td><td>52.0</td><td>88%</td><td>2e-06</td><td>41.64%</td><td>76</td><td>WP_259546164.1</td></tr></tbody></table> <p>Fonction inconnue chaque fois</p>	Description	Scientific Name	Max Score	Total Score	Query Cover	E	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_51 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	139	139	98%	6e-41	94.29%	70	QX013643.1	<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_51 [Microbacterium phage Nucci]	Microbacterium phage Nucci	137	137	98%	2e-40	91.43%	70	QX052444.1	<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_52 [Microbacterium phage Quartz]	Microbacterium phage Quartz	137	137	98%	3e-40	92.86%	70	LV055271.1	<input checked="" type="checkbox"/> hypothetical protein SEA_YUY_52 [Microbacterium phage YuuY]	Microbacterium phage YuuY	67.0	67.0	91%	1e-12	52.31%	65	QF35380.1	<input checked="" type="checkbox"/> hypothetical protein [Microbacterium sp. AK031]	Microbacterium sp. AK031	52.0	52.0	88%	2e-06	41.64%	76	WP_259546164.1
Description	Scientific Name	Max Score	Total Score	Query Cover	E	Per. Ident	Acc. Len	Accession																																															
<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_51 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	139	139	98%	6e-41	94.29%	70	QX013643.1																																															
<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_51 [Microbacterium phage Nucci]	Microbacterium phage Nucci	137	137	98%	2e-40	91.43%	70	QX052444.1																																															
<input checked="" type="checkbox"/> hypothetical protein SEA_QUARTZ_52 [Microbacterium phage Quartz]	Microbacterium phage Quartz	137	137	98%	3e-40	92.86%	70	LV055271.1																																															
<input checked="" type="checkbox"/> hypothetical protein SEA_YUY_52 [Microbacterium phage YuuY]	Microbacterium phage YuuY	67.0	67.0	91%	1e-12	52.31%	65	QF35380.1																																															
<input checked="" type="checkbox"/> hypothetical protein [Microbacterium sp. AK031]	Microbacterium sp. AK031	52.0	52.0	88%	2e-06	41.64%	76	WP_259546164.1																																															
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<p>Non</p> <table><thead><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr></thead><tbody><tr><td><input type="checkbox"/> 1</td><td>PF18340.5</td><td>; TraI_2B ; DNA relaxase TraI_2B/2B-like domain</td><td>54.27</td><td>59</td><td>18.75</td><td>3</td><td>52</td><td>80</td></tr><tr><td><input type="checkbox"/> 2</td><td>3HQI_A</td><td>Peptide methionine sulfoxide reductase; Methionine sulfoxide reductase B, Xanthomonas campestris, oxidized form, OXIDORE</td><td>44.63</td><td>98</td><td>20.61</td><td>3.3</td><td>60</td><td>154</td></tr><tr><td><input type="checkbox"/> 3</td><td>6VWY_EE</td><td>Related to ribosomal protein S5 (Mitochondrial); Neurospora crassa, translating Mitochondrion, tRNA, mRNA, m108, TRANSL</td><td>37.47</td><td>48</td><td>29.28</td><td>1.4</td><td>21</td><td>477</td></tr><tr><td><input type="checkbox"/> 4</td><td>6SYM_A</td><td>Peptide methionine sulfoxide reductase MsrB; Reductase, Zinc-binding, R-; Methionine sulfoxide reductase, Oxidative stress</td><td>36.09</td><td>120</td><td>18.96</td><td>2.6</td><td>57</td><td>137</td></tr></tbody></table> <p>Visualization</p>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/> 1	PF18340.5	; TraI_2B ; DNA relaxase TraI_2B/2B-like domain	54.27	59	18.75	3	52	80	<input type="checkbox"/> 2	3HQI_A	Peptide methionine sulfoxide reductase; Methionine sulfoxide reductase B, Xanthomonas campestris, oxidized form, OXIDORE	44.63	98	20.61	3.3	60	154	<input type="checkbox"/> 3	6VWY_EE	Related to ribosomal protein S5 (Mitochondrial); Neurospora crassa, translating Mitochondrion, tRNA, mRNA, m108, TRANSL	37.47	48	29.28	1.4	21	477	<input type="checkbox"/> 4	6SYM_A	Peptide methionine sulfoxide reductase MsrB; Reductase, Zinc-binding, R-; Methionine sulfoxide reductase, Oxidative stress	36.09	120	18.96	2.6	57	137									
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																															
<input type="checkbox"/> 1	PF18340.5	; TraI_2B ; DNA relaxase TraI_2B/2B-like domain	54.27	59	18.75	3	52	80																																															
<input type="checkbox"/> 2	3HQI_A	Peptide methionine sulfoxide reductase; Methionine sulfoxide reductase B, Xanthomonas campestris, oxidized form, OXIDORE	44.63	98	20.61	3.3	60	154																																															
<input type="checkbox"/> 3	6VWY_EE	Related to ribosomal protein S5 (Mitochondrial); Neurospora crassa, translating Mitochondrion, tRNA, mRNA, m108, TRANSL	37.47	48	29.28	1.4	21	477																																															
<input type="checkbox"/> 4	6SYM_A	Peptide methionine sulfoxide reductase MsrB; Reductase, Zinc-binding, R-; Methionine sulfoxide reductase, Oxidative stress	36.09	120	18.96	2.6	57	137																																															
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<p>Fonction inconnue.</p>																																																						

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

Non



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue

DECISION:

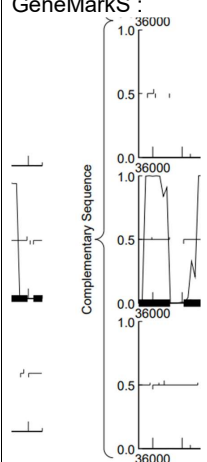
NFK

Student Gene Annotation Worksheet

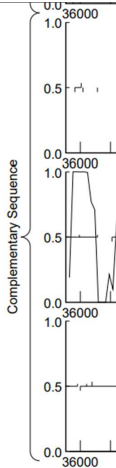
Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	52
Coordonnées du Stop	35919
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	Gap de 45pb
Coordonnées du Start retenu	36059
Fonction prédite	NFK

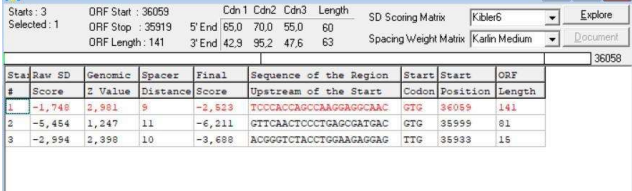
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS :</p>  <p>GeneMark :</p>

Commenté [14]: page entière

													
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<table><tr><td>Sequences producing significant alignments:</td><td>Score (bits)</td><td>E Value</td></tr><tr><td>Carostasia_Draft_52, function unknown, 46</td><td><u>100</u></td><td>1e-21</td></tr><tr><td>Quartz_53, function unknown, 46</td><td><u>98</u></td><td>6e-21</td></tr><tr><td>Mandalorian_52, function unknown, 46</td><td><u>98</u></td><td>6e-21</td></tr></table>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_52, function unknown, 46	<u>100</u>	1e-21	Quartz_53, function unknown, 46	<u>98</u>	6e-21	Mandalorian_52, function unknown, 46	<u>98</u>	6e-21
Sequences producing significant alignments:	Score (bits)	E Value											
Carostasia_Draft_52, function unknown, 46	<u>100</u>	1e-21											
Quartz_53, function unknown, 46	<u>98</u>	6e-21											
Mandalorian_52, function unknown, 46	<u>98</u>	6e-21											
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions												
DECISION:	OUI												

Décision #2 : Quel est le start du gène ?

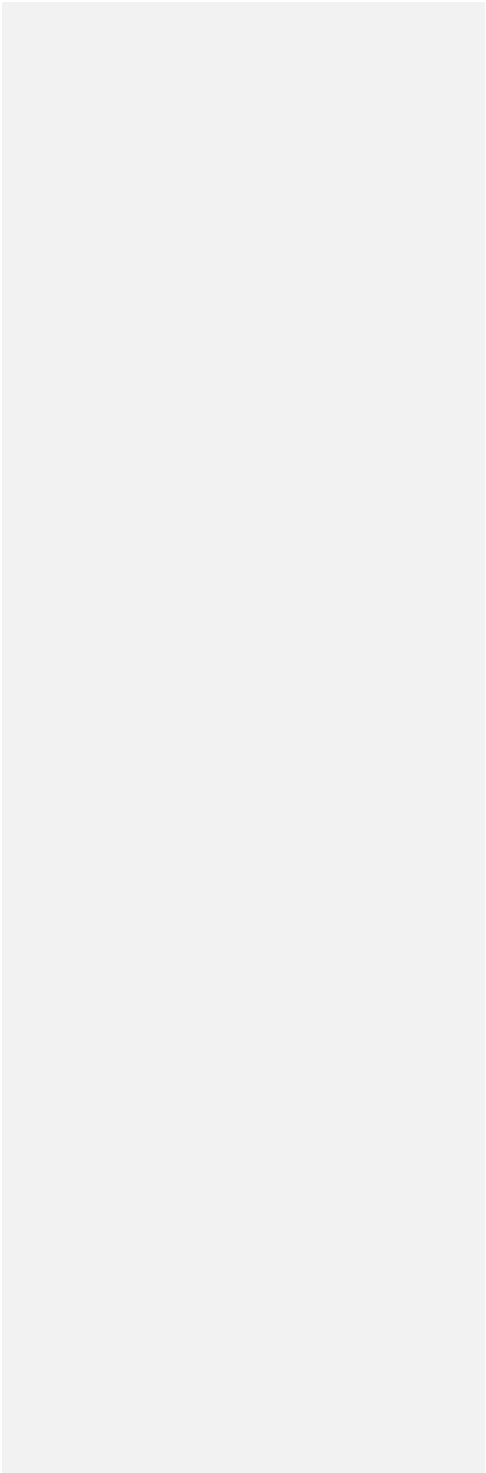
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 36059 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 36059
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <p>On choisit le premier start proposé</p>
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Start proposé : GTG, ORF le plus long, 141 pb. Correspond au départ du pouvoir codant.
Est-ce que le start est conservé chez les homologues voir Starterator ?	<ul style="list-style-type: none">Track 1 : Quartz_53, Manualation_53Track 2 : Carostasia_52, Yuyu_53
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	>Quartz_53, function unknown, 46 Length = 46 Score = 98.2 bits (243), Expect = 6e-21 Identities = 45/46 (97%), Positives = 46/46 (100%) Query: 1 MNDKAIWSATKKWFNSLSDDVQWGLTEHERTEARRVYLEEELQEG 46 MNDKAIWSATKKWFNSLSDDVQWGLTE+ERTEARRVYLEEELQEG Sbjct: 1 MNDKAIWSATKKWFNSLSDDVQWGLTEYERTEARRVYLEEELQEG 46
DECISION:	36059

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
------------------------------------	-------------

<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?</p>	<p>*** PhagesDB : (q#Quartz_53, function unknown, 46: s#98,2) : [97%] ; e-value : 3e-21</p> <p>>Quartz_53, function unknown, 46 Length = 46</p> <p>Score = 98.2 bits (243), Expect = 6e-21 Identities = 45/46 (97%), Positives = 46/46 (100%)</p> <p>Query: 1 MNDKAIWSATKKWFNSLSDVQWGDLTEHERTEARRVYLEEELQEG 46 MNDKAIWSATKKWFNSLSDVQWGDLTE+ERTEARRVYLEEELQEG Sbjct: 1 MNDKAIWSATKKWFNSLSDVQWGDLTEYERTEARRVYLEEELQEG 46</p> <p>*** nr : (q#hypothetical protein SEA_MANDALORIAN_52 [Microbacterium phage Mandalorian] : s#94) : [98%] ; e-value : 7e-24</p> <p>hypothetical protein SEA_MANDALORIAN_52 [Microbacterium phage Mandalorian] Sequence ID: OX013644.1 Length: 46 Number of Matches: 1 See 1 more title(s) ▾ See all Identical Proteins(IPG)</p> <p>Range 1: 1 to 46 GenPept Graphics ▼ Next Match ▲ Prev</p> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>94.0 bits(232)</td><td>7e-24</td><td>Compositional matrix adjust.</td><td>45/46(98%)</td><td>46/46(100%)</td><td>0/46(0%)</td></tr></table> <p>Query 1 MNDKAIWSATKKWFNSLSDVQWGDLTEHERTEARRVYLEEELQEG 46 MNDKAIWSATKKWFNSLSDVQWGDLTE+ERTEARRVYLEEELQEG Sbjct 1 MNDKAIWSATKKWFNSLSDVQWGDLTEYERTEARRVYLEEELQEG 46</p> <p>Les deux blastp donnent comme résultats des fonctions inconnues</p>	Score	Expect	Method	Identities	Positives	Gaps	94.0 bits(232)	7e-24	Compositional matrix adjust.	45/46(98%)	46/46(100%)	0/46(0%)
Score	Expect	Method	Identities	Positives	Gaps								
94.0 bits(232)	7e-24	Compositional matrix adjust.	45/46(98%)	46/46(100%)	0/46(0%)								
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?</p>	<p>HHPred : DUF5347 ; Family of unknown function (DUF5347) Hit PF17282.6 Probability 86,22 Database Pfam</p> <p>Pas de correspondance pertinente car l'alignement ne donne pas de fonction</p>												

	<div><div>DeepTMHMM - Most Likely Topology Type: Globular</div><div><div>Outside</div><div>Membrane</div><div>Inside</div></div><div><div>0</div><div>10</div><div>20</div><div>30</div><div>40</div></div></div> <div><div>DeepTMHMM - Posterior Probabilities</div><div><div>1.0</div><div>0.8</div><div>0.6</div><div>0.4</div><div>0.2</div><div>0.0</div></div><div>Probability</div><div><div>Inside</div></div><div><div>0</div><div>10</div><div>20</div><div>30</div><div>40</div></div><div>Sequence</div></div> <div><div>Phoebeius :</div><div>ID UNNAMED</div><div>FT TOPO_DOM</div><div>//</div><div>1</div><div>47</div><div>NON CYTOPLASMIC.</div></div> <div><div>Phoebeius posterior probabilities for UNNAMED</div><div><div>1</div><div>0.8</div><div>0.6</div><div>0.4</div><div>0.2</div><div>0</div></div><div>Posterior label probability</div><div><div>5</div><div>10</div><div>15</div><div>20</div><div>25</div><div>30</div><div>35</div><div>40</div><div>45</div></div><div>transmembrane cytoplasmic non cytoplasmic signal peptide</div></div>
--	---

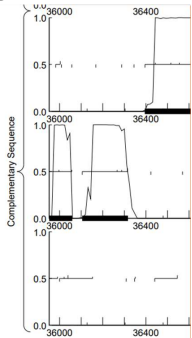


Student Gene Annotation Worksheet

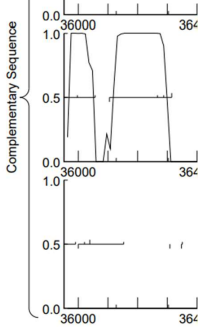
Carbone Melwin

Basic Phage Information	
Nom du Phage	Carostasia
Gène #	53
Coordonnées du Stop	36105
Direction (For/Rev)	Reverse
Gap/chevauchement avec un autre gène	Gap 77
Coordonnées du Start retenu	36317
Fonction prédite	NFK

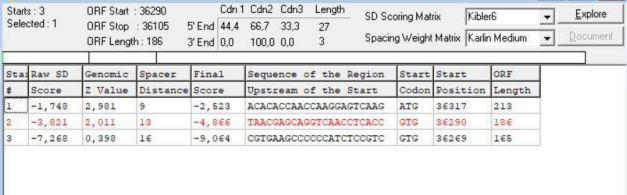
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS :</p>  <p>GeneMark :</p>

Commenté [15]: page entière

	 <p>Pouvoir codant stable.</p>												
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div>Sequences producing significant alignments:</div> <table><tr><th></th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_53, function unknown, 70</td><td>151</td><td>6e-37</td></tr><tr><td>Quartz_54, function unknown, 70</td><td>148</td><td>5e-36</td></tr><tr><td>Mandalorian_53, function unknown, 70</td><td>148</td><td>5e-36</td></tr></table>		Score (bits)	E Value	Carostasia_Draft_53, function unknown, 70	151	6e-37	Quartz_54, function unknown, 70	148	5e-36	Mandalorian_53, function unknown, 70	148	5e-36
	Score (bits)	E Value											
Carostasia_Draft_53, function unknown, 70	151	6e-37											
Quartz_54, function unknown, 70	148	5e-36											
Mandalorian_53, function unknown, 70	148	5e-36											
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	Pas de contradictions												
DECISION:	OUI												

Décision #2 : Quel est le start du gène ?

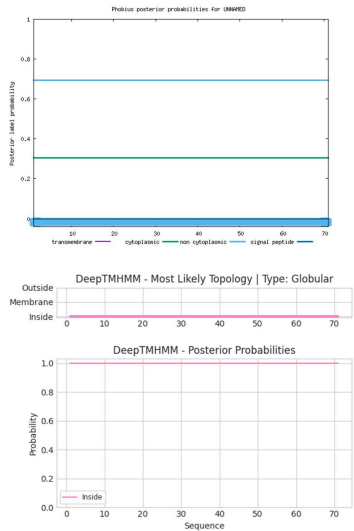
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 36317 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 36317
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <p>On choisit le premier start proposé</p>
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	ATG, ORF le plus long, 213pb. Pas de chevauchement.
Est-ce que le start est conservé chez les homologues voir Starterator ?	Track 10 : Quartz_54 • Track 11 : Quartz_54, Carostasia_53, Mandalorian_53
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	>Quartz_54, function unknown, 70 Length = 70 Score = 148 bits (373), Expect = 5e-36 Identities = 69/70 (98%), Positives = 69/70 (98%) Query: 1 MYNEQVNLTVKPPISVVRNHLENDWTTTEVDQFEGYFFRNVDGTTTTLFRQWEEAIY 60 MYNEQVNLTVKP ISVVRNHLENDWTTTEVDQFEGYFFRNVDGTTTTLFRQWEEAIY Sbjct: 1 MYNEQVNLTVKPTISVVRNHLENDWTTTEVDQFEGYFFRNVDGTTTTLFRQWEEAIY 60 Query: 61 AYSNEHFDLD 70 AYSNEHFDLD Sbjct: 61 AYSNEHFDLD 70
DECISION:	36317

Commenté [16]: justif

Décision #3 : Quelle est la fonction de la protéine putative ?

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

NON.



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue.

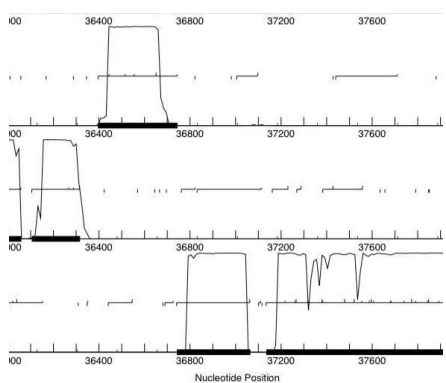
DECISION:

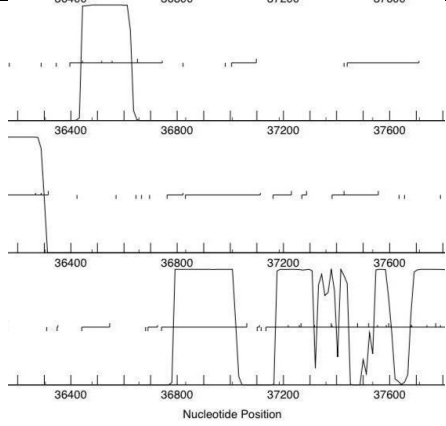
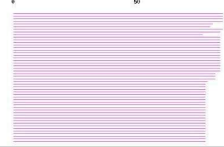
NFK

Student Gene Annotation Worksheet

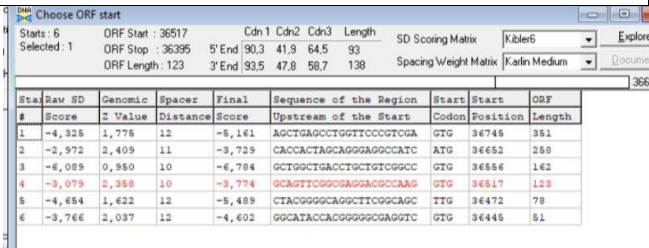

Basic Phage Information	
Nom du Phage	carostasia
Gène #	54
Coordonnées du Stop	36395
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Ove 4
Coordonnées du Start retenu	36745
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH (not the same start)
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenemarkS pouvoir codant élevé dans la region de l'ORF</p>  <p>The figure displays a GenemarkS codon potential plot. The x-axis represents the Nucleotide Position, ranging from 36400 to 37600. The y-axis represents the codon potential score, ranging from 0 to 100. A prominent, sharp peak is visible around position 36745, reaching a score of approximately 100. This indicates a high potential for coding in this region. The plot also shows several smaller peaks and troughs across the entire range.</p> <p>avec la sequence complementaire Genemark folio. Pareil</p>

																															
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div>Click here to show details and scores. Click to show alignments</div><div>Color Key for Alignment Scores</div><div>100% 50% 0%</div><div>100 50 0</div><div></div></div> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Quartz_55, function unknown, 116</td><td>172</td><td>3e-43</td></tr><tr><td>Carostasia_Draft_54, function unknown, 85</td><td>172</td><td>3e-43</td></tr><tr><td>Mandalorian_54, function unknown, 85</td><td>171</td><td>4e-43</td></tr><tr><td>YuuV_55, function unknown, 85</td><td>162</td><td>2e-40</td></tr><tr><td>Nucci_54, function unknown, 87</td><td>145</td><td>3e-35</td></tr><tr><td>Kauaia_48, function unknown, 81</td><td>144</td><td>8e-35</td></tr><tr><td>Juanyo_56, function unknown, 111</td><td>140</td><td>8e-34</td></tr><tr><td>Koji_48, function unknown, 85</td><td>135</td><td>4e-32</td></tr><tr><td>MonChoix_52, function unknown, 101</td><td>129</td><td>2e-30</td></tr></table> <p>Non annotés mais beaucoup d' homologues</p>	Sequences producing significant alignments:	Score (bits)	E Value	Quartz_55, function unknown, 116	172	3e-43	Carostasia_Draft_54, function unknown, 85	172	3e-43	Mandalorian_54, function unknown, 85	171	4e-43	YuuV_55, function unknown, 85	162	2e-40	Nucci_54, function unknown, 87	145	3e-35	Kauaia_48, function unknown, 81	144	8e-35	Juanyo_56, function unknown, 111	140	8e-34	Koji_48, function unknown, 85	135	4e-32	MonChoix_52, function unknown, 101	129	2e-30
Sequences producing significant alignments:	Score (bits)	E Value																													
Quartz_55, function unknown, 116	172	3e-43																													
Carostasia_Draft_54, function unknown, 85	172	3e-43																													
Mandalorian_54, function unknown, 85	171	4e-43																													
YuuV_55, function unknown, 85	162	2e-40																													
Nucci_54, function unknown, 87	145	3e-35																													
Kauaia_48, function unknown, 81	144	8e-35																													
Juanyo_56, function unknown, 111	140	8e-34																													
Koji_48, function unknown, 85	135	4e-32																													
MonChoix_52, function unknown, 101	129	2e-30																													
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION :</p> <ul style="list-style-type: none">-plus de 50pb-chevauchement inférieur à 30pb-direction similaire aux gènes qui l'entourent																														
DECISION:	OUI																														

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 36652</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 36745</i>
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<i>Le start 36745 conduit à une ORF de 351pb, la plus longue chevauchement -4</i> <i>Les autres starts conduisent à des ORFs nettement plus courtes</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	

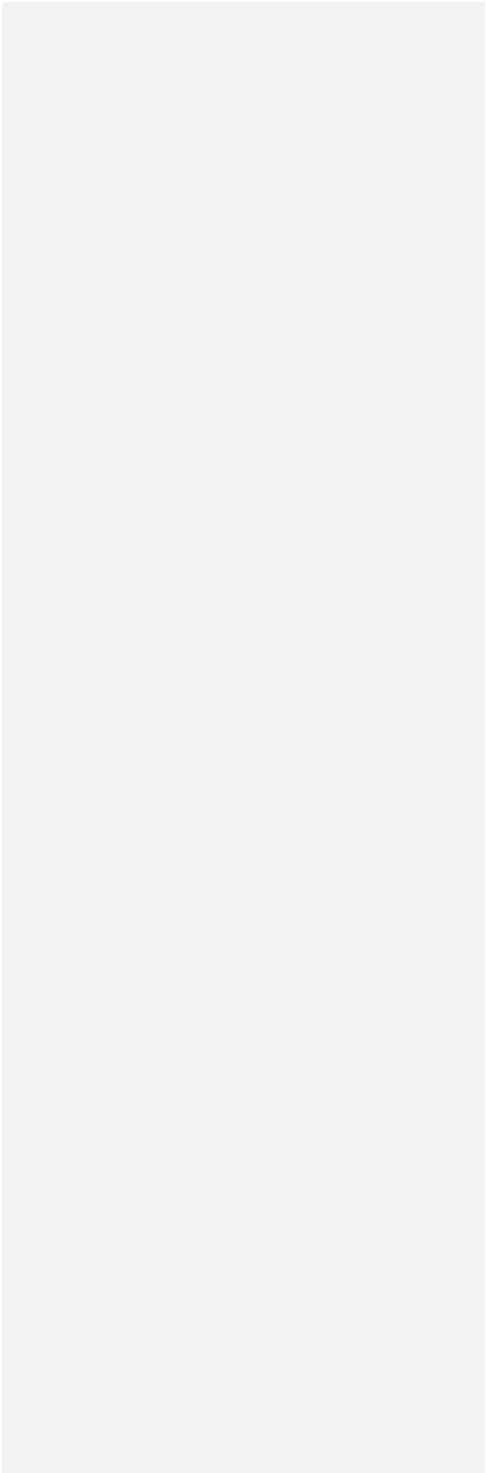
	<p>Start les plus conservés 36652 et 36745</p> <pre> ry: 1 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 60 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS ct: 32 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 91 ry: 61 LQAYHGGEVVISDAAPDDEDDDPYI 85 LQAYHGGEVVISDAAPDDEDDDPYI ct: 92 LQAYHGGEVVISDAAPDDEDDDPYI 116 rostasia_Draft_54, function unknown, 85 Length = 85 ore = 172 bits (435), Expect = 3e-43 entities = 85/85 (100%), Positives = 85/85 (100%) ry: 1 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 60 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS ct: 1 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 60 ry: 61 LQAYHGGEVVISDAAPDDEDDDPYI 85 LQAYHGGEVVISDAAPDDEDDDPYI ct: 61 LQAYHGGEVVISDAAPDDEDDDPYI 85 ndalorian_54, function unknown, 85 Length = 85 ore = 171 bits (434), Expect = 4e-43 entities = 84/85 (98%), Positives = 85/85 (100%) ry: 1 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 60 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS ct: 1 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 60 ry: 61 LQAYHGGEVVISDAAPDDEDDDPYI 85 LQAYHGGEVVISDAAPDDEDDDPYI ct: 61 LQAYHGGEVVISDAAPDDEDDDPYI 85 uY_55, function unknown, 85 Length = 85 ore = 162 bits (411), Expect = 2e-40 entities = 79/85 (92%), Positives = 84/85 (98%) ry: 1 MASTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVANGQRYGAGFGS 60 MA+TVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVV+NGQRYGAGFGS ct: 1 MANTVLIGVNDNQGYAPDQINTDVTLADLLSAVQDAIQQFGEDAKVVVNGQRYGAGFGS 60 ry: 61 LQAYHGGEVVISDAAPDDEDDDPYI 85 L++Y GGEVVISDAAPDDEDDDPY+ ct: 61 LESYMGGEVVISDAAPDDEDDDPYL 85 </pre> <p>Le Start 36652 est conservé (3x85=255)sauf pour le premier alignement avec un décalage de 32 AA (116x3=348)</p>
DECISION:	<p>36745: -bon score -bonne zvalue - conservé chez homologues -chevauchement < 30pb -permet d'éviter un long GAP -ORF>50pb et la 2e plus longue</p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction	<p>*** <i>PhagesDB</i> : (q#Quartz_55, function unknown, 116: s#235) : [100%] ; e-value : 3e-62</p>

<

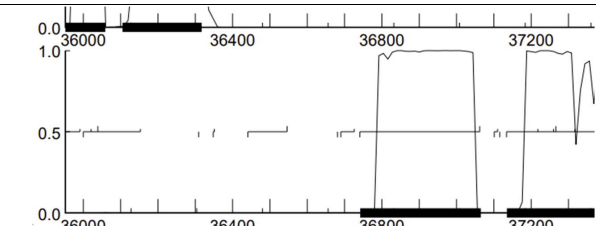
	<div><div><div>ID UNNAMED</div><div>FT TOPO_DOM</div><div>//</div></div><div><div>1</div><div>117</div></div><div><div>NON CYTOPLASMIC.</div></div></div> <div><p>Phobius posterior probabilities for UNNAMED</p><table border="1"><caption>Phobius Posterior Probabilities for UNNAMED</caption><thead><tr><th>Label</th><th>Probability</th></tr></thead><tbody><tr><td>transmembrane</td><td>~0.01</td></tr><tr><td>cytoplasmic</td><td>~0.01</td></tr><tr><td>non cytoplasmic</td><td>~0.25</td></tr><tr><td>signal peptide</td><td>~0.75</td></tr><tr><td>signal peptide</td><td>~0.75</td></tr></tbody></table></div>	Label	Probability	transmembrane	~0.01	cytoplasmic	~0.01	non cytoplasmic	~0.25	signal peptide	~0.75	signal peptide	~0.75
Label	Probability												
transmembrane	~0.01												
cytoplasmic	~0.01												
non cytoplasmic	~0.25												
signal peptide	~0.75												
signal peptide	~0.75												
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui</i>												
DECISION:	<i>NKF</i>												



Student Gene Annotation Worksheet

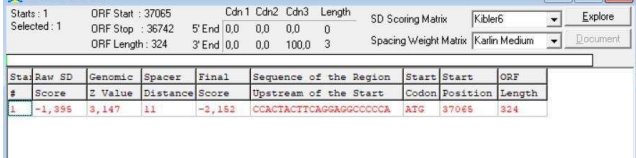
Basic Phage Information	
Nom du Phage	Carostasia
Gène #	55
Coordonnées du Stop	36742
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	Gap 69
Coordonnées du Start retenu	37065
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle																																				
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	Yes Both.																																				
Y-a-t-il des éléments supportant un potentiel codant ?																																					
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div><input checked="" type="checkbox"/> select all 7 sequences selected</div><div><div>GenPent</div><div>Graphics</div><div>Distance tree of results</div><div>Multiple alignment</div><div>MSA View</div></div><table><thead><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr></thead><tbody><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_55 (Microbacterium phage Mandalorian)</td><td>Microbacterium phage Mandalorian</td><td>214</td><td>214</td><td>100%</td><td>2e-69</td><td>100.00%</td><td>107</td><td>QX013847.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_55 (Microbacterium phage Nucci)</td><td>Microbacterium phage Nucci</td><td>202</td><td>202</td><td>100%</td><td>3e-65</td><td>92.52%</td><td>108</td><td>QX002448.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEA_JUANJO_57 (Microbacterium phage Juanyo)</td><td>Microbacterium phage Juanyo</td><td>197</td><td>197</td><td>100%</td><td>3e-63</td><td>91.59%</td><td>108</td><td>LSH44806.1</td></tr></tbody></table><div>Quartz_56, function unknown, 1072179e-57 Mandalorian_55, function unknown, 1072179e-57 Carostasia_Draft_55, function unknown, 1072179e-57 Nucci_55, function unknown, 1082053e-53</div></div> <div>Oui, mais ils ne sont pas annote.</div>	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_55 (Microbacterium phage Mandalorian)	Microbacterium phage Mandalorian	214	214	100%	2e-69	100.00%	107	QX013847.1	<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_55 (Microbacterium phage Nucci)	Microbacterium phage Nucci	202	202	100%	3e-65	92.52%	108	QX002448.1	<input checked="" type="checkbox"/> hypothetical protein SEA_JUANJO_57 (Microbacterium phage Juanyo)	Microbacterium phage Juanyo	197	197	100%	3e-63	91.59%	108	LSH44806.1
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																													
<input checked="" type="checkbox"/> hypothetical protein SEA_MANDALORIAN_55 (Microbacterium phage Mandalorian)	Microbacterium phage Mandalorian	214	214	100%	2e-69	100.00%	107	QX013847.1																													
<input checked="" type="checkbox"/> hypothetical protein SEA_NUCCI_55 (Microbacterium phage Nucci)	Microbacterium phage Nucci	202	202	100%	3e-65	92.52%	108	QX002448.1																													
<input checked="" type="checkbox"/> hypothetical protein SEA_JUANJO_57 (Microbacterium phage Juanyo)	Microbacterium phage Juanyo	197	197	100%	3e-63	91.59%	108	LSH44806.1																													

Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<i>Les genes avant et après sont de même direction. Aucune contradiction avec les principes d'annotation.</i>
DECISION:	<i>Oui</i>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Start 37065.</i>
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <p><i>Le start comment par ATG, il y a qu'un seul choix de RBS.</i></p>
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? <i>Si le plus long pas retenu, quel est l'écart intergène résultant ?</i>	<p><i>L'ORF fait 324 nucleotides. Ecart de 71 nucleotide avec l'ORF en amont.</i></p> <p><i>Le potentiel codant correspond au start choisi.</i></p>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<p>Start 4:</p> <ul style="list-style-type: none"> Found in 16 of 16 (100.0%) of genes in pham Manual Annotations of this start: 14 of 14 Called 100.0% of time when present Phage (with cluster) where this start called: Andromedas_59 (EA2), Carostasia_55 (EA10), ChikPic_59 (EA2), ColaCorta_59 (EA2), Eleri_59 (EA2), Glamour_58 (EA2), Juanyo_57 (EA10), MCubed_59 (EA2), Mandalorian_55 (EA10), Nucci_55 (EA10), Quartz_56 (EA10), Sansa_57 (EA2), Saratos_59 (EA2), Shamu_59 (EA2), Yuyu_56 (EA10), Zenitsu_59 (EA2). <p>Gene: Carostasia_55 Start: 37065, Stop: 36742, Start Num: 4 Candidate Starts for Carostasia_55: (Start: 4 @37065 has 14 MA's),</p>
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Query 1 MATIVTLSTKAGYPVTYLSSEVFNIAGIDGHIYVDLSGKYAAQGTASGKPEVSLIGVT 60 MATIVTLSTKAGYPVTYLSSEVFNIAGIDGHIYVDLSGKYAAQGTASGKPEVSLIGVT Sbjct 1 MATIVTLSTKAGYPVTYLSSEVFNIAGIDGHIYVDLSGKYAAQGTASGKPEVSLIGVT 60</p> <p>Query 1 MATIVTLSTKAGYPVTYLSSEVFNIAGIDGHIYVDLSGKYAAQGTASGKPEVSLIGVT 60 MATIVTLSTKAGYPVTYLSSEVFNIAGIDGHIYVDLSGKYAAQGTASGKPEVSLIGVT Sbjct 1 MATIVTLSTKAGYPVTYLSSEVFNIAGIDGHIYVDLSGKYAAQGTASGKPEVSLIGVT 60</p> <p><i>Start conserve.</i></p>
DECISION:	<p><i>start 37065</i></p> <p><i>Le premier codon start, commence par un ATG, et il n'en existe qu'un seul.</i></p>

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																													
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?</p>	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>(q#: s#) : [alignment] ; e-value :</p> <div><div>Sequences producing significant alignments:</div><div><div>Score</div><div>E</div><div>(bits) Value</div></div><div>Quartz_56, function unknown, 1072179e-57</div><div>Mandalorian_55, function unknown, 1072179e-57</div><div>Carostasia_Draft_55, function unknown, 1072179e-57</div><div>Nucci_55, function unknown, 1082053e-53</div><div>Juanyo_57, function unknown, 1082015e-52</div><div>YuuY_56, function unknown, 1081892e-48</div></div> <p>*** nr :</p> <p>(q#: s#) : [alignment] ; e-value :</p> <table><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cov</th><th>E value</th><th>Pos. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEFA_MANDALORIAN_55 [Mycobacterium phage Mandalorian]</td><td>Mycobacterium phage Mandalorian</td><td>214</td><td>214</td><td>99%</td><td>2e-68</td><td>100.00%</td><td>107</td><td>GH013847.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEFA_NUCCI_55 [Mycobacterium phage Nucci]</td><td>Mycobacterium phage Nucci</td><td>202</td><td>202</td><td>99%</td><td>3e-65</td><td>92.92%</td><td>108</td><td>GH013848.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEFA_JUANYO_57 [Mycobacterium phage Juanyo]</td><td>Mycobacterium phage Juanyo</td><td>199</td><td>199</td><td>100%</td><td>1e-63</td><td>98.74%</td><td>108</td><td>U0184895.1</td></tr><tr><td><input checked="" type="checkbox"/> hypothetical protein SEFA_YUUY_56 [Mycobacterium phage YuuY]</td><td>Mycobacterium phage YuuY</td><td>109</td><td>109</td><td>100%</td><td>9e-60</td><td>85.19%</td><td>108</td><td>GF295364.1</td></tr></table> <p>Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.</p>	Description	Scientific Name	Max Score	Total Score	Query Cov	E value	Pos. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/> hypothetical protein SEFA_MANDALORIAN_55 [Mycobacterium phage Mandalorian]	Mycobacterium phage Mandalorian	214	214	99%	2e-68	100.00%	107	GH013847.1	<input checked="" type="checkbox"/> hypothetical protein SEFA_NUCCI_55 [Mycobacterium phage Nucci]	Mycobacterium phage Nucci	202	202	99%	3e-65	92.92%	108	GH013848.1	<input checked="" type="checkbox"/> hypothetical protein SEFA_JUANYO_57 [Mycobacterium phage Juanyo]	Mycobacterium phage Juanyo	199	199	100%	1e-63	98.74%	108	U0184895.1	<input checked="" type="checkbox"/> hypothetical protein SEFA_YUUY_56 [Mycobacterium phage YuuY]	Mycobacterium phage YuuY	109	109	100%	9e-60	85.19%	108	GF295364.1
Description	Scientific Name	Max Score	Total Score	Query Cov	E value	Pos. Ident	Acc. Len	Accession																																						
<input checked="" type="checkbox"/> hypothetical protein SEFA_MANDALORIAN_55 [Mycobacterium phage Mandalorian]	Mycobacterium phage Mandalorian	214	214	99%	2e-68	100.00%	107	GH013847.1																																						
<input checked="" type="checkbox"/> hypothetical protein SEFA_NUCCI_55 [Mycobacterium phage Nucci]	Mycobacterium phage Nucci	202	202	99%	3e-65	92.92%	108	GH013848.1																																						
<input checked="" type="checkbox"/> hypothetical protein SEFA_JUANYO_57 [Mycobacterium phage Juanyo]	Mycobacterium phage Juanyo	199	199	100%	1e-63	98.74%	108	U0184895.1																																						
<input checked="" type="checkbox"/> hypothetical protein SEFA_YUUY_56 [Mycobacterium phage YuuY]	Mycobacterium phage YuuY	109	109	100%	9e-60	85.19%	108	GF295364.1																																						
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?</p>	<div><div>Nr</div><div>Hit</div><div>Name</div><div>Probability</div><div>E-value</div><div>Score</div><div>SS</div><div>Aligned cols</div><div>Target Length</div></div> <div><div><input type="checkbox"/> 1</div><div>PF03829.17</div><div>: PTS30A_gutA : PTS system glucitol/sorbitol-specific IIA component</div><div>55.87</div><div>57</div><div>24.07</div><div>3.9</div><div>34</div><div>113</div></div> <div><div><input type="checkbox"/> 2</div><div>2F9H_B</div><div>PTS system, IIA component: alpha-beta structure, beta-barrel, dimer; Structural Genomics, PSI, Protein Structure Initiat</div><div>52.78</div><div>64</div><div>24.51</div><div>3.8</div><div>34</div><div>129</div></div> <div><div><input type="checkbox"/> 3</div><div>8TXR_C</div><div>Exodeoxyribonuclease 7 large subunit; Exonuclease, Endonuclease, DNA repair; DNA BINDING PROTEIN;(Escherichia coli)</div><div>52.21</div><div>190</div><div>24.28</div><div>6.5</div><div>71</div><div>456</div></div> <div><div><input type="checkbox"/> 4</div><div>2K3M_A</div><div>Rv1761c PROTEIN, integral membrane protein, MEMBRANE PROTEIN; HET: MTN; NMR (Mycobacterium tuberculosis)</div><div>48.99</div><div>31</div><div>28.7</div><div>1.9</div><div>19</div><div>151</div></div>																																													

Visualization

Resubmit Section

59

92

PTS30A_gutA : PTS

2F9H_B

8TXR_C

2K3M_A

Phage IIR-15

006400 family 6

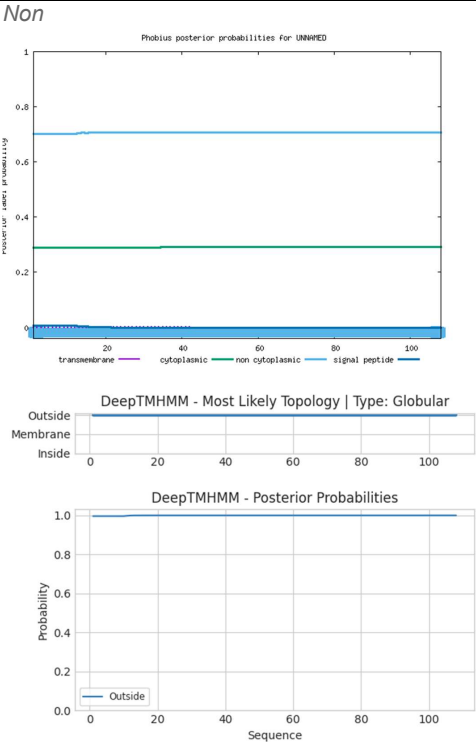
006400 family 6

006400 family 6

NON

Fonction inconnue.

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue

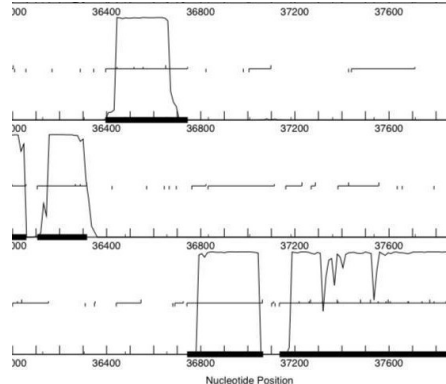
DECISION:

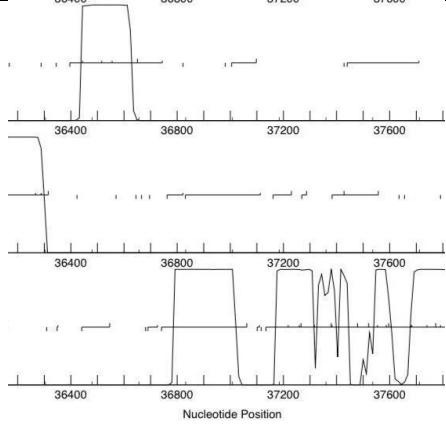
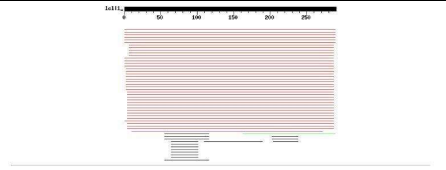
NFK

Student Gene Annotation Worksheet

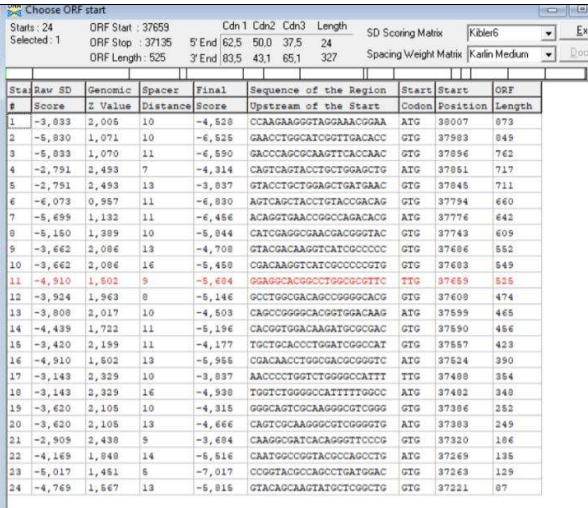
Basic Phage Information	
Nom du Phage	carostasia
Gène #	56
Coordonnées du Stop	37135
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	ove 4
Coordonnées du Start retenu	38007
Fonction prédite	Kinase


Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GenemarkS potentiel codant élevé dans la region de l'ORF</i></p>  <p><i>avec la sequence complementaire</i> <i>Genemark folio. Pareil</i></p>

																																											
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 <p>Sequences producing significant alignments:</p> <table><tr><th></th><th>Score</th><th>E Value</th></tr><tr><td>Carotassia_Draft_56, function unknown, 290</td><td>587</td><td>e-167</td></tr><tr><td>Mandalorian_56, function unknown, 290</td><td>573</td><td>e-163</td></tr><tr><td>Quartz_57, function unknown, 290</td><td>572</td><td>e-163</td></tr><tr><td>Nucl_56, function unknown, 290</td><td>570</td><td>e-162</td></tr><tr><td>Yuv_57, function unknown, 290</td><td>550</td><td>e-156</td></tr><tr><td>Juanyo_58, function unknown, 290</td><td>540</td><td>e-153</td></tr><tr><td>Morrigan_53, sGTP-lyase domain-containing protein, 289</td><td>463</td><td>e-132</td></tr><tr><td>Luna18_53, function unknown, 289</td><td>469</td><td>e-132</td></tr><tr><td>KatChan_52, function unknown, 289</td><td>463</td><td>e-132</td></tr><tr><td>Chepli_53, function unknown, 290</td><td>462</td><td>e-132</td></tr><tr><td>WilliamStrong_52, function unknown, 289</td><td>466</td><td>e-131</td></tr><tr><td>Neferthena_55, function unknown, 291</td><td>459</td><td>e-129</td></tr><tr><td>Alakazan_52, function unknown, 291</td><td>451</td><td>e-127</td></tr></table> <p>Non annotés sauf Morrigan mais beaucoup d' homologues</p>		Score	E Value	Carotassia_Draft_56, function unknown, 290	587	e-167	Mandalorian_56, function unknown, 290	573	e-163	Quartz_57, function unknown, 290	572	e-163	Nucl_56, function unknown, 290	570	e-162	Yuv_57, function unknown, 290	550	e-156	Juanyo_58, function unknown, 290	540	e-153	Morrigan_53, sGTP-lyase domain-containing protein, 289	463	e-132	Luna18_53, function unknown, 289	469	e-132	KatChan_52, function unknown, 289	463	e-132	Chepli_53, function unknown, 290	462	e-132	WilliamStrong_52, function unknown, 289	466	e-131	Neferthena_55, function unknown, 291	459	e-129	Alakazan_52, function unknown, 291	451	e-127
	Score	E Value																																									
Carotassia_Draft_56, function unknown, 290	587	e-167																																									
Mandalorian_56, function unknown, 290	573	e-163																																									
Quartz_57, function unknown, 290	572	e-163																																									
Nucl_56, function unknown, 290	570	e-162																																									
Yuv_57, function unknown, 290	550	e-156																																									
Juanyo_58, function unknown, 290	540	e-153																																									
Morrigan_53, sGTP-lyase domain-containing protein, 289	463	e-132																																									
Luna18_53, function unknown, 289	469	e-132																																									
KatChan_52, function unknown, 289	463	e-132																																									
Chepli_53, function unknown, 290	462	e-132																																									
WilliamStrong_52, function unknown, 289	466	e-131																																									
Neferthena_55, function unknown, 291	459	e-129																																									
Alakazan_52, function unknown, 291	451	e-127																																									
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION :</p> <ul style="list-style-type: none">-plus de 50pb-chevauchement inférieur à 30pb-direction similaire aux gènes qui l'entourent																																										
DECISION:	OUI																																										

Décision #2 : Quel est le start du gène ?

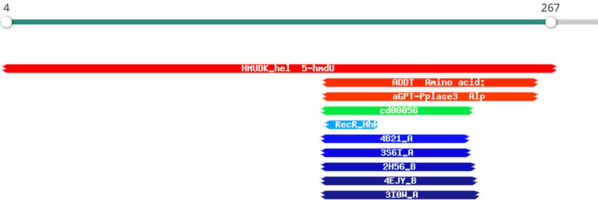
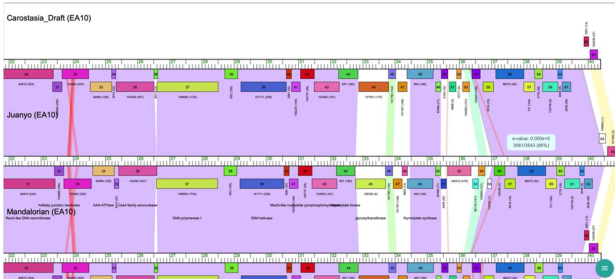
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 38007</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :38007</i>
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<i>Le start 38007 conduit à une ORF de 873pb, la plus longue avec un chevauchement -4</i> <i>Le start 37983 conduit à une ORF plus courte de 849pb avec un score beaucoup moins bon</i> <i>Les autres starts conduisent à des ORFs beaucoup plus courtes</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	

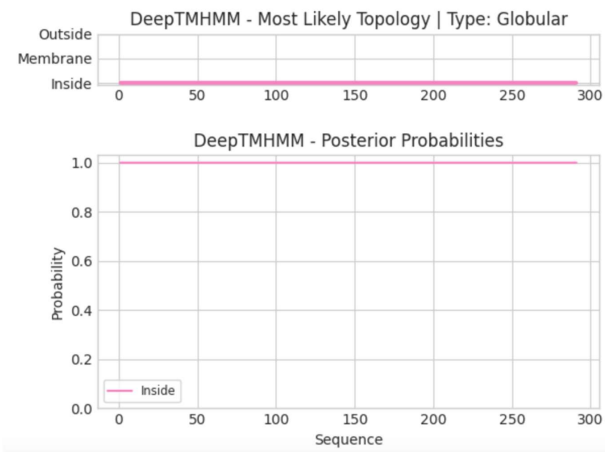
	<div></div> <p>Start le plus conservé est 38007</p>
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div><p>Identities = 290/290 (100%), Positives = 290/290 (100%)</p><p>Query: 1 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60 Sbjct: 1 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60</p><p>Query: 61 DRLAVSYLRQVNRPTMDIIEANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFINGA 120 DRLAVSYLRQVNRPTMDIIEANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFINGA 120 Sbjct: 61 DRLAVSYLRQVNRPTMDIIEANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFINGA 120</p><p>Query: 121 YIILIKPGDSRGTVDKMRDVFPAAAPWIGHVAEDNLATRVMLQETPGLGFFFLAMQIAT 180 YIILIKPGDSRGTVDKMRDVFPAAAPWIGHVAEDNLATRVMLQETPGLGFFFLAMQIAT 180 Sbjct: 121 YIILIKPGDSRGTVDKMRDVFPAAAPWIGHVAEDNLATRVMLQETPGLGFFFLAMQIAT 180</p><p>Query: 181 DLGYTRGEPDQENDYILAGPSRKGVGVMGKATPATPAEAVKAITGFPVERLPTLPESN 240 DLGYTRGEPDQENDYILAGPSRKGVGVMGKATPATPAEAVKAITGFPVERLPTLPESN 240 Sbjct: 181 DLGYTRGEPDQENDYILAGPSRKGVGVMGKATPATPAEAVKAITGFPVERLPTLPESN 240</p><p>Query: 241 GRASLMDVQNVFCEYSKYARLVQOGYKGTGQPTVRNGSFDLTIPEQFKK 290 GRASLMDVQNVFCEYSKYARLVQOGYKGTGQPTVRNGSFDLTIPEQFKK 290 Sbjct: 241 GRASLMDVQNVFCEYSKYARLVQOGYKGTGQPTVRNGSFDLTIPEQFKK 290</p><p>>Mandalorian_56, function unknown, 290 Length = 290 Score = 573 bits (1477), Expect = e-163 Identities = 283/290 (98%), Positives = 285/290 (98%)</p><p>Query: 1 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60 Sbjct: 1 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60</p><p>Query: 61 DRLAVSYLRQVNRPTMDIIEANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFINGA 120 DRLAVSYLRQVNRPTMDIIEANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFINGA 120 Sbjct: 61 DRLAVSYLRQVNRPTMDIIEANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFINGA 120</p><p>Query: 121 YIILIKPGDSRGTVDKMRDVFPAAAPWIGHVAEDNLATRVMLQETPGLGFFFLAMQIAT 180 YIILIKPGDSRGTVDKMRDVFPAAAPWIGHVAEDNLATRVMLQETPGLGFFFLAMQIAT 180 Sbjct: 121 YIILIKPGDSRGTVDKMRDVFPAAAPWIGHVAEDNLATRVMLQETPGLGFFFLAMQIAT 180</p><p>Query: 181 DLGYTRGEPDQENDYILAGPSRKGVGVMGKATPATPAEAVKAITGFPVERLPTLPESN 240 DLGYTRGEPDQENDYILAGPSRKGVGVMGKATPATP EAV IT FPVERLPTLP*SN Sbjct: 181 DLGYTRGEPDQENDYILAGPSRKGVGVMGKATPATPKEAVDVTSTFPVERLPTLPDSN 240</p><p>Query: 241 GRASLMDVQNVFCEYSKYARLVQOGYKGTGQPTVRNGSFDLTIPEQFKK 290 GRASLMDVQNVFCEYSKYARLVQOGYKGTGQPTVRNGSFDLTIPEQFKK 290 Sbjct: 241 GRASLMDVQNVFCEYSKYARLVQOGYKGTGQPTVRNGSFDLTIPEQFKK 290</p><p>>Quartz_57, function unknown, 290 Length = 290 Score = 572 bits (1475), Expect = e-163 Identities = 283/290 (98%), Positives = 285/290 (98%)</p><p>Query: 1 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60 Sbjct: 1 MNLASVDTVVKFAKARHRLMEGKTVKSPILTQRKFTNVRVIDRGSQYLLELMNVFSDPL 60</p></div> <p>Le Start 38007 est conservé (290x3=870)</p>
DECISION:	<p>Indiquez le start retenu et justifiez brièvement.</p> <p>38007:</p> <ul style="list-style-type: none">-bon score-bonne zvalue

	- tres conservé chez homologues -chevauchement inférieur à 30pb -ORF>50pb et la plus longue
--	---

Décision #3 : Quelle est la fonction de la protéine putative ?

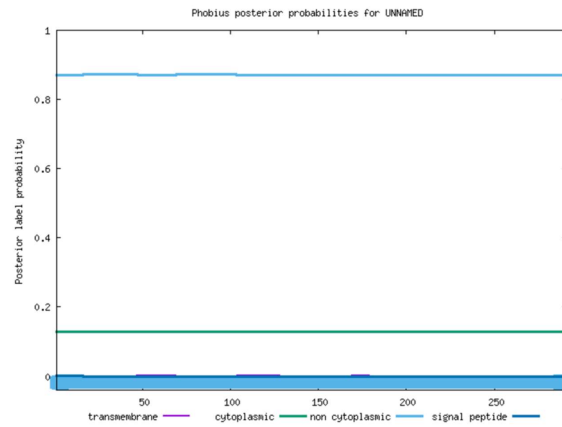
Collection des éléments de réponse	Rationnelle																																																																								
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>*** <i>PhagesDB</i> :</p> <p>(q#Mandalorian_56, function unknown, 290: s#573) : [97%] ; e-value : e-163</p> <p>>Mandalorian_56, function unknown, 290 Length = 290</p> <p>Score = 573 bits (1477), Expect = e-163 Identities = 283/290 (97%), Positives = 285/290 (98%)</p> <p>Query: 1 MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL 60 MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL Sbjct: 1 MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL 60</p> <p>Query: 61 DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA 120 DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA Sbjct: 61 DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA 120</p> <p>Query: 121 YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT 180 YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT Sbjct: 121 YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT 180</p> <p>Query: 181 DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATPAEAVKAITGFPVERLPTLPESN 240 DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATP EAV IT FPVERLPTLP+SN Sbjct: 181 DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATPKEADVITSFPVERLPTLPDSN 240</p> <p>Query: 241 GRASLMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290 GR+AS MDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK Sbjct: 241 GRHASWMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290</p> <p>*** <i>nr</i> :</p> <p>(q#hypothetical protein SEA_MANDALORIAN_56 [Microbacterium phage Mandalorian]: s#582) : [98%] ; e-value : 0.0</p> <p>hypothetical protein SEA_MANDALORIAN_56 [Microbacterium phage Mandalorian] Sequence ID: OXO13648.1 Length: 290 Number of Matches: 1</p> <p>Range 1: 1 to 290 GenPept Graphics ▼ Next Match ▲ Previous</p> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>582 bits(1501)</td><td>0.0</td><td>Compositional matrix adjust.</td><td>283/290(98%)</td><td>285/290(98%)</td><td>0/290(0%)</td></tr><tr><td>Query 1</td><td>MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL</td><td></td><td></td><td></td><td>60</td></tr><tr><td>Sbjct 1</td><td>MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL</td><td></td><td></td><td></td><td>60</td></tr><tr><td>Query 61</td><td>DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA</td><td></td><td></td><td></td><td>120</td></tr><tr><td>Sbjct 61</td><td>DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA</td><td></td><td></td><td></td><td>120</td></tr><tr><td>Query 121</td><td>YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT</td><td></td><td></td><td></td><td>180</td></tr><tr><td>Sbjct 121</td><td>YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT</td><td></td><td></td><td></td><td>180</td></tr><tr><td>Query 181</td><td>DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATPAEAVKAITGFPVERLPTLPESN</td><td></td><td></td><td></td><td>240</td></tr><tr><td>Sbjct 181</td><td>DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATP EAV IT FPVERLPTLP+SN</td><td></td><td></td><td></td><td>240</td></tr><tr><td>Query 241</td><td>GRASLMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290</td><td></td><td></td><td></td><td></td></tr><tr><td>Sbjct 241</td><td>GRHASWMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290</td><td></td><td></td><td></td><td></td></tr></table>	Score	Expect	Method	Identities	Positives	Gaps	582 bits(1501)	0.0	Compositional matrix adjust.	283/290(98%)	285/290(98%)	0/290(0%)	Query 1	MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL				60	Sbjct 1	MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL				60	Query 61	DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA				120	Sbjct 61	DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA				120	Query 121	YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT				180	Sbjct 121	YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT				180	Query 181	DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATPAEAVKAITGFPVERLPTLPESN				240	Sbjct 181	DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATP EAV IT FPVERLPTLP+SN				240	Query 241	GRASLMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290					Sbjct 241	GRHASWMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290				
Score	Expect	Method	Identities	Positives	Gaps																																																																				
582 bits(1501)	0.0	Compositional matrix adjust.	283/290(98%)	285/290(98%)	0/290(0%)																																																																				
Query 1	MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL				60																																																																				
Sbjct 1	MNLASVDTVVKFAKARHRLWEGKTVKSPILTQRKFTNVFRVIDRGSQYLLELMNVFSDPL				60																																																																				
Query 61	DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA				120																																																																				
Sbjct 61	DRLAVSYLYRQVNRPDTHADIIANDGYVPDADDIFNPKWYDKVIAPVVEARPGAFNGA				120																																																																				
Query 121	YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT				180																																																																				
Sbjct 121	YIILIKPGDSRGTVDKMRDVFPAAPWIGHVAEVDNLATRVMLLQETPGLGPFLAMQIAT				180																																																																				
Query 181	DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATPAEAVKAITGFPVERLPTLPESN				240																																																																				
Sbjct 181	DLGYTRGEPDQENDYILAGPSRKGVGVMKGATPATP EAV IT FPVERLPTLP+SN				240																																																																				
Query 241	GRASLMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290																																																																								
Sbjct 241	GRHASWMDVQNVFCEYSKYARLVQQGYKGTGQPYVRNGSFDLTIPEQFKK 290																																																																								

	Les deux blastp donnent des fonctions inconnues pour chaque alignement																																								
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<p>HHPred HMUDK_hel ; 5-hmdU DNA kinase, helical domain Probability 100% Hit PF18723.5 Database Pfam</p>  <table><tr><td><input type="checkbox"/></td><td>1</td><td>PF18723.5</td><td>; HMUDK_hel ; 5-hmdU DNA kinase, helical domain</td><td>100</td><td>3.5e-67</td><td>481.32</td><td>26.4</td><td>256</td><td>279</td></tr><tr><td><input type="checkbox"/></td><td>2</td><td>PF18724.5</td><td>; ADDT ; Amino acid:DNA transferase</td><td>96.31</td><td>0.032</td><td>52.2</td><td>7.2</td><td>92</td><td>234</td></tr><tr><td><input type="checkbox"/></td><td>3</td><td>PF18746.5</td><td>; aGPT-Ppase3 ; Alpha-glutamyl/putrescyl thymine pyrophosphorylase clade 3</td><td>95.17</td><td>0.12</td><td>50.05</td><td>6.2</td><td>94</td><td>280</td></tr><tr><td><input type="checkbox"/></td><td>4</td><td>cd00056</td><td>ENDO3c endonuclease III; includes endonuclease III (DNA/apurinic or apyrimidinic site) lyase, alkylbase DNA glycosida</td><td>65.9</td><td>70</td><td>26.79</td><td>6.8</td><td>64</td><td>158</td></tr></table>	<input type="checkbox"/>	1	PF18723.5	; HMUDK_hel ; 5-hmdU DNA kinase, helical domain	100	3.5e-67	481.32	26.4	256	279	<input type="checkbox"/>	2	PF18724.5	; ADDT ; Amino acid:DNA transferase	96.31	0.032	52.2	7.2	92	234	<input type="checkbox"/>	3	PF18746.5	; aGPT-Ppase3 ; Alpha-glutamyl/putrescyl thymine pyrophosphorylase clade 3	95.17	0.12	50.05	6.2	94	280	<input type="checkbox"/>	4	cd00056	ENDO3c endonuclease III; includes endonuclease III (DNA/apurinic or apyrimidinic site) lyase, alkylbase DNA glycosida	65.9	70	26.79	6.8	64	158
<input type="checkbox"/>	1	PF18723.5	; HMUDK_hel ; 5-hmdU DNA kinase, helical domain	100	3.5e-67	481.32	26.4	256	279																																
<input type="checkbox"/>	2	PF18724.5	; ADDT ; Amino acid:DNA transferase	96.31	0.032	52.2	7.2	92	234																																
<input type="checkbox"/>	3	PF18746.5	; aGPT-Ppase3 ; Alpha-glutamyl/putrescyl thymine pyrophosphorylase clade 3	95.17	0.12	50.05	6.2	94	280																																
<input type="checkbox"/>	4	cd00056	ENDO3c endonuclease III; includes endonuclease III (DNA/apurinic or apyrimidinic site) lyase, alkylbase DNA glycosida	65.9	70	26.79	6.8	64	158																																
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<p>Aucune synténie observée</p> 																																								
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	<p>No DeepTMHMM :</p>																																								



Phoebeius :

ID UNNAMED
FT TOPO_DOM 1 291 NON CYTOPLASMIC.
//

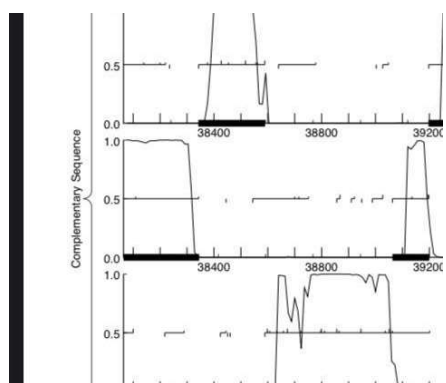


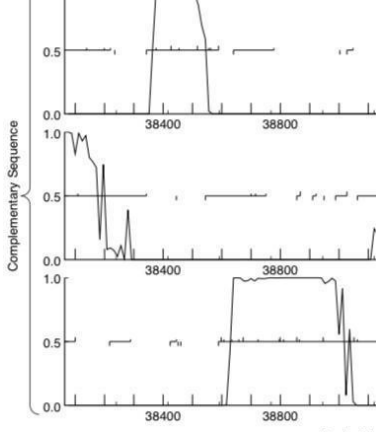
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<div><div>the following genes are part of a thymine hyper modification system: look for all parts before calling them:</div><div><div>ParB-like nuclease domain</div><div>Tet-like J-binding protein</div><div>aGPT-Pplase2 domain protein</div><div>5-hmU DNA kinase</div><div></div><div>Hypothetical Protein</div></div></div> <div><i>On n'a pas trouvé tous les gènes faisant partie de "thymine hyper modification". Donc notre gène est considéré comme une kinase.</i></div>
DECISION:	<i>kinase</i>

Student Gene Annotation Worksheet

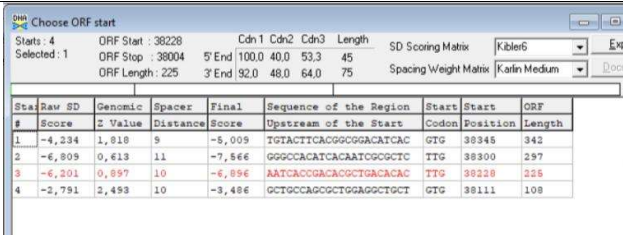
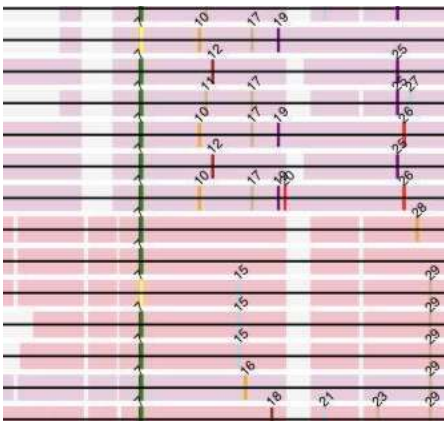
Basic Phage Information	
Nom du Phage	carostasia
Gène #	57
Coordonnées du Stop	38004
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Chev 4
Coordonnées du Start retenu	38345
Fonction prédite	NKF

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GenemarkS pouvoir codant élevé dans la région de l'ORF</i></p>  <p><i>avec la sequence complementaire Genemark folio. Pareil</i></p>

	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div><div>Click here to show details and scores. Click to show alignments</div><div>Color Key for Alignment Scores</div><div><div>0-10</div><div>10-20</div><div>20-30</div><div>30-40</div><div>40-50</div><div>50-60</div><div>60-70</div><div>70-80</div><div>80-90</div><div>90-100</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div><div>50%</div><div>40%</div><div>30%</div><div>20%</div><div>10%</div><div>0%</div></div><div><div>100%</div><div>90%</div><div>80%</div><div>70%</div><div>60%</div></div></div></div>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 38345</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :38345</i>																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table><tr><th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-4,234</td><td>1,010</td><td>9</td><td>-5,009</td><td>TGTACTTCACGGCGGACATCAC</td><td>GTG</td><td>38345</td><td>342</td></tr><tr><td>2</td><td>-6,809</td><td>0,613</td><td>11</td><td>-7,566</td><td>GGGCCACATCACAAATCGGCTC</td><td>TTG</td><td>38300</td><td>297</td></tr><tr><td>3</td><td>-6,201</td><td>0,897</td><td>10</td><td>-6,896</td><td>AATCACCGACACGCTGACACAC</td><td>TTG</td><td>38228</td><td>226</td></tr><tr><td>4</td><td>-2,791</td><td>2,493</td><td>10</td><td>-3,486</td><td>GCTGCCAGCGCTGGAGGCTGCT</td><td>GTG</td><td>38111</td><td>108</td></tr></table>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-4,234	1,010	9	-5,009	TGTACTTCACGGCGGACATCAC	GTG	38345	342	2	-6,809	0,613	11	-7,566	GGGCCACATCACAAATCGGCTC	TTG	38300	297	3	-6,201	0,897	10	-6,896	AATCACCGACACGCTGACACAC	TTG	38228	226	4	-2,791	2,493	10	-3,486	GCTGCCAGCGCTGGAGGCTGCT	GTG	38111	108
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																															
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																															
1	-4,234	1,010	9	-5,009	TGTACTTCACGGCGGACATCAC	GTG	38345	342																																															
2	-6,809	0,613	11	-7,566	GGGCCACATCACAAATCGGCTC	TTG	38300	297																																															
3	-6,201	0,897	10	-6,896	AATCACCGACACGCTGACACAC	TTG	38228	226																																															
4	-2,791	2,493	10	-3,486	GCTGCCAGCGCTGGAGGCTGCT	GTG	38111	108																																															
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<i>Le start 38345 conduit à une ORF de 342 pb, la plus longue chevauchement -4</i> <i>Les starts suivants conduisent à des ORFs beaucoup plus courtes</i>																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?	 <p><i>Start le plus conservé est 38345</i></p>																																																						

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<pre> Score = 426 bits (2/2), Expect = 2e-29 Identities = 113/113 (100%), Positives = 113/113 (100%) uery: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 bjct: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 uery: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 bjct: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 Mandalorian_57, function unknown, 113 Length = 113 Score = 224 bits (571), Expect = 6e-59 Identities = 112/113 (99%), Positives = 113/113 (100%) uery: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 bjct: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 uery: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 bjct: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 Quartz_58, function unknown, 113 Length = 113 Score = 223 bits (567), Expect = 2e-58 Identities = 111/113 (98%), Positives = 112/113 (99%) uery: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 bjct: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTS S 60 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTS 60 uery: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 bjct: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 Nucci_57, function unknown, 113 Length = 113 Score = 223 bits (567), Expect = 2e-58 Identities = 111/113 (98%), Positives = 112/113 (99%) uery: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTSAS 60 bjct: 1 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTS S 60 MNASRRQTGHITIALLECANSGTIGDEAHLAITDTLTHLQREEFHOLRAQEKYEHTS 60 uery: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 bjct: 61 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 NATQVAICRVLPALAAVDANNKDDFDTVIEQLRVAITADGKAKPKKGRKKR 113 Le Start 38345 est conservé (113x3=339) </pre>
DECISION:	<p>38345:</p> <ul style="list-style-type: none"> -bon score -bonne zvalue - tres conservé chez homologues -chevauchement inférieur à 30pb -ORF>50pb et la plus longue

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>(q#: s#) : [alignment] ; e-value :</p>

une e-value < 10⁻⁴ et une couverture acceptable ?

Sequences producing significant alignments:		Score (bits)	E Value					
Carostasia_Draft_57, function unknown, 113		226	2e-59					
Mandalorian_57, function unknown, 113		224	6e-59					
Quartz_58, function unknown, 113		223	2e-58					
Nucci_57, function unknown, 113		223	2e-58					
Juanyo_59, function unknown, 114		210	1e-54					
*** nr :								
(q#: s#) : [alignment] ; e-value :								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein SEFA_MANDALORIAN_57 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	231	231	99%	2e-76	99.12%	113	QX013649.1
<input checked="" type="checkbox"/> hypothetical protein SEFA_NUCCI_57 [Microbacterium phage Nucci]	Microbacterium phage Nucci	230	230	99%	1e-75	99.23%	113	QX024550.1
<input checked="" type="checkbox"/> hypothetical protein SEFA_JUANYO_59 [Microbacterium phage Juanyo]	Microbacterium phage Juanyo	217	217	100%	2e-70	99.35%	114	U5144808.1

Fonction inconnue chaque fois

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

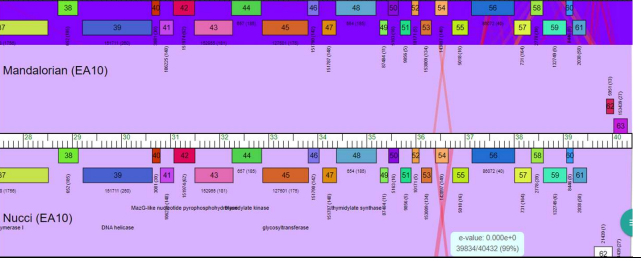
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/> 1	3ZDM_A	SMALL GLUTAMINE-RICH TETRATRICOPEPTIDE REPEAT-CONTAINING PROTEIN 2; CHAPERONE-SIGNALING PROTEIN COMPLEX; 1.803A (SACCHA	82.27	22	23.66	6.1	43	72
<input type="checkbox"/> 2	PF07023.16	: DUF1315; Protein of unknown function (DUF1315)	80.65	14	27.74	5.2	44	82
<input type="checkbox"/> 3	8OYY_A	De novo designed soluble GPCR-like protein; GPCR, solubilized, de novo designed, DE NOVO PROTEIN; 1.85A (synthetic const	76.88	95	27.46	11.3	86	242
<input type="checkbox"/> 4	PF07219.17	: HemY_N; HemY protein N-terminus	75.75	36	23.87	5.8	43	107



NON

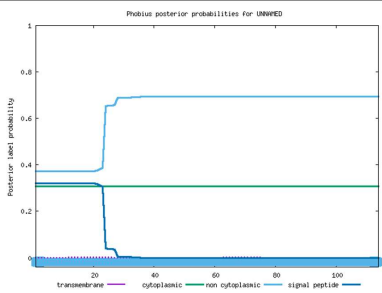
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

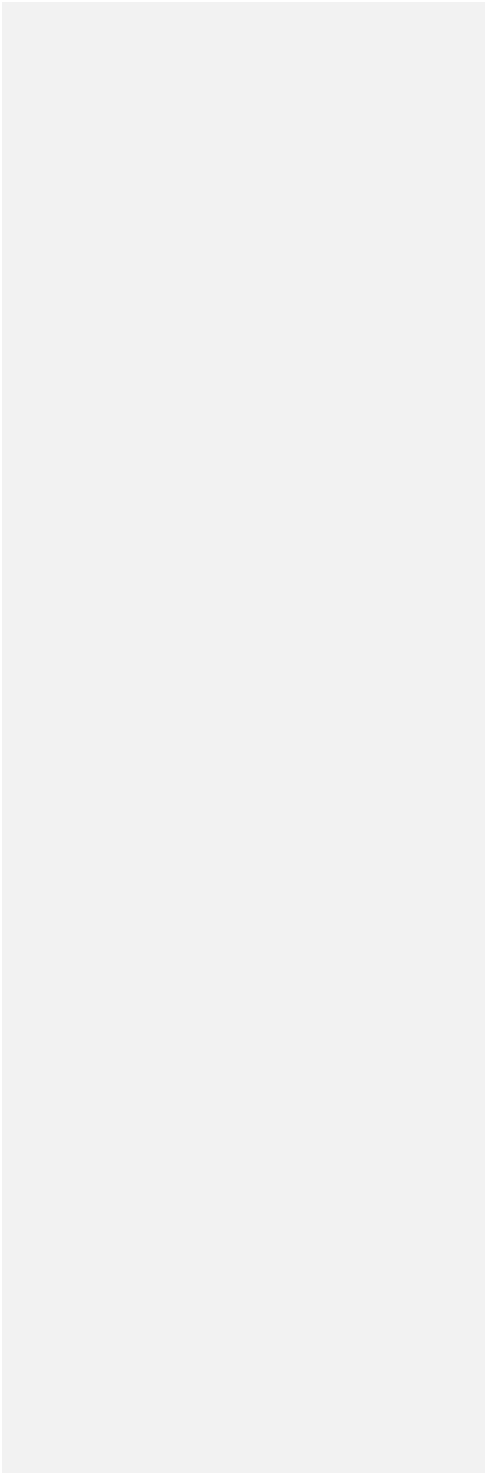
Fonction inconnue.



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

Non.

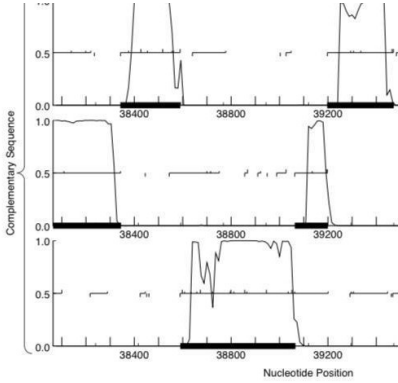
	
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Fonction inconnue.</i>
DECISION:	<i>NKF</i>

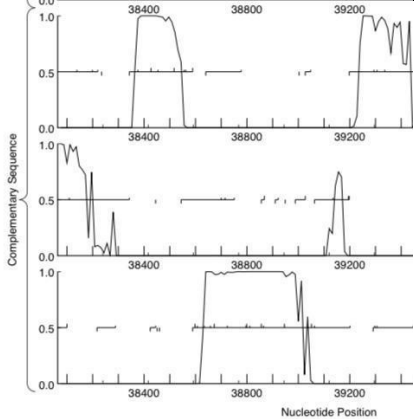
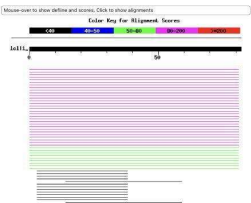


Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	58
Coordonnées du Stop	38342
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	ove 1
Coordonnées du Start retenu	38590
Fonction prédite	NFk

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>BOTH YES</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GenemarkS pouvoir codant élevé dans la région de l'ORF</i></p>  <p><i>Genemark folio. Pareil</i></p>

																									
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Distribution of 67 Blast Hits on the Query Sequence</p> <p>Click to show details and scores. Click to show alignments.</p> <p>Color Key for Blast Scores</p>  <p>Sequences producing significant alignments:</p> <table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Nucci_58, function unknown, 82</td><td>167</td><td>8e-42</td></tr><tr><td>Carostasia_Draft_58, function unknown, 82</td><td>167</td><td>8e-42</td></tr><tr><td>Mandalorian_58, function unknown, 82</td><td>166</td><td>2e-41</td></tr><tr><td>Quartz_59, function unknown, 82</td><td>163</td><td>1e-40</td></tr><tr><td>YuuV_59, function unknown, 82</td><td>139</td><td>2e-33</td></tr><tr><td>Juanyo_60, function unknown, 82</td><td>134</td><td>8e-32</td></tr><tr><td>Phlebotomus_58, function unknown, 82</td><td>124</td><td>1e-38</td></tr></tbody></table> <p>Non annotés mais beaucoup d' homologues</p>		Score (bits)	E Value	Nucci_58, function unknown, 82	167	8e-42	Carostasia_Draft_58, function unknown, 82	167	8e-42	Mandalorian_58, function unknown, 82	166	2e-41	Quartz_59, function unknown, 82	163	1e-40	YuuV_59, function unknown, 82	139	2e-33	Juanyo_60, function unknown, 82	134	8e-32	Phlebotomus_58, function unknown, 82	124	1e-38
	Score (bits)	E Value																							
Nucci_58, function unknown, 82	167	8e-42																							
Carostasia_Draft_58, function unknown, 82	167	8e-42																							
Mandalorian_58, function unknown, 82	166	2e-41																							
Quartz_59, function unknown, 82	163	1e-40																							
YuuV_59, function unknown, 82	139	2e-33																							
Juanyo_60, function unknown, 82	134	8e-32																							
Phlebotomus_58, function unknown, 82	124	1e-38																							
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION :</p> <ul style="list-style-type: none">-plus de 50pb-chevauchement inférieur à 30pb-direction similaire aux gènes qui l'entourent																								
DECISION:	OUI																								

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																	
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 38590</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 38590</i>																																																																																	
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Choose ORF start</div><div>Stats: 7 ORF Start : 38518 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kåbler6 Explor Selected: 1 ORF Stop : 38342 5' End 77.8 44.4 44.4 27 ORF Length: 177 3' End 87.5 50.0 81.3 48 Spacing Weight Matrix Karlin Medium Document</div><table><tr><th>Sta</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-2,377</td><td>2,687</td><td>13</td><td>-3,422</td><td>ACGTGGGAAGGATGCTCAACTG</td><td>ATG</td><td>38590</td><td>249</td></tr><tr><td>2</td><td>-3,821</td><td>2,011</td><td>10</td><td>-4,615</td><td>TCAGGCATACAGGTCTCTTTC</td><td>GTG</td><td>38563</td><td>232</td></tr><tr><td>3</td><td>-3,821</td><td>2,011</td><td>16</td><td>-5,617</td><td>ATACCAAGTCTCTCTGTGAA</td><td>GTG</td><td>38557</td><td>216</td></tr><tr><td>4</td><td>-5,244</td><td>1,299</td><td>12</td><td>-6,179</td><td>CGACCCGCGCGCTGTCCAG</td><td>ATG</td><td>38518</td><td>177</td></tr><tr><td>5</td><td>-4,654</td><td>1,622</td><td>11</td><td>-5,410</td><td>ACAGGAGCCAGGCTGACCAGG</td><td>GTG</td><td>38455</td><td>114</td></tr><tr><td>6</td><td>-3,264</td><td>2,272</td><td>19</td><td>-4,310</td><td>CCGTGACCCGGAACCCGCGTC</td><td>ATG</td><td>38429</td><td>87</td></tr><tr><td>7</td><td>-5,167</td><td>1,381</td><td>8</td><td>-6,389</td><td>CCGCGAGATCGACCCGAACACC</td><td>GTG</td><td>38377</td><td>96</td></tr></table></div>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-2,377	2,687	13	-3,422	ACGTGGGAAGGATGCTCAACTG	ATG	38590	249	2	-3,821	2,011	10	-4,615	TCAGGCATACAGGTCTCTTTC	GTG	38563	232	3	-3,821	2,011	16	-5,617	ATACCAAGTCTCTCTGTGAA	GTG	38557	216	4	-5,244	1,299	12	-6,179	CGACCCGCGCGCTGTCCAG	ATG	38518	177	5	-4,654	1,622	11	-5,410	ACAGGAGCCAGGCTGACCAGG	GTG	38455	114	6	-3,264	2,272	19	-4,310	CCGTGACCCGGAACCCGCGTC	ATG	38429	87	7	-5,167	1,381	8	-6,389	CCGCGAGATCGACCCGAACACC	GTG	38377	96
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																										
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																										
1	-2,377	2,687	13	-3,422	ACGTGGGAAGGATGCTCAACTG	ATG	38590	249																																																																										
2	-3,821	2,011	10	-4,615	TCAGGCATACAGGTCTCTTTC	GTG	38563	232																																																																										
3	-3,821	2,011	16	-5,617	ATACCAAGTCTCTCTGTGAA	GTG	38557	216																																																																										
4	-5,244	1,299	12	-6,179	CGACCCGCGCGCTGTCCAG	ATG	38518	177																																																																										
5	-4,654	1,622	11	-5,410	ACAGGAGCCAGGCTGACCAGG	GTG	38455	114																																																																										
6	-3,264	2,272	19	-4,310	CCGTGACCCGGAACCCGCGTC	ATG	38429	87																																																																										
7	-5,167	1,381	8	-6,389	CCGCGAGATCGACCCGAACACC	GTG	38377	96																																																																										
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<i>Le start 38590 conduit à une ORF de 249 pb, la plus longue chevauchement -1</i> <i>Les starts suivant conduisent à des ORFs plus courtes et avec de moins bons scores</i>																																																																																	
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																																																		

Décision #3 : Quelle est la fonction de la protéine putative ?

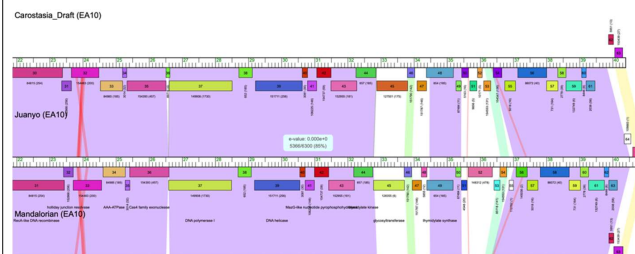
Collection des éléments de réponse	Rationnelle												
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>*** <i>PhagesDB</i> :</p> <p>(q#Nucci_58, function unknown, 82: s#167) : [100%] ; e-value : 8e-42</p> <p>>Nucci_58, function unknown, 82 Length = 82</p> <p>Score = 167 bits (423), Expect = 8e-42 Identities = 82/82 (100%), Positives = 82/82 (100%)</p> <p>Query: 1 MPQAYQVVFVEVGLYNHDPRGAVQMAQQSADQKFEALAQEARLTRVTRDPEPRVMTYPEV 60 MPQAYQVVFVEVGLYNHDPRGAVQMAQQSADQKFEALAQEARLTRVTRDPEPRVMTYPEV Sbjct: 1 MPQAYQVVFVEVGLYNHDPRGAVQMAQQSADQKFEALAQEARLTRVTRDPEPRVMTYPEV 60</p> <p>Query: 61 ANAFPQIDPNTVVGLYFTADIT 82 ANAFPQIDPNTVVGLYFTADIT Sbjct: 61 ANAFPQIDPNTVVGLYFTADIT 82</p> <p>*** <i>nr</i> :</p> <p>(q#hypothetical protein SEA_NUCCI_58 [Microbacterium phage Nucci]: s#168) : [100%] ; e-value : 3e-52</p> <p>hypothetical protein SEA_NUCCI_58 [Microbacterium phage Nucci] Sequence ID: QDK02451.1 Length: 82 Number of Matches: 1</p> <p>Range 1: 1 to 82 GenPept Graphics ▼ Next Match ▲ Prev</p> <table><tr><th>Score</th><th>Expect</th><th>Method</th><th>Identities</th><th>Positives</th><th>Gaps</th></tr><tr><td>168 bits(425)</td><td>3e-52</td><td>Compositional matrix adjust.</td><td>82/82(100%)</td><td>82/82(100%)</td><td>0/82(0%)</td></tr></table> <p>Query 1 MPQAYQVVFVEVGLYNHDPRGAVQMAQQSADQKFEALAQEARLTRVTRDPEPRVMTYPEV 60 MPQAYQVVFVEVGLYNHDPRGAVQMAQQSADQKFEALAQEARLTRVTRDPEPRVMTYPEV Sbjct 1 MPQAYQVVFVEVGLYNHDPRGAVQMAQQSADQKFEALAQEARLTRVTRDPEPRVMTYPEV 60</p> <p>Query 61 ANAFPQIDPNTVVGLYFTADIT 82 ANAFPQIDPNTVVGLYFTADIT Sbjct 61 ANAFPQIDPNTVVGLYFTADIT 82</p> <p>Tous les alignements donnent des fonctions inconnues</p>	Score	Expect	Method	Identities	Positives	Gaps	168 bits(425)	3e-52	Compositional matrix adjust.	82/82(100%)	82/82(100%)	0/82(0%)
Score	Expect	Method	Identities	Positives	Gaps								
168 bits(425)	3e-52	Compositional matrix adjust.	82/82(100%)	82/82(100%)	0/82(0%)								
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<p>HHPred : Mt_ATP-synt_D ; ATP synthase D chain, mitochondrial (ATP5H) Probability 70,49 Hit PF05873.16 Data base Pfam</p> 												

<input type="checkbox"/>	1	PF05873.16	MT_ATP-synt_D , ATP synthase D chain, mitochondrial (ATP5F3)	70.49	32	24.64	4.7	51	153
<input type="checkbox"/>	2	cd22344	PDDEXX_nuclease; uncharacterized PDDEXX nuclease may function as a restriction endonuclease. This family belongs to a su	69.66	13	28.75	2.9	38	225
<input type="checkbox"/>	3	2RAW_B	Borealin; Cell cycle, DasraB, chromosomal passender complex, IAP, BIR, Apoptosis, Cell division, Centromere, Chromosomal	62.65	33	20.25	3	57	68
<input type="checkbox"/>	4	2CLY_B	ATP SYNTHASE D CHAIN, MITOCHONDRIAL; MITOCHONDRIA, MITOCHONDRION, ION TRANSPORT, CF0, STATOR, TRANSPORT, ACETYLATION,	57.41	81	22.87	4.7	56	160

Pas de correspondance pertinente car la probabilité $< 90\%$

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

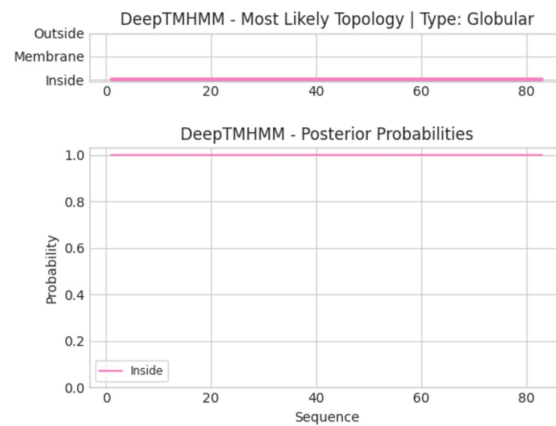
Aucune synténie observée



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

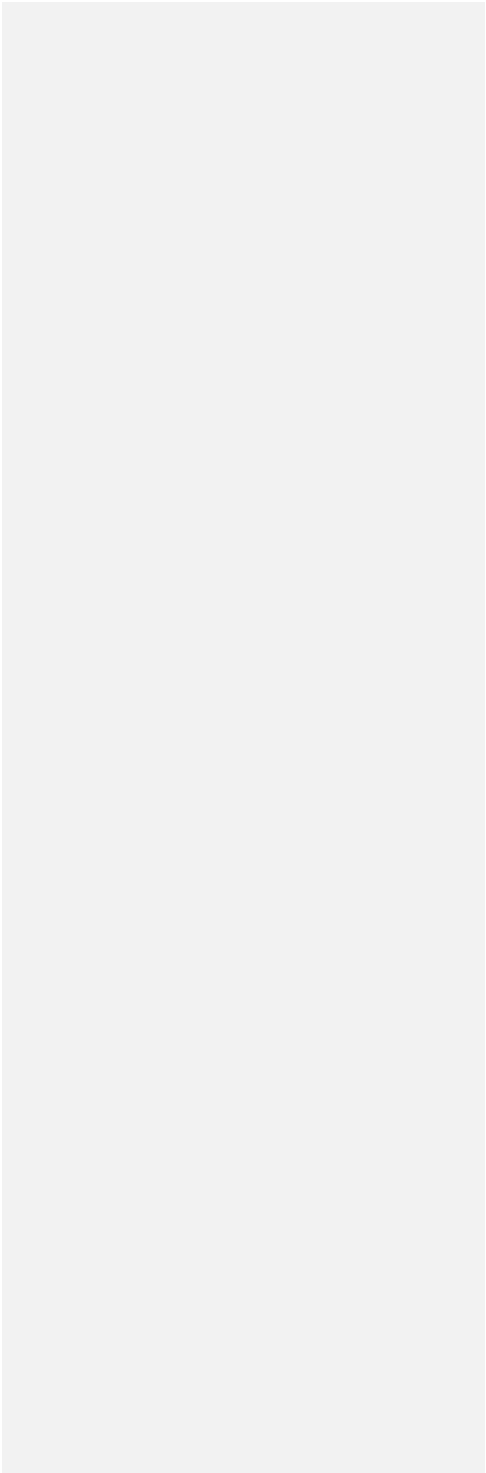
No

DeepTMHMM :



Phoebius :

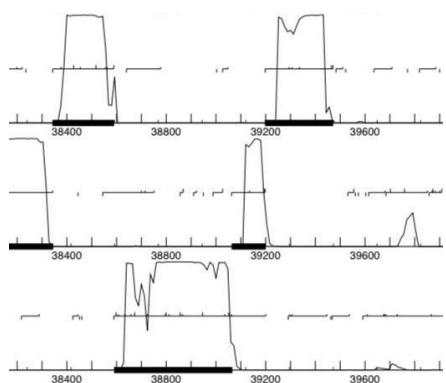
	<div><div><div>ID UNNAMED</div><div>FT TOPO_DOM</div><div>//</div></div><div><div>1</div><div>83</div><div>NON CYTOPLASMIC.</div></div></div> <div><p>Phobius posterior probabilities for UNNAMED</p><table><caption>Phobius Posterior Probabilities Data</caption><tr><th>Residue Range</th><th>Label</th><th>Posterior Probability</th></tr><tr><td>0 - 80</td><td>signal peptide</td><td>~0.7</td></tr><tr><td>0 - 80</td><td>cytoplasmic</td><td>~0.3</td></tr><tr><td>0 - 80</td><td>transmembrane</td><td>~0.05</td></tr><tr><td>0 - 80</td><td>non cytoplasmic</td><td>~0.05</td></tr></table></div>	Residue Range	Label	Posterior Probability	0 - 80	signal peptide	~0.7	0 - 80	cytoplasmic	~0.3	0 - 80	transmembrane	~0.05	0 - 80	non cytoplasmic	~0.05
Residue Range	Label	Posterior Probability														
0 - 80	signal peptide	~0.7														
0 - 80	cytoplasmic	~0.3														
0 - 80	transmembrane	~0.05														
0 - 80	non cytoplasmic	~0.05														
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui</i>															
DECISION:	<i>NKF</i>															

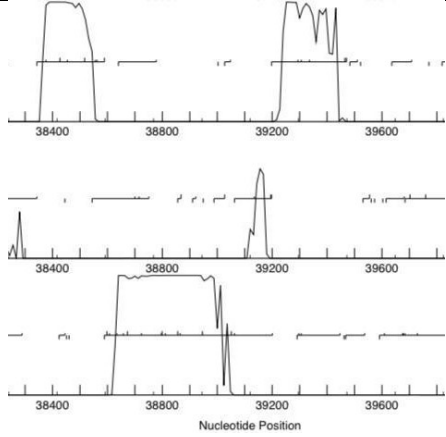
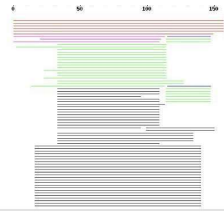


Student Gene Annotation Worksheet

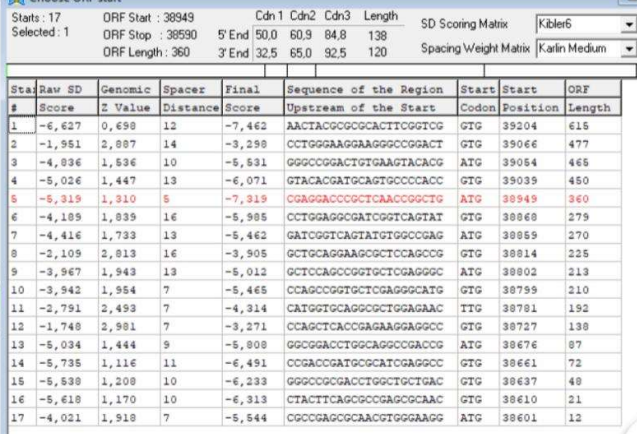
Basic Phage Information	
Nom du Phage	carostasia
Gène #	59
Coordonnées du Stop	38590
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Chev 4
Coordonnées du Start retenu	39066
Fonction prédite	NKF


Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	<i>BOTH YES</i>
Y-a-t-il des éléments supportant un potentiel codant ?	<p><i>GenemarkS pouvoir codant élevé dans la region de l'ORF</i></p>  <p><i>avec la sequence complementaire Genemark folio. Pareil</i></p>

																															
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 <p>Sequences producing significant alignments:</p> <table><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Carostasia_Draft_59, function unknown, 158</td><td>322</td><td>2e-88</td></tr><tr><td>Nucci_59, function unknown, 158</td><td>318</td><td>3e-87</td></tr><tr><td>Mandalorian_59, function unknown, 158</td><td>317</td><td>9e-87</td></tr><tr><td>Quartz_60, function unknown, 158</td><td>315</td><td>3e-86</td></tr><tr><td>Juanyo_61, function unknown, 158</td><td>294</td><td>7e-80</td></tr><tr><td>YuuY_60, function unknown, 160</td><td>268</td><td>5e-72</td></tr><tr><td>Ixel_58, function unknown, 164</td><td>85</td><td>8e-17</td></tr><tr><td>Alekazan_57, function unknown, 164</td><td>82</td><td>4e-16</td></tr><tr><td>Neferthena_60, function unknown, 157</td><td>81</td><td>9e-16</td></tr></tbody></table> <p>Non annotés mais beaucoup d' homologues de tailles différentes</p>		Score (bits)	E Value	Carostasia_Draft_59, function unknown, 158	322	2e-88	Nucci_59, function unknown, 158	318	3e-87	Mandalorian_59, function unknown, 158	317	9e-87	Quartz_60, function unknown, 158	315	3e-86	Juanyo_61, function unknown, 158	294	7e-80	YuuY_60, function unknown, 160	268	5e-72	Ixel_58, function unknown, 164	85	8e-17	Alekazan_57, function unknown, 164	82	4e-16	Neferthena_60, function unknown, 157	81	9e-16
	Score (bits)	E Value																													
Carostasia_Draft_59, function unknown, 158	322	2e-88																													
Nucci_59, function unknown, 158	318	3e-87																													
Mandalorian_59, function unknown, 158	317	9e-87																													
Quartz_60, function unknown, 158	315	3e-86																													
Juanyo_61, function unknown, 158	294	7e-80																													
YuuY_60, function unknown, 160	268	5e-72																													
Ixel_58, function unknown, 164	85	8e-17																													
Alekazan_57, function unknown, 164	82	4e-16																													
Neferthena_60, function unknown, 157	81	9e-16																													
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION :</p> <ul style="list-style-type: none">-plus de 50pb-chevauchement inférieur à 30pb-direction similaire aux gènes qui l'entourent																														
DECISION:	OUI																														

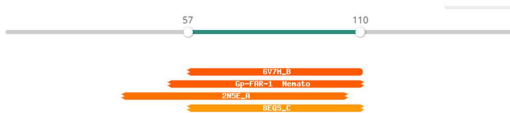
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 39066 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 39066
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start 39204 conduit à une ORF de 615pb, la plus longue chevauchement -142 Le start 39066 conduit à une ORF de 477pb chevauchement -4 Le start 39054 conduit à une ORF de 465pb avec un GAP de 8pb et un score moins bon Les starts suivants conduisent à des ORFs plus courtes avec des scores moins bons
Est-ce que le start est conservé chez les homologues voir Starterator ?	

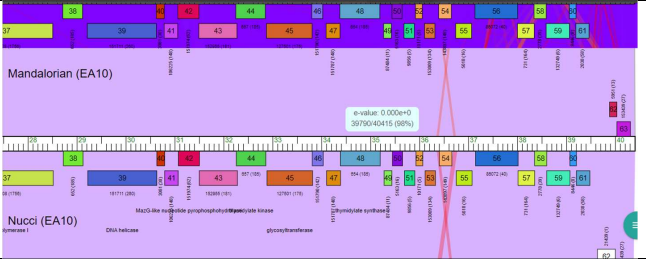
	 <p><i>Start le plus conservé est 39066</i></p>
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<pre>- Sbjct: 1 MKYTMQCPTVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL MKYTMQCPTVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60 Query: 61 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER 120 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER Sbjct: 61 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER 120 Query: 121 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN 158 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN Sbjct: 121 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN 158 >Nucci_59, function unknown, 158 Length = 158 Score = 318 bits (815), Expect = 3e-87 Identities = 156/158 (98%), Positives = 156/158 (98%) Query: 1 MKYTMQCPTVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60 MKYTMQCP VQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL Sbjct: 1 MKYTMQCPVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60 Query: 61 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER 120 EAIGQVVAEMIQQFI ALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER Sbjct: 61 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER 120 Query: 121 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN 158 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN Sbjct: 121 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN 158 >Mandalorian_59, function unknown, 158 Length = 158 Score = 317 bits (811), Expect = 9e-87 Identities = 155/158 (98%), Positives = 155/158 (98%) Query: 1 MKYTMQCPTVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60 MKYTMQCP VQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL Sbjct: 1 MKYTMQCPVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60 Query: 61 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER 120 EAIGQVVAEMIQQFI ALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEAVPPRAER Sbjct: 61 EAIGQVVAEMIQQFITALQ EALQPVLEGMVQALENLNHSIPEEVRAQLTEKEVPPRAER 120 Query: 121 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN 158 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN Sbjct: 121 QRAADLAGRPMRIEAVGRDLAADVAYFSAERNVGRMVN 158 >Quartz_60, function unknown, 158 Length = 158 Score = 315 bits (807), Expect = 3e-86 Identities = 154/158 (97%), Positives = 154/158 (97%) Query: 1 MKYTMQCPTVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60 MKYTMQCP VQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL Sbjct: 1 MKYTMQCPVQKHEGHTWQEP EEGELTC LGHTEDPLNRLMGVADTLHG GGLTPKEQADL 60</pre> <p><i>Le Start 39066 est conservé (158x3=474)</i></p>
DECISION:	<p>39066: -meilleur score</p>

	<ul style="list-style-type: none">-meilleure zvalue- tres conservé chez homologues-chevauchement inférieur à 30pb contrairement au start 1-ORF>50pb et 2e plus longue
--	---

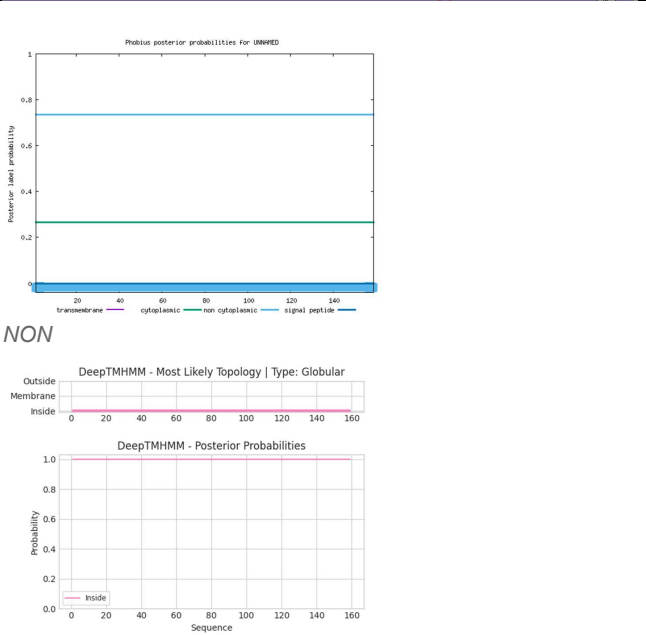
Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle																																																																																																			
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p> <table><tr><th>Sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr><tr><td>Carostasia_Draft_59, function unknown, 158</td><td>322</td><td>2e-88</td></tr><tr><td>Nucci_59, function unknown, 158</td><td>318</td><td>3e-87</td></tr><tr><td>Mandalorian_59, function unknown, 158</td><td>317</td><td>1e-86</td></tr><tr><td>Quartz_60, function unknown, 158</td><td>315</td><td>3e-86</td></tr><tr><td>Juanyo_61, function unknown, 158</td><td>294</td><td>7e-80</td></tr></table> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p> <table><tr><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per Ident</th><th>Acc Len</th><th>Accession</th></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_NUGCL_89 (Microbacterium phages.Nucci)</td><td>Microbacterium phages.Nucci</td><td>322</td><td>322</td><td>99%</td><td>1e-110</td><td>98.77%</td><td>158</td><td>GD030552.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_NUGCL_89 (Microbacterium phages.Nucci)</td><td>Microbacterium phages.Nucci</td><td>320</td><td>320</td><td>99%</td><td>5e-110</td><td>98.10%</td><td>158</td><td>GD030552.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_G04872_49 (Microbacterium phages.Quartz)</td><td>Microbacterium phages.Quartz</td><td>318</td><td>318</td><td>99%</td><td>2e-109</td><td>87.47%</td><td>158</td><td>UN058029.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_G04872_49 (Microbacterium phages.Quartz)</td><td>Microbacterium phages.Quartz</td><td>290</td><td>290</td><td>99%</td><td>3e-101</td><td>89.87%</td><td>158</td><td>UN058029.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)</td><td>Microbacterium phages.Juanyo</td><td>270</td><td>270</td><td>92%</td><td>6e-99</td><td>87.87%</td><td>158</td><td>GF295388.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)</td><td>Microbacterium phages.Juanyo</td><td>270</td><td>270</td><td>92%</td><td>6e-99</td><td>87.87%</td><td>158</td><td>GF295388.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)</td><td>Microbacterium phages.Juanyo</td><td>82.8</td><td>82.8</td><td>71%</td><td>2e-16</td><td>38.50%</td><td>154</td><td>GD080758.1</td></tr><tr><td><input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)</td><td>Microbacterium phages.Juanyo</td><td>81.3</td><td>81.3</td><td>69%</td><td>1e-16</td><td>36.26%</td><td>157</td><td>GF295388.1</td></tr></table> <p>Fonction inconnue chaque fois</p> <p>Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.</p>	Sequences producing significant alignments:	Score (bits)	E Value	Carostasia_Draft_59, function unknown, 158	322	2e-88	Nucci_59, function unknown, 158	318	3e-87	Mandalorian_59, function unknown, 158	317	1e-86	Quartz_60, function unknown, 158	315	3e-86	Juanyo_61, function unknown, 158	294	7e-80	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession	<input checked="" type="checkbox"/> Uncharacterized protein SEA_NUGCL_89 (Microbacterium phages.Nucci)	Microbacterium phages.Nucci	322	322	99%	1e-110	98.77%	158	GD030552.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_NUGCL_89 (Microbacterium phages.Nucci)	Microbacterium phages.Nucci	320	320	99%	5e-110	98.10%	158	GD030552.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_G04872_49 (Microbacterium phages.Quartz)	Microbacterium phages.Quartz	318	318	99%	2e-109	87.47%	158	UN058029.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_G04872_49 (Microbacterium phages.Quartz)	Microbacterium phages.Quartz	290	290	99%	3e-101	89.87%	158	UN058029.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	270	270	92%	6e-99	87.87%	158	GF295388.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	270	270	92%	6e-99	87.87%	158	GF295388.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	82.8	82.8	71%	2e-16	38.50%	154	GD080758.1	<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	81.3	81.3	69%	1e-16	36.26%	157	GF295388.1
Sequences producing significant alignments:	Score (bits)	E Value																																																																																																		
Carostasia_Draft_59, function unknown, 158	322	2e-88																																																																																																		
Nucci_59, function unknown, 158	318	3e-87																																																																																																		
Mandalorian_59, function unknown, 158	317	1e-86																																																																																																		
Quartz_60, function unknown, 158	315	3e-86																																																																																																		
Juanyo_61, function unknown, 158	294	7e-80																																																																																																		
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_NUGCL_89 (Microbacterium phages.Nucci)	Microbacterium phages.Nucci	322	322	99%	1e-110	98.77%	158	GD030552.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_NUGCL_89 (Microbacterium phages.Nucci)	Microbacterium phages.Nucci	320	320	99%	5e-110	98.10%	158	GD030552.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_G04872_49 (Microbacterium phages.Quartz)	Microbacterium phages.Quartz	318	318	99%	2e-109	87.47%	158	UN058029.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_G04872_49 (Microbacterium phages.Quartz)	Microbacterium phages.Quartz	290	290	99%	3e-101	89.87%	158	UN058029.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	270	270	92%	6e-99	87.87%	158	GF295388.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	270	270	92%	6e-99	87.87%	158	GF295388.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	82.8	82.8	71%	2e-16	38.50%	154	GD080758.1																																																																																												
<input checked="" type="checkbox"/> Uncharacterized protein SEA_V0147_61 (Microbacterium phages.Juanyo)	Microbacterium phages.Juanyo	81.3	81.3	69%	1e-16	36.26%	157	GF295388.1																																																																																												
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un cribre HHPred avec une proba >= 90% et une couverture acceptable ?	<p>OUI</p> <table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td><input type="checkbox"/> 1</td><td>6V7M_B</td><td>Apolipoprotein E: proteolysis, Alzheimers, lipid, lipoprotein, lipids, disease, LIPID TRANSPORT; 2.0A (Homo sapiens)</td><td>92.5</td><td>2.8</td><td>28.52</td><td>7</td><td>54</td><td>83</td></tr><tr><td><input type="checkbox"/> 2</td><td>PFO5823.16</td><td>: Gp-FAR-1; Nematode fatty acid retinoid binding protein (Gp-FAR-1)</td><td>92.24</td><td>3.4</td><td>31.2</td><td>7.8</td><td>60</td><td>149</td></tr><tr><td><input type="checkbox"/> 3</td><td>2N5E_A</td><td>Apolipoprotein A-1; nanodisc, HDL, lipoproteins, cardiovascular disease, LIPID BINDING PROTEIN; NMR (Homo sapiens)</td><td>90.38</td><td>4.7</td><td>30.96</td><td>7.1</td><td>66</td><td>167</td></tr></table> 	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/> 1	6V7M_B	Apolipoprotein E: proteolysis, Alzheimers, lipid, lipoprotein, lipids, disease, LIPID TRANSPORT; 2.0A (Homo sapiens)	92.5	2.8	28.52	7	54	83	<input type="checkbox"/> 2	PFO5823.16	: Gp-FAR-1; Nematode fatty acid retinoid binding protein (Gp-FAR-1)	92.24	3.4	31.2	7.8	60	149	<input type="checkbox"/> 3	2N5E_A	Apolipoprotein A-1; nanodisc, HDL, lipoproteins, cardiovascular disease, LIPID BINDING PROTEIN; NMR (Homo sapiens)	90.38	4.7	30.96	7.1	66	167																																																															
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																																																												
<input type="checkbox"/> 1	6V7M_B	Apolipoprotein E: proteolysis, Alzheimers, lipid, lipoprotein, lipids, disease, LIPID TRANSPORT; 2.0A (Homo sapiens)	92.5	2.8	28.52	7	54	83																																																																																												
<input type="checkbox"/> 2	PFO5823.16	: Gp-FAR-1; Nematode fatty acid retinoid binding protein (Gp-FAR-1)	92.24	3.4	31.2	7.8	60	149																																																																																												
<input type="checkbox"/> 3	2N5E_A	Apolipoprotein A-1; nanodisc, HDL, lipoproteins, cardiovascular disease, LIPID BINDING PROTEIN; NMR (Homo sapiens)	90.38	4.7	30.96	7.1	66	167																																																																																												
Ce gène est-il situé à côté de gènes de fonction connue et dans une région	Fonction inconnue.																																																																																																			

du génome qui montre une forte conservation de l'ordre des gènes ?



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue

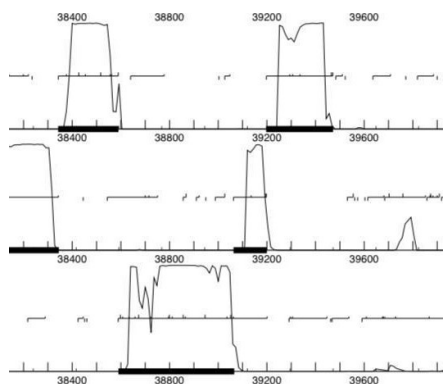
DECISION:

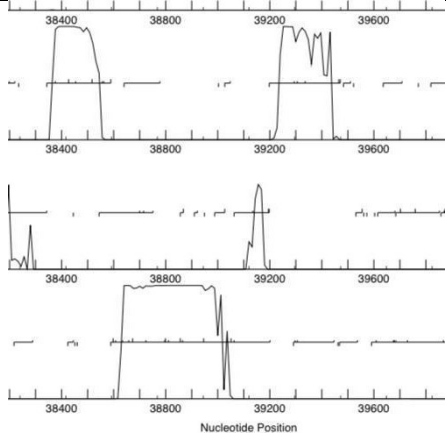

NKF

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	60
Coordonnées du Stop	39063
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	ove 4
Coordonnées du Start retenu	39200
Fonction prédite	NFk

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenemarkS : pouvoir codant élevé dans la region de l'ORF</p>  <p>avec la sequence complementaire Genemark folio. Pareil</p>

																													
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p><u>Distribution of 8 Blast Hits on the Query Sequence</u></p> <p>Mouse-over to show details and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <table><tr><td>0-10</td><td>10-20</td><td>20-30</td><td>30-40</td><td>40-50</td><td>50-60</td><td>60-70</td><td>70-80</td><td>80-90</td><td>90-100</td></tr></table>  <p>Producing significant alignments:</p> <table><tr><th>Accession</th><th>Score (bits)</th></tr><tr><td>raft_60, function unknown</td><td>45</td></tr><tr><td>unction unknown</td><td>44</td></tr><tr><td>unction unknown</td><td>44</td></tr><tr><td>60, function unknown</td><td>44</td></tr><tr><td>unction unknown</td><td>45</td></tr><tr><td>unction unknown</td><td>44</td></tr><tr><td>56, function unknown</td><td>44</td></tr><tr><td>57, function unknown</td><td>44</td></tr></table> <p>Non annotés mais quelques d' homologues</p>	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	Accession	Score (bits)	raft_60, function unknown	45	unction unknown	44	unction unknown	44	60, function unknown	44	unction unknown	45	unction unknown	44	56, function unknown	44	57, function unknown	44
0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100																				
Accession	Score (bits)																												
raft_60, function unknown	45																												
unction unknown	44																												
unction unknown	44																												
60, function unknown	44																												
unction unknown	45																												
unction unknown	44																												
56, function unknown	44																												
57, function unknown	44																												
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION :</p> <ul style="list-style-type: none">-plus de 50pb-chevauchement inférieur à 30pb-direction similaire aux gènes qui l'entourent																												
DECISION:	OUI																												

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 39200 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 39200																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div>  <table> <tr> <th>Start</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr> <tr> <th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr> <tr> <td>1</td><td>-6,188</td><td>0,904</td><td>9</td><td>-6,962</td><td>ACGCGCGCACTTCGGTGGTGC</td><td>ATG</td><td>39200</td><td>138</td></tr> <tr> <td>2</td><td>-6,491</td><td>1,136</td><td>8</td><td>-6,913</td><td>CGCGCACTTCGGTGGTGCATG</td><td>ATG</td><td>39197</td><td>135</td></tr> <tr> <td>3</td><td>-6,761</td><td>1,103</td><td>11</td><td>-6,818</td><td>CTGGGATCACTGCGACGTCAGC</td><td>TTC</td><td>39164</td><td>102</td></tr> <tr> <td>4</td><td>-6,161</td><td>1,384</td><td>9</td><td>-6,935</td><td>GGTCGACCCGAGACGCGCTACTTC</td><td>GTG</td><td>39137</td><td>75</td></tr> </table> </div>	Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-6,188	0,904	9	-6,962	ACGCGCGCACTTCGGTGGTGC	ATG	39200	138	2	-6,491	1,136	8	-6,913	CGCGCACTTCGGTGGTGCATG	ATG	39197	135	3	-6,761	1,103	11	-6,818	CTGGGATCACTGCGACGTCAGC	TTC	39164	102	4	-6,161	1,384	9	-6,935	GGTCGACCCGAGACGCGCTACTTC	GTG	39137	75
Start	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																															
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																															
1	-6,188	0,904	9	-6,962	ACGCGCGCACTTCGGTGGTGC	ATG	39200	138																																															
2	-6,491	1,136	8	-6,913	CGCGCACTTCGGTGGTGCATG	ATG	39197	135																																															
3	-6,761	1,103	11	-6,818	CTGGGATCACTGCGACGTCAGC	TTC	39164	102																																															
4	-6,161	1,384	9	-6,935	GGTCGACCCGAGACGCGCTACTTC	GTG	39137	75																																															
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Le start 39200 conduit à une ORF de 138pb, la plus longue chevauchement -4</p> <p>Le start 39197 conduit à une ORF de 135pb à peine plus courte chevauchement -1</p> <p>Les starts suivants conduisent à des ORFs bien plus courtes</p>																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?	<div>  </div> <p>Start le plus conservé est 39200 suivi par 39197</p>																																																						

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<pre> >re = 105 bits (262), Expect = 4e-23 ntities = 45/45 (100%), Positives = 45/45 (100%) :y: 1 MMGYLCDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45 MMGYLCDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL :t: 1 MMGYLCDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45 :ci_60, function unknown, 44 Length = 44 >re = 103 bits (257), Expect = 1e-22 ntities = 44/44 (100%), Positives = 44/44 (100%) :y: 2 MGYLECDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45 MGYLECDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL :t: 1 MGYLECDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 44 :rtz_61, function unknown, 44 Length = 44 >re = 68.6 bits (166), Expect = 5e-12 ntities = 28/44 (63%), Positives = 32/44 (72%) :y: 2 MGYLECDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45 M YLCDHC+ L +D NGYFV +SDC +D GHTWEGRAGL :t: 1 MPYLECDHCEEELHLDRNGYFVGRDNTSDCAEDPAGHTWEGRAGL 44 :idalorian_60, function unknown, 44 Length = 44 >re = 68.6 bits (166), Expect = 5e-12 ntities = 28/44 (63%), Positives = 32/44 (72%) :y: 2 MGYLECDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45 M YLCDHC+ L +D NGYFV +SDC +D GHTWEGRAGL :t: 1 MPYLECDHCEEELHLDRNGYFVGRDNTSDCAEDPAGHTWEGRAGL 44 :inyo_62, function unknown, 45 Length = 45 >re = 67.0 bits (162), Expect = 2e-11 ntities = 26/44 (59%), Positives = 32/44 (72%) :y: 2 MGYLECDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45 MGY CD C S+Q D NG+FV G+SDCP +GGHTWEG+ G+ :t: 1 MGYRCDSCHGSIQPDENGFFVGKDGTSDCPMSDGGHTWEGKEGI 44 </pre> <p>Les 2 Starts conservés</p>
DECISION:	<p>39200:</p> <ul style="list-style-type: none"> -bon score -meilleure zvalue - très conservé chez homologues -chevauchement inférieur à 30pb -ORF>50pb et la plus longue

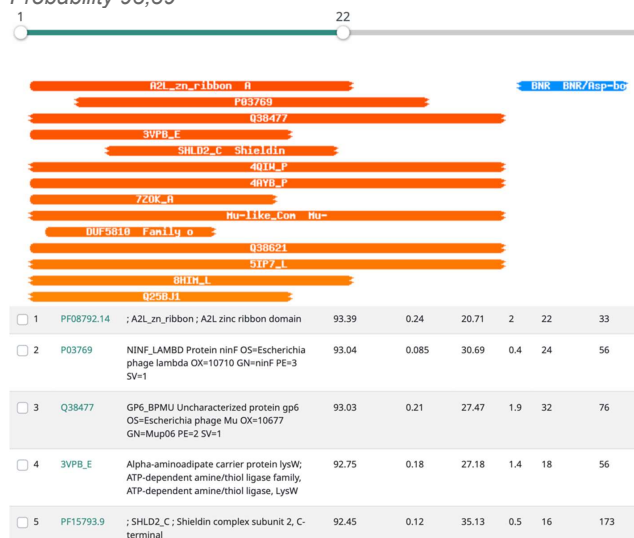
Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>*** <i>PhagesDB</i> :</p> <p>(q#Nucci_60, function unknown, 44: s#103) : [100%] ; e-value : 2e-22</p>

*** nr :
(q#hypothetical protein SEA_NUCCI_60 [Microbacterium phage Nucci]: #s93,6) : [100%]; e-value : 1e-23
hypothetical protein SEA_NUCCI_60 [Microbacterium phage Nucci]
Sequence ID: [QDK02453.1](#) Length: 44 Number of Matches: 1

Range 1: 1 to 44 GenPept Graphics ▼ Next Match ▲						
Score	Expect	Method	Identities	Positives	Gaps	
93.6 bits(231)	1e-23	Compositional matrix adjust.	44/44(100%)	44/44(100%)	0/44(0%)	
Query	2	MGYLCHDCHDVSQVDPNGFYVDVDSGSSDCPKDEGGHTWEGRAGL		45		
		MGYLCHDCHDVSQVDPNGFYVDVDSGSSDCPKDEGGHTWEGRAGL				
Sbjct	1	MGYLCHDCHDVSQVDPNGFYVDVDSGSSDCPKDEGGHTWEGRAGL		44		

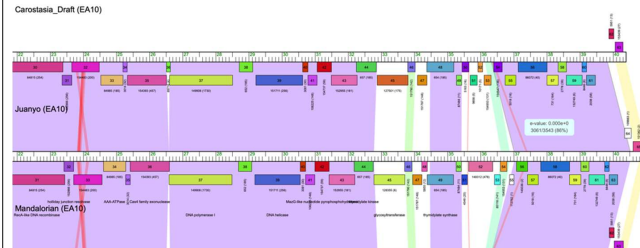
Les deux blastp ne donnent que des résultats sont fonctions
 HHPred A2L_zn_ribbon ; A2L zinc ribbon domain
 Database Pfam
 Hit [PF08792.14](#)
 Probability 93,39



Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

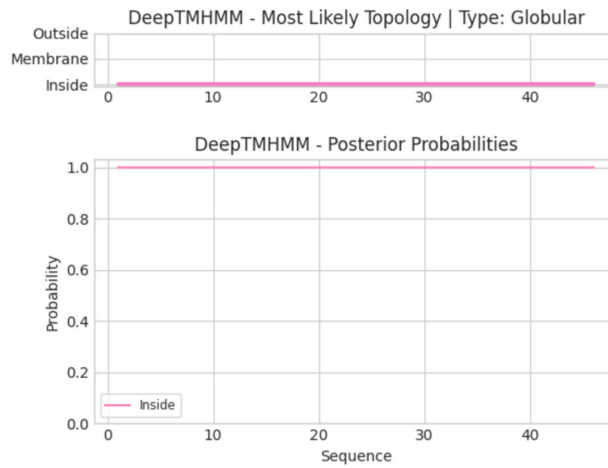
Aucune synténie observée



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

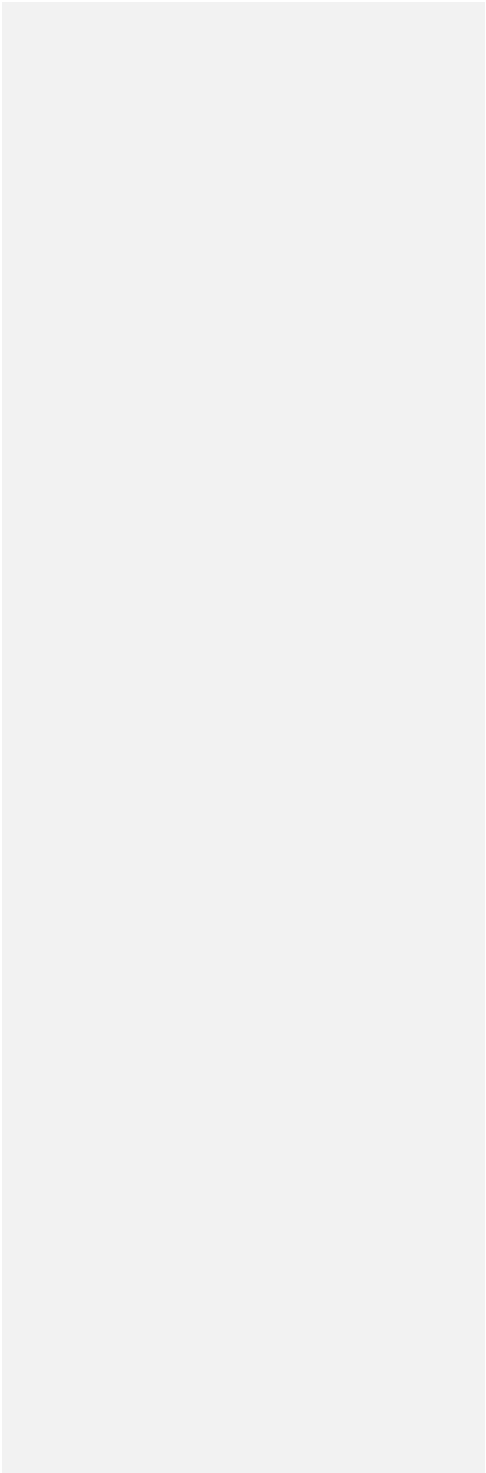
No

DeepTMHMM :



Phoebeius :

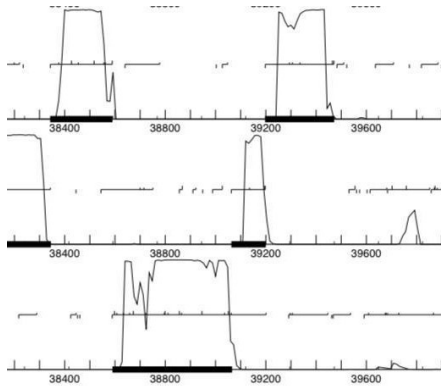
	<p>Prediction of UNNAMED</p> <p>ID UNNAMED FT TOPO_DOM 1 46 NON CYTOPLASMIC. //</p> <p>Phobius posterior probabilities for UNNAMED</p> <p>Posterior label probability</p> <p>transmembrane cytoplasmic non cytoplasmic signal peptide</p>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui.</i>
DECISION:	<i>NKF</i>

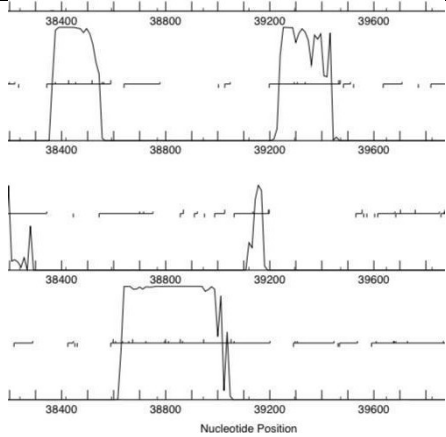


Student Gene Annotation Worksheet

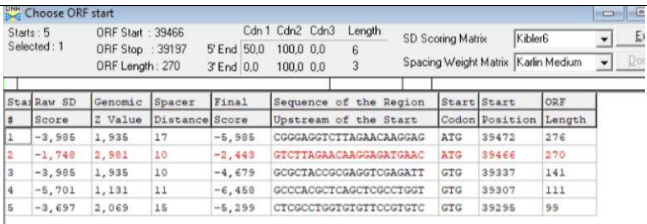
Basic Phage Information	
Nom du Phage	carostasia
Gène #	61
Coordonnées du Stop	39197
Direction (For/Rev)	rev
Gap/chevauchement avec un autre gène	Gap 191
Coordonnées du Start retenu	39466
Fonction prédite	NFK

Décision #1 : Est-ce un gène ?

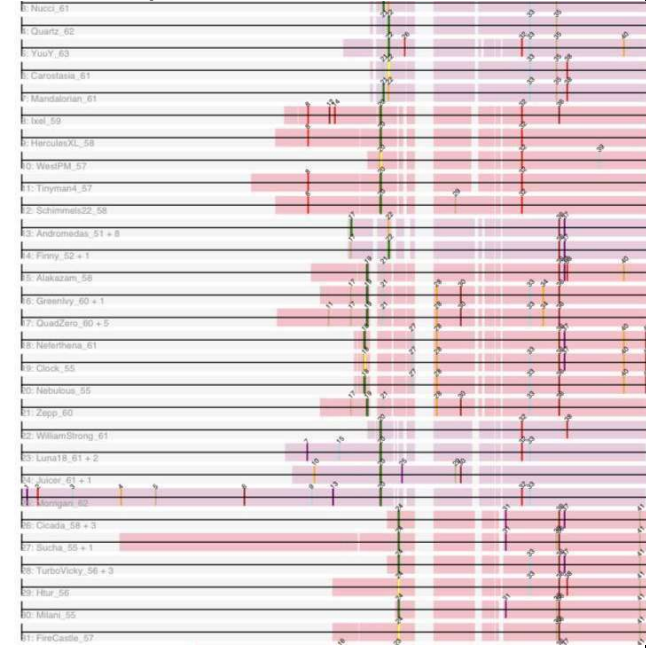
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH (same start)
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenemarkS pouvoir codant élevé dans la region de l'ORF</p>  <p>avec la sequence complementaire Genemark folio. Pareil</p>

	 <p>Nucleotide Position</p>
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><div>Mouse over to show details and scores. Click to show alignment.</div><div>Color Key for Alignment Scores</div><div>400 300 200 100 0</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><div>1000 500 250 125 62 31 16 8 4 2 1</div><</div>

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																						
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 39466 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 39466																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <p>The screenshot shows the 'Choose ORF start' tool interface. It includes input fields for 'Starts' (5), 'ORF Start' (39466), 'Cdn1', 'Cdn2', 'Cdn3', 'Length' (6), 'SD Scoring Matrix' (Kibler6), and 'Spacing Weight Matrix' (Karlin Medium). Below the interface is a table of results with columns: #, Raw SD, Genomic Z Value, Distance, Final Score, Sequence of the Region, Start Codon, Start Position, and ORF Length.</p> <table border="1"> <thead> <tr> <th>#</th> <th>Raw SD</th> <th>Genomic Z Value</th> <th>Distance</th> <th>Final Score</th> <th>Sequence of the Region</th> <th>Start Codon</th> <th>Start Position</th> <th>ORF Length</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>-3,985</td> <td>1,935</td> <td>17</td> <td>-5,985</td> <td>CGGAGGCTTTAGAACAAAGAG</td> <td>ATG</td> <td>39472</td> <td>276</td> </tr> <tr> <td>2</td> <td>-1,748</td> <td>2,981</td> <td>10</td> <td>-2,448</td> <td>GTCTTAGAACAGGAGATGAAC</td> <td>ATG</td> <td>39466</td> <td>270</td> </tr> <tr> <td>3</td> <td>-3,985</td> <td>1,935</td> <td>10</td> <td>-4,679</td> <td>GCGCTACCGCGAGGTCGAGATT</td> <td>GTG</td> <td>39337</td> <td>141</td> </tr> <tr> <td>4</td> <td>-5,701</td> <td>1,131</td> <td>11</td> <td>-6,458</td> <td>GCCACGCTCAGTCGCGCTGGT</td> <td>GTG</td> <td>39307</td> <td>111</td> </tr> <tr> <td>5</td> <td>-3,697</td> <td>2,069</td> <td>15</td> <td>-5,299</td> <td>CTCGCTGTGTGTTCTCGTGTC</td> <td>GTG</td> <td>39295</td> <td>99</td> </tr> </tbody> </table>	#	Raw SD	Genomic Z Value	Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length	1	-3,985	1,935	17	-5,985	CGGAGGCTTTAGAACAAAGAG	ATG	39472	276	2	-1,748	2,981	10	-2,448	GTCTTAGAACAGGAGATGAAC	ATG	39466	270	3	-3,985	1,935	10	-4,679	GCGCTACCGCGAGGTCGAGATT	GTG	39337	141	4	-5,701	1,131	11	-6,458	GCCACGCTCAGTCGCGCTGGT	GTG	39307	111	5	-3,697	2,069	15	-5,299	CTCGCTGTGTGTTCTCGTGTC	GTG	39295	99
#	Raw SD	Genomic Z Value	Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length																																															
1	-3,985	1,935	17	-5,985	CGGAGGCTTTAGAACAAAGAG	ATG	39472	276																																															
2	-1,748	2,981	10	-2,448	GTCTTAGAACAGGAGATGAAC	ATG	39466	270																																															
3	-3,985	1,935	10	-4,679	GCGCTACCGCGAGGTCGAGATT	GTG	39337	141																																															
4	-5,701	1,131	11	-6,458	GCCACGCTCAGTCGCGCTGGT	GTG	39307	111																																															
5	-3,697	2,069	15	-5,299	CTCGCTGTGTGTTCTCGTGTC	GTG	39295	99																																															
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<p>Le start 39472 conduit à une ORF de 276pb, la plus longue pas de chevauchement mais la distance entre l'ATG et Rbs est de 17>12</p> <p>Le start 39466 conduit à une ORF de 270 pb à peine plus courte</p> <p>Les autres starts conduisent à des ORFs bien plus courtes</p>																																																						
Est-ce que le start est conservé chez les homologues voir Starterator ?																																																							


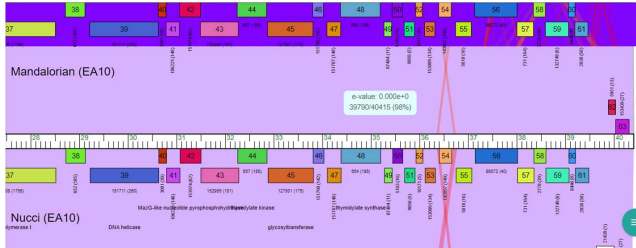
Start 39466 le plus conservé

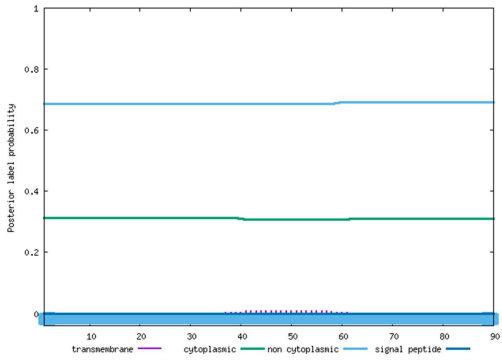
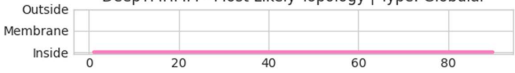
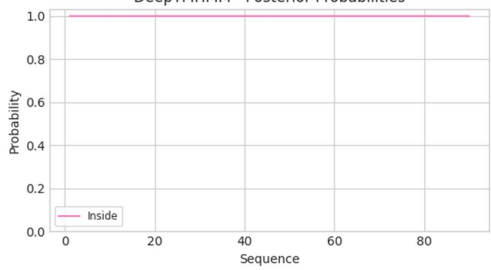


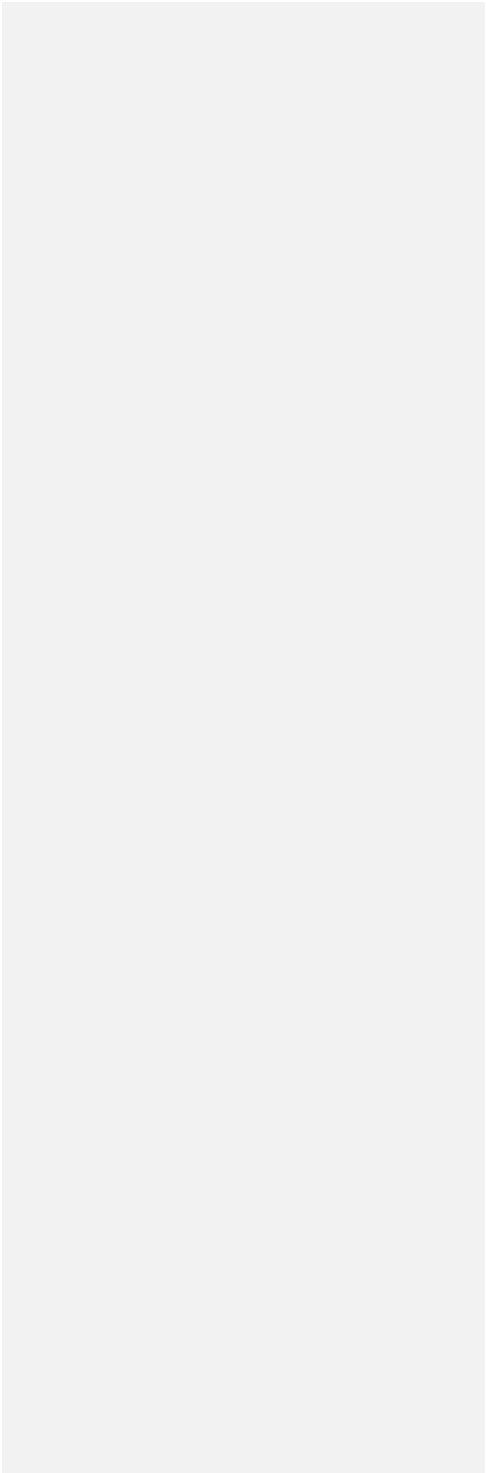
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>Score = 180 bits (457), Expect = 1e-45 Identities = 89/89 (100%), Positives = 89/89 (100%)</p> <p>Query: 1 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 Sbjct: 1 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60</p> <p>Query: 61 PKAPHTVGKLEFTRSKLIENYARTSVGA 89 PKAPHTVGKLEFTRSKLIENYARTSVGA Sbjct: 61 PKAPHTVGKLEFTRSKLIENYARTSVGA 89</p> <p>>Nucci_61, function unknown, 91 Length = 91</p> <p>Score = 180 bits (457), Expect = 1e-45 Identities = 89/89 (100%), Positives = 89/89 (100%)</p> <p>Query: 1 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 Sbjct: 3 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 62</p> <p>Query: 61 PKAPHTVGKLEFTRSKLIENYARTSVGA 89 PKAPHTVGKLEFTRSKLIENYARTSVGA Sbjct: 63 PKAPHTVGKLEFTRSKLIENYARTSVGA 91</p> <p>>Carostasia_Draft_61, function unknown, 89 Length = 89</p> <p>Score = 180 bits (457), Expect = 1e-45 Identities = 89/89 (100%), Positives = 89/89 (100%)</p> <p>Query: 1 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 Sbjct: 1 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60</p> <p>Query: 61 PKAPHTVGKLEFTRSKLIENYARTSVGA 89 PKAPHTVGKLEFTRSKLIENYARTSVGA Sbjct: 61 PKAPHTVGKLEFTRSKLIENYARTSVGA 89</p> <p>>Mandalorian_61, function unknown, 91 Length = 91</p> <p>Score = 180 bits (456), Expect = 1e-45 Identities = 88/89 (98%), Positives = 89/89 (100%)</p> <p>Query: 1 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 60 Sbjct: 3 MSINHSDIQVQGHWDVVRNTERFGKNGRKLAKPRYREVEIVALPTLSSPGVFRVKA 62</p> <p>Query: 61 PKAPHTVGKLEFTRSKLIENYARTSVGA 89 PKAPHTVGKLEFTRSKLIENYARTSVGA Sbjct: 63 PKAPHTVGKLEFTRSKLIENYARTSVGA 91</p> <p>>YusuY_63, function unknown, 85 Length = 85</p> <p>Les deux potentiels Starts semblent conservés mais parfois decalage de 3AA Remarque : si</p>
DECISION:	<p>39466:</p> <ul style="list-style-type: none"> -meilleur score -meilleur zvalue - conservé chez homologues -pas de chevauchement -ORF>50pb et 2e plus longue

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10 ⁻⁴ et une couverture acceptable ?	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB : (q#: s#) : [alignment] ; e-value :</p>

	<div>Sequences producing significant alignments:</div> <div><table><tr><th></th><th>Score</th><th>E</th></tr><tr><th></th><th>(bits)</th><th>Value</th></tr><tr><td>Quartz_62, function unknown, 89</td><td>180</td><td>1e-45</td></tr><tr><td>Nucci_61, function unknown, 91</td><td>180</td><td>1e-45</td></tr><tr><td>Carostasia_Draft_61, function unknown, 89</td><td>180</td><td>1e-45</td></tr><tr><td>Mandalorian_61, function unknown, 91</td><td>180</td><td>1e-45</td></tr><tr><td>YuuY_63, function unknown, 85</td><td>136</td><td>2e-32</td></tr><tr><td>Juanyo_63, function unknown, 85</td><td>131</td><td>7e-31</td></tr></table></div> <div>*** nr :</div> <div>(q#: s#) : [alignment] ; e-value :</div> <div><table><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr><tr><td>✓</td><td>hypothetical protein SEA_NUCC_61 [Microbacterium phage Nucci]</td><td>Microbacterium phage Nucci</td><td>184</td><td>184</td><td>98%</td><td>2e-58</td><td>100.00%</td><td>91</td><td>GD002454.1</td></tr><tr><td>✓</td><td>hypothetical protein SEA_QUARTZ_62 [Microbacterium phage Quartz]</td><td>Microbacterium phage Quartz</td><td>184</td><td>184</td><td>98%</td><td>3e-58</td><td>100.00%</td><td>89</td><td>LV055281.1</td></tr><tr><td>✓</td><td>hypothetical protein SEA_MANDALORIAN_61 [Microbacterium phage Mandalorian]</td><td>Microbacterium phage Mandalorian</td><td>184</td><td>184</td><td>98%</td><td>3e-58</td><td>98.88%</td><td>91</td><td>GX013653.1</td></tr><tr><td>✓</td><td>hypothetical protein SEA_YUYU_63 [Microbacterium phage YuuY]</td><td>Microbacterium phage YuuY</td><td>139</td><td>139</td><td>96%</td><td>2e-49</td><td>77.01%</td><td>85</td><td>GF296391.1</td></tr></table></div> <div>Fonction inconnue chaque fois</div>		Score	E		(bits)	Value	Quartz_62, function unknown, 89	180	1e-45	Nucci_61, function unknown, 91	180	1e-45	Carostasia_Draft_61, function unknown, 89	180	1e-45	Mandalorian_61, function unknown, 91	180	1e-45	YuuY_63, function unknown, 85	136	2e-32	Juanyo_63, function unknown, 85	131	7e-31		Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	✓	hypothetical protein SEA_NUCC_61 [Microbacterium phage Nucci]	Microbacterium phage Nucci	184	184	98%	2e-58	100.00%	91	GD002454.1	✓	hypothetical protein SEA_QUARTZ_62 [Microbacterium phage Quartz]	Microbacterium phage Quartz	184	184	98%	3e-58	100.00%	89	LV055281.1	✓	hypothetical protein SEA_MANDALORIAN_61 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	184	184	98%	3e-58	98.88%	91	GX013653.1	✓	hypothetical protein SEA_YUYU_63 [Microbacterium phage YuuY]	Microbacterium phage YuuY	139	139	96%	2e-49	77.01%	85	GF296391.1
	Score	E																																																																									
	(bits)	Value																																																																									
Quartz_62, function unknown, 89	180	1e-45																																																																									
Nucci_61, function unknown, 91	180	1e-45																																																																									
Carostasia_Draft_61, function unknown, 89	180	1e-45																																																																									
Mandalorian_61, function unknown, 91	180	1e-45																																																																									
YuuY_63, function unknown, 85	136	2e-32																																																																									
Juanyo_63, function unknown, 85	131	7e-31																																																																									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																																																		
✓	hypothetical protein SEA_NUCC_61 [Microbacterium phage Nucci]	Microbacterium phage Nucci	184	184	98%	2e-58	100.00%	91	GD002454.1																																																																		
✓	hypothetical protein SEA_QUARTZ_62 [Microbacterium phage Quartz]	Microbacterium phage Quartz	184	184	98%	3e-58	100.00%	89	LV055281.1																																																																		
✓	hypothetical protein SEA_MANDALORIAN_61 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	184	184	98%	3e-58	98.88%	91	GX013653.1																																																																		
✓	hypothetical protein SEA_YUYU_63 [Microbacterium phage YuuY]	Microbacterium phage YuuY	139	139	96%	2e-49	77.01%	85	GF296391.1																																																																		
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	<div></div> <div>Hitlist</div> <div>Show 25 Entries Search:</div> <div><table><tr><th>Nr</th><th>Hit</th><th>Name</th><th>Probability</th><th>E-value</th><th>Score</th><th>SS</th><th>Aligned cols</th><th>Target Length</th></tr><tr><td>1</td><td>PF19881.3</td><td>: DUF6354 Family of unknown function (DUF6354)</td><td>97.12</td><td>0.0031</td><td>46.39</td><td>5.3</td><td>66</td><td>117</td></tr><tr><td>2</td><td>cd10436</td><td>GIY-YIG_EndoII_Hpy188L-like: Catalytic GIY-YIG domain of coliphage T4 non-specific endonuclease II, type II restriction</td><td>29.21</td><td>90</td><td>18.09</td><td>1.2</td><td>31</td><td>99</td></tr><tr><td>3</td><td>PF06315.15</td><td>: AceK_kinase : Isocitrate dehydrogenase kinase/phosphatase (AceK) kinase domain</td><td>23.33</td><td>270</td><td>24.41</td><td>3.4</td><td>32</td><td>256</td></tr></table></div> <div>oui</div>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	1	PF19881.3	: DUF6354 Family of unknown function (DUF6354)	97.12	0.0031	46.39	5.3	66	117	2	cd10436	GIY-YIG_EndoII_Hpy188L-like: Catalytic GIY-YIG domain of coliphage T4 non-specific endonuclease II, type II restriction	29.21	90	18.09	1.2	31	99	3	PF06315.15	: AceK_kinase : Isocitrate dehydrogenase kinase/phosphatase (AceK) kinase domain	23.33	270	24.41	3.4	32	256																																						
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																																			
1	PF19881.3	: DUF6354 Family of unknown function (DUF6354)	97.12	0.0031	46.39	5.3	66	117																																																																			
2	cd10436	GIY-YIG_EndoII_Hpy188L-like: Catalytic GIY-YIG domain of coliphage T4 non-specific endonuclease II, type II restriction	29.21	90	18.09	1.2	31	99																																																																			
3	PF06315.15	: AceK_kinase : Isocitrate dehydrogenase kinase/phosphatase (AceK) kinase domain	23.33	270	24.41	3.4	32	256																																																																			
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	<div>Fonction inconnue.</div> <div></div>																																																																										
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	Non																																																																										

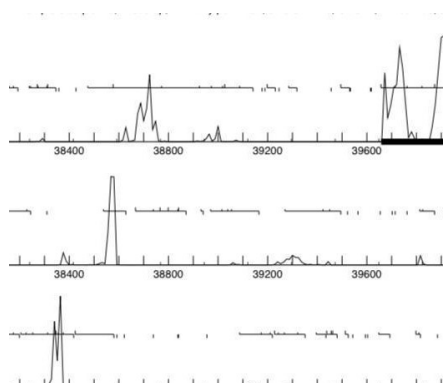
	<div><p>Phobius posterior probabilities for UNWIPED</p><p>DeepTMHMM - Most Likely Topology Type: Globular</p><p>DeepTMHMM - Posterior Probabilities</p></div>
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>Fonction inconnue.</i>
DECISION:	<i>NKF</i>

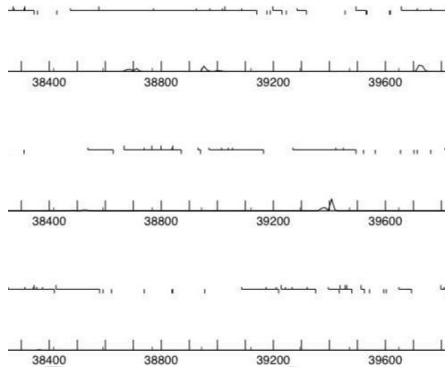



Student Gene Annotation Worksheet

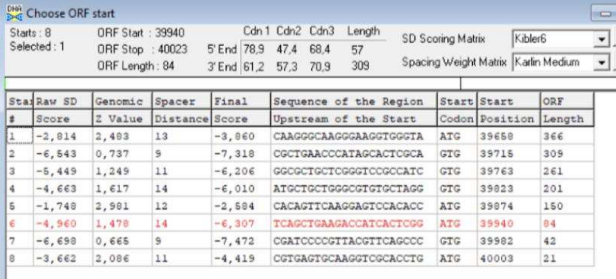
Basic Phage Information	
Nom du Phage	carostasia
Gène #	62
Coordonnées du Stop	40023
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 191
Coordonnées du Start retenu	39658
Fonction prédite	NFk

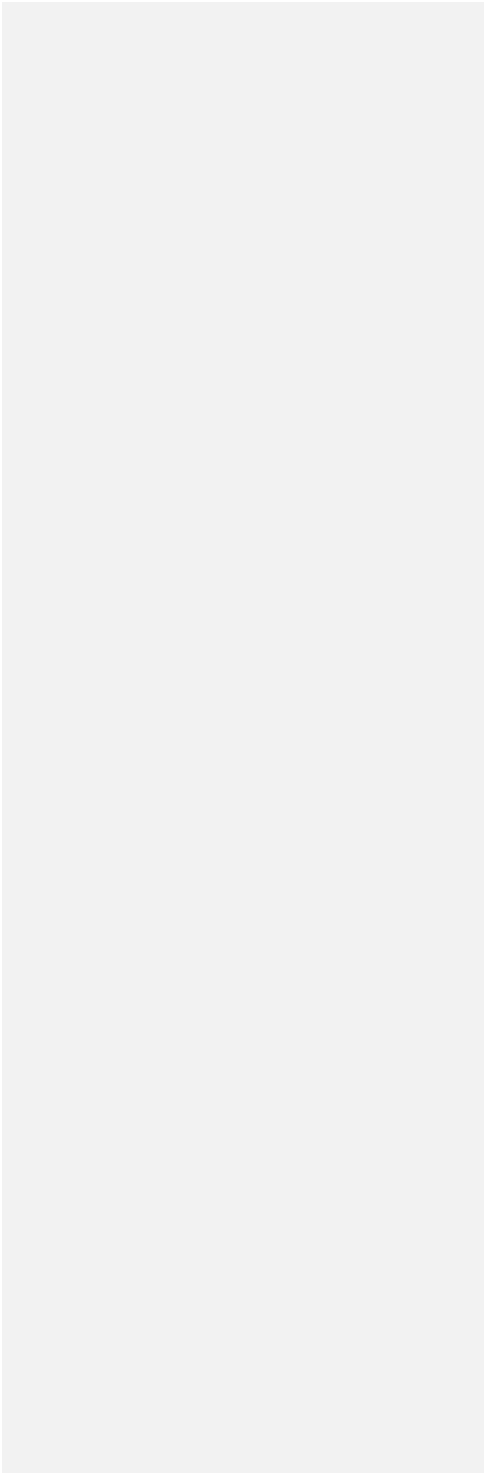
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH (same start)
Y-a-t-il des éléments supportant un potentiel codant ?	<p>Résultats GenemarkS : pouvoir codant assez élevé dans la</p>  <p>région de l'ORF</p> <p>Résultats Genemark folio : Pouvoir codant moyen dans la</p> <p>region de l'ORF</p>

																																														
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<div><p>Distribution of 14 Blast Hits on the Query Sequence</p><p>Mouse over to show details and scores. Click to show alignment.</p><p>Color Key For Alignment Scores</p><p>100 40-100 50-60 60-200 200-500</p><p>1000</p></div> <table><thead><tr><th>sequences producing significant alignments:</th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>artz_63, function unknown, 49</td><td>101</td><td>7e-2</td></tr><tr><td>cci_62, function unknown, 121</td><td>101</td><td>7e-2</td></tr><tr><td>rostasia_Draft_62, function unknown, 49</td><td>101</td><td>7e-2</td></tr><tr><td>ndalorian_62, function unknown, 49</td><td>99</td><td>4e-2</td></tr><tr><td>ratos_62, function unknown, 49</td><td>83</td><td>3e-1</td></tr><tr><td>laCorta_62, function unknown, 49</td><td>83</td><td>3e-1</td></tr><tr><td>dromedas_62, function unknown, 49</td><td>83</td><td>3e-1</td></tr><tr><td>nsa_60, function unknown, 49</td><td>82</td><td>4e-1</td></tr><tr><td>amu_Draft_62, function unknown, 49</td><td>77</td><td>1e-1</td></tr><tr><td>amour_61, function unknown, 49</td><td>77</td><td>1e-1</td></tr><tr><td>eri_62, function unknown, 49</td><td>77</td><td>1e-1</td></tr><tr><td>nitsu_62, function unknown, 49</td><td>77</td><td>1e-1</td></tr><tr><td>ubed_62, function unknown, 49</td><td>77</td><td>1e-1</td></tr><tr><td>ikPic_62, function unknown, 49</td><td>70</td><td>2e-1</td></tr></tbody></table> <p>Non annotés mais plusieurs homologues</p>	sequences producing significant alignments:	Score (bits)	E Value	artz_63, function unknown, 49	101	7e-2	cci_62, function unknown, 121	101	7e-2	rostasia_Draft_62, function unknown, 49	101	7e-2	ndalorian_62, function unknown, 49	99	4e-2	ratos_62, function unknown, 49	83	3e-1	laCorta_62, function unknown, 49	83	3e-1	dromedas_62, function unknown, 49	83	3e-1	nsa_60, function unknown, 49	82	4e-1	amu_Draft_62, function unknown, 49	77	1e-1	amour_61, function unknown, 49	77	1e-1	eri_62, function unknown, 49	77	1e-1	nitsu_62, function unknown, 49	77	1e-1	ubed_62, function unknown, 49	77	1e-1	ikPic_62, function unknown, 49	70	2e-1
sequences producing significant alignments:	Score (bits)	E Value																																												
artz_63, function unknown, 49	101	7e-2																																												
cci_62, function unknown, 121	101	7e-2																																												
rostasia_Draft_62, function unknown, 49	101	7e-2																																												
ndalorian_62, function unknown, 49	99	4e-2																																												
ratos_62, function unknown, 49	83	3e-1																																												
laCorta_62, function unknown, 49	83	3e-1																																												
dromedas_62, function unknown, 49	83	3e-1																																												
nsa_60, function unknown, 49	82	4e-1																																												
amu_Draft_62, function unknown, 49	77	1e-1																																												
amour_61, function unknown, 49	77	1e-1																																												
eri_62, function unknown, 49	77	1e-1																																												
nitsu_62, function unknown, 49	77	1e-1																																												
ubed_62, function unknown, 49	77	1e-1																																												
ikPic_62, function unknown, 49	70	2e-1																																												
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONTRADICTION ;</p> <p>-plus de 50pb</p> <p>-pas de chevauchement de plus de 30pb (-1)</p> <p>-direction différente du gène en amont (mais son écartés)</p> <p>mais similaire à celui en aval</p>																																													
DECISION:	OUI																																													

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	<i>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 39874</i> <i>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 39874</i>
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	<i>Le start 39658 conduit à une ORF de 366pb, la plus longue avec un chevauchement de -1>-30 avec le gene aval et un gap de 191pb</i> <i>Le start 39874 conduit à une ORF de 150 pb avec donc le même chevauchement mais un gap de 407 pb</i> <i>Les starts suivants suivent conduisent à des ORFs beaucoup plus courtes.</i>
Est-ce que le start est conservé chez les homologues voir Starterator ?	<i>Les deux starts sont relativement conservés</i>



Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<p>1 bits (251), Expect = 7e-22 = 49/49 (100%), Positives = 49/49 (100%)</p> <p>NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49 NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49</p> <p>unction unknown, 121 ngth = 121</p> <p>1 bits (251), Expect = 7e-22 = 49/49 (100%), Positives = 49/49 (100%)</p> <p>SNTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49 SNTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH SNTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 121</p> <p>Draft_62, function unknown, 49 ngth = 49</p> <p>1 bits (251), Expect = 7e-22 = 49/49 (100%), Positives = 49/49 (100%)</p> <p>NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49 NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49</p> <p>_62, function unknown, 49 ngth = 49</p> <p>0 bits (245), Expect = 4e-21 = 48/49 (97%), Positives = 48/49 (97%)</p> <p>NTLEKPVQTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49 NTLEKPV QTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH NTLEKPVMTIPVQLKITITRMGEVRPSDPRYVQPV SARSHLMNAGSH 49</p> <p>Start majoritairement le même avec 49 AA donc environ 150pb Mais aussi 121 AA donc environ 360pb</p>
DECISION:	<p>39658:</p> <ul style="list-style-type: none"> -bon score -bon zvalue -meilleure longueur -très conservé chez homologues -diminution du GAP -ORF>50pb

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP	<p>*** PhagesDB :</p> <p>(q#Nucci_62, function unknown, 121: s#252) : [100%] ; e-value : 2e-67</p>

(phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

>Nucci_62, function unknown, 121
Length = 121

Score = 252 bits (644), Expect = 2e-67
Identities = 121/121 (100%), Positives = 121/121 (100%)

Query: 1 MEGERPHTHDGAEPIALAVTNPRHAPEALLGSAIVTLIAALYCRACGCCWACARVGAAP 60
MEGERPHHTHDGAEPIALAVTNPRHAPEALLGSAIVTLIAALYCRACGCCWACARVGAAP
Sbjct: 1 MEGERPHTHDGAEPIALAVTNPRHAPEALLGSAIVTLIAALYCRACGCCWACARVGAAP 60

Query: 61 KAGRSTVQGVHTMSNTLEKPVQTIPVQLKTITRMGEVRFSDPRYVQPVARSHLMNAGS 120
KAGRSTVQGVHTMSNTLEKPVQTIPVQLKTITRMGEVRFSDPRYVQPVARSHLMNAGS
Sbjct: 61 KAGRSTVQGVHTMSNTLEKPVQTIPVQLKTITRMGEVRFSDPRYVQPVARSHLMNAGS 120

Query: 121 H 121
H
Sbjct: 121 H 121

*** nr :
(q#hypothetical protein SEA_NUCCI_60 [Microbacterium phage Nucci]: s#93,6) : [100%] ; e-value : 1e-23
hypothetical protein SEA_NUCCI_60 [Microbacterium phage Nucci]
Sequence ID: [QDK02453.1](#) Length: 44 Number of Matches: 1

Range 1: 1 to 44 [GenPept](#) [Graphics](#) [▼ Next Match ▲](#)

Score	Expect	Method	Identities	Positives	Gaps
93.6 bits(231)	1e-23	Compositional matrix adjust.	44/44(100%)	44/44(100%)	0/44(0%)

Query 2 MGYLCDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 45
MGYLCDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL
Sbjct 1 MGYLCDHCDVSLQVDPNGYFVDDSGSSDCPKDEGGHTWEGRAGL 44

Les alignements ne donnent pas de résultats de fonctions

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

HHPred : QueC ; Queuosine biosynthesis protein QueC
Probability 76,49
Hit [PF06508.17](#)
Database Pfam

4756

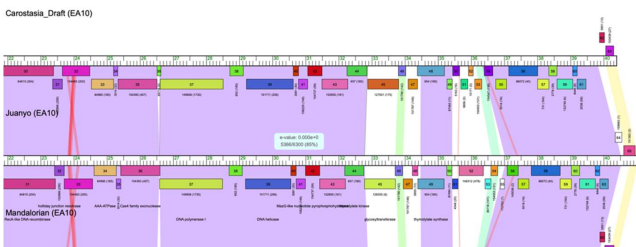
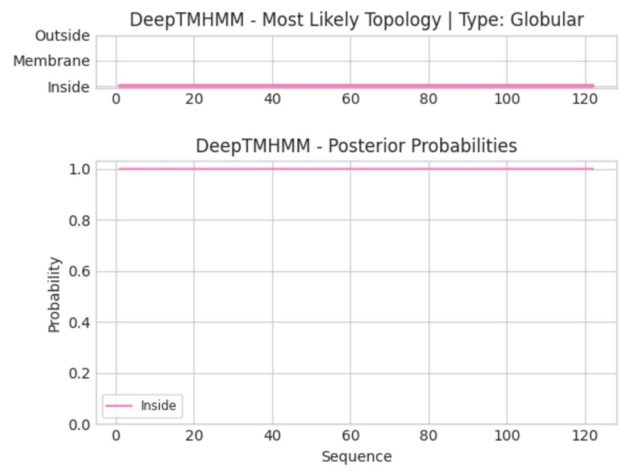
QueC

cd01995

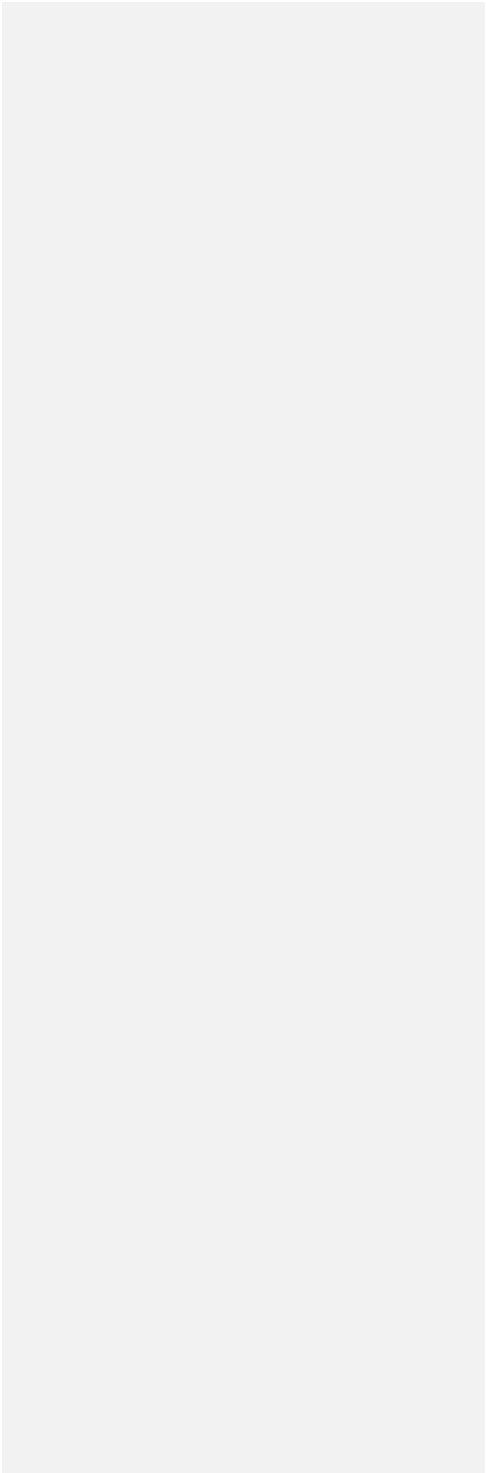
1xQ4_A

<input type="checkbox"/>	1	PF06508.17	; QueC ; Queuosine biosynthesis protein QueC	76.49	3.5	25.71	1	10	212
<input type="checkbox"/>	2	cd01995	ExsB; ExsB is a transcription regulator related protein. It is a subfamily of a Adenosine nucleotide binding superfamily	72.05	5.1	24.92	1	10	210
<input type="checkbox"/>	3	3BL5_C	Queuosine biosynthesis protein queC; preQ1 biosynthesis, RNA modification, tRNA, Hydrolase, Queuosine biosynthesis; HET;	61.98	10	26.07	0.9	9	219
<input type="checkbox"/>	4	Q9YMP9	CATV_NPVLD Viral cathepsin OS=Lymantria dispar multicapsid nuclear polyhedrosis virus OX=10449 GN=VCATH PE=3 SV=1	61.84	7.9	29.41	0.4	12	356
<input type="checkbox"/>	5	2COY_A	PROCATHEPSIN S; PROCATHEPSIN S, PROENZYME, PROTEINASE, HYDROLASE, THIOL PROTEASE, PROSEGMENT BINDING LOOP, GLYCOPROTEIN,	61.18	8.5	27.12	0.5	12	315

Pas de correspondance pertinente car probability < 80% et les divers résultats donnent des fonctions différentes

<p>Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?</p>	<p><i>Aucune synténie observée</i></p> 
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p><i>No</i></p> <p><i>DeepTMHMM :</i></p>  <p><i>Phoebeius :</i></p>

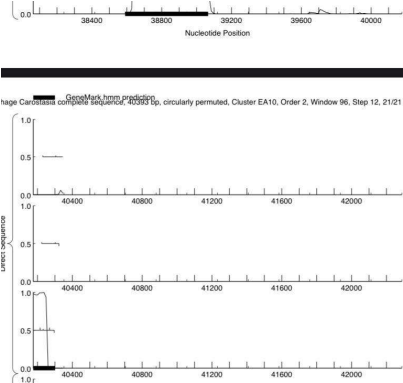
	<div><div><div>ID</div><div>FT</div><div>//</div></div><div>UNNAMED TOPO_DOM</div><div>146</div><div>NON CYTOPLASMIC.</div></div> <div><p>Phobius posterior probabilities for UNNAMED</p><table border="1"><caption>Phobius Posterior Probabilities for UNNAMED</caption><thead><tr><th>Label</th><th>Probability</th></tr></thead><tbody><tr><td>transmembrane</td><td>~0.02</td></tr><tr><td>cytoplasmic</td><td>~0.32</td></tr><tr><td>non cytoplasmic</td><td>~0.35</td></tr><tr><td>signal peptide</td><td>~0.68</td></tr><tr><td>Other</td><td>~0.02</td></tr></tbody></table></div>	Label	Probability	transmembrane	~0.02	cytoplasmic	~0.32	non cytoplasmic	~0.35	signal peptide	~0.68	Other	~0.02
Label	Probability												
transmembrane	~0.02												
cytoplasmic	~0.32												
non cytoplasmic	~0.35												
signal peptide	~0.68												
Other	~0.02												
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	<i>oui</i>												
DECISION:	<i>NKF</i>												

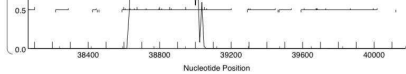
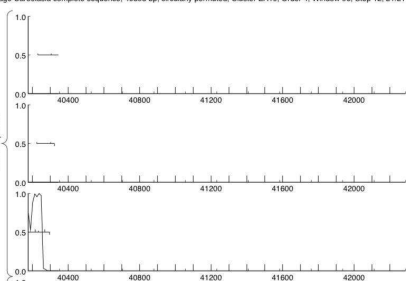
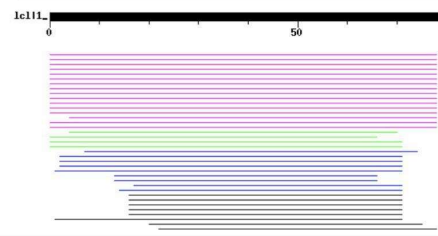


Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	carostasia
Gène #	63
Coordonnées du Stop	40301
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Chev 1
Coordonnées du Start retenu	40023
Fonction prédite	nfk

Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>Résultats GenemarkS:</p>  <p>Résultats Genemark folio:</p>

	 <p>age Carostasia complete sequence, 40393 bp, circularly permuted, Cluster EA10, Order 4, Window 96, Step 12, 21/21</p>  <p>-> Pic fin dans les deux cas entre 40050 et 40300</p>
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	<p>Distribution of 37 Blast Hits on the Query Sequence</p> <p>Mouse-over to show define and scores. Click to show alignments</p> <p>Color Key for Alignment Scores</p> <p><40 40-50 50-80 80-200 >200</p>  <p>Nombreux homologues non-annotés</p>
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	<p>PAS DE CONFLITS :</p> <ul style="list-style-type: none"> -ORF de plus de 50pb -pas de chevauchement de plus de 30pb (-1) -même direction que le gène amont
DECISION:	OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																					
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 40062 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 40023																																																																																																																					
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div><div>Choose ORF start</div><div>Stats: 11 ORF Start : 40134 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler6 Selected: 1 ORF Stop : 40301 5' End 69.2 46.2 76.9 39 ORF Length: 168 3' End 72.5 45.0 81.3 240 Spacing Weight Matrix Karlin Medium</div><table><tr><th>Stat</th><th>Raw SD</th><th>Genomic</th><th>Spacer</th><th>Final</th><th>Sequence of the Region</th><th>Start</th><th>Start</th><th>ORF</th></tr><tr><th>#</th><th>Score</th><th>Z Value</th><th>Distance</th><th>Score</th><th>Upstream of the Start</th><th>Codon</th><th>Position</th><th>Length</th></tr><tr><td>1</td><td>-3,857</td><td>1,994</td><td>11</td><td>-4,614</td><td>TGATGAACGCTGGGAGCCACTG</td><td>ATG</td><td>40023</td><td>279</td></tr><tr><td>2</td><td>-5,653</td><td>1,154</td><td>11</td><td>-6,410</td><td>TGGGATGTCGGGCGTTCGCAC</td><td>ATG</td><td>40062</td><td>240</td></tr><tr><td>3</td><td>-2,757</td><td>2,509</td><td>10</td><td>-3,452</td><td>GCACATGTCCAAGGACCCAGCAG</td><td>GTG</td><td>40080</td><td>222</td></tr><tr><td>4</td><td>-2,757</td><td>2,509</td><td>13</td><td>-3,803</td><td>CATGTCCAAGGACCCAGCAGTG</td><td>ATG</td><td>40083</td><td>219</td></tr><tr><td>5</td><td>-2,812</td><td>2,484</td><td>16</td><td>-4,608</td><td>CCAGCAGGTGATGGCTCGGGAG</td><td>ATG</td><td>40095</td><td>207</td></tr><tr><td>6</td><td>-5,791</td><td>1,089</td><td>5</td><td>-7,791</td><td>GCCGAAGTGGCTTGCCAGCGTC</td><td>ATG</td><td>40125</td><td>177</td></tr><tr><td>7</td><td>-3,642</td><td>2,095</td><td>7</td><td>-5,164</td><td>GCTTGCCAGCGTCATGGACCAG</td><td>GTG</td><td>40134</td><td>168</td></tr><tr><td>8</td><td>-4,495</td><td>1,696</td><td>13</td><td>-5,541</td><td>TGGTCCAAGGCTGATCCGGCT</td><td>GTG</td><td>40164</td><td>138</td></tr><tr><td>9</td><td>-7,021</td><td>0,514</td><td>8</td><td>-8,242</td><td>GGGGCCAGACCTTCGGCTGGGC</td><td>ATG</td><td>40218</td><td>84</td></tr><tr><td>10</td><td>-4,047</td><td>1,906</td><td>13</td><td>-5,092</td><td>GGGCATGGGGAGTGGCGTGCC</td><td>GTG</td><td>40236</td><td>66</td></tr><tr><td>11</td><td>-4,228</td><td>1,821</td><td>15</td><td>-5,830</td><td>CGAGGCAGGCATGTCCGCGGG</td><td>ATG</td><td>40272</td><td>30</td></tr></table></div>	Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length	1	-3,857	1,994	11	-4,614	TGATGAACGCTGGGAGCCACTG	ATG	40023	279	2	-5,653	1,154	11	-6,410	TGGGATGTCGGGCGTTCGCAC	ATG	40062	240	3	-2,757	2,509	10	-3,452	GCACATGTCCAAGGACCCAGCAG	GTG	40080	222	4	-2,757	2,509	13	-3,803	CATGTCCAAGGACCCAGCAGTG	ATG	40083	219	5	-2,812	2,484	16	-4,608	CCAGCAGGTGATGGCTCGGGAG	ATG	40095	207	6	-5,791	1,089	5	-7,791	GCCGAAGTGGCTTGCCAGCGTC	ATG	40125	177	7	-3,642	2,095	7	-5,164	GCTTGCCAGCGTCATGGACCAG	GTG	40134	168	8	-4,495	1,696	13	-5,541	TGGTCCAAGGCTGATCCGGCT	GTG	40164	138	9	-7,021	0,514	8	-8,242	GGGGCCAGACCTTCGGCTGGGC	ATG	40218	84	10	-4,047	1,906	13	-5,092	GGGCATGGGGAGTGGCGTGCC	GTG	40236	66	11	-4,228	1,821	15	-5,830	CGAGGCAGGCATGTCCGCGGG	ATG	40272	30
Stat	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																																																																																														
#	Score	Z Value	Distance	Score	Upstream of the Start	Codon	Position	Length																																																																																																														
1	-3,857	1,994	11	-4,614	TGATGAACGCTGGGAGCCACTG	ATG	40023	279																																																																																																														
2	-5,653	1,154	11	-6,410	TGGGATGTCGGGCGTTCGCAC	ATG	40062	240																																																																																																														
3	-2,757	2,509	10	-3,452	GCACATGTCCAAGGACCCAGCAG	GTG	40080	222																																																																																																														
4	-2,757	2,509	13	-3,803	CATGTCCAAGGACCCAGCAGTG	ATG	40083	219																																																																																																														
5	-2,812	2,484	16	-4,608	CCAGCAGGTGATGGCTCGGGAG	ATG	40095	207																																																																																																														
6	-5,791	1,089	5	-7,791	GCCGAAGTGGCTTGCCAGCGTC	ATG	40125	177																																																																																																														
7	-3,642	2,095	7	-5,164	GCTTGCCAGCGTCATGGACCAG	GTG	40134	168																																																																																																														
8	-4,495	1,696	13	-5,541	TGGTCCAAGGCTGATCCGGCT	GTG	40164	138																																																																																																														
9	-7,021	0,514	8	-8,242	GGGGCCAGACCTTCGGCTGGGC	ATG	40218	84																																																																																																														
10	-4,047	1,906	13	-5,092	GGGCATGGGGAGTGGCGTGCC	GTG	40236	66																																																																																																														
11	-4,228	1,821	15	-5,830	CGAGGCAGGCATGTCCGCGGG	ATG	40272	30																																																																																																														
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	Le start 40023 permet une ORF de 279pb avec un chevauchement de 1pb (<30) Les starts suivants conduisent à des ORFs plus courtes et souvent avec un moins bon score.																																																																																																																					
Est-ce que le start est conservé chez les homologues voir Starterator ?	<div>Pham 142756</div> <div></div>																																																																																																																					

	<i>Start conservé</i>
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	<div><div>20:07</div><div>Saher_65, HNH endonuclease, 389 Picard_41, DNA binding protein, 301 >Nucl_63, function unknown, 92 Length = 92 Score = 154 bits (389), Expect = 7e-38 Identities = 79/79 (100%), Positives = 79/79 (100%) Query: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Sbjct: 14 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Query: 61 DLVEAGIVAGHAKAGTTPR 79 DLVEAGIVAGHAKAGTTPR Sbjct: 74 DLVEAGIVAGHAKAGTTPR 92 >Carostasia_Draft_63, function unknown, 79 Length = 79 Score = 154 bits (389), Expect = 7e-38 Identities = 79/79 (100%), Positives = 79/79 (100%) Query: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Sbjct: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Query: 61 DLVEAGIVAGHAKAGTTPR 79 DLVEAGIVAGHAKAGTTPR Sbjct: 61 DLVEAGIVAGHAKAGTTPR 79 >Mandalorian_63, function unknown, 92 Length = 92 Score = 153 bits (386), Expect = 2e-37 Identities = 78/79 (98%), Positives = 79/79 (100%) Query: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Sbjct: 14 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Query: 61 DLVEAGIVAGHAKAGTTPR 79 DLVEAGIVAGHAKAGTTPR Sbjct: 74 DLVEAGIVAGHAKAGTTPR 92 >Quartz_64, function unknown, 92 Length = 92 Score = 152 bits (385), Expect = 2e-37 Identities = 78/79 (98%), Positives = 78/79 (98%) Query: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Sbjct: 14 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Query: 61 DLVEAGIVAGHAKAGTTPR 79 DLVEAGIVAGHAKAGTTPR Sbjct: 74 DLVEAGIVAGHAKAGTTPR 92 >Sanaa_61, function unknown, 103 Length = 103 Score = 152 bits (385), Expect = 2e-37 Identities = 78/79 (98%), Positives = 78/79 (98%) Query: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Sbjct: 14 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Query: 61 DLVEAGIVAGHAKAGTTPR 79 DLVEAGIVAGHAKAGTTPR Sbjct: 74 DLVEAGIVAGHAKAGTTPR 92 >Sanaa_61, function unknown, 103 Length = 103 Score = 152 bits (385), Expect = 2e-37 Identities = 78/79 (98%), Positives = 78/79 (98%) Query: 1 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Sbjct: 14 MSKDQVHARENSQPKSLASVNDQVGRSKADPAVEALALQIETGPDRLGHS Query: 61 DLVEAGIVAGHAKAGTTPR 79 DLVEAGIVAGHAKAGTTPR Sbjct: 74 DLVEAGIVAGHAKAGTTPR 92 à phagesdb.org</div><p><i>Start pas toujours le même, décalage de 14AA</i></p></div>
DECISION:	<p>40023:</p> <ul style="list-style-type: none">-bonne longueur-chevauchement de moins de 30pb (et donc proximité des ORF)-bon score et Z value-conservé chez des homologues

Décision #3 : Quelle est la fonction de la protéine putative ?

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction	

annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :

(q#: s#) : [alignment] ; e-value :

Sequences producing significant alignments:	Score (bits)	E Value
Nucci_63, function unknown, 92	180	1e-45
Mandalorian_63, function unknown, 92	179	3e-45
Quartz_64, function unknown, 92	178	4e-45
Carostasia_Draft_63, function unknown, 79	154	7e-38
Sansa_61, function unknown, 103	146	2e-35
ColaCorta_63, function unknown, 103	144	6e-35
Andromedas_63, function unknown, 103	144	6e-35

*** nr :

(q#: s#) : [alignment] ; e-value :

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Hypothetical protein SEA_NUCCI_63 [Microbacterium phage Nucci]	Microbacterium phage Nucci	182	182	98%	2e-57	100.00%	92	GD032456.1
<input checked="" type="checkbox"/> Hypothetical protein SEA_MANDALORIAN_63 [Microbacterium phage Mandalorian]	Microbacterium phage Mandalorian	181	181	98%	4e-57	98.91%	92	QX013655.1
<input checked="" type="checkbox"/> Hypothetical protein SEA_QUARTZ_64 [Microbacterium phage Quartz]	Microbacterium phage Quartz	181	181	98%	7e-57	98.91%	92	U9559283.1
<input checked="" type="checkbox"/> Hypothetical protein SEA_SANSA_61 [Microbacterium phage Sansa]	Microbacterium phage Sansa	148	148	98%	8e-44	73.79%	103	AX048634.1
<input checked="" type="checkbox"/> Hypothetical protein SEA_COLACORTA_63 [Microbacterium phage ColaCorta]	Microbacterium phage ColaCorta	145	145	98%	6e-43	71.84%	103	AX078616.1

Fonction inconnue chaque fois

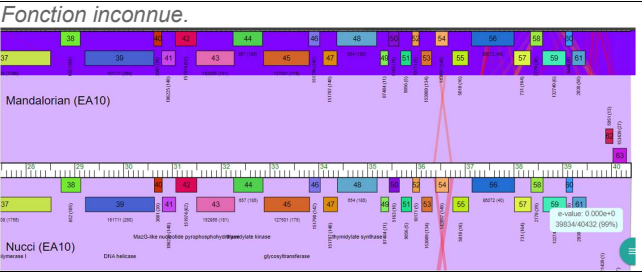
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un cribble HHPred avec une proba >= 90% et une couverture acceptable ?

non

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/> 1	7B2H_A	SuHc Archaea, Crenarchaea, Sulfolobus, DNA binding protein; NMR (Sulfolobus islandicus)	71.11	14	23.22	2.8	35	59
<input type="checkbox"/> 2	4AU1_A	PRECORRIN-BX METHYLMUTASE, ISOMERASE, PRECORRIN, HBA-VET-PEK, 1.45A (RHODOBACTER CAPSULATUS) SCOP: c.23.17.1	37.56	140	24.96	3.8	33	229
<input type="checkbox"/> 3	PF03713.17	: DUF305 : Domain of unknown function (DUF305)	37.23	150	19.63	3.3	29	148
<input type="checkbox"/> 4	1OU0_C	precorrin-BX methylmutase related protein; structural genomics, methylmutase, precorrin-BX, PSL, Protein Structure Initi	36.67	160	23.92	4	33	207

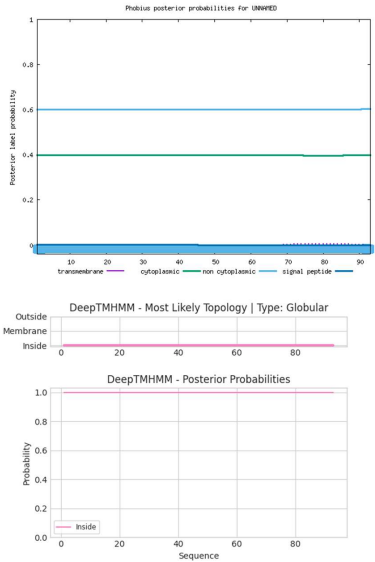
Visualization

Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?



Est-ce que ce gène code pour une protéine transmembranaire (TM) ?

Non.



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?

Fonction inconnue.

DECISION:

NKF