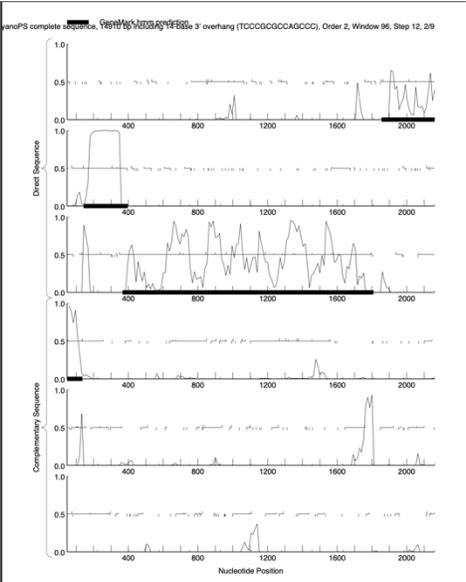


Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	1
Coordonnées du Stop	397
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 60
Coordonnées du Start retenu	77
Fonction prédite	terminase, small subunit

Décision #1 : Est-ce un gène ?

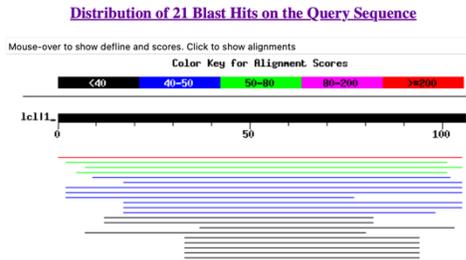
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS montre un potentiel codant.</p> 

NR:

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession
<input checked="" type="checkbox"/> hypothetical protein AS774_06175 [Corynebacterium sp. EPI-003-04-2554_SCH247...]	Corynebacterium sp. EPI-003-04-2554_S...	89.0	89.0	95%	6e-20	47.06%	124	CBAS4007.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	88.6	88.6	95%	1e-19	47.06%	124	WP_284849028.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium sp. EPI-003-04-2554_SCH2473622]	Corynebacterium sp. EPI-003-04-2554_S...	88.6	88.6	95%	1e-19	47.06%	138	WP_231892363.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium crocinosum]	Corynebacterium crocinosum	84.7	84.7	95%	3e-18	46.08%	125	WP_284594237.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	82.4	82.4	73%	3e-17	51.28%	124	WP_284587129.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	75.9	75.9	97%	8e-15	43.27%	113	WP_051255733.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	75.1	75.1	78%	1e-14	44.58%	106	WP_041627971.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	75.1	75.1	78%	1e-14	44.58%	106	WP_003852540.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	74.7	74.7	78%	2e-14	44.58%	114	WP_241879851.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	74.7	74.7	78%	2e-14	44.58%	113	WP_241880226.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	74.7	74.7	78%	2e-14	44.58%	115	WP_241879860.1

Contre la PhageDB on retrouve plusieurs hit avec des e value basses.

Query= CyranoPS_1
(106 letters)



Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_1, function unknown, 106	208	4e-54
Schiebs_1, terminase small subunit, 135	59	4e-09
SallySpecial_1, terminase small subunit, 131	59	4e-09
Emperor_1, terminase small subunit, 115	58	7e-09
RRH1_01, terminase small subunit, 126	49	6e-06
EpicDab_1, function unknown, 142	46	3e-05
Coeur_1, terminase small subunit, 148	45	8e-05
Rahul_1, terminase small subunit, 145	42	4e-04
McDonagall_1, function unknown, 120	42	5e-04
Jeanie_1, function unknown, 120	42	5e-04
GRU3_1, function unknown, 145	42	5e-04
GMA5_1, function unknown, 135	41	0.001
PFR2_01, function unknown, 131	37	0.022
PFR1_01, function unknown, 131	37	0.022

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

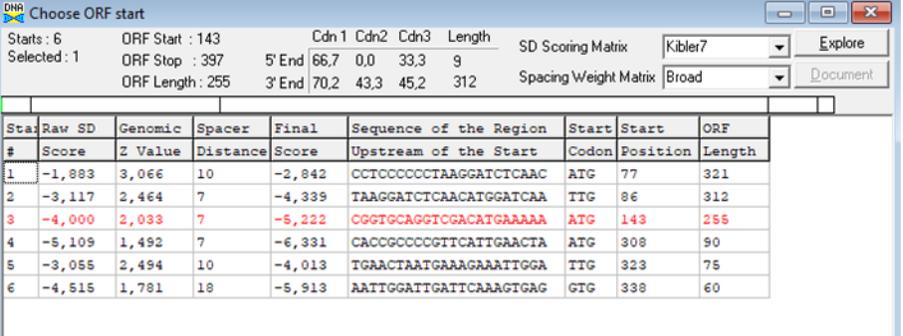
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

En fonction de ORF 1 : il faut un espace d'au moins 50pb avec le gene precedent qui est dans la direction opposée

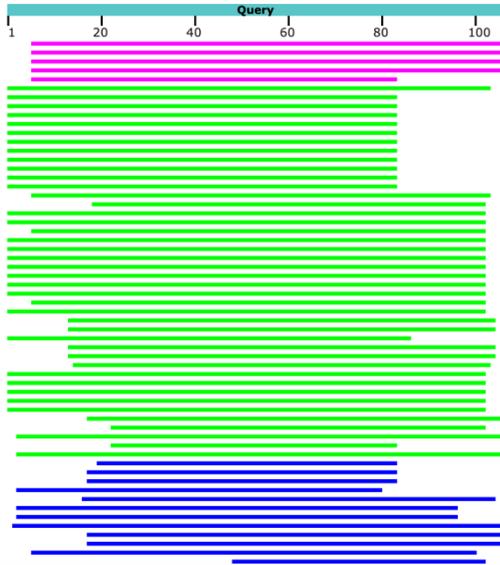
DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																								
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 77 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 143</p>																																																																								
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <table border="1"> <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>-1,883</td> <td>3,066</td> <td>10</td> <td>-2,842</td> <td>CCTCCCCCTAAGGATCTCAAC</td> <td>ATG</td> <td>77</td> <td>321</td> </tr> <tr> <td>2</td> <td>-3,117</td> <td>2,464</td> <td>7</td> <td>-4,339</td> <td>TAAGGATCTCAACATGGATCAA</td> <td>TTG</td> <td>86</td> <td>312</td> </tr> <tr> <td>3</td> <td>-4,000</td> <td>2,033</td> <td>7</td> <td>-5,222</td> <td>CGGTCCAGGTCCACATGAAAA</td> <td>ATG</td> <td>143</td> <td>255</td> </tr> <tr> <td>4</td> <td>-5,109</td> <td>1,492</td> <td>7</td> <td>-6,331</td> <td>CACGCCCCGTCATTGAACTA</td> <td>ATG</td> <td>308</td> <td>90</td> </tr> <tr> <td>5</td> <td>-3,055</td> <td>2,494</td> <td>10</td> <td>-4,013</td> <td>TGAACTAATGAAAGAAATTGGA</td> <td>TTG</td> <td>323</td> <td>75</td> </tr> <tr> <td>6</td> <td>-4,515</td> <td>1,781</td> <td>18</td> <td>-5,913</td> <td>AATGGATTGATCAAAAGTGAG</td> <td>GTG</td> <td>338</td> <td>60</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	-1,883	3,066	10	-2,842	CCTCCCCCTAAGGATCTCAAC	ATG	77	321	2	-3,117	2,464	7	-4,339	TAAGGATCTCAACATGGATCAA	TTG	86	312	3	-4,000	2,033	7	-5,222	CGGTCCAGGTCCACATGAAAA	ATG	143	255	4	-5,109	1,492	7	-6,331	CACGCCCCGTCATTGAACTA	ATG	308	90	5	-3,055	2,494	10	-4,013	TGAACTAATGAAAGAAATTGGA	TTG	323	75	6	-4,515	1,781	18	-5,913	AATGGATTGATCAAAAGTGAG	GTG	338	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	-1,883	3,066	10	-2,842	CCTCCCCCTAAGGATCTCAAC	ATG	77	321																																																																	
2	-3,117	2,464	7	-4,339	TAAGGATCTCAACATGGATCAA	TTG	86	312																																																																	
3	-4,000	2,033	7	-5,222	CGGTCCAGGTCCACATGAAAA	ATG	143	255																																																																	
4	-5,109	1,492	7	-6,331	CACGCCCCGTCATTGAACTA	ATG	308	90																																																																	
5	-3,055	2,494	10	-4,013	TGAACTAATGAAAGAAATTGGA	TTG	323	75																																																																	
6	-4,515	1,781	18	-5,913	AATGGATTGATCAAAAGTGAG	GTG	338	60																																																																	
<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>	<p>ATG 77 = ORF la plus longue 321pb ATG 143 = ORF de 255pb TTG 86 = ORF de 312 pb</p> <p>-> Favoriser celui de glimmer à 77 : ORF le plus long</p>																																																																								
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>Pas d'information sur Starterator</p> <table border="1"> <thead> <tr> <th>Starterator</th> <th>Orphan, no data</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Starterator	Orphan, no data																																																																						
Starterator	Orphan, no data																																																																								
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<p>Nr :</p> <div data-bbox="573 1381 1476 1648"> <p>Download GenPept Graphics Next Previous Descriptions</p> <p>hypothetical protein A5774_06175 [Corynebacterium sp. EPI-003-04-2554_SCH2473622] Sequence ID: OBAS4007.1 Length: 124 Number of Matches: 1</p> <p>Range 1: 11 to 112 GenPept Graphics Next Match Previous Match Related Information AlphaFold Structure - 3D structure displays</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>89.0 bits(219)</td> <td>6e-20</td> <td>Compositional matrix adjust.</td> <td>48/102(47%)</td> <td>67/102(65%)</td> <td>1/102(0%)</td> </tr> </tbody> </table> <p>Query 6 LFDPPPPQAGRHEKMKVRLKALAEAEETRGTLTAVDGAISLAIANAWALDEAEHEGQPF 65 LFD P+R AGRHE+ + +A+E A+ G + +D +SLA ANAWALDEAE G P+ Sbjct 11 LFDVADIPRRPAGRHEQQLDRAIEHAQQSGAVDTLDAGLVSLARANAWALDEAEANGPLY 70</p> <p>Query 66 AVAQITAPFIELMKEIGLIQSEVST-DDDKLALALQELSATD 106 AVA +T PF E++ E+GL + +T +D+ L AL +L A D Sbjct 71 AVANLTPPFREVLAEGLTPAARNANDNALNDALNKLLADD 112</p> </div> <div data-bbox="573 1675 1476 1953"> <p>Download GenPept Graphics Next Previous Descriptions</p> <p>hypothetical protein [Corynebacterium pseudodiphtheriticum] Sequence ID: WP_284849028.1 Length: 124 Number of Matches: 1 See 1 more title(s) See all Identical Proteins(IPG)</p> <p>Range 1: 11 to 112 GenPept Graphics Next Match Previous Match Related Information Identical Proteins - Identical proteins to WP_284849028.1</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>88.6 bits(218)</td> <td>1e-19</td> <td>Compositional matrix adjust.</td> <td>48/102(47%)</td> <td>67/102(65%)</td> <td>1/102(0%)</td> </tr> </tbody> </table> <p>Query 6 LFDPPPPQAGRHEKMKVRLKALAEAEETRGTLTAVDGAISLAIANAWALDEAEHEGQPF 65 LFD P+R AGRHE+ + +A+E A+ G + +D +SLA ANAWALDEAE G P+ Sbjct 11 LFDVSDIPRRPAGRHEQQLDRAIEHAQQSGAVDELDAGLVSLARANAWALDEAEANGPLY 70</p> <p>Query 66 AVAQITAPFIELMKEIGLIQSEVST-DDDKLALALQELSATD 106 AVA +T PF E++ E+GL + +T +D+ L AL +L A D Sbjct 71 AVANLTPPFREVLAEGLTPAARNANDNALNDALNKLLADD 112</p> </div>	Score	Expect	Method	Identities	Positives	Gaps	89.0 bits(219)	6e-20	Compositional matrix adjust.	48/102(47%)	67/102(65%)	1/102(0%)	Score	Expect	Method	Identities	Positives	Gaps	88.6 bits(218)	1e-19	Compositional matrix adjust.	48/102(47%)	67/102(65%)	1/102(0%)																																																
Score	Expect	Method	Identities	Positives	Gaps																																																																				
89.0 bits(219)	6e-20	Compositional matrix adjust.	48/102(47%)	67/102(65%)	1/102(0%)																																																																				
Score	Expect	Method	Identities	Positives	Gaps																																																																				
88.6 bits(218)	1e-19	Compositional matrix adjust.	48/102(47%)	67/102(65%)	1/102(0%)																																																																				

Distribution of the top 59 Blast Hits on 59 subject sequences



Phage DB :

```
>Schiebs_1, terminase small subunit, 135
Length = 135

Score = 58.9 bits (141), Expect = 4e-09
Identities = 33/100 (33%), Positives = 50/100 (50%), Gaps = 4/100 (4%)

Query: 8  DPPQPPQRGAGRHEKMRKALEEAETRGLTAVDGAISLAIANAWALDEAHEGQFFAV 67
DP PP GRH V +AL A G + VD A +++ A AW+LD E + QP+
Sbjct: 20  DPATPPP---GRHSAVERALGAARDAGLVDDVDEALLTIVTAGAWSLDTFERQNPYGP 76

Query: 68  AQITAPPIELMKEIGLIQSEVSTD--DDKLALALQELSATD 106
A++ P + ++E L T DD + L +L++ D
Sbjct: 77  AKLIEPMVNALREARLTPDARQTSVDDSIKELGLDASAD 116

>SallySpecial_1, terminase small subunit, 131
Length = 131

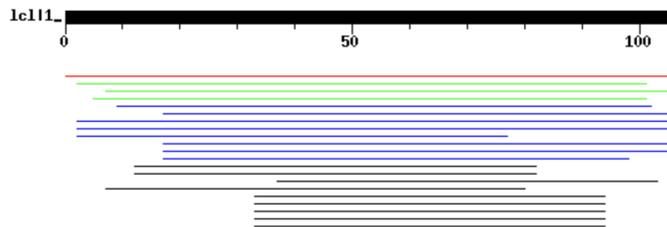
Score = 58.9 bits (141), Expect = 4e-09
Identities = 36/104 (34%), Positives = 56/104 (53%), Gaps = 5/104 (4%)

Query: 3  QLRLFPPOP---PQRGAGRHEKMRKALEEAETRGLTAVDGAISLAIANAWALDEAE 59
Q +FD +P P R GRH + +AL A+ +T VD A ++ A+AWALD E
Sbjct: 13  QAAMFDAGEPVATPPRH-GRHSEATERALAAAKAADLITDVEALAAVVRASAWALDRFE 71
```

Distribution of 21 Blast Hits on the Query Sequence

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

Color Key for Alignment Scores



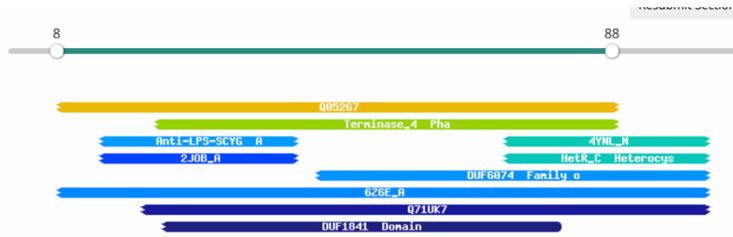
DECISION:

ATG 77 : meilleur score et ORF la plus longue .

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>(q6: s11) :</p> <p>>Schiebs_1, terminase small subunit, 135 Length = 135</p> <p>Score = 58.9 bits (141), Expect = 4e-09 Identities = 33/100 (33%), Positives = 50/100 (50%), Gaps = 4/100 (4%)</p> <p>Query: 8 DPPQPPQRGAGRHEKMRKALEEAETRGLTAVDGAISLAIANAWALDEAEHEGQPPFAV 67 DP PP GRH V +AL A G + VD A +++ A AW+LD E + QP+</p> <p>Sbjct: 20 DPATPPP---GRHSAVERALGAARDAGLVDDVEALLTIVTAGAWSLDTFRQNPYPG 76</p> <p>Query: 68 AQITAPFIELMKEIGLIQSEVSTD-DDKLALALQELSATD 106 A++ P + ++E L T DD + L +L++ D</p> <p>Sbjct: 77 AKLIEPMVNALREARLTPDARQTSVDDSIKELLDLASAD 116</p> <p>*** nr :</p> <p>hypothetical protein A5774_06175 [Corynebacterium sp. EPI-003-04-2554_SCH2473622] Sequence ID: OBAS4007.1 Length: 124 Number of Matches: 1</p> <p>Range 1: 11 to 112 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>89.0 bits(219)</td> <td>6e-20</td> <td>Compositional matrix adjust.</td> <td>48/102(47%)</td> <td>67/102(65%)</td> <td>1/102(0%)</td> </tr> </tbody> </table> <p>Query 6 LFDPPQPPQRGAGRHEKMRKALEEAETRGLTAVDGAISLAIANAWALDEAEHEGQPP 65 Sbjct 11 LFDVADIPRRPAGRHEQQLDRAIEHAQQSGAVDTLDAGLVSLARANAWALDEAEANGLPY 70</p> <p>Query 66 AVAQITAPFIELMKEIGLIQSEVSTD-DDKLALALQELSATD 106 Sbjct 71 AVANLTPPFREVLAEGLTPAARNTANDNALDNLKLADD 112</p> <table border="1"> <thead> <tr> <th>Accession</th> <th>Description</th> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>WP_258182534.1</td> <td>HNH endonuclease signature motif containing protein [Corynebacterium diptheriae]</td> <td>74.3</td> <td>74.3</td> <td>78%</td> <td>9e-14</td> <td>44.56%</td> <td>171</td> </tr> <tr> <td>WP_070736130.1</td> <td>hypothetical protein [unclassified Corynebacterium]</td> <td>70.9</td> <td>70.9</td> <td>92%</td> <td>4e-13</td> <td>40.40%</td> <td>103</td> </tr> <tr> <td>K087263.1</td> <td>Corynebacterium ulcerans</td> <td>70.9</td> <td>70.9</td> <td>79%</td> <td>7e-13</td> <td>43.53%</td> <td>108</td> </tr> <tr> <td>CAB0870127.1</td> <td>Corynebacterium diptheriae</td> <td>69.7</td> <td>69.7</td> <td>96%</td> <td>2e-12</td> <td>40.78%</td> <td>103</td> </tr> <tr> <td>CAB0618015.1</td> <td>Corynebacterium diptheriae</td> <td>69.3</td> <td>69.3</td> <td>96%</td> <td>2e-12</td> <td>39.81%</td> <td>103</td> </tr> <tr> <td>WP_070420691.1</td> <td>Corynebacterium</td> <td>69.3</td> <td>69.3</td> <td>91%</td> <td>2e-12</td> <td>41.84%</td> <td>106</td> </tr> <tr> <td>WP_014318089.1</td> <td>Corynebacterium diptheriae</td> <td>69.3</td> <td>69.3</td> <td>96%</td> <td>2e-12</td> <td>40.78%</td> <td>103</td> </tr> <tr> <td>WP_259344571.1</td> <td>Corynebacterium diptheriae</td> <td>69.7</td> <td>69.7</td> <td>96%</td> <td>2e-12</td> <td>40.78%</td> <td>115</td> </tr> <tr> <td>WP_106361648.1</td> <td>Corynebacterium diptheriae</td> <td>69.3</td> <td>69.3</td> <td>96%</td> <td>2e-12</td> <td>40.78%</td> <td>103</td> </tr> <tr> <td>WP_134884353.1</td> <td>Corynebacterium diptheriae</td> <td>68.9</td> <td>68.9</td> <td>96%</td> <td>2e-12</td> <td>40.78%</td> <td>103</td> </tr> <tr> <td>WP_197692073.1</td> <td>Corynebacterium diptheriae</td> <td>68.2</td> <td>68.2</td> <td>96%</td> <td>6e-12</td> <td>39.81%</td> <td>103</td> </tr> <tr> <td>WP_234917901.1</td> <td>Corynebacterium belfanti</td> <td>68.2</td> <td>68.2</td> <td>96%</td> <td>7e-12</td> <td>39.81%</td> <td>115</td> </tr> <tr> <td>CAB0524009.1</td> <td>Corynebacterium diptheriae</td> <td>67.8</td> <td>67.8</td> <td>96%</td> <td>7e-12</td> <td>39.81%</td> <td>103</td> </tr> <tr> <td>WP_187974182.1</td> <td>Corynebacterium poyangense</td> <td>67.8</td> <td>67.8</td> <td>91%</td> <td>9e-12</td> <td>41.00%</td> <td>113</td> </tr> <tr> <td>CAB0871381.1</td> <td>Corynebacterium diptheriae</td> <td>66.6</td> <td>66.6</td> <td>96%</td> <td>2e-11</td> <td>40.78%</td> <td>103</td> </tr> <tr> <td>WP_337890923.1</td> <td>Corynebacterium mastitidis</td> <td>64.7</td> <td>64.7</td> <td>85%</td> <td>1e-10</td> <td>39.13%</td> <td>108</td> </tr> <tr> <td>WP_018119280.1</td> <td>Corynebacterium mastitidis</td> <td>64.3</td> <td>64.3</td> <td>85%</td> <td>2e-10</td> <td>39.13%</td> <td>108</td> </tr> <tr> <td>CAB0659549.1</td> <td>Corynebacterium diptheriae</td> <td>63.9</td> <td>63.9</td> <td>81%</td> <td>2e-10</td> <td>41.86%</td> <td>103</td> </tr> <tr> <td>WP_284829931.1</td> <td>Corynebacterium mastitidis</td> <td>62.8</td> <td>62.8</td> <td>85%</td> <td>7e-10</td> <td>39.13%</td> <td>108</td> </tr> <tr> <td>WP_055121355.1</td> <td>Corynebacterium oculi</td> <td>62.4</td> <td>62.4</td> <td>85%</td> <td>1e-09</td> <td>39.13%</td> <td>108</td> </tr> <tr> <td>WP_302527865.1</td> <td>Corynebacterium accolens</td> <td>62.4</td> <td>62.4</td> <td>83%</td> <td>1e-09</td> <td>40.66%</td> <td>107</td> </tr> <tr> <td>CAB0612630.1</td> <td>Corynebacterium diptheriae</td> <td>59.7</td> <td>59.7</td> <td>96%</td> <td>1e-08</td> <td>36.89%</td> <td>103</td> </tr> <tr> <td>CAB0931246.1</td> <td>Corynebacterium diptheriae</td> <td>57.4</td> <td>57.4</td> <td>96%</td> <td>1e-07</td> <td>35.92%</td> <td>103</td> </tr> <tr> <td>CAB0963973.1</td> <td>Corynebacterium diptheriae</td> <td>56.6</td> <td>56.6</td> <td>96%</td> <td>2e-07</td> <td>35.92%</td> <td>103</td> </tr> <tr> <td>CAB0734388.1</td> <td>Corynebacterium diptheriae</td> <td>56.6</td> <td>56.6</td> <td>96%</td> <td>2e-07</td> <td>35.92%</td> <td>103</td> </tr> <tr> <td>CAB0884068.1</td> <td>Corynebacterium diptheriae</td> <td>55.5</td> <td>55.5</td> <td>96%</td> <td>5e-07</td> <td>34.95%</td> <td>103</td> </tr> <tr> <td>WP_253792491.1</td> <td>Nocardia amikacinitolerans</td> <td>55.8</td> <td>55.8</td> <td>83%</td> <td>6e-07</td> <td>35.56%</td> <td>133</td> </tr> <tr> <td>CAB1005534.1</td> <td>Corynebacterium diptheriae</td> <td>53.1</td> <td>53.1</td> <td>75%</td> <td>3e-06</td> <td>38.27%</td> <td>83</td> </tr> </tbody> </table>	Score	Expect	Method	Identities	Positives	Gaps	89.0 bits(219)	6e-20	Compositional matrix adjust.	48/102(47%)	67/102(65%)	1/102(0%)	Accession	Description	Score	Expect	Method	Identities	Positives	Gaps	WP_258182534.1	HNH endonuclease signature motif containing protein [Corynebacterium diptheriae]	74.3	74.3	78%	9e-14	44.56%	171	WP_070736130.1	hypothetical protein [unclassified Corynebacterium]	70.9	70.9	92%	4e-13	40.40%	103	K087263.1	Corynebacterium ulcerans	70.9	70.9	79%	7e-13	43.53%	108	CAB0870127.1	Corynebacterium diptheriae	69.7	69.7	96%	2e-12	40.78%	103	CAB0618015.1	Corynebacterium diptheriae	69.3	69.3	96%	2e-12	39.81%	103	WP_070420691.1	Corynebacterium	69.3	69.3	91%	2e-12	41.84%	106	WP_014318089.1	Corynebacterium diptheriae	69.3	69.3	96%	2e-12	40.78%	103	WP_259344571.1	Corynebacterium diptheriae	69.7	69.7	96%	2e-12	40.78%	115	WP_106361648.1	Corynebacterium diptheriae	69.3	69.3	96%	2e-12	40.78%	103	WP_134884353.1	Corynebacterium diptheriae	68.9	68.9	96%	2e-12	40.78%	103	WP_197692073.1	Corynebacterium diptheriae	68.2	68.2	96%	6e-12	39.81%	103	WP_234917901.1	Corynebacterium belfanti	68.2	68.2	96%	7e-12	39.81%	115	CAB0524009.1	Corynebacterium diptheriae	67.8	67.8	96%	7e-12	39.81%	103	WP_187974182.1	Corynebacterium poyangense	67.8	67.8	91%	9e-12	41.00%	113	CAB0871381.1	Corynebacterium diptheriae	66.6	66.6	96%	2e-11	40.78%	103	WP_337890923.1	Corynebacterium mastitidis	64.7	64.7	85%	1e-10	39.13%	108	WP_018119280.1	Corynebacterium mastitidis	64.3	64.3	85%	2e-10	39.13%	108	CAB0659549.1	Corynebacterium diptheriae	63.9	63.9	81%	2e-10	41.86%	103	WP_284829931.1	Corynebacterium mastitidis	62.8	62.8	85%	7e-10	39.13%	108	WP_055121355.1	Corynebacterium oculi	62.4	62.4	85%	1e-09	39.13%	108	WP_302527865.1	Corynebacterium accolens	62.4	62.4	83%	1e-09	40.66%	107	CAB0612630.1	Corynebacterium diptheriae	59.7	59.7	96%	1e-08	36.89%	103	CAB0931246.1	Corynebacterium diptheriae	57.4	57.4	96%	1e-07	35.92%	103	CAB0963973.1	Corynebacterium diptheriae	56.6	56.6	96%	2e-07	35.92%	103	CAB0734388.1	Corynebacterium diptheriae	56.6	56.6	96%	2e-07	35.92%	103	CAB0884068.1	Corynebacterium diptheriae	55.5	55.5	96%	5e-07	34.95%	103	WP_253792491.1	Nocardia amikacinitolerans	55.8	55.8	83%	6e-07	35.56%	133	CAB1005534.1	Corynebacterium diptheriae	53.1	53.1	75%	3e-06	38.27%	83
Score	Expect	Method	Identities	Positives	Gaps																																																																																																																																																																																																																																																
89.0 bits(219)	6e-20	Compositional matrix adjust.	48/102(47%)	67/102(65%)	1/102(0%)																																																																																																																																																																																																																																																
Accession	Description	Score	Expect	Method	Identities	Positives	Gaps																																																																																																																																																																																																																																														
WP_258182534.1	HNH endonuclease signature motif containing protein [Corynebacterium diptheriae]	74.3	74.3	78%	9e-14	44.56%	171																																																																																																																																																																																																																																														
WP_070736130.1	hypothetical protein [unclassified Corynebacterium]	70.9	70.9	92%	4e-13	40.40%	103																																																																																																																																																																																																																																														
K087263.1	Corynebacterium ulcerans	70.9	70.9	79%	7e-13	43.53%	108																																																																																																																																																																																																																																														
CAB0870127.1	Corynebacterium diptheriae	69.7	69.7	96%	2e-12	40.78%	103																																																																																																																																																																																																																																														
CAB0618015.1	Corynebacterium diptheriae	69.3	69.3	96%	2e-12	39.81%	103																																																																																																																																																																																																																																														
WP_070420691.1	Corynebacterium	69.3	69.3	91%	2e-12	41.84%	106																																																																																																																																																																																																																																														
WP_014318089.1	Corynebacterium diptheriae	69.3	69.3	96%	2e-12	40.78%	103																																																																																																																																																																																																																																														
WP_259344571.1	Corynebacterium diptheriae	69.7	69.7	96%	2e-12	40.78%	115																																																																																																																																																																																																																																														
WP_106361648.1	Corynebacterium diptheriae	69.3	69.3	96%	2e-12	40.78%	103																																																																																																																																																																																																																																														
WP_134884353.1	Corynebacterium diptheriae	68.9	68.9	96%	2e-12	40.78%	103																																																																																																																																																																																																																																														
WP_197692073.1	Corynebacterium diptheriae	68.2	68.2	96%	6e-12	39.81%	103																																																																																																																																																																																																																																														
WP_234917901.1	Corynebacterium belfanti	68.2	68.2	96%	7e-12	39.81%	115																																																																																																																																																																																																																																														
CAB0524009.1	Corynebacterium diptheriae	67.8	67.8	96%	7e-12	39.81%	103																																																																																																																																																																																																																																														
WP_187974182.1	Corynebacterium poyangense	67.8	67.8	91%	9e-12	41.00%	113																																																																																																																																																																																																																																														
CAB0871381.1	Corynebacterium diptheriae	66.6	66.6	96%	2e-11	40.78%	103																																																																																																																																																																																																																																														
WP_337890923.1	Corynebacterium mastitidis	64.7	64.7	85%	1e-10	39.13%	108																																																																																																																																																																																																																																														
WP_018119280.1	Corynebacterium mastitidis	64.3	64.3	85%	2e-10	39.13%	108																																																																																																																																																																																																																																														
CAB0659549.1	Corynebacterium diptheriae	63.9	63.9	81%	2e-10	41.86%	103																																																																																																																																																																																																																																														
WP_284829931.1	Corynebacterium mastitidis	62.8	62.8	85%	7e-10	39.13%	108																																																																																																																																																																																																																																														
WP_055121355.1	Corynebacterium oculi	62.4	62.4	85%	1e-09	39.13%	108																																																																																																																																																																																																																																														
WP_302527865.1	Corynebacterium accolens	62.4	62.4	83%	1e-09	40.66%	107																																																																																																																																																																																																																																														
CAB0612630.1	Corynebacterium diptheriae	59.7	59.7	96%	1e-08	36.89%	103																																																																																																																																																																																																																																														
CAB0931246.1	Corynebacterium diptheriae	57.4	57.4	96%	1e-07	35.92%	103																																																																																																																																																																																																																																														
CAB0963973.1	Corynebacterium diptheriae	56.6	56.6	96%	2e-07	35.92%	103																																																																																																																																																																																																																																														
CAB0734388.1	Corynebacterium diptheriae	56.6	56.6	96%	2e-07	35.92%	103																																																																																																																																																																																																																																														
CAB0884068.1	Corynebacterium diptheriae	55.5	55.5	96%	5e-07	34.95%	103																																																																																																																																																																																																																																														
WP_253792491.1	Nocardia amikacinitolerans	55.8	55.8	83%	6e-07	35.56%	133																																																																																																																																																																																																																																														
CAB1005534.1	Corynebacterium diptheriae	53.1	53.1	75%	3e-06	38.27%	83																																																																																																																																																																																																																																														
<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</p>	<p>"Pas de correspondance pertinente", les probabilités sont toutes inférieures à 90%.</p>																																																																																																																																																																																																																																																				

d'un crible HHPred avec une proba $\geq 90\%$ et une couverture acceptable ?



Hitlist

Show 25 Entries

Search:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	1	Q05267 VG05_BPML5 Gene 5 protein OS=Mycobacterium phage L5 OX=31757 GN=5 PE=4 SV=1	83.94	24	26.51	7.5	78	155
<input type="checkbox"/>	2	PF05119.16 ; Terminase_4 ; Phage terminase, small subunit	78.91	32	23.14	8.4	67	108
<input type="checkbox"/>	3	4YNL_N Heterocyst differentiation control protein; Heterocyst differentiation, Transcription factor, TRANSCRIPTION; 2.1A (Nosto	59.47	2.9	29.58	-2	30	90
<input type="checkbox"/>	4	PF18460.5 ; HetR_C ; Heterocyst differentiation regulator C-terminal Hood domain	59.08	3	28.85	-1.9	30	79

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 ?

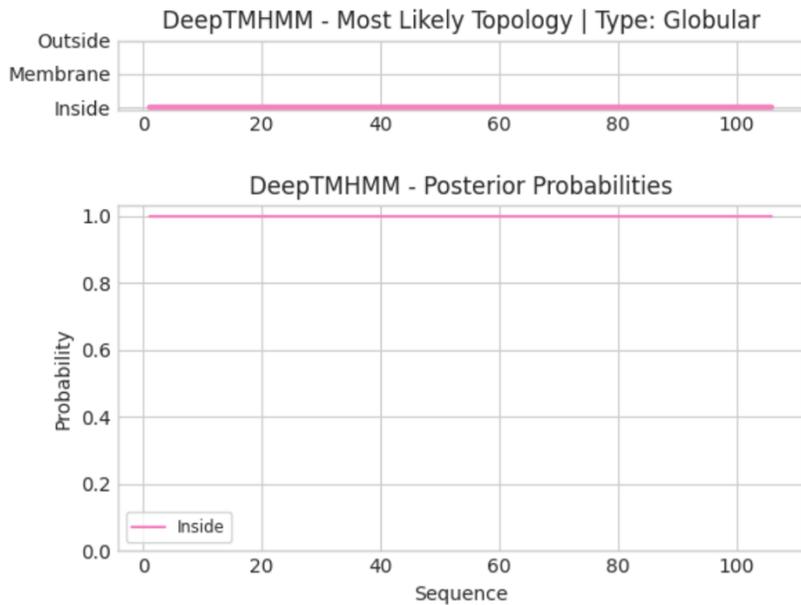
Oui, le gène suivant code pour une « terminase large subunit »

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

NON

DeepTMHMM - Predictions

Predicted topologies can be downloaded in [.gff3 format](#) and [.3line format](#)



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

OUI

DECISION:

terminase, small subunit

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	2
Coordonnées du Stop	1808
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove de 47

Coordonnées du Start retenu	351
Fonction prédite	Large terminase subunit

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>Il y a un potentiel codant prédit par la carte de Genemark</p> <p>The figure is a Genemark gene prediction plot. It displays two main sections: 'Direct Sequence' and 'Complementary Sequence'. Each section contains multiple tracks showing sequence alignment and predicted coding regions. A prominent thick black bar is drawn across the top of the 'Direct Sequence' section, indicating a predicted coding region. The x-axis represents 'Nucleotide Position' from 0 to 2000, and the y-axis represents sequence alignment scores from 0.0 to 1.0.</p>

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

Oui

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

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"/> terminase [unclassified Corynebacterium]	unclassified Corynebacterium	542	542	96%	0.0	61.41%	489	WP_246817218.1
<input checked="" type="checkbox"/> putative chape terminase protein [Corynebacterium striatum]	Corynebacterium striatum	541	541	96%	0.0	61.41%	489	CCD13955.1
<input checked="" type="checkbox"/> terminase large subunit [Corynebacterium]	Corynebacterium	521	521	93%	2e-179	58.33%	451	WP_168161724.1
<input checked="" type="checkbox"/> TPA_terminase [Corynebacterium striatum]	Corynebacterium striatum	521	521	93%	2e-179	58.33%	451	HCD1553127.1
<input checked="" type="checkbox"/> TPA_terminase [Corynebacterium striatum]	Corynebacterium striatum	520	520	91%	4e-179	62.02%	443	HCT5225604.1
<input checked="" type="checkbox"/> terminase [Corynebacterium sp. HMSC065A05]	Corynebacterium sp. HMSC065A05	519	519	91%	9e-179	61.80%	443	QFP20321.1
<input checked="" type="checkbox"/> terminase [Corynebacterium diophtheriae]	Corynebacterium diophtheriae	499	499	96%	1e-170	56.50%	480	CAB0972035.1
<input checked="" type="checkbox"/> terminase [Corynebacterium diophtheriae]	Corynebacterium diophtheriae	494	494	96%	3e-168	56.72%	480	CAB0871369.1
<input checked="" type="checkbox"/> terminase [Corynebacterium diophtheriae]	Corynebacterium diophtheriae	493	493	96%	4e-168	56.50%	480	CAB0734368.1
<input checked="" type="checkbox"/> terminase [Corynebacterium diophtheriae]	Corynebacterium diophtheriae	493	493	94%	5e-168	57.39%	480	CAB0564988.1
<input checked="" type="checkbox"/> terminase [Corynebacterium diophtheriae]	Corynebacterium diophtheriae	491	491	94%	5e-167	57.17%	481	WP_148263452.1
<input checked="" type="checkbox"/> terminase [Corynebacterium diophtheriae]	Corynebacterium diophtheriae	486	486	96%	4e-165	55.32%	490	CAB0565394.1

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:

Sequence	Score	E Value
CyranoPS_Draft_2, function unknown, 485	580	0.0
Empereor_2, terminase large subunit, 474	322	e-109
SallySpecial_2, terminase large subunit, 476	320	e-108
QMS_2, putative large terminase subunit, 464	343	1e-95
SpicDab_2, terminase, 439	342	1e-93
RRH1_02, terminase large subunit, 436	341	2e-93
Bahul_2, terminase large subunit, 466	317	3e-86
Comur_2, terminase large subunit, 444	305	2e-82
NOComagall_2, terminase large subunit, 461	293	1e-80
Jeanie_2, terminase large subunit, 461	293	1e-80
Schiebs_2, terminase large subunit, 445	226	1e-79
GRU3_2, putative large terminase subunit, 473	227	4e-74
SerialPhiller_2, terminase, 504	126	1e-43

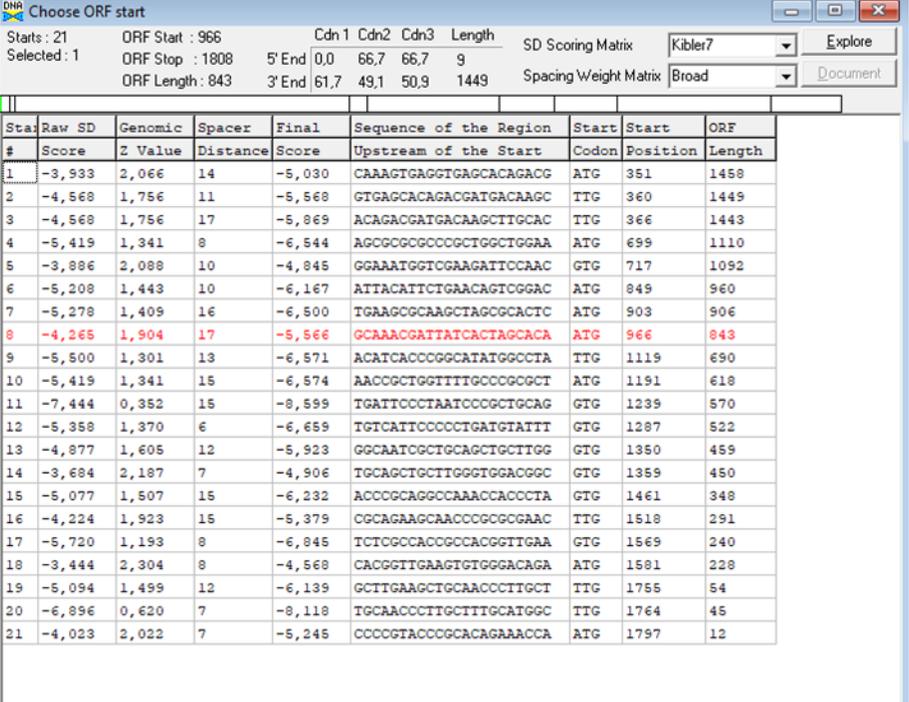
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

Suffisamment long, mais chevauchement de 47pb (en fonction du START)

DECISION:

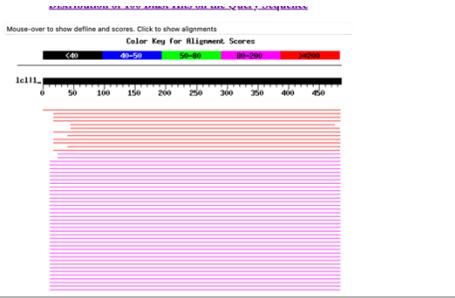
OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																															
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 351 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 366</p>																																																																																																																																																																																																															
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <p>DNA Choose ORF start</p> <p>Starts : 21 ORF Start : 966 Cdn1 Cdn2 Cdn3 Length Selected : 1 ORF Stop : 1808 5' End 0,0 66,7 66,7 9 SD Scoring Matrix Kibler7 Explore ORF Length : 843 3' End 61,7 49,1 50,9 1449 Spacing Weight Matrix Broad Document</p> <table border="1"> <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>-3,933</td><td>2,066</td><td>14</td><td>-5,030</td><td>CRAAGTGAAGGTGAGCCACAGACG</td><td>ATG</td><td>351</td><td>1458</td></tr> <tr><td>2</td><td>-4,568</td><td>1,756</td><td>11</td><td>-5,568</td><td>GTGAGCCACAGACGATGACAAAGC</td><td>TTG</td><td>360</td><td>1449</td></tr> <tr><td>3</td><td>-4,568</td><td>1,756</td><td>17</td><td>-5,869</td><td>ACAGACGATGACAAAGCTTGCAC</td><td>TTG</td><td>366</td><td>1443</td></tr> <tr><td>4</td><td>-5,419</td><td>1,341</td><td>8</td><td>-6,544</td><td>AGCGCGCGCCCGCTGGCTGGAA</td><td>ATG</td><td>699</td><td>1110</td></tr> <tr><td>5</td><td>-3,886</td><td>2,088</td><td>10</td><td>-4,845</td><td>GGAAATGGTCGAAGATTCCAAC</td><td>GTG</td><td>717</td><td>1092</td></tr> <tr><td>6</td><td>-5,208</td><td>1,443</td><td>10</td><td>-6,167</td><td>ATTACATTCGACAGCTGGGAC</td><td>ATG</td><td>849</td><td>960</td></tr> <tr><td>7</td><td>-5,278</td><td>1,409</td><td>16</td><td>-6,500</td><td>TGAAGCGCAAGCTAGCGCACCTC</td><td>ATG</td><td>903</td><td>906</td></tr> <tr><td>8</td><td>-4,265</td><td>1,904</td><td>17</td><td>-5,566</td><td>GCAACGATTACTACTAGCACA</td><td>ATG</td><td>966</td><td>843</td></tr> <tr><td>9</td><td>-5,500</td><td>1,301</td><td>13</td><td>-6,571</td><td>ACATCACCCGGCATATGGCCTA</td><td>TTG</td><td>1119</td><td>690</td></tr> <tr><td>10</td><td>-5,419</td><td>1,341</td><td>15</td><td>-6,574</td><td>AAOCGCTGGTTTTGCCCGGCT</td><td>ATG</td><td>1191</td><td>618</td></tr> <tr><td>11</td><td>-7,444</td><td>0,352</td><td>15</td><td>-8,599</td><td>TGATTCCCTAATCCCGCTGCAG</td><td>GTG</td><td>1239</td><td>570</td></tr> <tr><td>12</td><td>-5,358</td><td>1,370</td><td>6</td><td>-6,659</td><td>TGTCATTCCTCCCTGATGATTT</td><td>GTG</td><td>1287</td><td>522</td></tr> <tr><td>13</td><td>-4,877</td><td>1,605</td><td>12</td><td>-5,923</td><td>GGCAATGGCTGCAGCTGCTTGG</td><td>GTG</td><td>1350</td><td>459</td></tr> <tr><td>14</td><td>-3,684</td><td>2,187</td><td>7</td><td>-4,906</td><td>TGCAGCTGCTTGGGTGGAGGCG</td><td>GTG</td><td>1359</td><td>450</td></tr> <tr><td>15</td><td>-5,077</td><td>1,507</td><td>15</td><td>-6,232</td><td>ACCCGCGAGCCAAACCCACCTA</td><td>GTG</td><td>1461</td><td>348</td></tr> <tr><td>16</td><td>-4,224</td><td>1,923</td><td>15</td><td>-5,379</td><td>CCGAGAAGCAACCCCGCGGAAC</td><td>TTG</td><td>1518</td><td>291</td></tr> <tr><td>17</td><td>-5,720</td><td>1,193</td><td>8</td><td>-6,845</td><td>TCTGGCCACCGCCACGGTTGAA</td><td>GTG</td><td>1569</td><td>240</td></tr> <tr><td>18</td><td>-3,444</td><td>2,304</td><td>8</td><td>-4,568</td><td>CACGGTTGAAGTGTGGGACAGA</td><td>ATG</td><td>1581</td><td>228</td></tr> <tr><td>19</td><td>-5,094</td><td>1,499</td><td>12</td><td>-6,139</td><td>GCTTGAAGCTGCACCCCTTGT</td><td>TTG</td><td>1755</td><td>54</td></tr> <tr><td>20</td><td>-6,896</td><td>0,620</td><td>7</td><td>-8,118</td><td>TGCAACCCCTTGGCTTTCATGGC</td><td>TTG</td><td>1764</td><td>45</td></tr> <tr><td>21</td><td>-4,023</td><td>2,022</td><td>7</td><td>-5,245</td><td>CCCCGTACCCGACAGAAACCA</td><td>ATG</td><td>1797</td><td>12</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	-3,933	2,066	14	-5,030	CRAAGTGAAGGTGAGCCACAGACG	ATG	351	1458	2	-4,568	1,756	11	-5,568	GTGAGCCACAGACGATGACAAAGC	TTG	360	1449	3	-4,568	1,756	17	-5,869	ACAGACGATGACAAAGCTTGCAC	TTG	366	1443	4	-5,419	1,341	8	-6,544	AGCGCGCGCCCGCTGGCTGGAA	ATG	699	1110	5	-3,886	2,088	10	-4,845	GGAAATGGTCGAAGATTCCAAC	GTG	717	1092	6	-5,208	1,443	10	-6,167	ATTACATTCGACAGCTGGGAC	ATG	849	960	7	-5,278	1,409	16	-6,500	TGAAGCGCAAGCTAGCGCACCTC	ATG	903	906	8	-4,265	1,904	17	-5,566	GCAACGATTACTACTAGCACA	ATG	966	843	9	-5,500	1,301	13	-6,571	ACATCACCCGGCATATGGCCTA	TTG	1119	690	10	-5,419	1,341	15	-6,574	AAOCGCTGGTTTTGCCCGGCT	ATG	1191	618	11	-7,444	0,352	15	-8,599	TGATTCCCTAATCCCGCTGCAG	GTG	1239	570	12	-5,358	1,370	6	-6,659	TGTCATTCCTCCCTGATGATTT	GTG	1287	522	13	-4,877	1,605	12	-5,923	GGCAATGGCTGCAGCTGCTTGG	GTG	1350	459	14	-3,684	2,187	7	-4,906	TGCAGCTGCTTGGGTGGAGGCG	GTG	1359	450	15	-5,077	1,507	15	-6,232	ACCCGCGAGCCAAACCCACCTA	GTG	1461	348	16	-4,224	1,923	15	-5,379	CCGAGAAGCAACCCCGCGGAAC	TTG	1518	291	17	-5,720	1,193	8	-6,845	TCTGGCCACCGCCACGGTTGAA	GTG	1569	240	18	-3,444	2,304	8	-4,568	CACGGTTGAAGTGTGGGACAGA	ATG	1581	228	19	-5,094	1,499	12	-6,139	GCTTGAAGCTGCACCCCTTGT	TTG	1755	54	20	-6,896	0,620	7	-8,118	TGCAACCCCTTGGCTTTCATGGC	TTG	1764	45	21	-4,023	2,022	7	-5,245	CCCCGTACCCGACAGAAACCA	ATG	1797	12
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	-3,933	2,066	14	-5,030	CRAAGTGAAGGTGAGCCACAGACG	ATG	351	1458																																																																																																																																																																																																								
2	-4,568	1,756	11	-5,568	GTGAGCCACAGACGATGACAAAGC	TTG	360	1449																																																																																																																																																																																																								
3	-4,568	1,756	17	-5,869	ACAGACGATGACAAAGCTTGCAC	TTG	366	1443																																																																																																																																																																																																								
4	-5,419	1,341	8	-6,544	AGCGCGCGCCCGCTGGCTGGAA	ATG	699	1110																																																																																																																																																																																																								
5	-3,886	2,088	10	-4,845	GGAAATGGTCGAAGATTCCAAC	GTG	717	1092																																																																																																																																																																																																								
6	-5,208	1,443	10	-6,167	ATTACATTCGACAGCTGGGAC	ATG	849	960																																																																																																																																																																																																								
7	-5,278	1,409	16	-6,500	TGAAGCGCAAGCTAGCGCACCTC	ATG	903	906																																																																																																																																																																																																								
8	-4,265	1,904	17	-5,566	GCAACGATTACTACTAGCACA	ATG	966	843																																																																																																																																																																																																								
9	-5,500	1,301	13	-6,571	ACATCACCCGGCATATGGCCTA	TTG	1119	690																																																																																																																																																																																																								
10	-5,419	1,341	15	-6,574	AAOCGCTGGTTTTGCCCGGCT	ATG	1191	618																																																																																																																																																																																																								
11	-7,444	0,352	15	-8,599	TGATTCCCTAATCCCGCTGCAG	GTG	1239	570																																																																																																																																																																																																								
12	-5,358	1,370	6	-6,659	TGTCATTCCTCCCTGATGATTT	GTG	1287	522																																																																																																																																																																																																								
13	-4,877	1,605	12	-5,923	GGCAATGGCTGCAGCTGCTTGG	GTG	1350	459																																																																																																																																																																																																								
14	-3,684	2,187	7	-4,906	TGCAGCTGCTTGGGTGGAGGCG	GTG	1359	450																																																																																																																																																																																																								
15	-5,077	1,507	15	-6,232	ACCCGCGAGCCAAACCCACCTA	GTG	1461	348																																																																																																																																																																																																								
16	-4,224	1,923	15	-5,379	CCGAGAAGCAACCCCGCGGAAC	TTG	1518	291																																																																																																																																																																																																								
17	-5,720	1,193	8	-6,845	TCTGGCCACCGCCACGGTTGAA	GTG	1569	240																																																																																																																																																																																																								
18	-3,444	2,304	8	-4,568	CACGGTTGAAGTGTGGGACAGA	ATG	1581	228																																																																																																																																																																																																								
19	-5,094	1,499	12	-6,139	GCTTGAAGCTGCACCCCTTGT	TTG	1755	54																																																																																																																																																																																																								
20	-6,896	0,620	7	-8,118	TGCAACCCCTTGGCTTTCATGGC	TTG	1764	45																																																																																																																																																																																																								
21	-4,023	2,022	7	-5,245	CCCCGTACCCGACAGAAACCA	ATG	1797	12																																																																																																																																																																																																								
<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>	<p>ORF la plus longue : 1458pb - ATG 351 mais chevauchement de 47pb, meilleur score TTG 366 - > ORF = 1443 pb, chevauchement de 32pb TTG 360 ORF = 1449 pb ; chevauchement de 38pb</p>																																																																																																																																																																																																															
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>---</p>																																																																																																																																																																																																															
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<p>PhageDB.</p> <p>>Emperor_2, terminase large subunit, 474 Length = 474</p> <p>Score = 392 bits (1008), Expect = e-109 Identities = 219/475 (46%), Positives = 298/475 (62%), Gaps = 14/475 (2%)</p> <p>Query: 18 PRFITEW-PDAPSVGSKVAKVQRSLGSEPLPWQVLAHAVGARNPDGTPRPFVFLVSVPR 76 PR+TT PD S G +A+ LG +PWQ AA +G + +G +P V+++V R</p> <p>Sbjct: 4 PRYLTPRNPDRRSFGPNARTSALLGRPFMPWQHAAAGVIGEVDAEGRFAYPLVIMTVQR 63</p>																																																																																																																																																																																																															

	<p>Nr. MULTISPECIES: terminase large subunit [Corynebacterium] Sequence ID: WP_168161724.1 Length: 451 Number of Matches: 1 See 2 more title(s) See all Identical Proteins(IPG)</p> <p>Range 1: 1 to 450 GenPept Graphics Next Match Previous Matc</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>521 bits(1343)</td> <td>2e-179</td> <td>Compositional matrix adjust.</td> <td>266/456(58%)</td> <td>337/456(73%)</td> <td>7/456(1%)</td> </tr> </tbody> </table> <pre> Query 30 VGSKVAKVQRSLGSEPLPWQVLAHAVGARNPDGTPRFPFVLVSVPQAGKTRASWSWLY 89 +GSK+ +V + LG P+PWQ LAAH +GAR P+G P++PF++++VPRQ+GKT + + Sbjct 1 MGSKIEQVSKYLGRTMPWQRLAAHIIGARLPNGRPKWPFFIVITVPRQSGKTTICSAVQF 60 </pre>	Score	Expect	Method	Identities	Positives	Gaps	521 bits(1343)	2e-179	Compositional matrix adjust.	266/456(58%)	337/456(73%)	7/456(1%)
Score	Expect	Method	Identities	Positives	Gaps								
521 bits(1343)	2e-179	Compositional matrix adjust.	266/456(58%)	337/456(73%)	7/456(1%)								
DECISION:	351 : meilleur ORF, score bon, ATG												

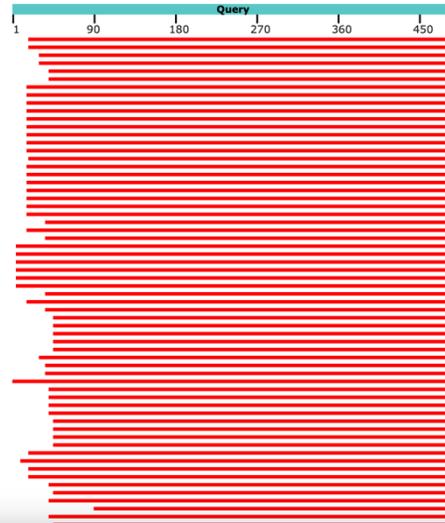
Décision #3 : Quelle est la fonction de la protéine putative ?

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 : (q#: s#) : [alignment] ; e-value :</p>  <p>Sequences producing significant alignments:</p> <table border="1"> <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>Cyran08_Draft_2, function unknown, 485</td> <td>280</td> <td>0.0</td> </tr> <tr> <td>Emperor_2, terminase large subunit, 474</td> <td>232</td> <td>e-109</td> </tr> <tr> <td>SallySpecial_2, terminase large subunit, 476</td> <td>230</td> <td>e-108</td> </tr> <tr> <td>OMAS_2, putative large terminase subunit, 464</td> <td>212</td> <td>1e-95</td> </tr> <tr> <td>Epic0ab_2, terminase, 429</td> <td>212</td> <td>1e-93</td> </tr> <tr> <td>NR81_02, terminase large subunit, 436</td> <td>211</td> <td>2e-93</td> </tr> <tr> <td>Rah1_2, terminase large subunit, 446</td> <td>211</td> <td>3e-86</td> </tr> <tr> <td>Cooc_2, terminase large subunit, 444</td> <td>205</td> <td>2e-82</td> </tr> <tr> <td>McGonagall_2, terminase large subunit, 461</td> <td>233</td> <td>1e-80</td> </tr> </tbody> </table> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p>		Score	E		(bits)	Value	Cyran08_Draft_2, function unknown, 485	280	0.0	Emperor_2, terminase large subunit, 474	232	e-109	SallySpecial_2, terminase large subunit, 476	230	e-108	OMAS_2, putative large terminase subunit, 464	212	1e-95	Epic0ab_2, terminase, 429	212	1e-93	NR81_02, terminase large subunit, 436	211	2e-93	Rah1_2, terminase large subunit, 446	211	3e-86	Cooc_2, terminase large subunit, 444	205	2e-82	McGonagall_2, terminase large subunit, 461	233	1e-80
	Score	E																																
	(bits)	Value																																
Cyran08_Draft_2, function unknown, 485	280	0.0																																
Emperor_2, terminase large subunit, 474	232	e-109																																
SallySpecial_2, terminase large subunit, 476	230	e-108																																
OMAS_2, putative large terminase subunit, 464	212	1e-95																																
Epic0ab_2, terminase, 429	212	1e-93																																
NR81_02, terminase large subunit, 436	211	2e-93																																
Rah1_2, terminase large subunit, 446	211	3e-86																																
Cooc_2, terminase large subunit, 444	205	2e-82																																
McGonagall_2, terminase large subunit, 461	233	1e-80																																

terminase [Corynebacterium diphtheriae]

No putative conserved domains have been detected

Distribution of the top 100 Blast Hits on 100 subject sequences



<input checked="" type="checkbox"/>	putative phage terminase protein [Corynebacterium striatum]	Corynebacterium striatum	541	541	96%	0.0	61.41%	489	CQD13855.1
<input checked="" type="checkbox"/>	terminase large subunit [Corynebacterium]	Corynebacterium	521	521	93%	2e-179	58.33%	451	WP_168161724.1
<input checked="" type="checkbox"/>	TPA_terminase [Corynebacterium striatum]	Corynebacterium striatum	521	521	93%	2e-179	58.33%	451	HCD1553127.1
<input checked="" type="checkbox"/>	TPA_terminase [Corynebacterium striatum]	Corynebacterium striatum	520	520	91%	4e-179	62.02%	443	HCT5225804.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium sp. HMSC065A05]	Corynebacterium sp. HMSC065A05	519	519	91%	9e-179	61.80%	443	QFP20321.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	499	499	96%	1e-170	56.50%	480	CAB0972035.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	494	494	96%	3e-168	56.72%	480	CAB0871369.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	493	493	96%	4e-168	56.50%	480	CAB0734368.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	493	493	94%	5e-168	57.39%	480	CAB0564988.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	491	491	94%	5e-167	57.17%	481	WP_148263452.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	486	486	96%	4e-165	55.32%	490	CAB0565394.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	483	483	94%	4e-164	58.04%	481	CAB0659529.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	483	483	96%	5e-164	56.93%	480	CAB0919103.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	482	482	96%	1e-163	56.50%	480	CAB0733954.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	481	481	96%	2e-163	56.84%	480	CAB0519177.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	481	481	96%	2e-163	56.72%	480	WP_241008567.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	481	481	96%	3e-163	56.72%	480	CAB0913794.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	480	480	96%	5e-163	56.93%	480	WP_342351212.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	480	480	96%	8e-163	56.29%	480	CAB0519844.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	480	480	96%	9e-163	56.29%	480	CAB1005517.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	479	479	96%	1e-162	56.72%	481	CAB0884063.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	478	478	96%	5e-162	56.72%	481	CAB0612854.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	474	474	92%	5e-161	56.70%	445	RLP10252.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	474	474	96%	2e-160	56.08%	480	CAB0567705.1
<input checked="" type="checkbox"/>	hypothetical protein CDBH8_1793 [Corynebacterium diphtheriae BH8]	Corynebacterium diphtheriae BH8	470	470	92%	2e-159	56.25%	445	AEX49311.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	468	468	98%	6e-158	54.98%	481	WP_258174498.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	468	468	98%	8e-158	54.98%	485	PSA85850.1
<input checked="" type="checkbox"/>	terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	466	466	98%	2e-157	54.77%	481	WP_226813674.1

MULTISPECIES: terminase [unclassified Corynebacterium]

Sequence ID: [WP_246817218.1](#) Length: 489 Number of Matches: 1

Range 1: 26 to 488 [GenPept](#) [Graphics](#)

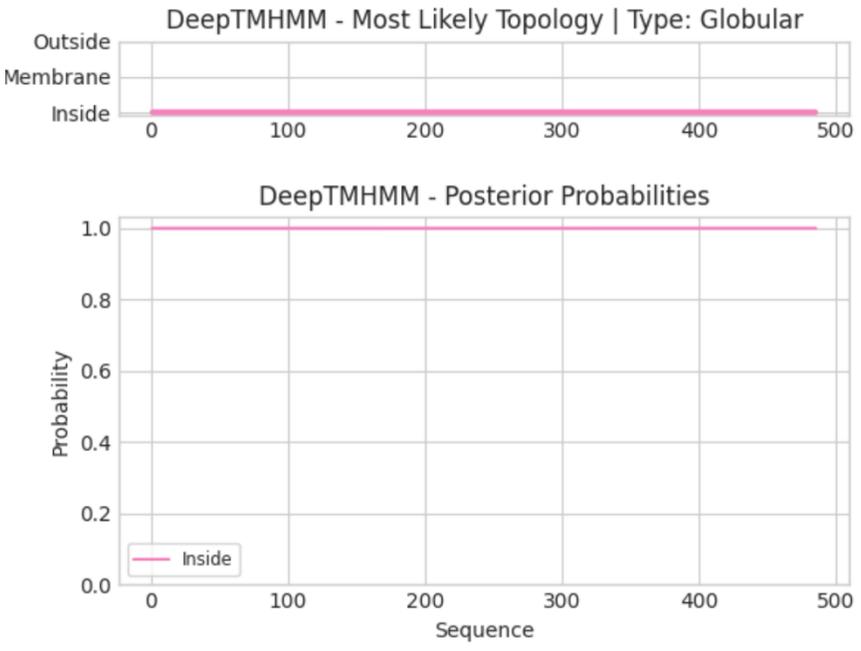
[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
542 bits(1396)	0.0	Compositional matrix adjust.	288/469(61%)	344/469(73%)	8/469(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 $\geq 90\%$ et une couverture acceptable ?

26 479

Vis	Hits	Aln	Select All	Forward	Forward Query A3M	Model using selection	Download HHR	Color Seqs	Wrap Seqs	
<input type="checkbox"/>	1	6Z6D_A		Terminase large subunit; genome packaging, bacteriophage, ATPase, nuclease, VIRAL PROTEIN; HET: BR; 2.2A {Enterobacteria	100	3.6e-39	315.48	48.8	429	514
<input type="checkbox"/>	2	Q05219		VG13_BPML5 Gene 13 protein OS=Mycobacterium phage L5 OX=31757 GN=13 PE=3 SV=1	100	1.6e-38	315.62	50.1	417	593
<input type="checkbox"/>	3	P59217		TERL_BPSF5 Putative terminase large subunit OS=Shigella phage SFV OX=55884 GN=2 PE=3 SV=1	100	2e-38	314.1	45.9	428	577
<input type="checkbox"/>	4	O21870		TERL_BPLSK Terminase large subunit OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1	100	6.6e-38	308.53	44.9	434	540
<input type="checkbox"/>	5	A7XXB7		TERL_BP234 Terminase, large subunit OS=Thermus virus P23-45 OX=466051 GN=P23p85 PE=3 SV=1	100	2.9e-35	286.01	42.4	400	485
<input type="checkbox"/>	6	P27753		TERL_BPP1 Probable terminase, large subunit OS=Escherichia phage P1 OX=10678 GN=pacB PE=3 SV=1	100	4.8e-35	285.12	37.2	410	494
<input type="checkbox"/>	7	Q9T1W6		TERL_BPMU Probable terminase, large subunit gp28 OS=Escherichia phage Mu OX=10677 GN=Mup28 PE=2 SV=1	100	6e-35	288.44	36.7	443	551
<input type="checkbox"/>	8	5OE8_A		Large subunit terminase; large terminase, VIRAL PROTEIN; 2.2A {Deep-sea thermophilic phage D6E}	100	5.6e-34	272.58	37.1	393	430
<input type="checkbox"/>	9	Q6QGD2		TERL_BPT5 Terminase, large subunit OS=Escherichia phage T5 OX=10726 GN=ORF144 PE=1 SV=1	100	2.1e-33	269.31	40.1	388	438
<input type="checkbox"/>	10	P17312		TERL_BPT4 Terminase, large subunit OS=Enterobacteria phage T4 OX=10665	100	3.3e-33	278.83	41.8	406	610

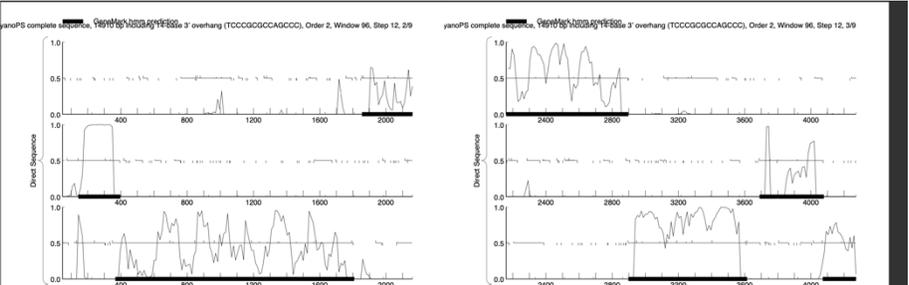
	<p>PDB : Terminase large subunit; genome packaging, bacteriophage, ATPase, nuclease, VIRAL PROTEIN; HET: BR; 2.2A {Enterobacteria phage HK97}</p> <p>UNIPROT : Q05219 VG13_BPML5 Gene 13 protein</p> <p>OS=Mycobacterium phage L5 OX=31757 GN=13 PE=3 SV=1</p> <p>Pfam : TerL_nuclease ; Terminase large subunit, endonuclease domain</p> <p>CD : 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.</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>Le gène avant code une terminase small subunit. Et le gene apres code une « portal protein »</p>
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	
<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>Terminase, large subunit</p>

Student Gene Annotation Worksheet

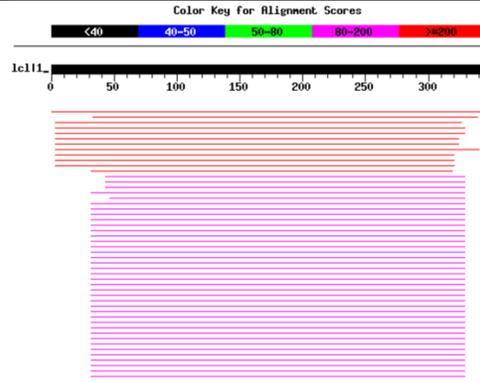
Basic Phage Information

Nom du Phage	CyranoPS
Gène #	3
Coordonnées du Stop	2898
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 55
Coordonnées du Start retenu	1864
Fonction prédite	portal 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 montre un potentiel codant :</p> 

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



Sequences producing significant alignments:

	Score (bits)	E Value
CyranoPS_Draft_3, function unknown, 347	679	0.0
EpicDab_4, portal protein, 358	257	3e-68
Schiebs_4, portal protein, 347	235	1e-61
Rahul_4, portal protein, 368	231	2e-60
Coeur_4, portal protein, 368	230	3e-60
Emperor_6, portal protein, 385	227	4e-59
RRH1_04, portal protein, 372	223	4e-58
SallySpecial_4, portal protein, 386	219	6e-57
GMA5_4, putative portal protein, 363	208	1e-53
McGonagall_4, portal protein, 364	207	3e-53
Jeanie_4, portal protein, 364	207	3e-53
GRU3_4, putative portal protein, 358	202	8e-52
SerialPhiller_4, portal protein, 381	158	2e-38
Kels_4, portal protein, 381	158	2e-38

Download GenPept Graphics Next Previous Descriptions

phage portal protein [Corynebacterium sp. HMSC06C06]

Sequence ID: WP_070420693.1 Length: 368 Number of Matches: 1

See 1 more title(s) See all Identical Proteins (IPG)

Range 1: 1 to 337 GenPept Graphics

Next Match Previous Match

Related Information
Identical Proteins - Identical proteins to WP_070420693.1

Score	Expect	Method	Identities	Positives	Gaps
371 bits(952)	1e-123	Compositional matrix adjust.	198/343(58%)	242/343(70%)	8/343(2%)
Query 4	MSLLDRIRKLGQLPQLLATGAPLSVWAGD-HLEQVTSPOVDYLPVTRSQAMAIPAVVRA	62			
Sbjct 1	M L+DR+R+L LPQ A S WA HL PD + + R AM IPA++RA	59			
Query 63	RTLIVTTVARCPVSDDEITPGWLHGIAPLGGHPQTEFHRLLNTADOLLFHGSAAWAE	122			
Sbjct 60	R L+VTT+AR P+ + + P W+ G P G P QTEF+R+L TADOL F+G +AWAIE	118			
Query 123	RDELGWTAAVHPRSLWERTADGTIIVDGVPAQAQEVAVFDGHHGILTHGARALRDAS	182			
Sbjct 119	RD G T A+HIP +LW +A+G +IVDG A EV F GIH GIL HGA ALRDA	178			
Query 183	NILSAAARVADTPAALTELRTNDAVMTREDIKALIQGYVDARRGKNGVSYSSGVEVH	242			
Sbjct 179	+ A+ARVAD PAAL ELRQTNDAAA ++I ALL YV AR+G+ GVSYSS GVEV	238			
Query 243	EHSLAPENLLIEGRNAAVDVARLLGVPAFIDATVGGTSLSYENASRMTLITFGVAP	302			
Sbjct 239	EHSLAPENLLIAGRNAATVDIARLANLPAPFIDATVGGTSLSYENASRMTLITAFGLSP	298			
Query 303	LAAISSRLNLSDLTPVG-TVSFDTAQIENISELLPTNMTN	344			
Sbjct 299	LAAI++RLNL+d+ P G + FDT +I+++I PT T N	337			

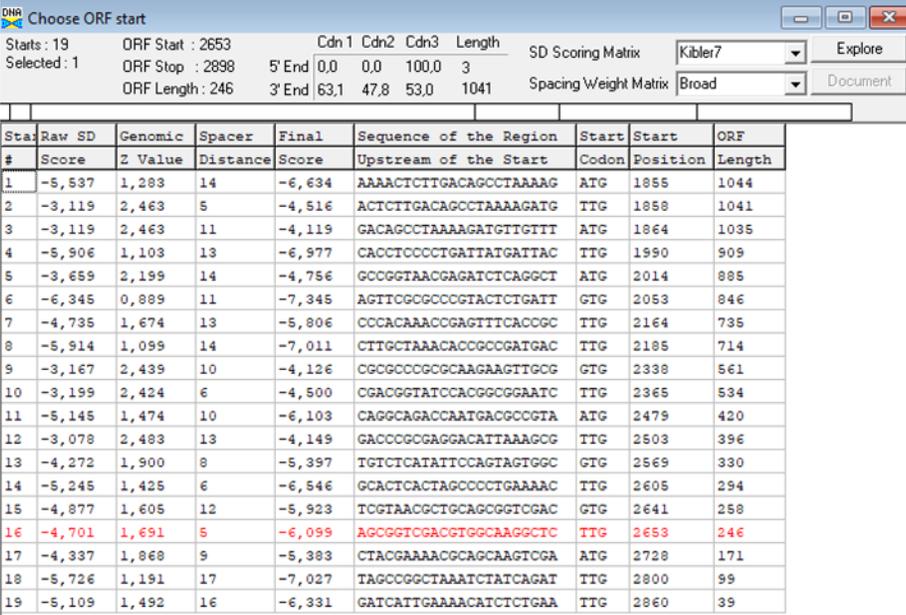
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

Ove > 30 bp

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																					
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 1855 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 1855</p>																																																																																																																																																																																					
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <p>DNA Choose ORF start</p> <p>Starts: 19 ORF Start: 2653 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler7 Explore Selected: 1 ORF Stop: 2898 5' End 0,0 0,0 100,0 3 Spacing Weight Matrix Broad Document ORF Length: 246 3' End 63,1 47,8 53,0 1041</p> <table border="1"> <thead> <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> </thead> <tbody> <tr><td>1</td><td>-5,537</td><td>1,283</td><td>14</td><td>-6,634</td><td>AAAACCTTTGACAGCCTAAAAG</td><td>ATG</td><td>1855</td><td>1044</td></tr> <tr><td>2</td><td>-3,119</td><td>2,463</td><td>5</td><td>-4,516</td><td>ACTCTTGACAGCCTAAAAGATG</td><td>TTG</td><td>1858</td><td>1041</td></tr> <tr><td>3</td><td>-3,119</td><td>2,463</td><td>11</td><td>-4,119</td><td>GACAGCCTAAAAGATGTTGTTT</td><td>ATG</td><td>1864</td><td>1035</td></tr> <tr><td>4</td><td>-5,906</td><td>1,103</td><td>13</td><td>-6,977</td><td>CACCTCCCTGATTATGATTAC</td><td>TTG</td><td>1950</td><td>909</td></tr> <tr><td>5</td><td>-3,659</td><td>2,199</td><td>14</td><td>-4,756</td><td>GCCGGTAACGAGATCTCAGGCT</td><td>ATG</td><td>2014</td><td>885</td></tr> <tr><td>6</td><td>-6,345</td><td>0,889</td><td>11</td><td>-7,345</td><td>AGTTCGGCCCGTACTCTGATT</td><td>GTG</td><td>2053</td><td>846</td></tr> <tr><td>7</td><td>-4,735</td><td>1,674</td><td>13</td><td>-5,806</td><td>CCCACAAACCGAGTTTACCCGC</td><td>TTG</td><td>2164</td><td>735</td></tr> <tr><td>8</td><td>-5,914</td><td>1,099</td><td>14</td><td>-7,011</td><td>CTTGTAAACACCCCGGATGAC</td><td>TTG</td><td>2185</td><td>714</td></tr> <tr><td>9</td><td>-3,167</td><td>2,439</td><td>10</td><td>-4,126</td><td>CGGCCCGCCGAAGAAGTTGCG</td><td>GTG</td><td>2338</td><td>561</td></tr> <tr><td>10</td><td>-3,199</td><td>2,424</td><td>6</td><td>-4,500</td><td>CGACGGTATCCACGGCGGAATC</td><td>TTG</td><td>2365</td><td>534</td></tr> <tr><td>11</td><td>-5,145</td><td>1,474</td><td>10</td><td>-6,103</td><td>CAGGCAGACCAATGACGCCGTA</td><td>ATG</td><td>2479</td><td>420</td></tr> <tr><td>12</td><td>-3,078</td><td>2,483</td><td>13</td><td>-4,149</td><td>GACCCGCGAGGACATTAAGCG</td><td>TTG</td><td>2503</td><td>396</td></tr> <tr><td>13</td><td>-4,272</td><td>1,900</td><td>8</td><td>-5,397</td><td>TGTCCTAATTCAGTAGTGGC</td><td>GTG</td><td>2569</td><td>330</td></tr> <tr><td>14</td><td>-5,245</td><td>1,425</td><td>6</td><td>-6,546</td><td>GCACTCACTAGCCCTGAAAAC</td><td>TTG</td><td>2605</td><td>294</td></tr> <tr><td>15</td><td>-4,877</td><td>1,605</td><td>12</td><td>-5,923</td><td>TCGTAACGCTGCAGCGGTCGAC</td><td>GTG</td><td>2641</td><td>258</td></tr> <tr><td>16</td><td>-4,701</td><td>1,691</td><td>5</td><td>-6,099</td><td>AGCGTCCAGCTGGCAAGGCTC</td><td>TTG</td><td>2653</td><td>246</td></tr> <tr><td>17</td><td>-4,337</td><td>1,868</td><td>9</td><td>-5,383</td><td>CTACGAAAACGCGCAAGTCGA</td><td>ATG</td><td>2728</td><td>171</td></tr> <tr><td>18</td><td>-5,726</td><td>1,191</td><td>17</td><td>-7,027</td><td>TAGCCGGCTAAATCTATCAGAT</td><td>TTG</td><td>2800</td><td>99</td></tr> <tr><td>19</td><td>-5,109</td><td>1,492</td><td>16</td><td>-6,331</td><td>GATCATTGAAAACATCTCGAA</td><td>TTG</td><td>2860</td><td>39</td></tr> </tbody> </table>	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	-5,537	1,283	14	-6,634	AAAACCTTTGACAGCCTAAAAG	ATG	1855	1044	2	-3,119	2,463	5	-4,516	ACTCTTGACAGCCTAAAAGATG	TTG	1858	1041	3	-3,119	2,463	11	-4,119	GACAGCCTAAAAGATGTTGTTT	ATG	1864	1035	4	-5,906	1,103	13	-6,977	CACCTCCCTGATTATGATTAC	TTG	1950	909	5	-3,659	2,199	14	-4,756	GCCGGTAACGAGATCTCAGGCT	ATG	2014	885	6	-6,345	0,889	11	-7,345	AGTTCGGCCCGTACTCTGATT	GTG	2053	846	7	-4,735	1,674	13	-5,806	CCCACAAACCGAGTTTACCCGC	TTG	2164	735	8	-5,914	1,099	14	-7,011	CTTGTAAACACCCCGGATGAC	TTG	2185	714	9	-3,167	2,439	10	-4,126	CGGCCCGCCGAAGAAGTTGCG	GTG	2338	561	10	-3,199	2,424	6	-4,500	CGACGGTATCCACGGCGGAATC	TTG	2365	534	11	-5,145	1,474	10	-6,103	CAGGCAGACCAATGACGCCGTA	ATG	2479	420	12	-3,078	2,483	13	-4,149	GACCCGCGAGGACATTAAGCG	TTG	2503	396	13	-4,272	1,900	8	-5,397	TGTCCTAATTCAGTAGTGGC	GTG	2569	330	14	-5,245	1,425	6	-6,546	GCACTCACTAGCCCTGAAAAC	TTG	2605	294	15	-4,877	1,605	12	-5,923	TCGTAACGCTGCAGCGGTCGAC	GTG	2641	258	16	-4,701	1,691	5	-6,099	AGCGTCCAGCTGGCAAGGCTC	TTG	2653	246	17	-4,337	1,868	9	-5,383	CTACGAAAACGCGCAAGTCGA	ATG	2728	171	18	-5,726	1,191	17	-7,027	TAGCCGGCTAAATCTATCAGAT	TTG	2800	99	19	-5,109	1,492	16	-6,331	GATCATTGAAAACATCTCGAA	TTG	2860	39
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	-5,537	1,283	14	-6,634	AAAACCTTTGACAGCCTAAAAG	ATG	1855	1044																																																																																																																																																																														
2	-3,119	2,463	5	-4,516	ACTCTTGACAGCCTAAAAGATG	TTG	1858	1041																																																																																																																																																																														
3	-3,119	2,463	11	-4,119	GACAGCCTAAAAGATGTTGTTT	ATG	1864	1035																																																																																																																																																																														
4	-5,906	1,103	13	-6,977	CACCTCCCTGATTATGATTAC	TTG	1950	909																																																																																																																																																																														
5	-3,659	2,199	14	-4,756	GCCGGTAACGAGATCTCAGGCT	ATG	2014	885																																																																																																																																																																														
6	-6,345	0,889	11	-7,345	AGTTCGGCCCGTACTCTGATT	GTG	2053	846																																																																																																																																																																														
7	-4,735	1,674	13	-5,806	CCCACAAACCGAGTTTACCCGC	TTG	2164	735																																																																																																																																																																														
8	-5,914	1,099	14	-7,011	CTTGTAAACACCCCGGATGAC	TTG	2185	714																																																																																																																																																																														
9	-3,167	2,439	10	-4,126	CGGCCCGCCGAAGAAGTTGCG	GTG	2338	561																																																																																																																																																																														
10	-3,199	2,424	6	-4,500	CGACGGTATCCACGGCGGAATC	TTG	2365	534																																																																																																																																																																														
11	-5,145	1,474	10	-6,103	CAGGCAGACCAATGACGCCGTA	ATG	2479	420																																																																																																																																																																														
12	-3,078	2,483	13	-4,149	GACCCGCGAGGACATTAAGCG	TTG	2503	396																																																																																																																																																																														
13	-4,272	1,900	8	-5,397	TGTCCTAATTCAGTAGTGGC	GTG	2569	330																																																																																																																																																																														
14	-5,245	1,425	6	-6,546	GCACTCACTAGCCCTGAAAAC	TTG	2605	294																																																																																																																																																																														
15	-4,877	1,605	12	-5,923	TCGTAACGCTGCAGCGGTCGAC	GTG	2641	258																																																																																																																																																																														
16	-4,701	1,691	5	-6,099	AGCGTCCAGCTGGCAAGGCTC	TTG	2653	246																																																																																																																																																																														
17	-4,337	1,868	9	-5,383	CTACGAAAACGCGCAAGTCGA	ATG	2728	171																																																																																																																																																																														
18	-5,726	1,191	17	-7,027	TAGCCGGCTAAATCTATCAGAT	TTG	2800	99																																																																																																																																																																														
19	-5,109	1,492	16	-6,331	GATCATTGAAAACATCTCGAA	TTG	2860	39																																																																																																																																																																														
<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>	<p>. ATG 1855 ☑ ORF la plus longue de 1044pb , gap de 46pb ATG 1864 ☑ ORF de 1035pb, gap de 55pb, meilleur score TTG 1858 ☑ ORF de 1041pb, gap de 49pb, bon score</p>																																																																																																																																																																																					
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>—</p>																																																																																																																																																																																					

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

START 1855 pas conservé :
phage portal protein [Corynebacterium striatum]

Sequence ID: [HCD1553128.1](#) Length: 368 Number of Matches: 1
[See 4 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 337 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
367 bits(942)	3e-122	Compositional matrix adjust.	197/343(57%)	242/343(70%)	8/343(2%)
Query 4	MSLLDRIRKLGQLPQLLATGTAPLSVWAGD-HLEQVTSPTYDYLPVTRSQAMAIPAVVRA	62	M L++R+R+L LPQ LA S WA HL PD + + R AM IPA++RA		
Sbjct 1	MGLINRLRRLTSLPQLAAAQYG-SPWANSAPHLAIADFPDLEDSAIHRDYAMQIPALLRA	59			
>EpicDab_4, portal protein, 358 Length = 358					
Score = 257 bits (657), Expect = 3e-68 Identities = 141/318 (44%), Positives = 198/318 (62%), Gaps = 22/318 (6%)					
Query: 34	HLEQVTSPTYDYL----PVTRSQAMAIPAVVRARTLIVTTVARCPIVS---DTDEITPGW	86	HLE + D L PVTR++AMA+PA+ RAR L+ T+A P+ + + +++ W		
Sbjct: 26	HLEPIVWADVAGLATHQPVTRAEAMAVPALARARNLLAGTIAGIPLRAVRGEGEDVPLSW	85			
Query: 87	LH---GIAPLGGHPQTEFHRLNTADDLLFHGSAAWAIERDELGNVTAHVHIPSRLWERT	143	L GIA + F R+L+T DDLF+G + W + RD GNV A H+P S W		
Sbjct: 86	LDRSDGIA-----SPFQRLHTTDDLLFYGISLWGLRRDFGNVVRAAHVPMRSRWTDD	138			
Query: 144	ADGTIIVDGVPAQAQEVAVFDGIHGGILTHGARALRDASNILSAAARVADTPAALIELRQ	203	+G I++DG P ++EV V GIH GILT+G +R A N++ A + + P A IEL Q		
Sbjct: 139	TEGYIILIDGEPVSSREVCVIPGIEGILTYGRTTIRHAGNLRHADKAVENPNAYIELHQ	198			
Query: 204	TNDAVMTREDIKALIQGYVDARRGKNGVSYSSSGVEVHEHSLAPENLLIEGRNAAADV	263	+DA MT DI+ L+ +V ARR NGGV+++ G+ V EH E LL EGRNAAA+D+		
Sbjct: 199	ESDAPMTEPDIEKLVNRWVKARRGDNGGVAFVTKGIRVIEHGSPEKELTEGRNAAADI	258			
Query: 264	ARLLGVPAPFIDATVGGTSLSYENAASRMTELITFGVAPLLAAISSRLNLSDLTPVGT-V	322	ARL G+PA +DA GTS++Y+N+ +RM EL+TFG+AP++ AI +RL D+ GT V		
Sbjct: 259	ARLTGLPASLVDAQSGTSTVYQNSQARMAELVTFGLAPIMGAIVARLQDDIVARGTRV	318			
Query: 323	SFDTAQIIENISELLPTN 340		FD +EN LPT+		
Sbjct: 319	EFD----LENALALPTS 332				

>Schiebs_4, portal protein, 347
Length = 347

Score = 235 bits (599), Expect = 1e-61
Identities = 132/326 (40%), Positives = 194/326 (59%), Gaps = 4/326 (1%)

Query: 4	MSLLDRIRKLGQLPQLLATGTAPLSVWA-GDHLEQVTSPTYDYLPVTRSQAMAIPAVVRA	62	MSLL ++R LP + ATGTA S W+ +HL+ + P +TR AM +P+V RA
Sbjct: 1	MSLLAKVRDALALPYMAATGTA--SPWSYPNHLQPLPWPSGAEGTITRDAAMRVPSVARA	58	

START 1864, le plus conservé ! :

phage portal protein [Corynebacterium striatum]

Sequence ID: [HCD1553128.1](#) Length: 368 Number of Matches: 1
[See 4 more title\(s\)](#) [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 337 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
367 bits(941)	5e-122	Compositional matrix adjust.	197/343(57%)	242/343(70%)	8/343(2%)
Query 1	MSLLDRIRKLGQLPQLLATGTAPLSVWAGD-HLEQVTSPTYDYLPVTRSQAMAIPAVVRA	59	M L++R+R+L LPQ LA S WA HL PD + + R AM IPA++RA		
Sbjct 1	MGLINRLRRLTSLPQLAAAQYG-SPWANSAPHLAIADFPDLEDSAIHRDYAMQIPALLRA	59			

DECISION:

1864 : ATG ; bon score ; le plus conservé ; ORf de grande taille. .

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

Collection des éléments de réponse

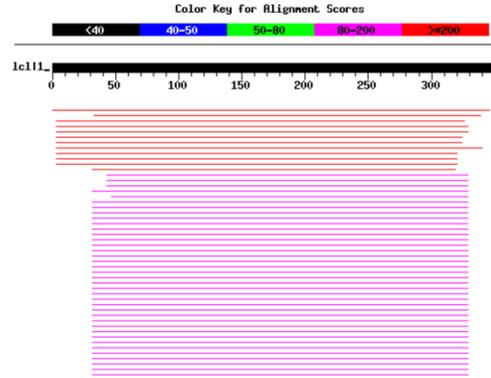
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10⁻⁴ et une couverture acceptable ?

Rationnelle

Listez le meilleur hit Blastp pour chaque source :

*** PhagesDB :

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



Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_3, function unknown, 347	579	0.0
EpicDab_4, portal protein, 358	257	3e-68
Schiebs_4, portal protein, 347	235	1e-61
Rahul_4, portal protein, 368	231	2e-60
Coeur_4, portal protein, 368	230	3e-60
Emperor_6, portal protein, 385	227	4e-59
RRH1_04, portal protein, 372	223	4e-58
SallySpecial_4, portal protein, 386	219	6e-57
GMA5_4, putative portal protein, 363	208	1e-53
McGonagall_4, portal protein, 364	207	3e-53
Jeanie_4, portal protein, 364	207	3e-53
GRU3_4, putative portal protein, 358	202	8e-52
SerialPhiller_4, portal protein, 381	158	2e-38
Kels_4, portal protein, 381	158	2e-38

*** nr :

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

100 sequences selected Putative conserved domains have been detected, click on the image below for detailed results.

Query	Subject	Score	Identical	Positives	Bits	E-Value	Accession
✓ phage_portal protein [Corynebacterium sp. HMSC06C06]	Corynebacterium sp. HMSC06C06	371	371	98%	1e-123	57.73%	368 WP_070420693.1
✓ phage_portal protein [Corynebacterium striatum]	Corynebacterium striatum	369	369	98%	5e-123	57.43%	368 WP_201806831.1
✓ TPA_phage_portal protein [Corynebacterium striatum]	Corynebacterium striatum	367	367	98%	3e-122	57.43%	368 HCD1553128.1
✓ phage_portal protein [unclassified Corynebacterium]	unclassified Corynebacterium	366	366	95%	4e-122	57.83%	347 WP_070736128.1
✓ conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	364	364	95%	1e-121	57.53%	347 CQD13958.1
✓ phage_portal protein [Corynebacterium sp. EPI-003-04-2554_SCH2473622]	Corynebacterium sp. EPI-003-04-255...	332	332	94%	8e-109	54.24%	342 WP_064833798.1
✓ phage_portal protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	332	332	94%	2e-108	53.64%	351 WP_284849030.1
✓ phage_portal protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	331	331	94%	2e-108	53.94%	342 WP_284587132.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	308	308	95%	2e-99	50.88%	349 WP_010935359.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	308	308	95%	3e-99	50.88%	349 WP_196975735.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	305	305	82%	3e-98	55.33%	329 WP_179148619.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	305	305	95%	3e-98	50.88%	349 CAB0923839.1
✓ phage_portal protein [Corynebacterium propinquum]	Corynebacterium propinquum	305	305	98%	4e-98	49.57%	342 WP_284594231.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	305	305	82%	4e-98	55.33%	349 WP_088259518.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	305	305	82%	4e-98	55.33%	343 WP_003852544.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	302	302	95%	9e-97	51.71%	357 CAB0913807.1
✓ phage_portal protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	301	301	95%	1e-96	51.41%	357 CAB0519145.1
✓ phage_portal protein [Corynebacterium belfantii]	Corynebacterium belfantii	300	300	95%	3e-96	51.59%	353 WP_197692071.1

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 $\geq 90\%$ et une couverture acceptable ?

Visualization



Nr	Hit	Name	Probability	E-value	Score	SS	Acc
1	3KDR_C	HK97 Family Phage Portal Protein; Phage, HK97 family, Portal, Corynebacterium, diphtheriae, PSI, MCSG, Structural Genom	100	2.6e-41	308.35	38.7	2:
2	8FQL_D	Portal protein; Prohead L, icosahedral symmetry, HK97, phage, capsid, VIRUS; 3.6A (Escherichia phage HK97)	100	1.5e-34	274.4	34.3	2:
3	Q9ZX82	PORTL_BPPHC Probable portal protein OS=Streptomyces phage phiC31 OX=10719 GN=34 PE=3 SV=1	100	4.1e-31	253.64	36.3	3:
4	8FXR_AJ	Portal protein, gp7; Myophage, redox trigger, VIRUS; 4.5A (Agrobacterium phage Milano)	100	2.2e-32	260.06	25.3	3:

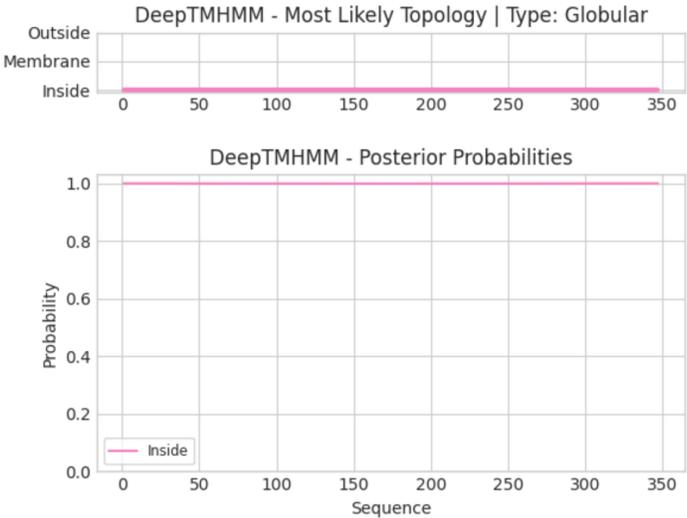
PDB : HK97 Family Phage Portal Protein; Phage, HK97 family, Portal, Corynebacterium, diphtheriae, PSI, MCSG, Structural Genomics, Protein Structure Initiative, Midwest Center for; HET: PO4, MSE, SO4, GOL; 2.9A {Corynebacterium diphtheriae}

Uniprot : PORTL_BPPHC Probable portal protein OS=Streptomyces phage phiC31 OX=10719 GN=34 PE=3 SV=1

Pfam : Phage_portal_2 ; Phage portal protein, lambda family
CD : pas de hit

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
Le gene avant code une Terminase large subunit
Le gene apres pour une endolysine

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	 <p>DeepTMHMM - Most Likely Topology Type: Globular</p> <p>DeepTMHMM - Posterior Probabilities</p>
<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>portal protein</p>

Student Gene Annotation Worksheet

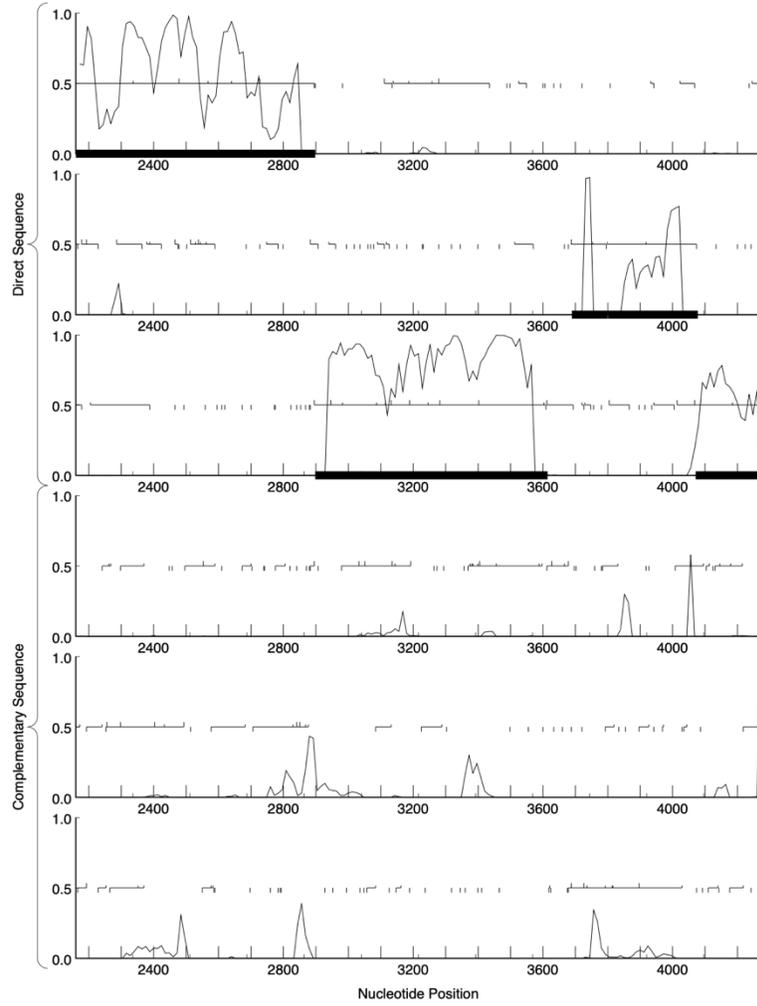
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	4
Coordonnées du Stop	3614
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove 1
Coordonnées du Start retenu	2898
Fonction prédite	Endolysine

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 GeneMark only

Oui il y a un potentiel codant avec GeneMarksS

GeneMark hmmprediction
yanoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGGCCAGCCC), Order 2, Window 96, Step 12, 3/9

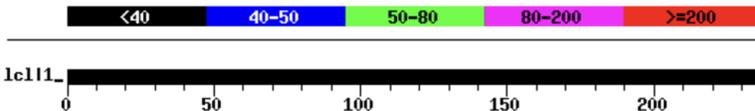


Y-a-t-il des éléments supportant un potentiel codant ?

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



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

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

N-acetylmuramoyl-L-alanine amidase [*Corynebacterium matruchotii*]

Sequence ID: [VEI99905.1](#) Length: 226 Number of Matches: 1

Range 1: 2 to 223 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
265 bits(677) 6e-86 Compositional matrix adjust. 136/234(58%) 161/234(68%) 12/234(5%)					
Query 3		RYKVOPGVL TENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEPIE	62		
Sbjct 2		RY+ Q GV +ENGWPM ++ VS VVPAAK+VPL AG+VATILNAW+I +NREVEPI RYQAQIGVPSSENGWPMCSSASCSEMVVPAAKRVPLRAGDVATILNAWILYNREVEPI	61		
Query 63		SQVWGHASADNDVWNSNHMSGTAIDIGAPKYFPFKRVMPAATKAKVRALLDKFNGVYVWGA	122		
Sbjct 62		SQVWGHASADNDVWNSNHMSGTA+DIGAPKYP+G+R MP ATKAKVRALL KF GV++WGA SQVWGHASADNDVWNSNHMSGTAVDIGAPKYFPGQRTMPATKAKVRALLKFEVIFWGA	121		
Query 123		YWSPDEMHPQIGLPPSNAKVAELANKLGGYLNIVGDTTTEPESTSMSLTAQQDRLNE	182		
Sbjct 122		W YPDEMH+QIGLPPS+ +V A +LN GYL Y DT P+ SM+ D L + DWDYPDEMHYQIGLPPSDPRVHAFERLNNGLGAY -ADTPAPKGNM-- -DDDLYLRD	177		
Query 183		ILENTKTIKNQNEILKQLGQGGWPGGGRTVYDLISAVAEIEGVPNTRDTLA	236		
Sbjct 178		+ LG GWPQ GGRV D ++A+ +P D A VKAQLTGAPG-----LGAYPGWPQLGGRTVVDALAAIGAALKIPGFADLKA	223		

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

MAG TPA: Chitinase A [*Siphoviridae* sp. ctCCv12]

Sequence ID: [DAD89994.1](#) Length: 230 Number of Matches: 1

Range 1: 1 to 227 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
259 bits(662) 2e-83 Compositional matrix adjust. 134/236(57%) 160/236(67%) 9/236(3%)					
Query 1		MTRYKVPQGVLTENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEP	60		
Sbjct 1		MTRY+ Q GV +ENGWPM ++ VS VVPA+ VPL AG+VATILNAW+I +NR VEP MTRYQAQIGVPSSENGWPMCSSAACVSETVVAARAVPLRAGDVATILNAWILYNRLVEP	60		

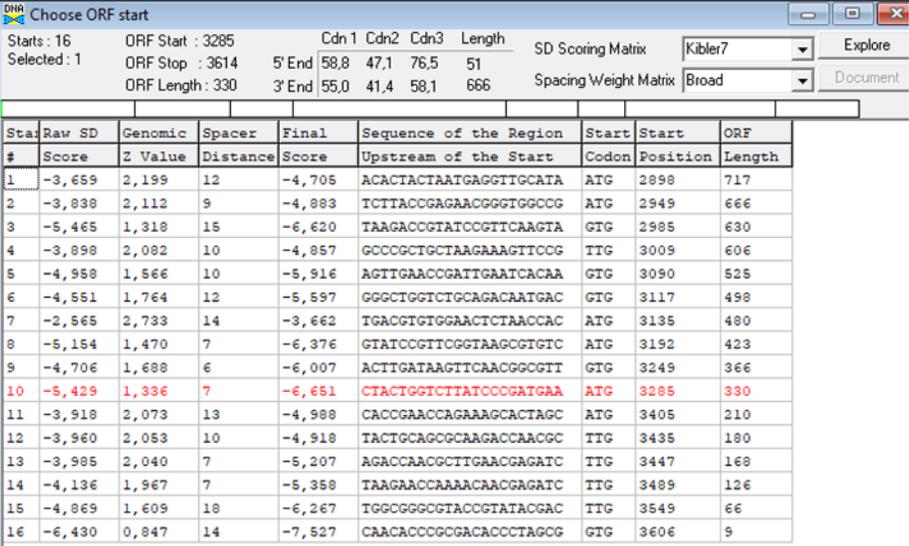
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

non

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																		
Quel start Glimmer et GeneMark suggèrent-ils ?	2898																																																																																																																																																																		
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <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>-3,659</td> <td>2,199</td> <td>12</td> <td>-4,705</td> <td>ACACTACTAATGAGGTTGCATA</td> <td>ATG</td> <td>2898</td> <td>717</td> </tr> <tr> <td>2</td> <td>-3,838</td> <td>2,112</td> <td>9</td> <td>-4,883</td> <td>TCTTACCGAGAACGGGTGGCCG</td> <td>ATG</td> <td>2949</td> <td>666</td> </tr> <tr> <td>3</td> <td>-5,465</td> <td>1,318</td> <td>15</td> <td>-6,620</td> <td>TAAGACCGTATCCGTTCAAGTA</td> <td>GTG</td> <td>2985</td> <td>630</td> </tr> <tr> <td>4</td> <td>-3,898</td> <td>2,082</td> <td>10</td> <td>-4,857</td> <td>GCCCGCTGCTAAGAAAGTTCCG</td> <td>TTG</td> <td>3009</td> <td>606</td> </tr> <tr> <td>5</td> <td>-4,958</td> <td>1,566</td> <td>10</td> <td>-5,916</td> <td>AGTTGAACCGATTGAATCACAA</td> <td>GTG</td> <td>3090</td> <td>525</td> </tr> <tr> <td>6</td> <td>-4,551</td> <td>1,764</td> <td>12</td> <td>-5,597</td> <td>GGGCTGGTCTGCAGACAATGAC</td> <td>GTG</td> <td>3117</td> <td>498</td> </tr> <tr> <td>7</td> <td>-2,565</td> <td>2,733</td> <td>14</td> <td>-3,662</td> <td>TGACGTGTGGAAGTCTAACCAC</td> <td>ATG</td> <td>3135</td> <td>480</td> </tr> <tr> <td>8</td> <td>-5,154</td> <td>1,470</td> <td>7</td> <td>-6,376</td> <td>GTATCCGTTCCGTAAGCCGTGTC</td> <td>ATG</td> <td>3192</td> <td>423</td> </tr> <tr> <td>9</td> <td>-4,706</td> <td>1,688</td> <td>6</td> <td>-6,007</td> <td>ACTTGATAAGTTCAACGGCGTIT</td> <td>GTG</td> <td>3249</td> <td>366</td> </tr> <tr> <td>10</td> <td>-5,429</td> <td>1,336</td> <td>7</td> <td>-6,651</td> <td>CTACTGGTCTTATCCCGATGAA</td> <td>ATG</td> <td>3285</td> <td>330</td> </tr> <tr> <td>11</td> <td>-3,918</td> <td>2,073</td> <td>13</td> <td>-4,988</td> <td>CACCGAACCCAGAAAGCACTAGC</td> <td>ATG</td> <td>3405</td> <td>210</td> </tr> <tr> <td>12</td> <td>-3,960</td> <td>2,053</td> <td>10</td> <td>-4,918</td> <td>TACTGCAGCGCAGAACCAACGC</td> <td>TTG</td> <td>3435</td> <td>180</td> </tr> <tr> <td>13</td> <td>-3,985</td> <td>2,040</td> <td>7</td> <td>-5,207</td> <td>AGACCAACGCTTGAACGAGATC</td> <td>TTG</td> <td>3447</td> <td>168</td> </tr> <tr> <td>14</td> <td>-4,136</td> <td>1,967</td> <td>7</td> <td>-5,358</td> <td>TAAGAACCAAAACAACGAGATC</td> <td>TTG</td> <td>3489</td> <td>126</td> </tr> <tr> <td>15</td> <td>-4,869</td> <td>1,609</td> <td>18</td> <td>-6,267</td> <td>TGGCGGGCGTACCGTATACGAC</td> <td>TTG</td> <td>3549</td> <td>66</td> </tr> <tr> <td>16</td> <td>-6,430</td> <td>0,847</td> <td>14</td> <td>-7,527</td> <td>CAACACCCCGGACACCOCTAGCG</td> <td>GTG</td> <td>3606</td> <td>9</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	-3,659	2,199	12	-4,705	ACACTACTAATGAGGTTGCATA	ATG	2898	717	2	-3,838	2,112	9	-4,883	TCTTACCGAGAACGGGTGGCCG	ATG	2949	666	3	-5,465	1,318	15	-6,620	TAAGACCGTATCCGTTCAAGTA	GTG	2985	630	4	-3,898	2,082	10	-4,857	GCCCGCTGCTAAGAAAGTTCCG	TTG	3009	606	5	-4,958	1,566	10	-5,916	AGTTGAACCGATTGAATCACAA	GTG	3090	525	6	-4,551	1,764	12	-5,597	GGGCTGGTCTGCAGACAATGAC	GTG	3117	498	7	-2,565	2,733	14	-3,662	TGACGTGTGGAAGTCTAACCAC	ATG	3135	480	8	-5,154	1,470	7	-6,376	GTATCCGTTCCGTAAGCCGTGTC	ATG	3192	423	9	-4,706	1,688	6	-6,007	ACTTGATAAGTTCAACGGCGTIT	GTG	3249	366	10	-5,429	1,336	7	-6,651	CTACTGGTCTTATCCCGATGAA	ATG	3285	330	11	-3,918	2,073	13	-4,988	CACCGAACCCAGAAAGCACTAGC	ATG	3405	210	12	-3,960	2,053	10	-4,918	TACTGCAGCGCAGAACCAACGC	TTG	3435	180	13	-3,985	2,040	7	-5,207	AGACCAACGCTTGAACGAGATC	TTG	3447	168	14	-4,136	1,967	7	-5,358	TAAGAACCAAAACAACGAGATC	TTG	3489	126	15	-4,869	1,609	18	-6,267	TGGCGGGCGTACCGTATACGAC	TTG	3549	66	16	-6,430	0,847	14	-7,527	CAACACCCCGGACACCOCTAGCG	GTG	3606	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	-3,659	2,199	12	-4,705	ACACTACTAATGAGGTTGCATA	ATG	2898	717																																																																																																																																																											
2	-3,838	2,112	9	-4,883	TCTTACCGAGAACGGGTGGCCG	ATG	2949	666																																																																																																																																																											
3	-5,465	1,318	15	-6,620	TAAGACCGTATCCGTTCAAGTA	GTG	2985	630																																																																																																																																																											
4	-3,898	2,082	10	-4,857	GCCCGCTGCTAAGAAAGTTCCG	TTG	3009	606																																																																																																																																																											
5	-4,958	1,566	10	-5,916	AGTTGAACCGATTGAATCACAA	GTG	3090	525																																																																																																																																																											
6	-4,551	1,764	12	-5,597	GGGCTGGTCTGCAGACAATGAC	GTG	3117	498																																																																																																																																																											
7	-2,565	2,733	14	-3,662	TGACGTGTGGAAGTCTAACCAC	ATG	3135	480																																																																																																																																																											
8	-5,154	1,470	7	-6,376	GTATCCGTTCCGTAAGCCGTGTC	ATG	3192	423																																																																																																																																																											
9	-4,706	1,688	6	-6,007	ACTTGATAAGTTCAACGGCGTIT	GTG	3249	366																																																																																																																																																											
10	-5,429	1,336	7	-6,651	CTACTGGTCTTATCCCGATGAA	ATG	3285	330																																																																																																																																																											
11	-3,918	2,073	13	-4,988	CACCGAACCCAGAAAGCACTAGC	ATG	3405	210																																																																																																																																																											
12	-3,960	2,053	10	-4,918	TACTGCAGCGCAGAACCAACGC	TTG	3435	180																																																																																																																																																											
13	-3,985	2,040	7	-5,207	AGACCAACGCTTGAACGAGATC	TTG	3447	168																																																																																																																																																											
14	-4,136	1,967	7	-5,358	TAAGAACCAAAACAACGAGATC	TTG	3489	126																																																																																																																																																											
15	-4,869	1,609	18	-6,267	TGGCGGGCGTACCGTATACGAC	TTG	3549	66																																																																																																																																																											
16	-6,430	0,847	14	-7,527	CAACACCCCGGACACCOCTAGCG	GTG	3606	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 ?	ORF la plus longue = ATG 2898, bon 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 ?	START 2898, retrouvé chez certains homologues sur la nr :																																																																																																																																																																		

	<p>MAG TPA: Chitinase A [Siphoviridae sp. ctCCv12] Sequence ID: DAD89994.1 Length: 230 Number of Matches: 1</p> <p>Range 1: 1 to 227 GenPept Graphics ▼ Next Match ▲ Previous</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>259 bits(662)</td> <td>2e-83</td> <td>Compositional matrix adjust.</td> <td>134/236(57%)</td> <td>160/236(67%)</td> <td>9/236(3%)</td> </tr> </tbody> </table> <p>Query 1 MTRYKVVQPGVLTENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEP 60 MTRY+ Q GV +ENGWPM ++ VS VVPAA+ VPL AG+VATILNAW+I +NR VEP Sbjct 1 MTRYQAQIGVPSSENGWPMCSSAACVSETVVPAAARAVPLRAGDVATILNAWLILYNRLVEP 60</p> <p>Query 61 IESQVWGSADNDVWNSNHMSGTAIDIGAPKYPPFGKRVMPAATKAKVRALLDKFNGVVY 120 I SQVWGSADNDVWNSNHMSGTA+DIGAPKY+G+R MP ATKAKVR LL KF GVV+W Sbjct 61 ITSQVWGSADNDVWNSNHMSGTAVDIGAPKYPWGQRTMPPATKAKVRTLLAKFEGVVFW 120</p> <p>Query 121 GAYWSYPDEMHHQIGLPPSNAKVAELANKLNGGYLNIYGTDTTEPESTSMLTAAQDQRL 180 GA W YPDEM+QIGLPPS+ +V A +LN GYL Y D +A + + Sbjct 121 GADWDYPDEMHHQIGLPPSDPRVHFAERLNNGYLGAYPADDAP-----SAPKGNPM 172</p> <p>Query 181 NEILENTKTIKNQNEILKQLGQPGGWPGGGRTVYDLISAVAEIEGVPNTRDTLA 236 N+ + +K Q LG GWPQ GGRTV D ++A+ +P D A Sbjct 173 NDDDLYLRLDKAQMTGA-PGLGAYPGWPQLGGRTVVDALAAIGAKLEIPGFADLKA 227</p> <p>>RRH1_05, putative lysin, 244 Length = 244</p> <p>Score = 184 bits (467), Expect = 2e-46 Identities = 85/148 (57%), Positives = 108/148 (72%)</p> <p>Query: 12 TENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEPIESQVWGSAD 71 +ENGW M N D+TV+V+VVP + P+ G ATILNAW+I ++R VEPI SQVWGS D Sbjct: 12 SENGWRMCNHDETVTVKVVPGSVGAPIRRGAPATILNAWLEIYDRHVEPITSQVWGSRD 71</p> <p>Query: 72 NDVWNSNHMSGTAIDIGAPKYPPFGKRVMPAATKAKVRALLDKFNGVVYWGAYWSYPDEM 131 NDV NSNH++GTA+DI APKY+G+RVMP +VRA L++F G V+WGA WS DEMH Sbjct: 72 NDVANSNHLAGTAVDINAPKYPWGRRVMPREIRLVRRAALNRFEGSVFVGADWSRADEM 131</p> <p>Query: 132 FQIGLPPSNAKVAELANKLNGGYLNIY 159 +Q+ + +V A KL GG+L IYG Sbjct: 132 YQLAWREGDPRVDAFARKLEGGHLGIY 159</p> <p>>Rahul_5, lysin A, L-Ala-D-Glu peptidase domain, 260 Length = 260</p> <p>Score = 178 bits (452), Expect = 9e-45 Identities = 104/245 (42%), Positives = 134/245 (54%), Gaps = 18/245 (7%)</p> <p>Query: 12 TENGWPMVNADKTVSV---QVVPAAKKVPLLAGNVATILNAWIIAFNREVEPIESQVWGW 68 +ENGW MVN D V +P P+ AG +TIL AW+ ++R VEP+ S VWGW Sbjct: 12 SENGWRMVNRDGCVVAGPSMGLPYTNTAPIRAGVASTILVAWLAWYHRNVEPLASPVWGW 71</p>	Score	Expect	Method	Identities	Positives	Gaps	259 bits(662)	2e-83	Compositional matrix adjust.	134/236(57%)	160/236(67%)	9/236(3%)
Score	Expect	Method	Identities	Positives	Gaps								
259 bits(662)	2e-83	Compositional matrix adjust.	134/236(57%)	160/236(67%)	9/236(3%)								
DECISION:	2898, conserve chez certains homologues, bon score ATG												

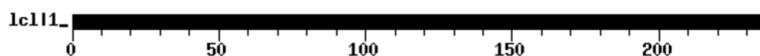
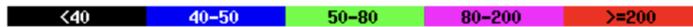
Décision #3 : Quelle est la fonction de la protéine putative ?

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 ?	Listez le meilleur hit Blastp pour chaque source : *** PhagesDB : (q#: s#) : [alignment] ; e-value :

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 (bits)	E Value
CyranoPS_Draft_4, function unknown, 238	488	e-138
RRH1_05, putative lysin, 244	184	2e-46
Rahul_5, lysin A, L-Ala-D-Glu peptidase domain, 260	178	9e-45
Dardanus_25, lysin A, 342	177	3e-44
Coeur_5, lysin A, L-Ala-D-Glu peptidase domain, 260	176	3e-44
Huffy_25, lysin A, 341	172	8e-43
DinoDaryn_25, lysin B, 341	172	8e-43
Jace_18, lysin A, 341	164	1e-40
Toil_31, putative endolysin, 227	154	1e-37
GRU3_5, putative lysin, 262	146	5e-35
Zolita_6, lysin A, 304	140	2e-33
UnionJack_7, lysin A, 221	140	2e-33
SydNat_7, lysin A, 304	140	2e-33
Stephig9_7, lysin A, 226	140	2e-33
Roary_7, lysin A, 226	140	2e-33
Ghoulboy_7, lysin A, 304	140	2e-33
Dixon_7, lysin A, 226	140	2e-33
Discoknowium_7, lysin A, 304	140	2e-33
Aragog_7, lysin A, 304	140	2e-33
Tiger_7, lysin A, 304	140	3e-33
Smeadley_7, lysin A, 226	140	3e-33
PickleBack_7, lysin A, 304	140	3e-33
Phlorence_7, lysin A, 304	140	3e-33
Phillis_7, lysin A, 225	140	3e-33
Lev2_7, lysin A, 304	140	3e-33
Groundhog_6, lysin A, 226	140	3e-33
ForGetIt_7, lysin A, 304	140	3e-33
Expelliarmus_6, lysin A, 226	140	3e-33
Danforth_6, lysin A, 226	140	3e-33
Conspiracy_7, LysA, 304	140	3e-33

	<p>*** nr :</p> <table border="1"> <tr><td><input checked="" type="checkbox"/></td><td>N-acetylmuramoyl-L-alanine amidase [Corynebacterium matrucholii]</td><td>Corynebacterium matrucholii</td><td>265</td><td>265</td><td>98%</td><td>6e-86</td><td>58.12%</td><td>226</td><td>VEI99905_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: Chitinase A [Siphoviridae sp. ctCv12]</td><td>Siphoviridae sp. ctCv12</td><td>259</td><td>259</td><td>99%</td><td>2e-83</td><td>56.78%</td><td>230</td><td>DAD89994_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: Chitinase A [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>258</td><td>258</td><td>98%</td><td>4e-83</td><td>57.69%</td><td>228</td><td>DAO68537_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: Chitinase A [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>254</td><td>254</td><td>97%</td><td>2e-81</td><td>55.41%</td><td>230</td><td>DAD60081_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>249</td><td>249</td><td>77%</td><td>4e-79</td><td>66.31%</td><td>252</td><td>WP_284571880_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>248</td><td>248</td><td>91%</td><td>6e-79</td><td>56.65%</td><td>252</td><td>WP_082148289_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: Chitinase A [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>247</td><td>247</td><td>91%</td><td>2e-78</td><td>56.65%</td><td>252</td><td>DAX64374_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>246</td><td>246</td><td>77%</td><td>7e-78</td><td>65.78%</td><td>252</td><td>WP_239210962_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium matrucholii]</td><td>Corynebacterium matrucholii</td><td>243</td><td>243</td><td>92%</td><td>1e-77</td><td>57.53%</td><td>210</td><td>WP_232022508_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>241</td><td>241</td><td>73%</td><td>3e-76</td><td>69.14%</td><td>237</td><td>MDK4326170_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium]</td><td>Corynebacterium</td><td>235</td><td>235</td><td>86%</td><td>7e-74</td><td>56.31%</td><td>243</td><td>WP_239688266_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: Chitinase A [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>233</td><td>233</td><td>92%</td><td>3e-73</td><td>55.71%</td><td>210</td><td>DAU29188_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>233</td><td>233</td><td>76%</td><td>6e-73</td><td>62.90%</td><td>258</td><td>WP_147579339_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium pseudodiphthericum]</td><td>Corynebacterium pseudodiphth...</td><td>233</td><td>233</td><td>76%</td><td>7e-73</td><td>62.90%</td><td>256</td><td>WP_081703804_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>233</td><td>233</td><td>76%</td><td>1e-72</td><td>62.90%</td><td>258</td><td>WP_144736461_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>D-alanyl-D-alanine carboxypeptidase-like protein [Williamsia marianensis]</td><td>Williamsia marianensis</td><td>218</td><td>218</td><td>94%</td><td>3e-67</td><td>51.84%</td><td>248</td><td>PVY33004_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein N579_07895 [Corynebacterium pseudodiphthericum.090104]</td><td>Corynebacterium pseudodiphth...</td><td>206</td><td>206</td><td>69%</td><td>1e-62</td><td>61.54%</td><td>239</td><td>ERJ44230_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Williamsia marianensis]</td><td>Williamsia marianensis</td><td>194</td><td>194</td><td>87%</td><td>5e-58</td><td>50.43%</td><td>232</td><td>WP_243404337_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>peptidoglycan-binding protein [Tsukamurella pulmonis]</td><td>Tsukamurella pulmonis</td><td>190</td><td>190</td><td>84%</td><td>2e-56</td><td>46.50%</td><td>236</td><td>WP_114652310_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus spongicola]</td><td>Rhodococcus spongicola</td><td>189</td><td>189</td><td>90%</td><td>3e-56</td><td>48.62%</td><td>220</td><td>WP_127945131_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Corynebacterium propinquum]</td><td>Corynebacterium propinquum</td><td>187</td><td>187</td><td>60%</td><td>2e-55</td><td>65.28%</td><td>204</td><td>WP_284589709_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus sp. AG1013]</td><td>Rhodococcus sp. AG1013</td><td>186</td><td>186</td><td>94%</td><td>9e-55</td><td>43.75%</td><td>218</td><td>WP_114724430_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus sp. FXJ9_536]</td><td>Rhodococcus sp. FXJ9_536</td><td>185</td><td>185</td><td>73%</td><td>2e-54</td><td>50.57%</td><td>224</td><td>WP_255969710_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus qingshengii]</td><td>Rhodococcus qingshengii</td><td>183</td><td>183</td><td>89%</td><td>1e-53</td><td>44.95%</td><td>236</td><td>WP_058226364_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus pyridinivorans]</td><td>Rhodococcus pyridinivorans</td><td>183</td><td>183</td><td>76%</td><td>2e-53</td><td>48.62%</td><td>235</td><td>WP_006564338_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus qingshengii]</td><td>Rhodococcus qingshengii</td><td>183</td><td>183</td><td>89%</td><td>2e-53</td><td>44.95%</td><td>236</td><td>WP_336874228_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>endolysin [Rhodococcus phage RRH1]</td><td>Rhodococcus phage RRH1</td><td>183</td><td>183</td><td>61%</td><td>3e-53</td><td>57.14%</td><td>244</td><td>YP_005087029_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>M15 family metallopeptidase [Rhodococcus aetherivorans]</td><td>Rhodococcus aetherivorans</td><td>180</td><td>180</td><td>65%</td><td>4e-53</td><td>55.13%</td><td>189</td><td>WP_006945204_1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>peptidoglycan-binding protein [Tsukamurella paurometabola]</td><td>Tsukamurella paurometabola</td><td>182</td><td>182</td><td>84%</td><td>6e-53</td><td>44.00%</td><td>239</td><td>WP_126197209_1</td></tr> </table>	<input checked="" type="checkbox"/>	N-acetylmuramoyl-L-alanine amidase [Corynebacterium matrucholii]	Corynebacterium matrucholii	265	265	98%	6e-86	58.12%	226	VEI99905_1	<input checked="" type="checkbox"/>	TPA: Chitinase A [Siphoviridae sp. ctCv12]	Siphoviridae sp. ctCv12	259	259	99%	2e-83	56.78%	230	DAD89994_1	<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	258	258	98%	4e-83	57.69%	228	DAO68537_1	<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	254	254	97%	2e-81	55.41%	230	DAD60081_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	249	249	77%	4e-79	66.31%	252	WP_284571880_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	248	248	91%	6e-79	56.65%	252	WP_082148289_1	<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	247	247	91%	2e-78	56.65%	252	DAX64374_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	246	246	77%	7e-78	65.78%	252	WP_239210962_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium matrucholii]	Corynebacterium matrucholii	243	243	92%	1e-77	57.53%	210	WP_232022508_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	241	241	73%	3e-76	69.14%	237	MDK4326170_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium]	Corynebacterium	235	235	86%	7e-74	56.31%	243	WP_239688266_1	<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	233	233	92%	3e-73	55.71%	210	DAU29188_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	233	233	76%	6e-73	62.90%	258	WP_147579339_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphth...	233	233	76%	7e-73	62.90%	256	WP_081703804_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	233	233	76%	1e-72	62.90%	258	WP_144736461_1	<input checked="" type="checkbox"/>	D-alanyl-D-alanine carboxypeptidase-like protein [Williamsia marianensis]	Williamsia marianensis	218	218	94%	3e-67	51.84%	248	PVY33004_1	<input checked="" type="checkbox"/>	hypothetical protein N579_07895 [Corynebacterium pseudodiphthericum.090104]	Corynebacterium pseudodiphth...	206	206	69%	1e-62	61.54%	239	ERJ44230_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Williamsia marianensis]	Williamsia marianensis	194	194	87%	5e-58	50.43%	232	WP_243404337_1	<input checked="" type="checkbox"/>	peptidoglycan-binding protein [Tsukamurella pulmonis]	Tsukamurella pulmonis	190	190	84%	2e-56	46.50%	236	WP_114652310_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus spongicola]	Rhodococcus spongicola	189	189	90%	3e-56	48.62%	220	WP_127945131_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	187	187	60%	2e-55	65.28%	204	WP_284589709_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus sp. AG1013]	Rhodococcus sp. AG1013	186	186	94%	9e-55	43.75%	218	WP_114724430_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus sp. FXJ9_536]	Rhodococcus sp. FXJ9_536	185	185	73%	2e-54	50.57%	224	WP_255969710_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus qingshengii]	Rhodococcus qingshengii	183	183	89%	1e-53	44.95%	236	WP_058226364_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus pyridinivorans]	Rhodococcus pyridinivorans	183	183	76%	2e-53	48.62%	235	WP_006564338_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus qingshengii]	Rhodococcus qingshengii	183	183	89%	2e-53	44.95%	236	WP_336874228_1	<input checked="" type="checkbox"/>	endolysin [Rhodococcus phage RRH1]	Rhodococcus phage RRH1	183	183	61%	3e-53	57.14%	244	YP_005087029_1	<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus aetherivorans]	Rhodococcus aetherivorans	180	180	65%	4e-53	55.13%	189	WP_006945204_1	<input checked="" type="checkbox"/>	peptidoglycan-binding protein [Tsukamurella paurometabola]	Tsukamurella paurometabola	182	182	84%	6e-53	44.00%	239	WP_126197209_1
<input checked="" type="checkbox"/>	N-acetylmuramoyl-L-alanine amidase [Corynebacterium matrucholii]	Corynebacterium matrucholii	265	265	98%	6e-86	58.12%	226	VEI99905_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	TPA: Chitinase A [Siphoviridae sp. ctCv12]	Siphoviridae sp. ctCv12	259	259	99%	2e-83	56.78%	230	DAD89994_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	258	258	98%	4e-83	57.69%	228	DAO68537_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	254	254	97%	2e-81	55.41%	230	DAD60081_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	249	249	77%	4e-79	66.31%	252	WP_284571880_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	248	248	91%	6e-79	56.65%	252	WP_082148289_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	247	247	91%	2e-78	56.65%	252	DAX64374_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	246	246	77%	7e-78	65.78%	252	WP_239210962_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium matrucholii]	Corynebacterium matrucholii	243	243	92%	1e-77	57.53%	210	WP_232022508_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	241	241	73%	3e-76	69.14%	237	MDK4326170_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium]	Corynebacterium	235	235	86%	7e-74	56.31%	243	WP_239688266_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	233	233	92%	3e-73	55.71%	210	DAU29188_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	233	233	76%	6e-73	62.90%	258	WP_147579339_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphth...	233	233	76%	7e-73	62.90%	256	WP_081703804_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	233	233	76%	1e-72	62.90%	258	WP_144736461_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	D-alanyl-D-alanine carboxypeptidase-like protein [Williamsia marianensis]	Williamsia marianensis	218	218	94%	3e-67	51.84%	248	PVY33004_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	hypothetical protein N579_07895 [Corynebacterium pseudodiphthericum.090104]	Corynebacterium pseudodiphth...	206	206	69%	1e-62	61.54%	239	ERJ44230_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Williamsia marianensis]	Williamsia marianensis	194	194	87%	5e-58	50.43%	232	WP_243404337_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	peptidoglycan-binding protein [Tsukamurella pulmonis]	Tsukamurella pulmonis	190	190	84%	2e-56	46.50%	236	WP_114652310_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus spongicola]	Rhodococcus spongicola	189	189	90%	3e-56	48.62%	220	WP_127945131_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	187	187	60%	2e-55	65.28%	204	WP_284589709_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus sp. AG1013]	Rhodococcus sp. AG1013	186	186	94%	9e-55	43.75%	218	WP_114724430_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus sp. FXJ9_536]	Rhodococcus sp. FXJ9_536	185	185	73%	2e-54	50.57%	224	WP_255969710_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus qingshengii]	Rhodococcus qingshengii	183	183	89%	1e-53	44.95%	236	WP_058226364_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus pyridinivorans]	Rhodococcus pyridinivorans	183	183	76%	2e-53	48.62%	235	WP_006564338_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus qingshengii]	Rhodococcus qingshengii	183	183	89%	2e-53	44.95%	236	WP_336874228_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	endolysin [Rhodococcus phage RRH1]	Rhodococcus phage RRH1	183	183	61%	3e-53	57.14%	244	YP_005087029_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	M15 family metallopeptidase [Rhodococcus aetherivorans]	Rhodococcus aetherivorans	180	180	65%	4e-53	55.13%	189	WP_006945204_1																																																																																																																																																																																																																																																																																										
<input checked="" type="checkbox"/>	peptidoglycan-binding protein [Tsukamurella paurometabola]	Tsukamurella paurometabola	182	182	84%	6e-53	44.00%	239	WP_126197209_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>	<p>Pfam : Peptidase_M15_4 ; D-alanyl-D-alanine carboxypeptidase CD : L-Ala-D-Glu_peptidase_like; L-Ala-D-Glu peptidase, also known as L-alanyl-D-glutamate endopeptidase. This L-Ala-D-Glu peptidase family includes L-alanyl-D-glutamate peptidase (bacteriophage T5) (also known as L-alanoyl-D-glutamate endopeptidase), and Ply118 and Ply500 L-Ala-D-Glu peptidase. Uniport : ENLYS_BPMD2 Endolysin A OS=Mycobacterium phage D29 OX=28369 GN=10 PE=1 SV=1 PDB : L-alanyl-D-glutamate peptidase; l-alanoyl-d-glutamate peptidase, bacteriophage T5, endolysin, Zn2+ and Ca2+ containing form, ANTIMICROBIAL PROTEIN; HET: ZN, CA; NMR {Escherichia phage T5} Cter HHPRED hits PDB N-acetylmuramoyl-L-alanine amidase domain</p>																																																																																																																																																																																																																																																																																																		

Vis	Hits	Aln	Select All	Forward	Forward Query A3M	Model using selection	Download HHR	Color Seqs	Wrap Seqs		
<input type="checkbox"/>	1	O64203		ENLYS_BPMD2 Endolysin A OS=Mycobacterium phage D29 OX=28369 GN=10 PE=1 SV=1		99.38	1.2e-12	123.02	6.1	123	493
<input type="checkbox"/>	2	PF13539.10		; Peptidase_M15_4; D-alanyl-D-alanine carboxypeptidase		99.12	3.2e-10	80.22	5.8	58	81
<input type="checkbox"/>	3	Q37976		AEPE_BPA18 L-alanyl-D-glutamate peptidase OS=Listeria phage A118 OX=40521 GN=ply PE=2 SV=2		98.77	1.8e-7	81.14	11.1	63	281
<input type="checkbox"/>	4	Q6QGP7		ENLYS_BPT5 L-alanyl-D-glutamate peptidase OS=Escherichia phage T5 OX=10726 GN=lys PE=1 SV=1		98.68	1.4e-7	73.7	6.8	59	137
<input type="checkbox"/>	5	8P3A_A		L-alanyl-D-glutamate peptidase; l-alanoyl- d-glutamate peptidase, bacteriophage T5, endolysin, Zn2+ and Ca2+ containing f		98.56	2.6e-7	71.88	5.2	61	137
<input type="checkbox"/>	6	5OPZ_A		ChiX; L-Ala D-Glu endopeptidase Serratia marcescens chitinase secretion anomalous dispersion Zinc enzyme, HYDROLASE; 1.3		98.49	0.0000015	67.42	8	62	137
<input type="checkbox"/>	7	Q05285		ENLYS_BPMLS Endolysin A OS=Mycobacterium phage L5 OX=31757 GN=10 PE=3 SV=1		98.3	8.6e-7	78.88	2.6	133	292
<input type="checkbox"/>	8	Q37979		AEPE_BPA50 L-alanyl-D-glutamate peptidase OS=Listeria phage A500 OX=40522 GN=ply PE=1 SV=1		98.24	0.000003	75.02	4.8	62	289
<input type="checkbox"/>	9	cd14845		L-Ala-D-Glu_peptidase_like; L-Ala-D-Glu peptidase, also known as L-alanyl-D- glutamate endopeptidase. This L-Ala-D- Glu pe		97.88	0.000079	55.4	5.7	59	129

Visualization

Resubmit Section

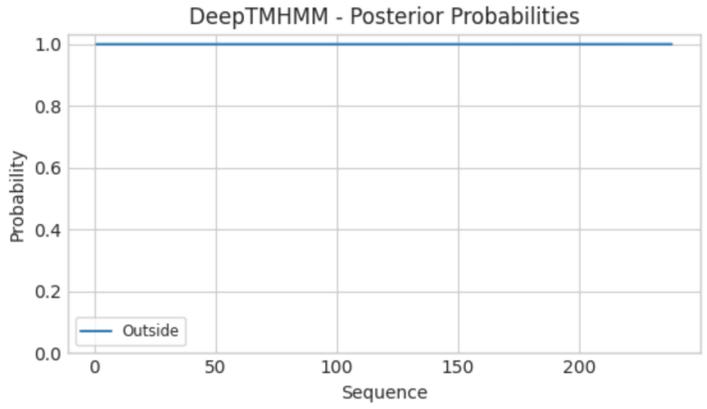
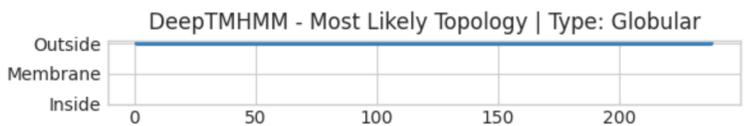


Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

Oui le gene avant code pour une portal protein

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

Non



Phobius prediction

Prediction of UNNAMED

```
ID UNNAMED
FT TOPO_DOM 1 238 NON CYTOPLASMIC.
//
```

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

OUI

DECISION:

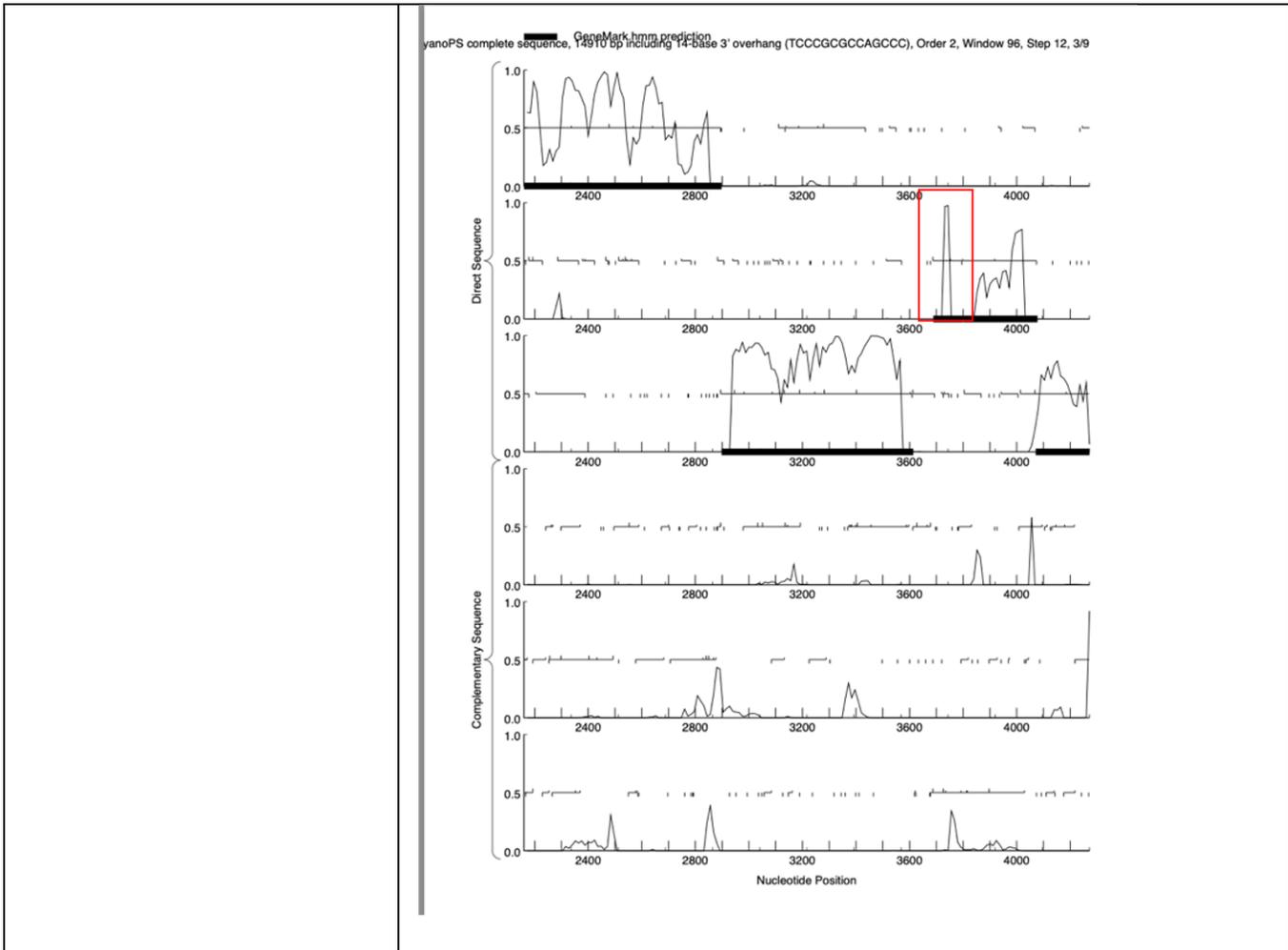
Endolysine

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	5
Coordonnées du Stop	3799
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap de 74
Coordonnées du Start retenu	3689
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>Original GeneMark call @bp 3689</i>
Y-a-t-il des éléments supportant un potentiel codant ?	GeneMarkS : potentiel codant détecté



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

NON

⚠ No significant similarity found. For reasons why [click here](#)

Query= CyranoPS_5
 (36 letters)

Distribution of 2 Blast Hits on the Query Sequence



Sequences producing significant alignments:

Sequence	Score (bits)	E Value
CyranoPS_Draft_5, function unknown, 36	68	9e-12
Camille_35, function unknown, 126	22	4.7

>CyranoPS_Draft_5, function unknown, 36
 Length = 36

Score = 67.8 bits (164), Expect = 9e-12
 Identities = 36/36 (100%), Positives = 36/36 (100%)

Query: 1 MALDLDLALFVASAIGGTVSAVAVLTAVALMLLKEIT 36
 MALDLDLALFVASAIGGTVSAVAVLTAVALMLLKEIT

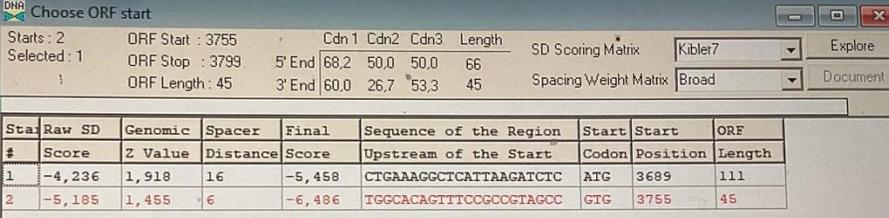
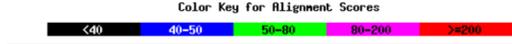
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

Conflit au niveau du potentiel codant chez GeneMark et Glimmer

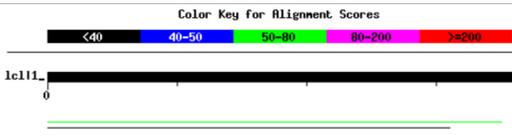
DECISION:

si on retire ce gene le gap entre les ORF 5 et ORF 7 sera énorme, alors on le garde

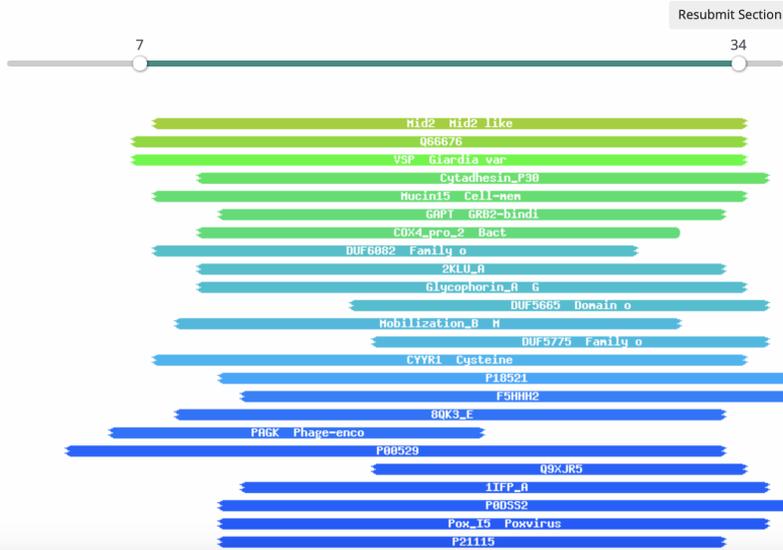
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle									
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</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>									
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>										
<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>	<p>ORF : ATG3689 la plus longue = 111pb meilleur score , gap de 74pb</p>									
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>NA</p>									
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	 <p>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th></th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>CyranoPS_Draft_5, function unknown, 36</td> <td>68</td> <td>9e-12</td> </tr> <tr> <td>Camille_35, function unknown, 126</td> <td>29</td> <td>4.7</td> </tr> </tbody> </table> <p>>CyranoPS_Draft_5, function unknown, 36 Length = 36</p> <p>Score = 67.8 bits (164), Expect = 9e-12 Identities = 36/36 (100%), Positives = 36/36 (100%)</p> <p>Query: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT Sbjct: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36</p> <p>>Camille_35, function unknown, 126 Length = 126</p> <p>Score = 28.9 bits (63), Expect = 4.7 Identities = 15/32 (46%), Positives = 23/32 (71%)</p> <p>Query: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLL 32 +AL L+A ++ SA G VAV++AVALW++ Sbjct: 29 VALILIAGYLESAPLGAALWVAVILAVALLWIV 60</p>		Score (bits)	E Value	CyranoPS_Draft_5, function unknown, 36	68	9e-12	Camille_35, function unknown, 126	29	4.7
	Score (bits)	E Value								
CyranoPS_Draft_5, function unknown, 36	68	9e-12								
Camille_35, function unknown, 126	29	4.7								
<p>DECISION:</p>	<p>Indiquez le start retenu et justifiez brièvement.</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>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th></th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>CyranoPS_Draft_5, function unknown, 36</td> <td>68</td> <td>9e-12</td> </tr> <tr> <td>Camille_35, function unknown, 126</td> <td>29</td> <td>4.7</td> </tr> </tbody> </table> <p>>CyranoPS_Draft_5, function unknown, 36 Length = 36</p> <p>Score = 67.8 bits (164), Expect = 9e-12 Identities = 36/36 (100%), Positives = 36/36 (100%)</p> <p>Query: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT Sbjct: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36</p> <p>>Camille_35, function unknown, 126 Length = 126</p> <p>Score = 28.9 bits (63), Expect = 4.7 Identities = 15/32 (46%), Positives = 23/32 (71%)</p> <p>Query: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLL 32 +AL L+A ++ SA G VAV++AVALW++ Sbjct: 29 VALLLIAGYLESAPLGAALWVAVILAVALWIV 60</p> <p>*** nr :</p> <p>RIEN</p> <div style="background-color: #fff9c4; padding: 10px; border: 1px solid #ccc; display: flex; align-items: center;">  <p style="font-size: 1.2em; font-weight: bold;">No significant similarity found.</p> </div>		Score (bits)	E Value	CyranoPS_Draft_5, function unknown, 36	68	9e-12	Camille_35, function unknown, 126	29	4.7
	Score (bits)	E Value								
CyranoPS_Draft_5, function unknown, 36	68	9e-12								
Camille_35, function unknown, 126	29	4.7								
<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>"Pas de correspondance pertinente"</p>									

Visualization

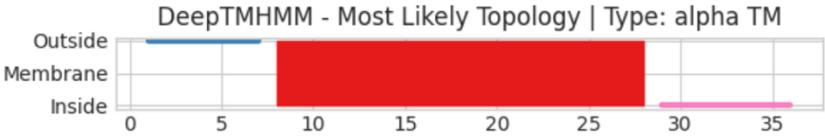
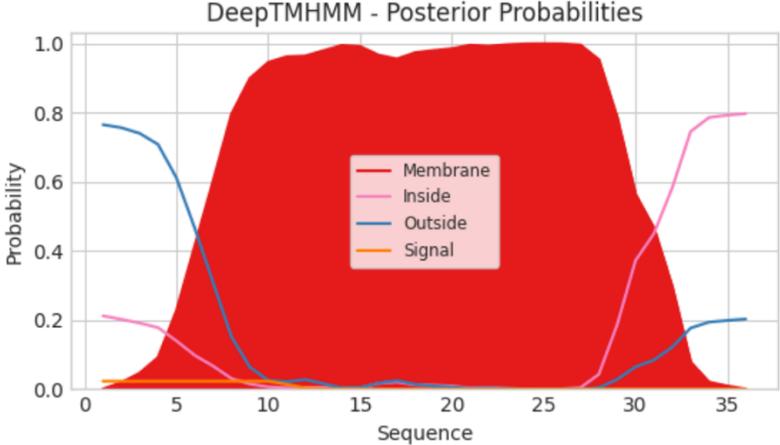


PFAM proba de 79%. evalue de 13 : Mid2 ; Mid2 like cell wall stress sensor

Uniprot proba de 77% evalue de 18 : VGE9_EHV2 Uncharacterized gene E9 protein OS=Equine herpesvirus 2 (strain 86/87) OX=82831 GN=E9 PE=3 SV=1

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 le gène avant est une endolysine

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>OUI</p> <p>DeepTMHMM - Predictions</p> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p>  
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>non</p>
<p>DECISION:</p>	<p>Membrane protein</p>

Student Gene Annotation Worksheet

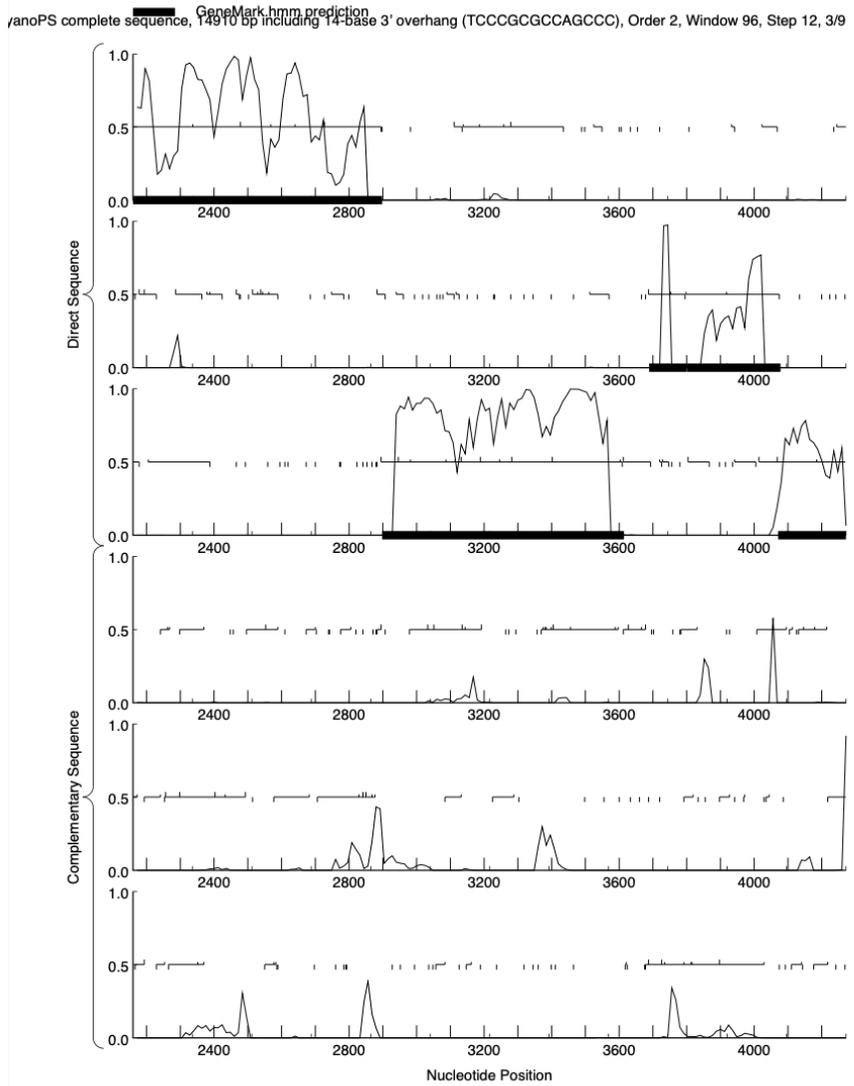
Basic Phage Information

Nom du Phage	CyranoPS
Gène #	6
Coordonnées du Stop	4078
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap de 0 si on conserve le gène 6 (Si on retire le 6 ; GAP de 185pb)
Coordonnées du Start retenu	3800
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

GeneMarkS : il ya du potentiel codant tout le long



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

Distribution of 75 Blast Hits on the Query Sequence

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

Color Key for Alignment Scores



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

Sequences producing significant alignments:

Sequence	Score (bits)	E Value
CyanoPS_Draft_6, function unknown, 52	109	2e-24
RRH1_07, caudovirus prohead protease family protein, 661	50	1e-06
Schiebs_7, function unknown, 73	43	2e-04

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

Décision #2 : Quel est le start du gène ?

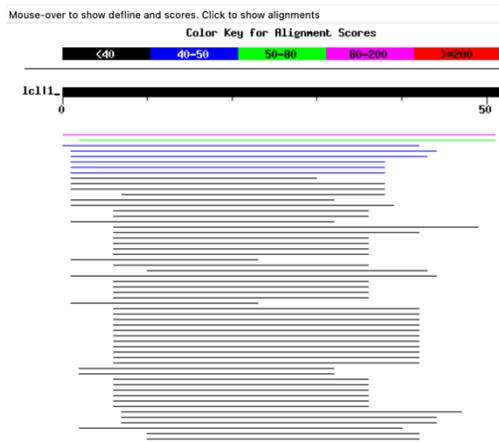
Collection des éléments de réponse	Rationnelle																																				
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 3920 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 3800																																				
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	<div style="border: 1px solid black; padding: 5px;"> <p>Starts: 2 ORF Start : 3920 Cdn 1 Cdn 2 Cdn 3 Length SD Scoring Matrix Kibler7 Explore Selected: 1 ORF Stop : 4078 5' End 65,0 47,5 52,5 120 ORF Length: 159 3' End 60,4 45,3 45,3 159 Spacing Weight Matrix Broad Document</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <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>-1,518</td> <td>3,244</td> <td>13</td> <td>-2,589</td> <td>GCTACTAAAGGAAATCAGTAA</td> <td>GTG</td> <td>3800</td> <td>279</td> </tr> <tr> <td>2</td> <td>-6,367</td> <td>0,878</td> <td>13</td> <td>-7,437</td> <td>CATAGCCCGCGAGTCCGGCTTA</td> <td>GTG</td> <td>3920</td> <td>159</td> </tr> </tbody> </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	-1,518	3,244	13	-2,589	GCTACTAAAGGAAATCAGTAA	GTG	3800	279	2	-6,367	0,878	13	-7,437	CATAGCCCGCGAGTCCGGCTTA	GTG	3920	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	-1,518	3,244	13	-2,589	GCTACTAAAGGAAATCAGTAA	GTG	3800	279																													
2	-6,367	0,878	13	-7,437	CATAGCCCGCGAGTCCGGCTTA	GTG	3920	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 ?	ORf la plus longue ☑ 279pb GTG 3800																																				
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 ?	non																																				
DECISION:	GTG 3800																																				

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	*** PhagesDB :

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

Distribution of 73 Blast Hits on the Query Sequence

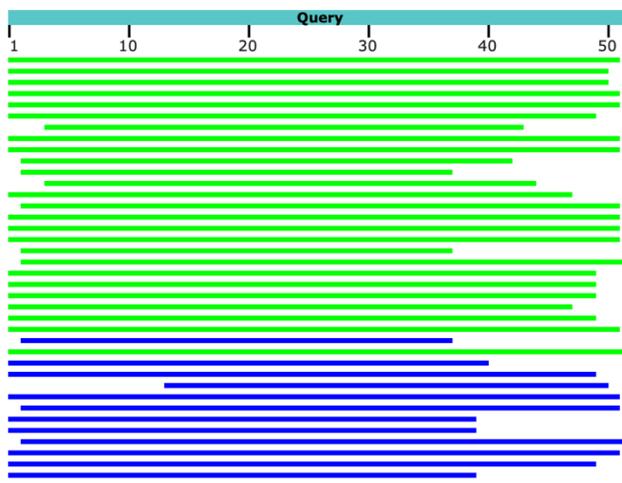


Sequences producing significant alignments:

	Score (bits)	E Value
CyranoPS_Draft_6, function unknown, 52	109	2e-24
RRH1_07, caudovirus prohead protease family protein, 661	50	1e-06
Schiebs_7, function unknown, 73	43	2e-04

*** nr : il n'y a que des hypothetical protein meme plus bas que sur la capture d'écran

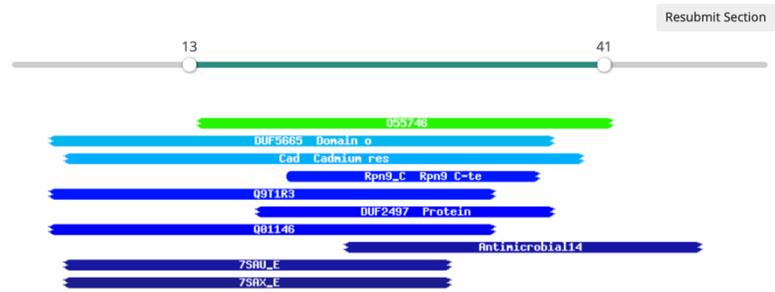
Distribution of the top 99 Blast Hits on 99 subject sequences



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium sp. EPI-003-04-2554_SCH2473...	Corynebacterium sp. EPI-003-04-2554_SCH2...	60.1	60.1	98%	9e-10	56.60%	90	WP_064833794.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	58.2	58.2	96%	4e-09	52.00%	90	WP_095275390.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	58.2	58.2	96%	6e-09	54.00%	93	DAI73899.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	57.8	57.8	98%	8e-09	54.72%	91	WP_284587138.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	57.8	57.8	98%	8e-09	54.72%	91	WP_284849032.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	54.3	54.3	94%	2e-07	46.94%	91	WP_150114276.1
<input checked="" type="checkbox"/> hypothetical protein [Propionivibrio sp.]	Propionivibrio sp.	53.9	53.9	76%	2e-07	57.50%	96	MBK7565154.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	53.9	53.9	98%	3e-07	49.02%	102	HDV6349140.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	53.9	53.9	98%	3e-07	49.02%	102	HCG2978162.1
<input checked="" type="checkbox"/> hypothetical protein U2A4042570135 [Corynebacterium striatum]	Corynebacterium striatum	52.4	52.4	78%	9e-07	60.98%	97	CQD13966.1
<input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	52.4	52.4	69%	1e-06	66.67%	97	WP_083314906.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	52.0	52.0	78%	1e-06	58.54%	100	DAX19516.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	51.6	51.6	90%	2e-06	53.06%	89	WP_035111211.1
<input checked="" type="checkbox"/> hypothetical protein [Nocardioides massiliensis]	Nocardioides massiliensis	52.4	52.4	96%	2e-06	48.00%	132	WP_220138296.1
<input checked="" type="checkbox"/> hypothetical protein [Tomitella bifornata]	Tomitella bifornata	51.6	51.6	98%	2e-06	42.31%	94	WP_024794453.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	50.4	50.4	98%	3e-06	41.18%	68	HAT1137148.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	50.1	50.1	98%	4e-06	41.18%	68	HAT1153063.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	50.8	50.8	69%	4e-06	63.89%	97	HCT5226543.1

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 $\geq 90\%$ et une couverture acceptable ?

Visualization



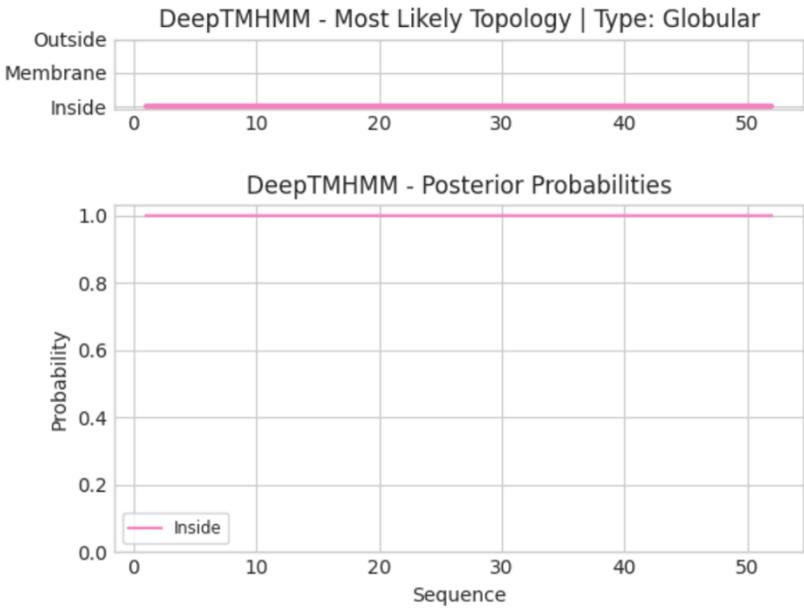
Hitlist

Show 25 Entries Search:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	
<input type="checkbox"/>	1	O55746	140L_IIV6 Uncharacterized protein 140L OS=Invertebrate iridescent virus 6 OX=176652 GN=IIV6-140L PE=4 SV=1	72.11	21	21.47	3.5	29	64
<input type="checkbox"/>	2	PF18910.4	; DUF5665 ; Domain of unknown function (DUF5665)	55.86	76	18.44	3.6	35	58
<input type="checkbox"/>	3	PF03596.17	; Cad ; Cadmium resistance transporter	54.7	70	22.09	3.9	36	191
<input type="checkbox"/>	4	PF18261.5	; Rpn9_C ; Rpn9 C-terminal helix	42.08	36	16.23	0.8	18	33
<input type="checkbox"/>	5	Q9T1R3	VP35_BPAPS Putative DNA transfer protein p35 OS=Acyrtosiphon pisum secondary endosymbiont phage 1 OX=67571 GN=35 PE=3 S	41.79	140	24.08	4.3	31	625

Uniprot proba de 72% : 140L_IIV6 Uncharacterized protein 140L OS=Invertebrate iridescent virus 6 OX=176652 GN=IIV6-140L PE=4 SV=1

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>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>DeepTMHMM - Predictions</p> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p> 
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>Non</p>
<p>DECISION:</p>	<p>NKF</p>

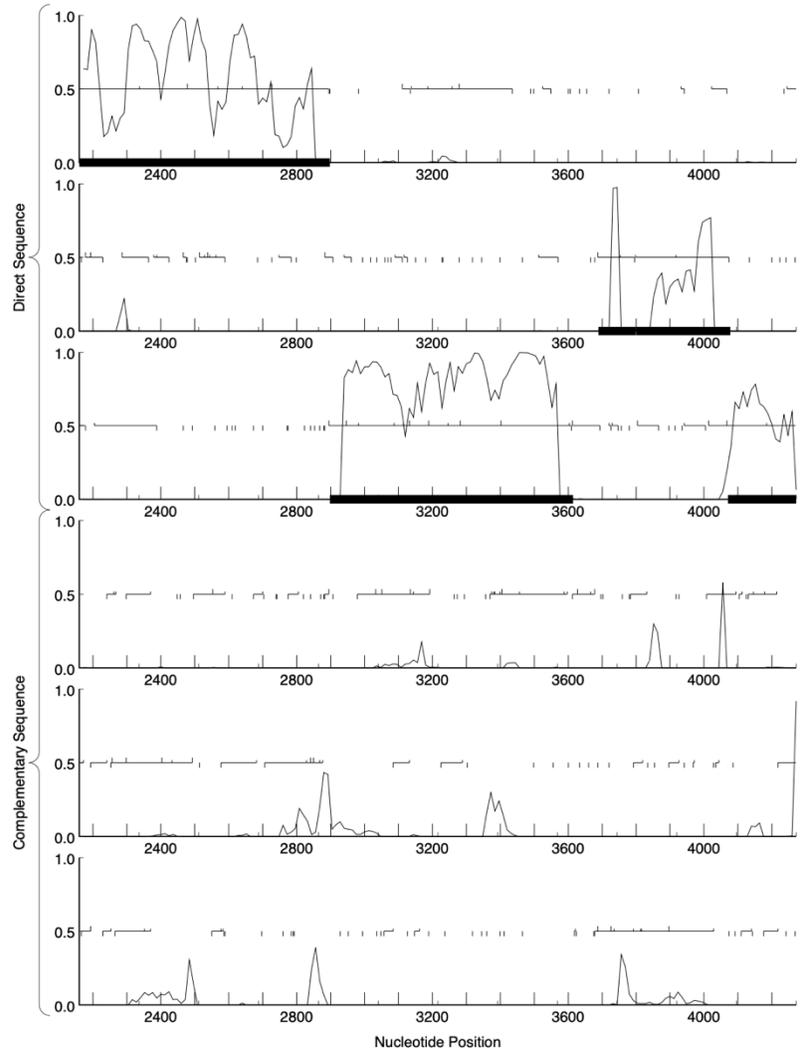
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	7
Coordonnées du Stop	5504
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove 8
Coordonnées du Start retenu	4071
Fonction prédite	major capsid and protease fusion protein

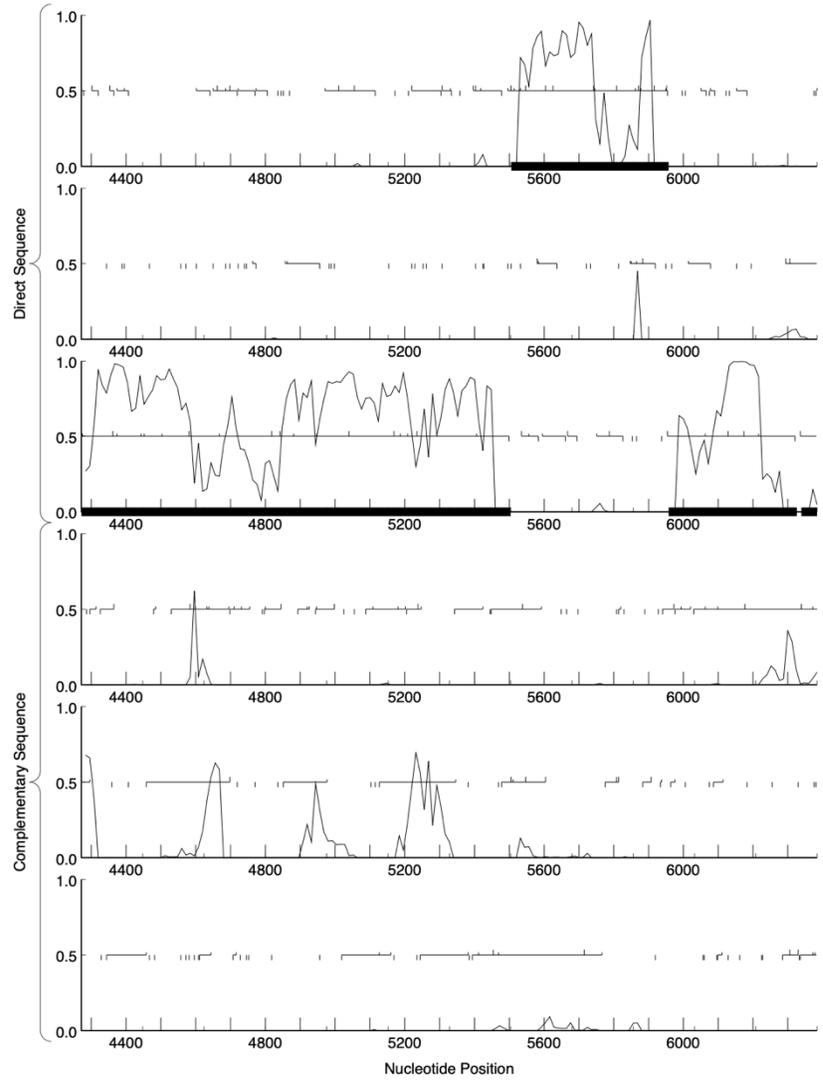
Décision #1 : Est-ce un gène ?

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 ?	GeneMarkS :

GeneMark hmmprediction
yanoPS complete sequence, 14910 bp, including 14-base 3' overhang (TCCC GCGCCAGCCC), Order 2, Window 96, Step 12, 3/9

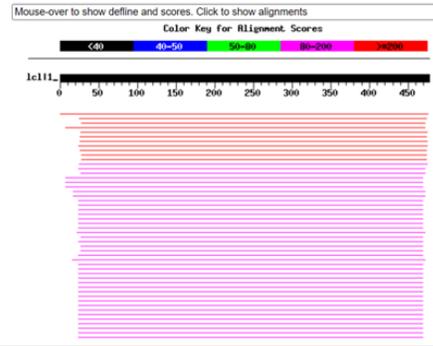


GenoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 4/9



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

Distribution of 102 Blast Hits on the Query Sequence



Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyranopS_Draft_7, function unknown, 477	531	0.0
EpiDab_9, major capsid and protease fusion protein, 628	387	e-107
SallySpecial_8, major capsid and capsid maturation protease, 635	350	9e-99
Emperor_10, capsid and capsid maturation protease, 619	356	6e-98
Schiebs_8, major capsid and protease fusion protein, 539	306	7e-83
RRH1_07, caudovirus prohead protease family protein, 661	225	1e-79
Coeur_7, major capsid protein and capsid maturation protease, 683	243	7e-64
Rahu1_7, major capsid protein and protease fusion protein, 689	251	2e-63

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

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 [Corynebacterium sp. EPI-003-04-2554_SCH2473622]	Corynebacterium sp. EPI-003-04-2554_SCH2...	546	546	98%	0.0	57.50%	487	WP_064833792.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	543	543	98%	0.0	57.08%	487	WP_284587137.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	540	540	96%	0.0	57.87%	480	WP_284849033.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium ericinoum]	Corynebacterium ericinoum	536	536	98%	0.0	56.34%	488	WP_302524537.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium ericinoum]	Corynebacterium ericinoum	535	535	98%	0.0	56.76%	488	WP_284594222.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	533	533	96%	0.0	53.65%	523	HCD4134811.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	533	533	96%	0.0	54.06%	521	WP_201806828.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	530	530	96%	0.0	52.71%	532	HCD1553130.1
<input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	512	512	96%	2e-175	53.88%	492	WP_070736124.1
<input checked="" type="checkbox"/> conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	483	483	97%	1e-163	49.81%	518	CGD13968.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium ocaii]	Corynebacterium ocaii	412	412	93%	1e-135	45.77%	523	WP_055121350.1
<input checked="" type="checkbox"/> hypothetical protein FRC0474_52082 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	410	410	98%	1e-135	48.09%	476	CAB0972127.1

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																																																																																																																																																																																																																																																																																																																											
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 4071</p> <p>Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 4071</p>																																																																																																																																																																																																																																																																																																																											
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	<table border="1" style="width: 100%; border-collapse: collapse; font-size: small;"> <thead> <tr> <th>Sta</th> <th>Raw 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> </thead> <tbody> <tr><td>1</td><td>-5,105</td><td>1,494</td><td>12</td><td>-6,151</td><td>GTCGAAAACCTGGCTAAACAAAT</td><td>ATG</td><td>4017</td><td>1488</td></tr> <tr><td>2</td><td>-2,310</td><td>2,857</td><td>8</td><td>-3,435</td><td>GGCAAAACATCGAAAGGACGATA</td><td>ATG</td><td>4071</td><td>1434</td></tr> <tr><td>3</td><td>-6,430</td><td>0,847</td><td>14</td><td>-7,527</td><td>GGACGCCCCGGAACCTCACTGGT</td><td>TTG</td><td>4185</td><td>1320</td></tr> <tr><td>4</td><td>-5,105</td><td>1,494</td><td>6</td><td>-6,406</td><td>CGCCCGGGAACCTCACTGGTITG</td><td>GTG</td><td>4188</td><td>1317</td></tr> <tr><td>5</td><td>-5,105</td><td>1,494</td><td>9</td><td>-6,151</td><td>CCGCGAACCTCACTGGTITGGTG</td><td>TTG</td><td>4191</td><td>1314</td></tr> <tr><td>6</td><td>-3,422</td><td>2,315</td><td>13</td><td>-4,492</td><td>GATCCCGAAGATCTTAAGCGC</td><td>GTG</td><td>4278</td><td>1227</td></tr> <tr><td>7</td><td>-2,729</td><td>2,653</td><td>16</td><td>-3,951</td><td>AATTAAGAGATGGCCTATAC</td><td>ATG</td><td>4365</td><td>1140</td></tr> <tr><td>8</td><td>-3,784</td><td>2,138</td><td>11</td><td>-4,784</td><td>TGGCCTATACATGAGTTTCOGI</td><td>GTG</td><td>4377</td><td>1128</td></tr> <tr><td>9</td><td>-5,953</td><td>1,080</td><td>14</td><td>-7,050</td><td>GGTATCCGAGACGCTCTTTCC</td><td>GTG</td><td>4446</td><td>1059</td></tr> <tr><td>10</td><td>-3,043</td><td>2,500</td><td>7</td><td>-4,264</td><td>AGACGCTCTTTCCGTTGGAACG</td><td>GTG</td><td>4455</td><td>1050</td></tr> <tr><td>11</td><td>-5,062</td><td>1,515</td><td>13</td><td>-6,132</td><td>CACCGCAGCGCAGCTTTCCGCC</td><td>GTG</td><td>4506</td><td>999</td></tr> <tr><td>12</td><td>-5,029</td><td>1,531</td><td>18</td><td>-6,427</td><td>GCTCGAGCCGCTAAGTCTGTC</td><td>ATG</td><td>4584</td><td>921</td></tr> <tr><td>13</td><td>-6,616</td><td>0,756</td><td>16</td><td>-7,838</td><td>CGCTAAGTCTGTGATGTCOCGA</td><td>TTG</td><td>4593</td><td>912</td></tr> <tr><td>14</td><td>-5,145</td><td>1,474</td><td>13</td><td>-6,215</td><td>TGSCAACACGACACCTCCACC</td><td>TTG</td><td>4641</td><td>864</td></tr> <tr><td>15</td><td>-5,219</td><td>1,438</td><td>15</td><td>-6,374</td><td>ACTTCGTGAAGTCTCCATGCT</td><td>GTG</td><td>4668</td><td>837</td></tr> <tr><td>16</td><td>-5,219</td><td>1,438</td><td>18</td><td>-6,617</td><td>TCGTGAAGTCTCCATGCTTTG</td><td>GTG</td><td>4671</td><td>834</td></tr> <tr><td>17</td><td>-5,715</td><td>1,196</td><td>18</td><td>-7,113</td><td>GAACGGTCAACCCCTACACCCGC</td><td>TTG</td><td>4806</td><td>699</td></tr> <tr><td>18</td><td>-5,447</td><td>1,326</td><td>14</td><td>-6,544</td><td>CACCCGCTTGATTTGCCAACCT</td><td>ATG</td><td>4821</td><td>684</td></tr> <tr><td>19</td><td>-4,780</td><td>1,652</td><td>15</td><td>-5,935</td><td>GGGAACCTGCAACCCCTAACCGAC</td><td>ATG</td><td>4845</td><td>660</td></tr> <tr><td>20</td><td>-4,147</td><td>1,961</td><td>17</td><td>-5,448</td><td>GCGGTGGGTAAACCCGCCCCGAG</td><td>GTG</td><td>4884</td><td>621</td></tr> <tr><td>21</td><td>-5,049</td><td>1,521</td><td>13</td><td>-6,119</td><td>CCCACGACCCGAGCCTTACACC</td><td>GTG</td><td>4950</td><td>555</td></tr> <tr><td>22</td><td>-6,135</td><td>0,991</td><td>7</td><td>-7,357</td><td>TGGGGAATTTAATCAGTGTGAT</td><td>TTG</td><td>5034</td><td>471</td></tr> <tr style="color: red;"><td>23</td><td>-6,135</td><td>0,991</td><td>16</td><td>-7,357</td><td>TATTCAGTCTACTTGGCGAGA</td><td>ATG</td><td>5043</td><td>462</td></tr> <tr><td>24</td><td>-4,763</td><td>1,660</td><td>7</td><td>-5,985</td><td>AGAGTGGACCTCACAGGCGCGAC</td><td>TTG</td><td>5139</td><td>366</td></tr> <tr><td>25</td><td>-3,846</td><td>2,108</td><td>13</td><td>-4,917</td><td>TGCAGGTAAAGGCACGCCAGATT</td><td>ATG</td><td>5172</td><td>333</td></tr> <tr><td>26</td><td>-4,439</td><td>1,819</td><td>16</td><td>-5,661</td><td>GATTATGAGCGCAACCCGCGGC</td><td>GTG</td><td>5190</td><td>315</td></tr> <tr><td>27</td><td>-3,043</td><td>2,500</td><td>16</td><td>-4,264</td><td>GCGGTGGAACCAACCCACTAT</td><td>TTG</td><td>5208</td><td>297</td></tr> <tr><td>28</td><td>-6,594</td><td>0,767</td><td>15</td><td>-7,748</td><td>CGTGGAAACCAACCACTAATTTG</td><td>GTG</td><td>5211</td><td>294</td></tr> <tr><td>29</td><td>-5,033</td><td>1,529</td><td>16</td><td>-6,254</td><td>CTCCGATGATCTGTTGATCTC</td><td>ATG</td><td>5238</td><td>267</td></tr> <tr><td>30</td><td>-5,208</td><td>1,443</td><td>13</td><td>-6,279</td><td>GCAGCTTGAAAGCCCTGAATTT</td><td>TTG</td><td>5271</td><td>234</td></tr> <tr><td>31</td><td>-4,503</td><td>1,787</td><td>7</td><td>-5,725</td><td>AAAGCCCTGAATTTTGGATCTC</td><td>TTG</td><td>5280</td><td>225</td></tr> <tr><td>32</td><td>-1,983</td><td>3,017</td><td>7</td><td>-3,205</td><td>AATCGCTTACCACAAGGAAGCG</td><td>TTG</td><td>5358</td><td>147</td></tr> <tr><td>33</td><td>-5,619</td><td>1,243</td><td>7</td><td>-6,841</td><td>TATCCGAGTCAATGCGGATGAC</td><td>GTG</td><td>5409</td><td>96</td></tr> <tr><td>34</td><td>-6,633</td><td>0,748</td><td>9</td><td>-7,679</td><td>GTTTGGCTACTACGGCCATAC</td><td>TTG</td><td>5460</td><td>45</td></tr> </tbody> </table>	Sta	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length	1	-5,105	1,494	12	-6,151	GTCGAAAACCTGGCTAAACAAAT	ATG	4017	1488	2	-2,310	2,857	8	-3,435	GGCAAAACATCGAAAGGACGATA	ATG	4071	1434	3	-6,430	0,847	14	-7,527	GGACGCCCCGGAACCTCACTGGT	TTG	4185	1320	4	-5,105	1,494	6	-6,406	CGCCCGGGAACCTCACTGGTITG	GTG	4188	1317	5	-5,105	1,494	9	-6,151	CCGCGAACCTCACTGGTITGGTG	TTG	4191	1314	6	-3,422	2,315	13	-4,492	GATCCCGAAGATCTTAAGCGC	GTG	4278	1227	7	-2,729	2,653	16	-3,951	AATTAAGAGATGGCCTATAC	ATG	4365	1140	8	-3,784	2,138	11	-4,784	TGGCCTATACATGAGTTTCOGI	GTG	4377	1128	9	-5,953	1,080	14	-7,050	GGTATCCGAGACGCTCTTTCC	GTG	4446	1059	10	-3,043	2,500	7	-4,264	AGACGCTCTTTCCGTTGGAACG	GTG	4455	1050	11	-5,062	1,515	13	-6,132	CACCGCAGCGCAGCTTTCCGCC	GTG	4506	999	12	-5,029	1,531	18	-6,427	GCTCGAGCCGCTAAGTCTGTC	ATG	4584	921	13	-6,616	0,756	16	-7,838	CGCTAAGTCTGTGATGTCOCGA	TTG	4593	912	14	-5,145	1,474	13	-6,215	TGSCAACACGACACCTCCACC	TTG	4641	864	15	-5,219	1,438	15	-6,374	ACTTCGTGAAGTCTCCATGCT	GTG	4668	837	16	-5,219	1,438	18	-6,617	TCGTGAAGTCTCCATGCTTTG	GTG	4671	834	17	-5,715	1,196	18	-7,113	GAACGGTCAACCCCTACACCCGC	TTG	4806	699	18	-5,447	1,326	14	-6,544	CACCCGCTTGATTTGCCAACCT	ATG	4821	684	19	-4,780	1,652	15	-5,935	GGGAACCTGCAACCCCTAACCGAC	ATG	4845	660	20	-4,147	1,961	17	-5,448	GCGGTGGGTAAACCCGCCCCGAG	GTG	4884	621	21	-5,049	1,521	13	-6,119	CCCACGACCCGAGCCTTACACC	GTG	4950	555	22	-6,135	0,991	7	-7,357	TGGGGAATTTAATCAGTGTGAT	TTG	5034	471	23	-6,135	0,991	16	-7,357	TATTCAGTCTACTTGGCGAGA	ATG	5043	462	24	-4,763	1,660	7	-5,985	AGAGTGGACCTCACAGGCGCGAC	TTG	5139	366	25	-3,846	2,108	13	-4,917	TGCAGGTAAAGGCACGCCAGATT	ATG	5172	333	26	-4,439	1,819	16	-5,661	GATTATGAGCGCAACCCGCGGC	GTG	5190	315	27	-3,043	2,500	16	-4,264	GCGGTGGAACCAACCCACTAT	TTG	5208	297	28	-6,594	0,767	15	-7,748	CGTGGAAACCAACCACTAATTTG	GTG	5211	294	29	-5,033	1,529	16	-6,254	CTCCGATGATCTGTTGATCTC	ATG	5238	267	30	-5,208	1,443	13	-6,279	GCAGCTTGAAAGCCCTGAATTT	TTG	5271	234	31	-4,503	1,787	7	-5,725	AAAGCCCTGAATTTTGGATCTC	TTG	5280	225	32	-1,983	3,017	7	-3,205	AATCGCTTACCACAAGGAAGCG	TTG	5358	147	33	-5,619	1,243	7	-6,841	TATCCGAGTCAATGCGGATGAC	GTG	5409	96	34	-6,633	0,748	9	-7,679	GTTTGGCTACTACGGCCATAC	TTG	5460	45
Sta	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length																																																																																																																																																																																																																																																																																																																				
1	-5,105	1,494	12	-6,151	GTCGAAAACCTGGCTAAACAAAT	ATG	4017	1488																																																																																																																																																																																																																																																																																																																				
2	-2,310	2,857	8	-3,435	GGCAAAACATCGAAAGGACGATA	ATG	4071	1434																																																																																																																																																																																																																																																																																																																				
3	-6,430	0,847	14	-7,527	GGACGCCCCGGAACCTCACTGGT	TTG	4185	1320																																																																																																																																																																																																																																																																																																																				
4	-5,105	1,494	6	-6,406	CGCCCGGGAACCTCACTGGTITG	GTG	4188	1317																																																																																																																																																																																																																																																																																																																				
5	-5,105	1,494	9	-6,151	CCGCGAACCTCACTGGTITGGTG	TTG	4191	1314																																																																																																																																																																																																																																																																																																																				
6	-3,422	2,315	13	-4,492	GATCCCGAAGATCTTAAGCGC	GTG	4278	1227																																																																																																																																																																																																																																																																																																																				
7	-2,729	2,653	16	-3,951	AATTAAGAGATGGCCTATAC	ATG	4365	1140																																																																																																																																																																																																																																																																																																																				
8	-3,784	2,138	11	-4,784	TGGCCTATACATGAGTTTCOGI	GTG	4377	1128																																																																																																																																																																																																																																																																																																																				
9	-5,953	1,080	14	-7,050	GGTATCCGAGACGCTCTTTCC	GTG	4446	1059																																																																																																																																																																																																																																																																																																																				
10	-3,043	2,500	7	-4,264	AGACGCTCTTTCCGTTGGAACG	GTG	4455	1050																																																																																																																																																																																																																																																																																																																				
11	-5,062	1,515	13	-6,132	CACCGCAGCGCAGCTTTCCGCC	GTG	4506	999																																																																																																																																																																																																																																																																																																																				
12	-5,029	1,531	18	-6,427	GCTCGAGCCGCTAAGTCTGTC	ATG	4584	921																																																																																																																																																																																																																																																																																																																				
13	-6,616	0,756	16	-7,838	CGCTAAGTCTGTGATGTCOCGA	TTG	4593	912																																																																																																																																																																																																																																																																																																																				
14	-5,145	1,474	13	-6,215	TGSCAACACGACACCTCCACC	TTG	4641	864																																																																																																																																																																																																																																																																																																																				
15	-5,219	1,438	15	-6,374	ACTTCGTGAAGTCTCCATGCT	GTG	4668	837																																																																																																																																																																																																																																																																																																																				
16	-5,219	1,438	18	-6,617	TCGTGAAGTCTCCATGCTTTG	GTG	4671	834																																																																																																																																																																																																																																																																																																																				
17	-5,715	1,196	18	-7,113	GAACGGTCAACCCCTACACCCGC	TTG	4806	699																																																																																																																																																																																																																																																																																																																				
18	-5,447	1,326	14	-6,544	CACCCGCTTGATTTGCCAACCT	ATG	4821	684																																																																																																																																																																																																																																																																																																																				
19	-4,780	1,652	15	-5,935	GGGAACCTGCAACCCCTAACCGAC	ATG	4845	660																																																																																																																																																																																																																																																																																																																				
20	-4,147	1,961	17	-5,448	GCGGTGGGTAAACCCGCCCCGAG	GTG	4884	621																																																																																																																																																																																																																																																																																																																				
21	-5,049	1,521	13	-6,119	CCCACGACCCGAGCCTTACACC	GTG	4950	555																																																																																																																																																																																																																																																																																																																				
22	-6,135	0,991	7	-7,357	TGGGGAATTTAATCAGTGTGAT	TTG	5034	471																																																																																																																																																																																																																																																																																																																				
23	-6,135	0,991	16	-7,357	TATTCAGTCTACTTGGCGAGA	ATG	5043	462																																																																																																																																																																																																																																																																																																																				
24	-4,763	1,660	7	-5,985	AGAGTGGACCTCACAGGCGCGAC	TTG	5139	366																																																																																																																																																																																																																																																																																																																				
25	-3,846	2,108	13	-4,917	TGCAGGTAAAGGCACGCCAGATT	ATG	5172	333																																																																																																																																																																																																																																																																																																																				
26	-4,439	1,819	16	-5,661	GATTATGAGCGCAACCCGCGGC	GTG	5190	315																																																																																																																																																																																																																																																																																																																				
27	-3,043	2,500	16	-4,264	GCGGTGGAACCAACCCACTAT	TTG	5208	297																																																																																																																																																																																																																																																																																																																				
28	-6,594	0,767	15	-7,748	CGTGGAAACCAACCACTAATTTG	GTG	5211	294																																																																																																																																																																																																																																																																																																																				
29	-5,033	1,529	16	-6,254	CTCCGATGATCTGTTGATCTC	ATG	5238	267																																																																																																																																																																																																																																																																																																																				
30	-5,208	1,443	13	-6,279	GCAGCTTGAAAGCCCTGAATTT	TTG	5271	234																																																																																																																																																																																																																																																																																																																				
31	-4,503	1,787	7	-5,725	AAAGCCCTGAATTTTGGATCTC	TTG	5280	225																																																																																																																																																																																																																																																																																																																				
32	-1,983	3,017	7	-3,205	AATCGCTTACCACAAGGAAGCG	TTG	5358	147																																																																																																																																																																																																																																																																																																																				
33	-5,619	1,243	7	-6,841	TATCCGAGTCAATGCGGATGAC	GTG	5409	96																																																																																																																																																																																																																																																																																																																				
34	-6,633	0,748	9	-7,679	GTTTGGCTACTACGGCCATAC	TTG	5460	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 ?

2eme ORF le plus long

Avec un ATG

Score de -3 tandis que celui qui même au plus long orf (qui a aussi un ATG) est de -6

BlastP avec le premier start proposé :

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

hypothetical protein [Corynebacterium sp. EPI-003-04-2554_SCH2473622]

Sequence ID: [WP_064833792.1](#) Length: 487 Number of Matches: 1

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

Range 1: 3 to 477 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
546 bits(1406)	0.0	Compositional matrix adjust.	276/480(58%)	350/480(72%)	14/480(2%)
Query 24	SIL	TASQSVTATRYIQPETLTAS-----ADE	DARELTGLVLPFGKPGRTSRGKLSV	74	
	+++TAS	++I P LTAS	+E+AREL+GLVLPF KPG+T+RG SV		
Sbjct 3	NLITASAHNQPKFILPADLTASQATDDDDTTTEENARELSGLVLPFNKPGKTRNGVFSV	62			
Query 75	TASALQIPEDLKRKLYRDHSDVGGTPVGYATAAEIKEDGLYMSFRVGATPDGDAALVDV	134			
	A+++PEDL RVKLYRDHS+ GTPVG ATAAEIK+DG++MSF VGAT DGDAAL DV				
Sbjct 63	NKDATELPEDLGRVLYRDHSNDNGTPVGRATAAEIKDDGIHMSFHVGATADGDAALADV	122			
Query 135	SEGIRDALSVELVSPQISGSQITAAQLSAVAIVAVPAYEDARVQSTPKPAAAKSVMSRLV	194			
	+EGIRDALSVEL P+I+ +TAA+L+AVA+VA+PAY+DARV I +++ S +				
Sbjct 123	TEGIRDALSVELDDPEITDGIIVTAARLTAVALVALPAYDDARV--TAAAGKPRNLASTVT	180			
Query 195	TRDVI TAGNNDSTLTLREVSNALVARLQGDSTHEDLTAALGEITVSGNPKVTSPTWLGE	254			
	T I A ++ T+ L + +N +V + + H D+TAALGEIT + NP V + WLGE				
Sbjct 181	TH--INASKHEEKTMNLTKAANTIVTGMTS--ADHSDITAAALGEITNANNPAVDNVQWLGE	237			
Query 255	LWNGQPYTRLIVPTMGATLTDMLTSGWRWTRPEVDDYAGNLAEIPTNTPTTEPYTVEA	314			
	LW+G + R I+PT+ + LT + L GWRW T P VDDYAGNL EIPTN+P TE VEA				
Sbjct 238	LWNGARFQRQIIPTLSSQPLTGIKLRGWRWKTLPVDDYAGNLQEIPNTSPATEAIEVEA	297			
Query 315	KRLAGGHKLRKYIDFPNAEFIQSVLREMARSYAEKTDARAATYIVGESTTVLESTSQAD	374			
	KR+AGGHKLRKYIDFP+A FIQSVL EM SYA K+D +AA ++V E+ E T Q +				
Sbjct 298	KRIAGGHKLRKYIDFPDAGFIQSVLAEMTNSYAMKSDEKAADFVVAEAKKKKTKQPN	357			

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

—

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<pre> Range: 1-560 477 Score Expect Method Identities Positives Gaps 546 bits(1408) 0.0 Compositional matrix adjust. 276/480(58%) 350/480(72%) 14/480(2%) Query 6 SILTASQSVTATRYIQPETLTAS-----ADEDARELTGLVLPFGKPGRTSRGKLSV 56 +++TAS ++I P LTAS +E+AREL+GLVLPF KPG+T+RG SV Sbjct 3 NLITASAHNQPKFILPADLTASQATDDDDTTTEENARELSGLVLPFNKPGKTNRGVFSV 62 Query 57 TASALQIPEDLKRKLYRDHSDVGGTPVGYATAAEIKEDGLYMSFRVGPATPDGDAALVDV 116 A+++PEDL RVKLYRDHS+ GTPVG ATAAEIK+DG++MSF VGAT DGDAAL DV Sbjct 63 NKDAIELPEDLGRVKLYRDHSNDNGTPVGRATAAEIKDDGIHMSFHVGTADGDAALADV 122 Query 117 SEGIRDALSVELSPQISGSQITAAQLSAVAIVAVPAYEDARVQSTPKPAAAKSVMSRLV 176 +EGIRDALSVEL P+I+ +TAA+L+AVA+VA+PAY+DARV T +++ S + Sbjct 123 TEGIRDALSVELDDPEITDGIIVTAARLTAVALVALPAYDDARV - -TAAAGKPRNLASTVT 180 Query 177 TRDVITAGNNDTSTLTLREVSNALVARLQGDSTHEDLTAALGEITVSGNPKVTSPTWLGE 236 T I A ++ T+ L + +N +V + + H D+TAALGEIT + NP V + WLGE Sbjct 181 TH--INASKHEEKTMLTKAANTIVTGMTS-ADHSDITAAALGEITNANPAVDNVQWLGE 237 Query 237 LWNGQPYTRLIVPTMGATLTDMTLSGWRWTRPEVDDYAGNLAEIPTNTPTTEPYTVEA 296 LW+G + R I+PT+ + LT + L GWRW T P VDDYAGNL EIPTN+P TE VEA Sbjct 238 LWSGARFQRQIPTLSSQPLTGIKLRGWRWKTLPVDDYAGNLQEIPTNSPATEAIEVEA 297 Query 297 KRLAGGHKLDKRYIDFPNAEFIQSYLREMARSYAEKTDARAATYIVGESSTVLESTSQAD 356 KR+AGGHKLDKRYIDFP+A FIQSYL EM SYA K+D +AA ++V E+ E T Q + Sbjct 298 KRIAGGHKLDKRYIDFPDAGFIQSYLAEMTNSYAMKSDEKAADFVVAEAKKKEKTKQPN 357 Query 357 LLRAAGKARQIMKRTARVEPTTYLVNSDDLFDLMDITQLEKPEFLDL LGVDPAAFVDDPA 416 LL AA KARQI+K+ R+EPT +LVN DDLF LMDIT ++ PE+L LLGV+P F+ D Sbjct 358 LLHAAAKARQIIKKQVRMEPTAFLVNPDDLFLKMDITMDNPEYLKLLGVEPEKFIADDK 417 Query 417 VPAGRVIAHYHKEALKFGELPGSPIRVNADDVARGGTDNALFGYYAPYLEDSRGLVAVPFG 476 VPA VIAYHK+A+ FGELPGSPIRVNA DVARGG+D+ALFGY+A ++ D+RG+V V F Sbjct 418 VPANSVIAHYKDAITFGELPGSPIRVNAVVDVARGGSDHALFGYWAAVFSVDARGIVQVEFN 477 </pre>
	<p>DECISION: 4071</p>

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

<p>Collection des éléments de réponse</p>	<p>Rationnelle</p>																		
<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>Distribution of 102 Blast Hits on the Query Sequence</p> <p>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th>Sequence</th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>CyranoS_Draft_7, function unknown, 477</td> <td>931</td> <td>0.0</td> </tr> <tr> <td>EpicDab_9, major capsid and protease fusion protein, 628</td> <td>387</td> <td>e-107</td> </tr> <tr> <td>SallySpecial_8, major capsid and capsid maturation protease, 635</td> <td>352</td> <td>9e-99</td> </tr> <tr> <td>Emperor_10, capsid and capsid maturation protease, 619</td> <td>356</td> <td>6e-98</td> </tr> <tr> <td>Schlake_2, major capsid and protease fusion protein, 630</td> <td>306</td> <td>7e-83</td> </tr> </tbody> </table> <p>*** nr : (q#: s#) : [alignment] ; e-value :</p>	Sequence	Score (bits)	E Value	CyranoS_Draft_7, function unknown, 477	931	0.0	EpicDab_9, major capsid and protease fusion protein, 628	387	e-107	SallySpecial_8, major capsid and capsid maturation protease, 635	352	9e-99	Emperor_10, capsid and capsid maturation protease, 619	356	6e-98	Schlake_2, major capsid and protease fusion protein, 630	306	7e-83
Sequence	Score (bits)	E Value																	
CyranoS_Draft_7, function unknown, 477	931	0.0																	
EpicDab_9, major capsid and protease fusion protein, 628	387	e-107																	
SallySpecial_8, major capsid and capsid maturation protease, 635	352	9e-99																	
Emperor_10, capsid and capsid maturation protease, 619	356	6e-98																	
Schlake_2, major capsid and protease fusion protein, 630	306	7e-83																	

<input checked="" type="checkbox"/>	Ig-like domain-containing protein [Rhodococcus sp. JEGM 1318]	Rhodococcus sp. JEGM 1318	331	331	92%	4e-103	40.87%	591	WP_317825412.1
<input checked="" type="checkbox"/>	hypothetical protein [Gordonia westfalica]	Gordonia westfalica	331	331	93%	8e-103	43.63%	619	WP_074850014.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium matruchotii]	Corynebacterium matruchotii	323	323	95%	1e-101	40.13%	480	WP_126299854.1
<input checked="" type="checkbox"/>	TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	321	321	95%	1e-100	39.70%	480	DAK55032.1
<input checked="" type="checkbox"/>	TPA: prohead serine protease [Siphoviridae sp. cICv12]	Siphoviridae sp. cICv12	319	319	95%	6e-100	39.70%	482	DAD90001.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	318	318	93%	4e-99	38.56%	498	HAT8525545.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	313	313	94%	2e-97	39.75%	485	WP_290246076.1
<input checked="" type="checkbox"/>	TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	311	311	98%	6e-97	39.58%	481	DAU29195.1
<input checked="" type="checkbox"/>	TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	310	310	95%	3e-96	39.44%	482	DAD60100.1
<input checked="" type="checkbox"/>	major head protein [Gordonia phage Sally/Special]	Gordonia phage Sally/Special	314	314	93%	3e-96	43.80%	635	YP_010674629.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium amycolatum]	Corynebacterium amycolatum	301	301	60%	2e-95	51.74%	303	MCQ9175172.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium amycolatum]	Corynebacterium amycolatum	303	303	64%	3e-95	49.68%	360	MCQ9177229.1
<input checked="" type="checkbox"/>	hypothetical protein [Nocardia salmonicida]	Nocardia salmonicida	298	298	93%	4e-91	38.01%	524	WP_328659909.1
<input checked="" type="checkbox"/>	hypothetical protein VH15_05935 [Corynebacterium ulcerans]	Corynebacterium ulcerans	290	290	79%	1e-89	42.15%	389	KKO87266.1
<input checked="" type="checkbox"/>	TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	289	289	93%	5e-88	37.17%	512	DAI73898.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	285	285	94%	9e-87	34.76%	493	WP_201816512.1
<input checked="" type="checkbox"/>	TPA: prohead protease [Caudoviricetes sp.]	Caudoviricetes sp.	279	409	84%	1e-83	49.10%	544	DAL07352.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	276	399	84%	2e-82	48.75%	559	WP_284900640.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	273	395	84%	3e-81	48.03%	547	WP_284895768.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium auriscanis]	Corynebacterium auriscanis	271	271	97%	7e-81	35.14%	510	WP_282939128.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	269	326	71%	2e-80	46.95%	481	WP_302526986.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	267	392	84%	2e-79	49.46%	536	HAT1137147.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	267	392	84%	3e-79	49.46%	536	HAT8642677.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	265	390	84%	2e-78	49.10%	536	HAT1171110.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	261	261	93%	5e-77	34.58%	504	WP_284790713.1
<input checked="" type="checkbox"/>	hypothetical protein FRC0370_01633 [Corynebacterium diptheriae]	Corynebacterium diptheriae	259	259	93%	9e-77	36.36%	493	CAB0869439.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium rouxii]	Corynebacterium rouxii	259	259	93%	1e-76	35.54%	493	WP_155873811.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium ulcerans]	Corynebacterium ulcerans	259	259	93%	1e-76	35.23%	487	WP_013912151.1

Descriptions Graphic Summary Alignments Taxonomy

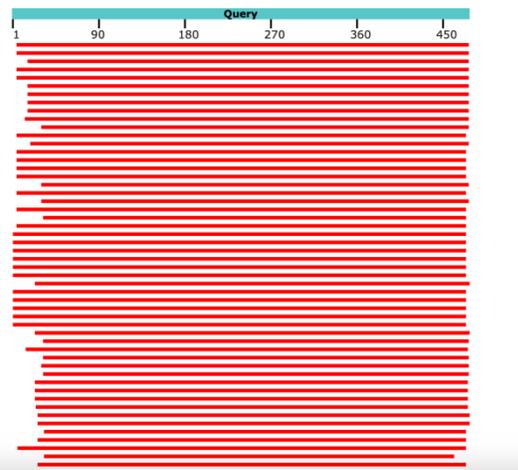
Sequences producing significant alignments Download Select columns Show 100

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium sp. EPI-003-04-255...]	Corynebacterium sp. EPI-003-04-255...	546	546	98%	0.0	57.50%	487	WP_064833792.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphtheriolum]	Corynebacterium pseudodiphtheriolum	543	543	98%	0.0	57.08%	487	WP_284587137.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphtheriolum]	Corynebacterium pseudodiphtheriolum	540	540	96%	0.0	57.87%	480	WP_284849033.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	536	536	98%	0.0	56.34%	488	WP_302524537.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	535	535	98%	0.0	56.76%	488	WP_284594222.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	533	533	96%	0.0	53.65%	523	HCD4134811.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	533	533	96%	0.0	54.06%	521	WP_201806828.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	530	530	96%	0.0	52.71%	532	HCD1553130.1
<input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	512	512	96%	2e-175	63.88%	492	WP_070736124.1
<input checked="" type="checkbox"/> conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	483	483	97%	1e-163	49.61%	518	CGQ13968.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium oculi]	Corynebacterium oculi	412	412	93%	1e-135	45.77%	523	WP_055121350.1
<input checked="" type="checkbox"/> hypothetical protein FRC0474_02082 [Corynebacterium diptheriae]	Corynebacterium diptheriae	410	410	98%	1e-135	48.00%	476	CAB0972127.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	411	411	95%	2e-135	45.97%	524	WP_337890918.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium belfanti]	Corynebacterium belfanti	409	409	98%	4e-135	47.88%	476	WP_197892068.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diptheriae]	Corynebacterium diptheriae	409	409	98%	5e-135	48.09%	476	WP_182001105.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diptheriae]	Corynebacterium diptheriae	408	408	98%	7e-135	47.88%	476	CAB0972127.1

100 sequences selected Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of the top 107 Blast Hits on 100 subject sequences



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 ?

Visualization

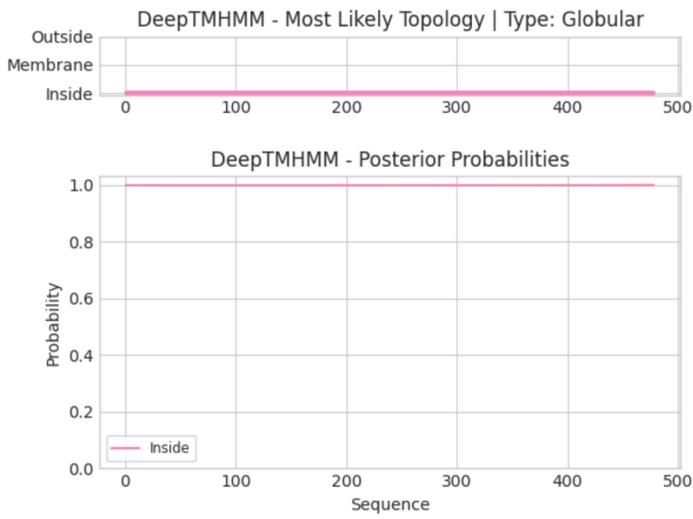


Pfam : L_lactis_ph-MCP ; Lactococcus lactis bacteriophage major capsid protein

<input type="checkbox"/>	1	PF0550	CAPSD_BPH41 Capsid polyprotein OS=Pseudomonas phage PAJL2 DK-504346 PE=1 SP=2	100	6-36	319.23	41.5	441	667
<input type="checkbox"/>	2	B0Z5F3	CAPSD_BPH41 Major capsid protein OS=Halomonas phage pho46P1 isolate - /Strf of Mesorhizobium C2021 D04-128337 GN=HalArgos PE=-	100	3.44-34	302.56	34.9	423	602
<input type="checkbox"/>	3	Q38300	CAPSD_BPHC2 Major capsid protein OS=Lactococcus phage c2 DK-31537 GN=H5 PE=4 SP=1	99.98	8.8e-30	261.83	28.7	409	480
<input type="checkbox"/>	4	P49861	CAPSD_BPHK7 Major capsid protein OS=Streptococcus phage HK37 DK-37554 GN=5 PE=1 SP=1	99.91	2.2e-22	201.26	21.6	260	385
<input type="checkbox"/>	5	TRWZ_C	Major capsid protein; HK97-like fold, capsid size reduction, major capsid protein, VIRUS: 4.0A (Staphylococcus aureus)	99.8	3.2e-22	201.55	20.2	277	402
<input type="checkbox"/>	6	3B85_D	major capsid protein; acne, bacteriophage, HK97-like, VIRUS: 3.7A (Propionibacterium phage P46)	99.9	4.4e-22	193.71	19.4	258	315
<input type="checkbox"/>	7	10HC_B	MAJOR CAPSID PROTEIN; VIRUS, VIRUS COAT PROTEIN, VIRUS/VIRAL PROTEIN, BACTERIOPHAGE, CAPSID, AUTO-CATALYTIC CROSS-LINK	99.9	1.5e-21	185.88	21.8	257	282
<input type="checkbox"/>	8	8TQK_F	Scaffolding domain delta; Prohead L icosahedral symmetry; HK97 phage, capsid, VIRUS: 3.5A (Escherichia phage HK97)	99.89	1.5e-21	195.43	21.4	260	385
<input type="checkbox"/>	9	6TSU_F4	Major capsid protein Rcd1687; "capsid", "jelly roll", "spike", "HK97", VIRUS: 3.42A (Rhodobacter capsulatus DS442)	99.88	1.2e-21	196.26	17.6	259	386

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 precedent est de fonction inconnue
gene suivant est une head to tail adaptator

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>DeepTMHMM - Predictions</p> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p>  <p>You can download the probabilities used to generate this plot here</p>
<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>major capsid and protease fusion protein</p>

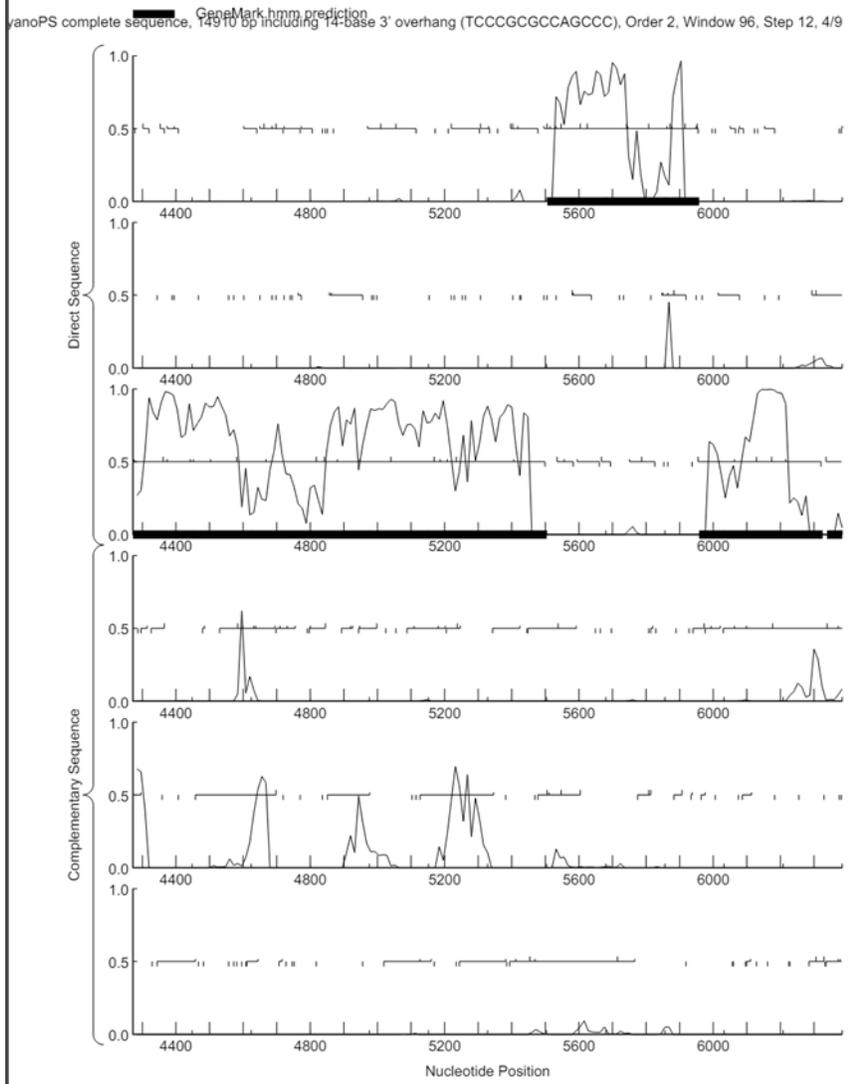
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	8
Coordonnées du Stop	5958
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 1
Coordonnées du Start retenu	5506
Fonction prédite	head-to-tail adaptor

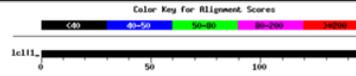
Décision #1 : Est-ce un gène ?

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

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



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



Sequences producing significant alignments:

Sequence	Score	E Value
CyanoP5_Draft_8, function unknown, 150	111	4e-85
Esperor_11, function unknown, 109	82	4e-16
SallySpecial_9, head-to-tail connector complex protein, 112	77	1e-14
Cowar_9, head-to-tail adaptor, 114	72	6e-13
Schiebs_9, head-to-tail adaptor, 111	71	8e-13
Rahul_8, head-to-tail adaptor, 114	70	1e-12
GRU2_7, function unknown, 113	70	2e-12
GMAS_7, function unknown, 113	67	1e-11
RRH1_08, function unknown, 118	64	1e-10
Adair2_6, head-to-tail adaptor, 114	62	7e-10
SerialPhiller_6, head-to-tail adaptor, 111	57	2e-08
McGonagall_8, function unknown, 107	67	2e-08

Sequences producing significant alignments

Download Select columns Show 100

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	110	110	67%	6e-28	53.47%	103	WP_003852555.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	108	108	68%	3e-27	53.40%	100	WP_134884367.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium belfanti]	Corynebacterium belfanti	107	107	68%	6e-27	52.43%	100	WP_197692067.1
<input checked="" type="checkbox"/> hypothetical protein FRC0497_01833 [Corynebacterium dichtherae]	Corynebacterium dichtherae	107	107	68%	6e-27	52.43%	100	CAB1005428.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	107	107	68%	1e-26	52.43%	100	WP_014318095.1
<input checked="" type="checkbox"/> hypothetical protein CIP107527_01916 [Corynebacterium dichtherae]	Corynebacterium dichtherae	107	107	68%	1e-26	52.43%	100	CAB0565497.1
<input checked="" type="checkbox"/> hypothetical protein CIP107524_01827 [Corynebacterium dichtherae]	Corynebacterium dichtherae	106	106	68%	2e-26	52.43%	100	CAB0565169.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	106	106	68%	2e-26	52.43%	100	WP_088298561.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	105	105	68%	5e-26	51.46%	100	WP_106361643.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	105	105	68%	5e-26	51.46%	100	WP_182001104.1
<input checked="" type="checkbox"/> hypothetical protein FRC0474_02083 [Corynebacterium dichtherae]	Corynebacterium dichtherae	105	105	68%	5e-26	51.46%	100	CAB09272152.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	105	105	68%	6e-26	50.49%	100	WP_205917741.1
<input checked="" type="checkbox"/> hypothetical protein FRC0370_01714 [Corynebacterium dichtherae]	Corynebacterium dichtherae	104	104	68%	1e-25	51.46%	100	CAB0870082.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	Corynebacterium	100	100	68%	5e-24	47.57%	102	WP_064833790.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodichtherricum]	Corynebacterium pseudodichtherricum	100	100	68%	9e-24	47.57%	102	WP_284587138.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium olauoum]	Corynebacterium olauoum	100	100	53%	9e-24	60.00%	103	WP_301925782.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium erooinoum]	Corynebacterium erooinoum	99.8	99.8	58%	1e-23	56.82%	103	WP_239210970.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	99.4	99.4	68%	1e-23	53.85%	103	WP_095275388.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium erooinoum]	Corynebacterium erooinoum	99.0	99.0	58%	2e-23	55.68%	103	WP_284571883.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	Corynebacterium	99.0	99.0	59%	2e-23	55.06%	103	WP_070420698.1

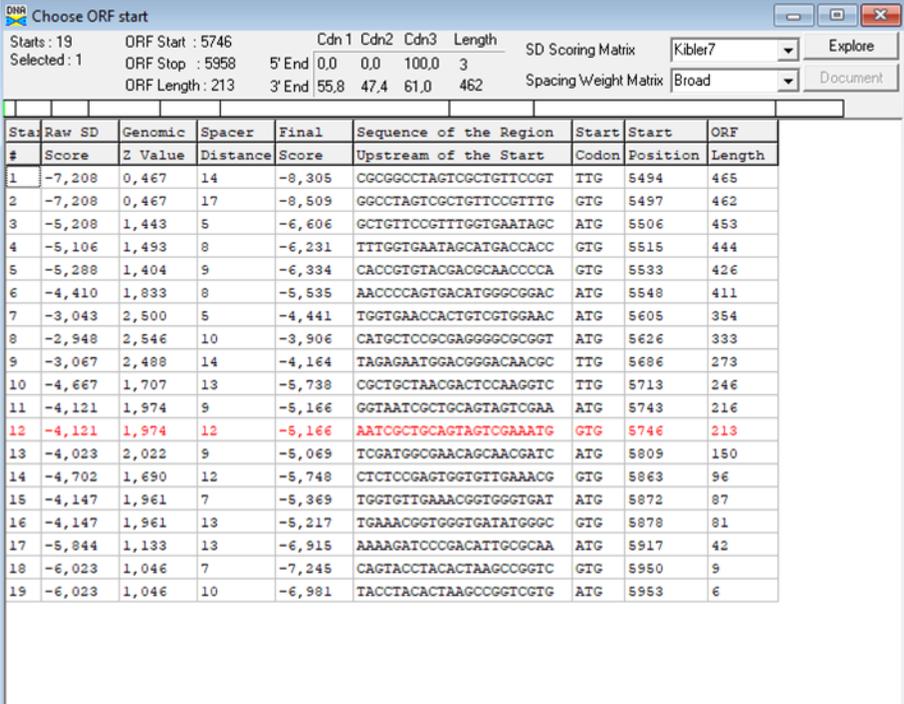
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																				
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 5506 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 5506																																																																																																																																																																																				
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <thead> <tr> <th>Sta</th> <th>Raw 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> </thead> <tbody> <tr><td>1</td><td>-7,208</td><td>0,467</td><td>14</td><td>-8,305</td><td>CGCGGCOCTAGTCGCTGTTCCGT</td><td>TTG</td><td>5494</td><td>465</td></tr> <tr><td>2</td><td>-7,208</td><td>0,467</td><td>17</td><td>-8,509</td><td>GGCCTAGTCGCTGTTCCGTTTG</td><td>GTG</td><td>5497</td><td>462</td></tr> <tr><td>3</td><td>-5,208</td><td>1,443</td><td>5</td><td>-6,606</td><td>GCTGTTCCGTTTGGTGAATAGC</td><td>ATG</td><td>5506</td><td>453</td></tr> <tr><td>4</td><td>-5,106</td><td>1,493</td><td>8</td><td>-6,231</td><td>TTTGGTGAATAGCATGACCACC</td><td>GTG</td><td>5515</td><td>444</td></tr> <tr><td>5</td><td>-5,288</td><td>1,404</td><td>9</td><td>-6,334</td><td>CACCGTGTACGACGCAACCCCA</td><td>GTG</td><td>5533</td><td>426</td></tr> <tr><td>6</td><td>-4,410</td><td>1,833</td><td>8</td><td>-5,535</td><td>AACCCAGTGCACATGGGCGGAC</td><td>ATG</td><td>5548</td><td>411</td></tr> <tr><td>7</td><td>-3,043</td><td>2,500</td><td>5</td><td>-4,441</td><td>TGGTGAACCACTGTCGTGGAAC</td><td>ATG</td><td>5605</td><td>354</td></tr> <tr><td>8</td><td>-2,948</td><td>2,546</td><td>10</td><td>-3,906</td><td>CATGTCGCGGAGGGGGCGGSI</td><td>ATG</td><td>5626</td><td>333</td></tr> <tr><td>9</td><td>-3,067</td><td>2,488</td><td>14</td><td>-4,164</td><td>TAGAGAATGGAAGGACACACGC</td><td>TTG</td><td>5686</td><td>273</td></tr> <tr><td>10</td><td>-4,667</td><td>1,707</td><td>13</td><td>-5,738</td><td>CGCTGCTAACGACTCCAGGTC</td><td>TTG</td><td>5713</td><td>246</td></tr> <tr><td>11</td><td>-4,121</td><td>1,974</td><td>9</td><td>-5,166</td><td>GGTAATCGCTGCACTAGTCGAA</td><td>ATG</td><td>5743</td><td>216</td></tr> <tr><td>12</td><td>-4,121</td><td>1,974</td><td>12</td><td>-5,166</td><td>AAICCGCTGCAGTAGTCGAAATG</td><td>ATG</td><td>5746</td><td>213</td></tr> <tr><td>13</td><td>-4,023</td><td>2,022</td><td>9</td><td>-5,069</td><td>TCGATGGGGAACAGCAACGATC</td><td>ATG</td><td>5809</td><td>150</td></tr> <tr><td>14</td><td>-4,702</td><td>1,690</td><td>12</td><td>-5,748</td><td>CTCTCCGAGTGGTGTGAAACG</td><td>GTG</td><td>5863</td><td>96</td></tr> <tr><td>15</td><td>-4,147</td><td>1,961</td><td>7</td><td>-5,369</td><td>TGGTGTGAAACGGTGGGTGAT</td><td>ATG</td><td>5872</td><td>87</td></tr> <tr><td>16</td><td>-4,147</td><td>1,961</td><td>13</td><td>-5,217</td><td>TGAAACGGTGGGTGATATGGGC</td><td>GTG</td><td>5878</td><td>81</td></tr> <tr><td>17</td><td>-5,844</td><td>1,133</td><td>13</td><td>-6,915</td><td>AAAAGATCCCGACATTGCGCAA</td><td>ATG</td><td>5917</td><td>42</td></tr> <tr><td>18</td><td>-6,023</td><td>1,046</td><td>7</td><td>-7,245</td><td>CAGTACCTACACTAAGCCGGTC</td><td>GTG</td><td>5950</td><td>9</td></tr> <tr><td>19</td><td>-6,023</td><td>1,046</td><td>10</td><td>-6,981</td><td>TACCTACACTAAGCCGGTCTGTG</td><td>ATG</td><td>5953</td><td>6</td></tr> </tbody> </table>	Sta	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length	1	-7,208	0,467	14	-8,305	CGCGGCOCTAGTCGCTGTTCCGT	TTG	5494	465	2	-7,208	0,467	17	-8,509	GGCCTAGTCGCTGTTCCGTTTG	GTG	5497	462	3	-5,208	1,443	5	-6,606	GCTGTTCCGTTTGGTGAATAGC	ATG	5506	453	4	-5,106	1,493	8	-6,231	TTTGGTGAATAGCATGACCACC	GTG	5515	444	5	-5,288	1,404	9	-6,334	CACCGTGTACGACGCAACCCCA	GTG	5533	426	6	-4,410	1,833	8	-5,535	AACCCAGTGCACATGGGCGGAC	ATG	5548	411	7	-3,043	2,500	5	-4,441	TGGTGAACCACTGTCGTGGAAC	ATG	5605	354	8	-2,948	2,546	10	-3,906	CATGTCGCGGAGGGGGCGGSI	ATG	5626	333	9	-3,067	2,488	14	-4,164	TAGAGAATGGAAGGACACACGC	TTG	5686	273	10	-4,667	1,707	13	-5,738	CGCTGCTAACGACTCCAGGTC	TTG	5713	246	11	-4,121	1,974	9	-5,166	GGTAATCGCTGCACTAGTCGAA	ATG	5743	216	12	-4,121	1,974	12	-5,166	AAICCGCTGCAGTAGTCGAAATG	ATG	5746	213	13	-4,023	2,022	9	-5,069	TCGATGGGGAACAGCAACGATC	ATG	5809	150	14	-4,702	1,690	12	-5,748	CTCTCCGAGTGGTGTGAAACG	GTG	5863	96	15	-4,147	1,961	7	-5,369	TGGTGTGAAACGGTGGGTGAT	ATG	5872	87	16	-4,147	1,961	13	-5,217	TGAAACGGTGGGTGATATGGGC	GTG	5878	81	17	-5,844	1,133	13	-6,915	AAAAGATCCCGACATTGCGCAA	ATG	5917	42	18	-6,023	1,046	7	-7,245	CAGTACCTACACTAAGCCGGTC	GTG	5950	9	19	-6,023	1,046	10	-6,981	TACCTACACTAAGCCGGTCTGTG	ATG	5953	6
Sta	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length																																																																																																																																																																													
1	-7,208	0,467	14	-8,305	CGCGGCOCTAGTCGCTGTTCCGT	TTG	5494	465																																																																																																																																																																													
2	-7,208	0,467	17	-8,509	GGCCTAGTCGCTGTTCCGTTTG	GTG	5497	462																																																																																																																																																																													
3	-5,208	1,443	5	-6,606	GCTGTTCCGTTTGGTGAATAGC	ATG	5506	453																																																																																																																																																																													
4	-5,106	1,493	8	-6,231	TTTGGTGAATAGCATGACCACC	GTG	5515	444																																																																																																																																																																													
5	-5,288	1,404	9	-6,334	CACCGTGTACGACGCAACCCCA	GTG	5533	426																																																																																																																																																																													
6	-4,410	1,833	8	-5,535	AACCCAGTGCACATGGGCGGAC	ATG	5548	411																																																																																																																																																																													
7	-3,043	2,500	5	-4,441	TGGTGAACCACTGTCGTGGAAC	ATG	5605	354																																																																																																																																																																													
8	-2,948	2,546	10	-3,906	CATGTCGCGGAGGGGGCGGSI	ATG	5626	333																																																																																																																																																																													
9	-3,067	2,488	14	-4,164	TAGAGAATGGAAGGACACACGC	TTG	5686	273																																																																																																																																																																													
10	-4,667	1,707	13	-5,738	CGCTGCTAACGACTCCAGGTC	TTG	5713	246																																																																																																																																																																													
11	-4,121	1,974	9	-5,166	GGTAATCGCTGCACTAGTCGAA	ATG	5743	216																																																																																																																																																																													
12	-4,121	1,974	12	-5,166	AAICCGCTGCAGTAGTCGAAATG	ATG	5746	213																																																																																																																																																																													
13	-4,023	2,022	9	-5,069	TCGATGGGGAACAGCAACGATC	ATG	5809	150																																																																																																																																																																													
14	-4,702	1,690	12	-5,748	CTCTCCGAGTGGTGTGAAACG	GTG	5863	96																																																																																																																																																																													
15	-4,147	1,961	7	-5,369	TGGTGTGAAACGGTGGGTGAT	ATG	5872	87																																																																																																																																																																													
16	-4,147	1,961	13	-5,217	TGAAACGGTGGGTGATATGGGC	GTG	5878	81																																																																																																																																																																													
17	-5,844	1,133	13	-6,915	AAAAGATCCCGACATTGCGCAA	ATG	5917	42																																																																																																																																																																													
18	-6,023	1,046	7	-7,245	CAGTACCTACACTAAGCCGGTC	GTG	5950	9																																																																																																																																																																													
19	-6,023	1,046	10	-6,981	TACCTACACTAAGCCGGTCTGTG	ATG	5953	6																																																																																																																																																																													
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 465pb -> TTG 5494, chevauchement de 11 L'ORF pour l'ATG 5506 est de 453pb, gap de 1 ATG 5548 on a une ORF de 411pb, gap de 43pb ATG 5605 on a une ORF de 354pb, gap de 100pb																																																																																																																																																																																				
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 ?	Les START 5506 et 5548 ne sont pas conservés (les autres n'ont pas été testés)																																																																																																																																																																																				
DECISION:	ATG 5506																																																																																																																																																																																				

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> <div data-bbox="873 415 1399 886"> <p>Distribution of 68 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>lc111</p> </div> <hr/> <table border="0"> <thead> <tr> <th>Sequences producing significant alignments:</th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>CyranoPS_Draft_8, function unknown, 150</td> <td>311</td> <td>4e-85</td> </tr> <tr> <td>Emperor_11, function unknown, 109</td> <td>82</td> <td>4e-16</td> </tr> <tr> <td>SallySpecial_9, head-to-tail connector complex protein, 112</td> <td>77</td> <td>1e-14</td> </tr> <tr> <td>Coeur_8, head-to-tail adaptor, 114</td> <td>72</td> <td>6e-13</td> </tr> <tr> <td>Schiebs_9, head-to-tail adaptor, 111</td> <td>71</td> <td>8e-13</td> </tr> </tbody> </table>	Sequences producing significant alignments:	Score (bits)	E Value	CyranoPS_Draft_8, function unknown, 150	311	4e-85	Emperor_11, function unknown, 109	82	4e-16	SallySpecial_9, head-to-tail connector complex protein, 112	77	1e-14	Coeur_8, head-to-tail adaptor, 114	72	6e-13	Schiebs_9, head-to-tail adaptor, 111	71	8e-13
Sequences producing significant alignments:	Score (bits)	E Value																	
CyranoPS_Draft_8, function unknown, 150	311	4e-85																	
Emperor_11, function unknown, 109	82	4e-16																	
SallySpecial_9, head-to-tail connector complex protein, 112	77	1e-14																	
Coeur_8, head-to-tail adaptor, 114	72	6e-13																	
Schiebs_9, head-to-tail adaptor, 111	71	8e-13																	
	<p>*** nr :</p> <p>(q#: s#) : [alignment] ; e-value :</p>																		

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 100 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Views

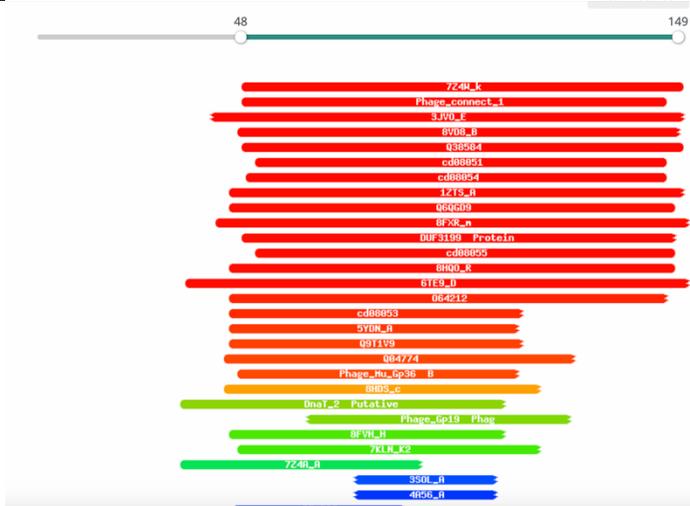
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	110	110	67%	6e-28	53.47%	103	WP_003852555.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	108	108	68%	3e-27	53.40%	100	WP_134884367.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium belfanti]	Corynebacterium belfanti	107	107	68%	6e-27	52.43%	100	WP_197692067.1
<input checked="" type="checkbox"/> hypothetical protein FRC0497_01833 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	107	107	68%	6e-27	52.43%	100	CAB1005428.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	107	107	68%	1e-26	52.43%	100	WP_014318095.1
<input checked="" type="checkbox"/> hypothetical protein CIP107527_01916 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	107	107	68%	1e-26	52.43%	100	CAB0565497.1
<input checked="" type="checkbox"/> hypothetical protein CIP107524_01827 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	106	106	68%	2e-26	52.43%	100	CAB0565169.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	106	106	68%	2e-26	52.43%	100	WP_088298561.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	105	105	68%	5e-26	51.46%	100	WP_106361843.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	105	105	68%	5e-26	51.46%	100	WP_182001104.1
<input checked="" type="checkbox"/> hypothetical protein FRC0474_02083 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	105	105	68%	5e-26	51.46%	100	CAB0972152.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	105	105	68%	6e-26	50.49%	100	WP_205917741.1
<input checked="" type="checkbox"/> hypothetical protein FRC0370_01714 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	104	104	68%	1e-25	51.46%	100	CAB0870082.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	Corynebacterium	100	100	68%	5e-24	47.57%	102	WP_064833790.1

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

100 sequences selected No putative conserved domains have been detected

Distribution of the top 100 Blast Hits on 100 subject sequences

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 ?



PFam : Phage_connect_1 ; Phage gp6-like head-tail connector protein

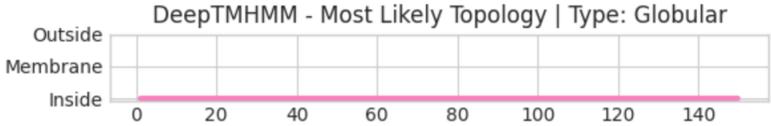
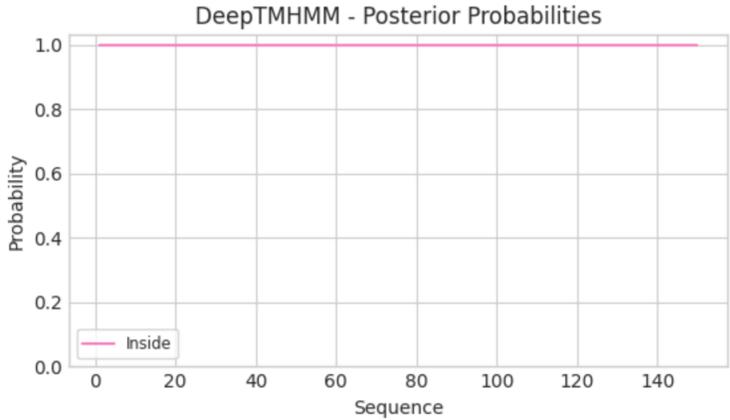
Cd : gp6_gp15_like; Head-Tail Connector Proteins gp6 and gp15, and similar proteins. Members of this family include the proteins gp6 and gp15 from bacteriophage HK97 and SPP1, respectively.

PDB : Head completion protein gp15; Bacteriophage, SPP1, Portal Protein, Head completion proteins, Connector Complex, DNA Channel, VIRAL PROTEIN; 2.7A {Bacillus subtilis}
Uniprot : HCP15_BPSP Head completion protein gp15 OS=Bacillus phage SPP1 OX=10724 GN=15 PE=1 SV=1

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	P85500	CAPSD_BPPAJ Capsid polyprotein OS=Pseudomonas phage PAJU2 OX=504346 PE=1 SV=2	100	6e-36	319.23	41.5	441	667
<input type="checkbox"/>	B0ZSF3	CAPSD_BPHA1 Major capsid protein OS=Halomonas phage phiHAP-1 (isolate - /Gulf of Mexico/-/2001) OX=1283337 GN=HAPgp05 PE=	100	3.4e-34	302.56	34.9	423	602
<input type="checkbox"/>	Q38300	CAPSD_BPLC2 Major capsid protein OS=Lactococcus phage c2 OX=31537 GN=I5 PE=4 SV=1	99.98	8.8e-30	261.83	28.7	409	480
<input type="checkbox"/>	P49861	CAPSD_BPHK7 Major capsid protein OS=Enterobacteria phage HK97 OX=37554 GN=5 PE=1 SV=1	99.91	2.2e-22	201.26	21.6	260	385
<input type="checkbox"/>	7RWZ_C	Major capsid protein; HK97-like fold, capsid size redirection, major capsid protein, VIRUS; 4.0A {Staphylococcus aureus}	99.9	3.2e-22	201.56	20.2	277	402
<input type="checkbox"/>	3JB5_D	major capsid protein; acne, bacteriophage, HK97-like, VIRUS; 3.7A {Propionibacterium phage PA6}	99.9	4.4e-22	193.71	19.4	258	315
<input type="checkbox"/>	1OHG_B	MAJOR CAPSID PROTEIN; VIRUS, VIRUS COAT PROTEIN, VIRUS/VIRAL PROTEIN, BACTERIOPHAGE, CAPSID, AUTO- CATALYTIC CROSS-LINK,	99.9	1.5e-21	185.88	21.8	257	282
<input type="checkbox"/>	8FQK_F	Scaffolding domain delta; Prohead I, icosahedral symmetry, HK97, phage, capsid, VIRUS; 3.5A {Escherichia phage HK97}	99.89	1.5e-21	195.43	21.4	260	385

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 le gene avant est une major capsid

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<div style="text-align: right;">  </div>   <p>Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebius par ex).</p>
<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>head-to-tail adaptor</p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	9
Coordonnées du Stop	6326
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove de 1

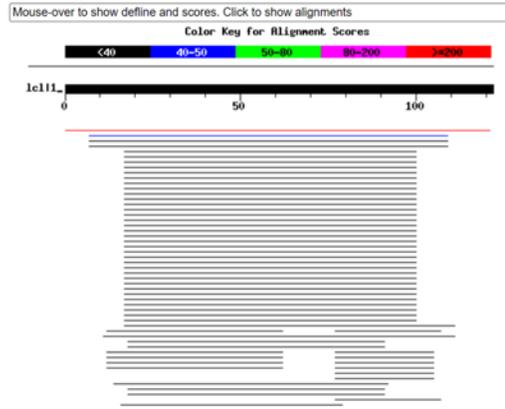
Coordonnées du Start retenu	5958
Fonction prédite	Tail terminator

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>anoPS complete sequence, 145 PD bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 4/9</p> <p>The figure displays a GeneMark hmmp prediction plot. It consists of eight vertically stacked tracks. The top four tracks are labeled 'Direct Sequence' and the bottom four are 'Complementary Sequence'. The x-axis represents 'Nucleotide Position' from 4400 to 6000. The y-axis represents signal intensity from 0.0 to 1.0. A thick black horizontal bar is drawn across the tracks, indicating a predicted coding region between approximately 5500 and 6000 nucleotide positions. The plot shows various peaks and troughs across the sequences, with the most significant activity occurring within the highlighted coding region.</p>

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

Distribution of 58 Blast Hits on the Query Sequence



Sequences producing significant alignments:

	Score (bits)	E Value
CyranoPS_Draft_9, function unknown, 122	248	3e-66
GMA5_8, function unknown, 113	48	0.002
Taj14_9, function unknown, 118	36	0.029
Saphira_9, function unknown, 118	36	0.029
Rahu_9, tail terminator, 113	36	0.029
Coeur_9, function unknown, 113	36	0.029
Yank_9, function unknown, 118	36	0.038
TymAbreu_9, function unknown, 118	36	0.038

Sequences producing significant alignments

Download Select columns Show 100

select all 29 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> conserved_hv00theical_protein [Corynebacterium striatum]	Corynebacterium striatum	86.7	86.7	89%	9e-19	40.37%	116	CGQ13973.1
<input checked="" type="checkbox"/> hv00theical_protein [unclassified Corynebacterium]	unclassified Corynebacterium	85.9	85.9	89%	1e-18	40.37%	116	WP_070736120.1
<input checked="" type="checkbox"/> TPA_hv00theical_protein [Corynebacterium striatum]	Corynebacterium striatum	85.9	85.9	89%	2e-18	40.37%	116	HCT5225723.1
<input checked="" type="checkbox"/> TPA_hv00theical_protein [Corynebacterium striatum]	Corynebacterium striatum	75.9	75.9	89%	1e-14	36.70%	116	HCD1553132.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium]	Corynebacterium	73.9	73.9	89%	7e-14	36.70%	116	WP_070420699.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium crocinoum]	Corynebacterium crocinoum	73.2	73.2	92%	1e-13	35.96%	117	WP_239210980.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium diththeriae]	Corynebacterium diththeriae	72.4	72.4	100%	4e-13	30.33%	124	WP_003852557.1
<input checked="" type="checkbox"/> hv00theical_protein FR0431_01840 [Corynebacterium diththeriae]	Corynebacterium diththeriae	72.0	72.0	100%	5e-13	30.33%	124	CAB0923853.1
<input checked="" type="checkbox"/> hv00theical_protein NS79_07875 [Corynebacterium oseudodiththerioum 090104]	Corynebacterium oseudodiththerioum 090104	69.3	69.3	80%	5e-12	34.69%	116	ERJ44226.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium oseudodiththerioum]	Corynebacterium oseudodiththerioum	69.7	69.7	90%	5e-12	34.23%	131	WP_284587139.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium]	Corynebacterium	69.3	69.3	92%	5e-12	35.09%	117	WP_049167470.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium hadale]	Corynebacterium hadale	69.3	69.3	84%	7e-12	34.95%	123	WP_095275387.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium crocinoum]	Corynebacterium crocinoum	68.9	68.9	80%	7e-12	34.69%	116	WP_144736465.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium oseudodiththerioum]	Corynebacterium oseudodiththerioum	68.6	68.6	90%	1e-11	34.23%	131	WP_284849034.1
<input checked="" type="checkbox"/> hv00theical_protein CIP107524_01828 [Corynebacterium diththeriae]	Corynebacterium diththeriae	68.2	68.2	99%	1e-11	28.93%	121	CAB0565189.1
<input checked="" type="checkbox"/> hv00theical_protein [Corynebacterium diththeriae]	Corynebacterium diththeriae	67.0	67.0	99%	4e-11	28.10%	121	WP_088298563.1
<input checked="" type="checkbox"/> Hv00theical_protein PIP1930 [Corynebacterium diththeriae]	Corynebacterium diththeriae	66.6	66.6	86%	4e-11	31.13%	128	CAB50300.1

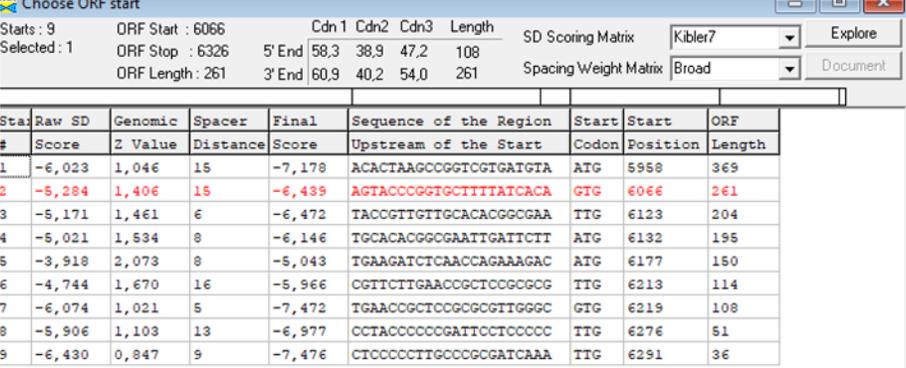
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

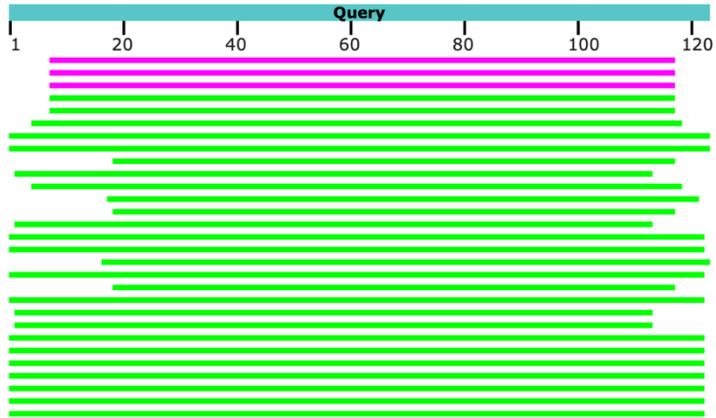
Collection des éléments de réponse	Rationnelle																																																																																																			
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 5958 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 5958</p>																																																																																																			
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <table border="1" data-bbox="565 604 1344 861"> <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>-6,023</td> <td>1,046</td> <td>15</td> <td>-7,178</td> <td>ACACTAAGCCGGTCGTGATGTA</td> <td>ATG</td> <td>5958</td> <td>369</td> </tr> <tr> <td>2</td> <td>-5,284</td> <td>1,406</td> <td>15</td> <td>-6,439</td> <td>AGTACCCGGTCTTTTATCACA</td> <td>GTG</td> <td>6066</td> <td>261</td> </tr> <tr> <td>3</td> <td>-5,171</td> <td>1,461</td> <td>6</td> <td>-6,472</td> <td>TACCGTTGTTCACACGGGGA</td> <td>TTG</td> <td>6123</td> <td>204</td> </tr> <tr> <td>4</td> <td>-5,021</td> <td>1,534</td> <td>8</td> <td>-6,146</td> <td>TGCACACGGGGAATTGATTCIT</td> <td>ATG</td> <td>6132</td> <td>195</td> </tr> <tr> <td>5</td> <td>-3,918</td> <td>2,073</td> <td>8</td> <td>-5,043</td> <td>TGAAGATCTCAACCAGAAAGAC</td> <td>ATG</td> <td>6177</td> <td>150</td> </tr> <tr> <td>6</td> <td>-4,744</td> <td>1,670</td> <td>16</td> <td>-5,966</td> <td>CGTTCITGAAACCGCTCCGCGCG</td> <td>TTG</td> <td>6213</td> <td>114</td> </tr> <tr> <td>7</td> <td>-6,074</td> <td>1,021</td> <td>5</td> <td>-7,472</td> <td>TGAACCGCTCCGCGGTTGGGC</td> <td>GTG</td> <td>6219</td> <td>108</td> </tr> <tr> <td>8</td> <td>-5,906</td> <td>1,103</td> <td>13</td> <td>-6,977</td> <td>CCTACCCCGGATTCCTCCCCC</td> <td>TTG</td> <td>6276</td> <td>51</td> </tr> <tr> <td>9</td> <td>-6,430</td> <td>0,847</td> <td>9</td> <td>-7,476</td> <td>CTCCCCCTGCCCCGATCAA</td> <td>TTG</td> <td>6291</td> <td>36</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	-6,023	1,046	15	-7,178	ACACTAAGCCGGTCGTGATGTA	ATG	5958	369	2	-5,284	1,406	15	-6,439	AGTACCCGGTCTTTTATCACA	GTG	6066	261	3	-5,171	1,461	6	-6,472	TACCGTTGTTCACACGGGGA	TTG	6123	204	4	-5,021	1,534	8	-6,146	TGCACACGGGGAATTGATTCIT	ATG	6132	195	5	-3,918	2,073	8	-5,043	TGAAGATCTCAACCAGAAAGAC	ATG	6177	150	6	-4,744	1,670	16	-5,966	CGTTCITGAAACCGCTCCGCGCG	TTG	6213	114	7	-6,074	1,021	5	-7,472	TGAACCGCTCCGCGGTTGGGC	GTG	6219	108	8	-5,906	1,103	13	-6,977	CCTACCCCGGATTCCTCCCCC	TTG	6276	51	9	-6,430	0,847	9	-7,476	CTCCCCCTGCCCCGATCAA	TTG	6291	36
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	-6,023	1,046	15	-7,178	ACACTAAGCCGGTCGTGATGTA	ATG	5958	369																																																																																												
2	-5,284	1,406	15	-6,439	AGTACCCGGTCTTTTATCACA	GTG	6066	261																																																																																												
3	-5,171	1,461	6	-6,472	TACCGTTGTTCACACGGGGA	TTG	6123	204																																																																																												
4	-5,021	1,534	8	-6,146	TGCACACGGGGAATTGATTCIT	ATG	6132	195																																																																																												
5	-3,918	2,073	8	-5,043	TGAAGATCTCAACCAGAAAGAC	ATG	6177	150																																																																																												
6	-4,744	1,670	16	-5,966	CGTTCITGAAACCGCTCCGCGCG	TTG	6213	114																																																																																												
7	-6,074	1,021	5	-7,472	TGAACCGCTCCGCGGTTGGGC	GTG	6219	108																																																																																												
8	-5,906	1,103	13	-6,977	CCTACCCCGGATTCCTCCCCC	TTG	6276	51																																																																																												
9	-6,430	0,847	9	-7,476	CTCCCCCTGCCCCGATCAA	TTG	6291	36																																																																																												
<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>	<p><i>ORF la plus longue</i></p> <p><i>ORF la plus longue avec un ATG</i></p> <p><i>Score de -7.170</i></p>																																																																																																			
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>—</p>																																																																																																			

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<div data-bbox="630 121 1474 451"> <p>Download GenPept Graphics</p> <p>conserved hypothetical protein [Corynebacterium striatum]</p> <p>Sequence ID: CQD13973.1 Length: 116 Number of Matches: 1</p> <p>Range 1: 7 to 114 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>86.7 bits(213)</td> <td>9e-19</td> <td>Compositional matrix adjust.</td> <td>44/109(40%)</td> <td>66/109(60%)</td> <td>1/109(0%)</td> </tr> </tbody> </table> <pre> Query 8 DWF CSTLNGLGI ECTADQRDIQVPGAFITVTKLSDWSIDLTTVAHGELILMTRDIGGRE 67 D + L+ GI T D R++ PG F+TVTK+SD ++D + V G+++ + RD+GG Sbjct 7 DQVLTVLDDAGIAATIDARNLSAPGCFVTVTKVSDLALDGS AKV -TGDIVAVVRDLGGAA 65 Query 68 DLNQKDMLLKAVLEPLRALGVVVTIRPTEQALPPDSSSPLPAIKLEWSL 116 D++ LL V+EPL+A V +T I EQA PP LPA +L ++L Sbjct 66 DIDNLSALLDDVVEPLKAANVEITDIITNEQATPPSGGTLPAARLYTYL 114 </pre> </div> <div data-bbox="630 493 1474 846"> <p>Download GenPept Graphics</p> <p>MULTISPECIES: hypothetical protein [unclassified Corynebacterium]</p> <p>Sequence ID: WP_070736120.1 Length: 116 Number of Matches: 1</p> <p>See 2 more title(s) See all Identical Proteins(IPG)</p> <p>Range 1: 7 to 114 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>85.9 bits(211)</td> <td>1e-18</td> <td>Compositional matrix adjust.</td> <td>44/109(40%)</td> <td>66/109(60%)</td> <td>1/109(0%)</td> </tr> </tbody> </table> <pre> Query 8 DWF CSTLNGLGI ECTADQRDIQVPGAFITVTKLSDWSIDLTTVAHGELILMTRDIGGRE 67 D + L+ GI T D R++ PG F+TVTK+SD ++D + V G+++ + RD+GG Sbjct 7 DQVLTVLDDAGIAATIDARNLSAPGCFVTVTKISDLALDGS AKV -TGDIVAVVRDLGGAA 65 Query 68 DLNQKDMLLKAVLEPLRALGVVVTIRPTEQALPPDSSSPLPAIKLEWSL 116 D++ LL V+EPL+A V +T I EQA PP LPA +L ++L Sbjct 66 DIDNLSALLDDVVEPLQAANVEITDIITNEQATPPSGGTLPAARLYTYL 114 </pre> </div>	Score	Expect	Method	Identities	Positives	Gaps	86.7 bits(213)	9e-19	Compositional matrix adjust.	44/109(40%)	66/109(60%)	1/109(0%)	Score	Expect	Method	Identities	Positives	Gaps	85.9 bits(211)	1e-18	Compositional matrix adjust.	44/109(40%)	66/109(60%)	1/109(0%)
Score	Expect	Method	Identities	Positives	Gaps																				
86.7 bits(213)	9e-19	Compositional matrix adjust.	44/109(40%)	66/109(60%)	1/109(0%)																				
Score	Expect	Method	Identities	Positives	Gaps																				
85.9 bits(211)	1e-18	Compositional matrix adjust.	44/109(40%)	66/109(60%)	1/109(0%)																				
<p>DECISION:</p>	<p>5958</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 : RIEN</p> <p>*** nr :</p>

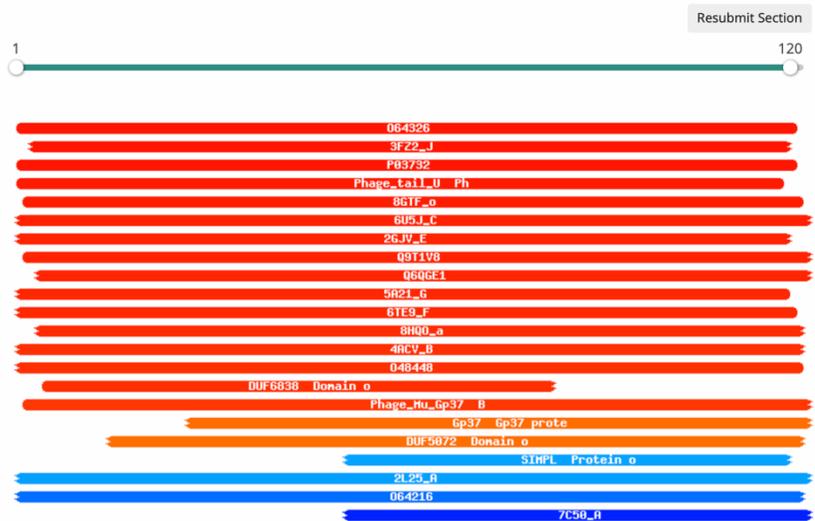
Distribution of the top 29 Blast Hits on 29 subject sequences



<input checked="" type="checkbox"/>	conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	86.7	86.7	89%	9e-19	40.37%	116	CQD13973.1
<input checked="" type="checkbox"/>	hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	85.9	85.9	89%	1e-18	40.37%	116	WP_070736120.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	85.9	85.9	89%	2e-18	40.37%	116	HCT5225723.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	75.9	75.9	89%	1e-14	36.70%	116	HCD1553132.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium]	Corynebacterium	73.9	73.9	89%	7e-14	36.70%	116	WP_070420699.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	73.2	73.2	92%	1e-13	35.96%	117	WP_239210980.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	72.4	72.4	100%	4e-13	30.33%	124	WP_003852557.1
<input checked="" type="checkbox"/>	hypothetical protein FRC0431_01840 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	72.0	72.0	100%	5e-13	30.33%	124	CAB0923853.1
<input checked="" type="checkbox"/>	hypothetical protein N579_07875 [Corynebacterium pseudodiphtheriticum 090104]	Corynebacterium pseudodiphtheriticum 090104	69.3	69.3	80%	5e-12	34.69%	116	ERJ44226.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	69.7	69.7	90%	5e-12	34.23%	131	WP_284587139.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium]	Corynebacterium	69.3	69.3	92%	5e-12	35.09%	117	WP_049167470.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	69.3	69.3	84%	7e-12	34.95%	123	WP_095275387.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	68.9	68.9	80%	7e-12	34.69%	116	WP_144736465.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	68.6	68.6	90%	1e-11	34.23%	131	WP_284849034.1
<input checked="" type="checkbox"/>	hypothetical protein CIP107524_01828 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	68.2	68.2	99%	1e-11	28.93%	121	CAB0585189.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	67.0	67.0	99%	4e-11	28.10%	121	WP_088298563.1
<input checked="" type="checkbox"/>	Hypothetical protein DIP1830 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	66.6	66.6	86%	4e-11	31.13%	108	CAE50360.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	67.0	67.0	99%	5e-11	28.10%	121	WP_342351207.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	66.2	66.2	80%	7e-11	33.67%	116	WP_284596343.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium belfanti]	Corynebacterium belfanti	66.2	66.2	99%	1e-10	28.10%	121	WP_197692066.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	66.2	66.2	90%	1e-10	33.33%	131	WP_284594218.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium sp. EPI-003-04-2554_SCH2473622]	Corynebacterium sp. EPI-003-04-2554_SCH24...	66.2	66.2	90%	1e-10	33.33%	131	WP_064833788.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	65.1	65.1	99%	2e-10	27.27%	121	WP_106361642.1
<input checked="" type="checkbox"/>	hypothetical protein FRC0474_02084 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	64.7	64.7	99%	4e-10	27.27%	121	CAB0972182.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	64.3	64.3	99%	4e-10	26.45%	121	WP_014318096.1
<input checked="" type="checkbox"/>	hypothetical protein FRC0087_01817 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	64.3	64.3	99%	5e-10	26.45%	121	CAB0733833.1
<input checked="" type="checkbox"/>	hypothetical protein FRC0497_01832 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.9	63.9	99%	7e-10	26.45%	121	CAB1005424.1
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.5	63.5	99%	8e-10	26.45%	121	WP_205917719.1
<input checked="" type="checkbox"/>	hypothetical protein CIP107559_01881 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.5	63.5	99%	8e-10	26.45%	121	CAB0612778.1

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 $\geq 90\%$ et une couverture acceptable ?

Visualization



<input type="checkbox"/>	1	O64326	TTTP_BPN15 Tail tube terminator protein OS=Escherichia phage N15 OX=40631 GN=gene 12 PE=3 SV=1	98.11	0.0013	44.42	14.1	114	132
<input type="checkbox"/>	2	3FZ2_J	Minor tail protein U; Mixed Alpha-Beta fold, VIRAL PROTEIN; HET: MSE, SO4; 2.7A {Enterobacteria phage lambda} SCOP: d.32	98.08	0.00088	44.87	12.8	113	134
<input type="checkbox"/>	3	P03732	TTTP_LAMBD Tail tube terminator protein OS=Escherichia phage lambda OX=10710 GN=U PE=1 SV=1	98.06	0.0011	44.72	13	114	131
<input type="checkbox"/>	4	PF06141.15	; Phage_tail_U ; Phage minor tail protein U	97.71	0.0062	40.55	12	115	130
<input type="checkbox"/>	5	8GTF_o	Terminator protein; Marine bacteriophage, Cryo-EM, Siphophage, Stopper protein, Terminator protein, Head-to-tail interfa	97.36	0.042	37.8	13.6	119	140
<input type="checkbox"/>	6	6U5J_C	Collar PA0615; bacteriocin, pyocin, ANTIMICROBIAL PROTEIN; 3.5A {Pseudomonas aeruginosa (strain ATCC 15692 / DSM 22644 /	97.16	0.086	37.46	13.7	116	171
<input type="checkbox"/>	7	2GJV_E	putative cytoplasmic protein; Structural genomics, Unknown function, PSI, Protein Structure Initiative, Midwest Center f	97.02	0.084	37.79	11.7	113	175
<input type="checkbox"/>	8	Q9T1V8	TRP_BPMU Probable tail terminator protein OS=Escherichia phage Mu OX=10677 GN=Mup37 PE=2 SV=1	96.97	0.073	38.24	11.1	119	182
<input type="checkbox"/>	9	Q6QGE1	TTTP_BPT5 Tail tube terminator protein OS=Escherichia phage T5 OX=10726 GN=ORF135 PE=2 SV=1	96.83	0.18	35.7	14	119	161
<input type="checkbox"/>	10	5A21_G	TAIL-TO-HEAD JOINING PROTEIN GP17; VIRAL PROTEIN, VIRAL INFECTION, TAILED BACTERIOPHAGE SIPHOPHAGE CDD1	96.66	0.18	33.41	12.4	116	134

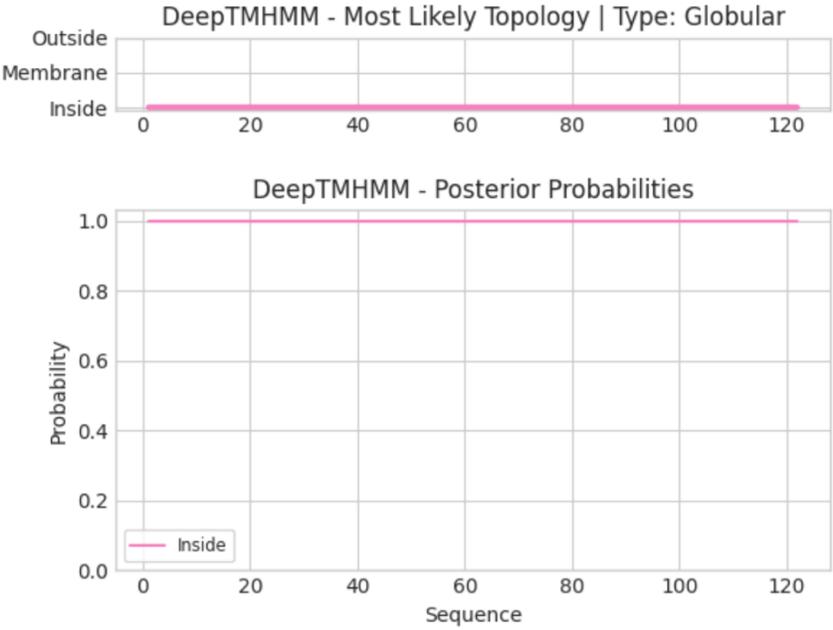
Uniprot : TTTP_BPN15 Tail tube terminator protein OS=Escherichia phage N15 OX=40631 GN=gene 12 PE=3 SV=1

PDB : Minor tail protein U; Mixed Alpha-Beta fold, VIRAL PROTEIN; HET: MSE, SO4; 2.7A {Enterobacteria phage lambda} SCOP: d.323.1.1, l.1.1.1

Pfam : Phage_tail_U ; Phage minor tail protein U

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

Oui

<p>qui montre une forte conservation de l'ordre des gènes ?</p>	
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>non</p> <h3>DeepTMHMM - Predictions</h3> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p> 
<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>Tail terminator</p>

Student Gene Annotation Worksheet

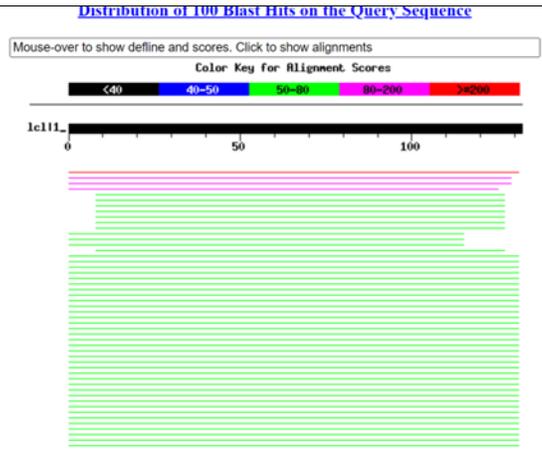
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	10
Coordonnées du Stop	6737
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 12

Coordonnées du Start retenu	6339
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)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GenoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGGCCAGCCC), Order 2, Window 96, Step 12, 5/9</p> <p>GeneMark hmm prediction</p> <p>Direct Sequence</p> <p>Complementary Sequence</p> <p>Nucleotide Position</p>

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



Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyranoPS_Draft_10, function unknown, 132	265	4e-71
Emperor_13, major tail protein, 146	88	9e-18
SallySpecial_11, major tail protein, 148	83	2e-16
EpicDab_12, major tail protein, 145	81	8e-16
Schiebs_11, major tail protein, 149	79	4e-15
Rahul_10, major tail protein, 153	72	6e-13
Coeur_10, major tail protein, 151	72	6e-13
GMAS_9, function unknown, 153	70	2e-12
RRH1_10, putative tail protein, 143	67	2e-11
McGonagall_10, major tail protein, 154	64	1e-10
TPA, hypothetical protein [Corynebacterium striatum]	64	1e-10

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 [Corynebacterium belfanti]	Corynebacterium belfanti	122	122	96%	3e-32	52.21%	139	WP_197692065.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	121	121	96%	5e-32	52.21%	139	WP_014318097.1
<input checked="" type="checkbox"/> hypothetical protein FRC0086_01805 [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	120	120	96%	1e-31	52.21%	139	CAB0734176.1
<input checked="" type="checkbox"/> hypothetical protein CIP107559_01882 [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	119	119	96%	2e-31	51.47%	139	CAB0612796.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	119	119	96%	3e-31	51.47%	139	WP_088298565.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	119	119	96%	4e-31	50.74%	139	WP_134884349.1
<input checked="" type="checkbox"/> hypothetical protein FRC0087_01816 [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	119	119	96%	4e-31	51.47%	139	CAB0733813.1
<input checked="" type="checkbox"/> hypothetical protein CIP107524_01829 [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	119	119	96%	4e-31	51.47%	139	CAB0665201.1
<input checked="" type="checkbox"/> hypothetical protein FRC0405_01824 [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	117	117	96%	3e-30	50.74%	139	CAB0884021.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	115	115	93%	5e-30	53.12%	135	HCT5225724.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	115	115	96%	9e-30	49.24%	135	HCD1553133.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium sp. HMSC065A05]	Corynebacterium sp. HMSC065A05	114	114	93%	2e-29	52.34%	136	WP_070776423.1
<input checked="" type="checkbox"/> putative tail protein [Corynebacterium striatum]	Corynebacterium striatum	114	114	93%	3e-29	52.34%	135	CQ019976.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	Corynebacterium	114	114	96%	3e-29	48.48%	135	WP_070420700.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium sp. HMSC077D10]	Corynebacterium sp. HMSC077D10	114	114	93%	4e-29	52.34%	136	WP_070736118.1
<input checked="" type="checkbox"/> TPA, tail tube protein [Cauloviricetes sp.]	Cauloviricetes sp.	112	112	100%	2e-28	45.99%	137	QAX64369.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	Corynebacterium	110	110	100%	1e-27	45.26%	137	WP_049167469.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium dichtheriae]	Corynebacterium dichtheriae	108	108	96%	3e-27	46.72%	140	WP_003852559.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium crocinosum]	Corynebacterium crocinosum	108	108	100%	7e-27	44.53%	137	WP_239210982.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	Corynebacterium	106	106	100%	4e-26	43.80%	137	WP_021353241.1
<input checked="" type="checkbox"/> hypothetical protein [Dietzia sp. oral taxon 368]	Dietzia sp. oral taxon 368	103	103	100%	5e-25	48.57%	143	WP_106385482.1

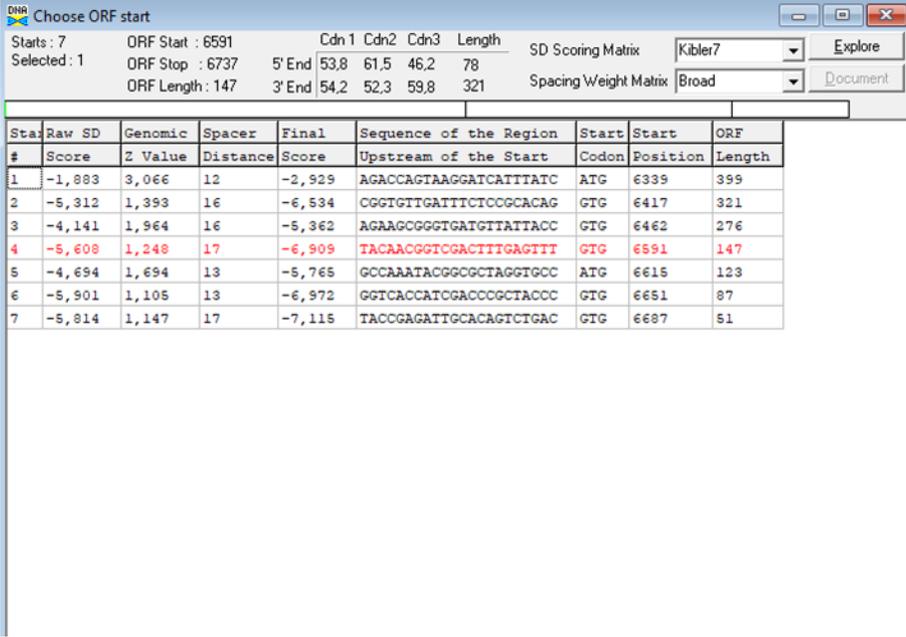
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																	
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 6339 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 6339</p>																																																																																	
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <table border="1"> <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>-1,883</td> <td>3,066</td> <td>12</td> <td>-2,929</td> <td>AGACCAGTAAGGATCAITTTATC</td> <td>ATG</td> <td>6339</td> <td>399</td> </tr> <tr> <td>2</td> <td>-5,312</td> <td>1,393</td> <td>16</td> <td>-6,534</td> <td>CGGTGTGATTTCTCCGCACAG</td> <td>GTG</td> <td>6417</td> <td>321</td> </tr> <tr> <td>3</td> <td>-4,141</td> <td>1,964</td> <td>16</td> <td>-5,362</td> <td>AGTAGCGGGTGATGTTATTACC</td> <td>GTG</td> <td>6462</td> <td>276</td> </tr> <tr> <td>4</td> <td>-5,608</td> <td>1,248</td> <td>17</td> <td>-6,909</td> <td>TACAACGGTCCGACTTTGAGTTT</td> <td>GTG</td> <td>6591</td> <td>147</td> </tr> <tr> <td>5</td> <td>-4,694</td> <td>1,694</td> <td>13</td> <td>-5,765</td> <td>GCCAAATACGGCGCTAGGTGCC</td> <td>ATG</td> <td>6615</td> <td>123</td> </tr> <tr> <td>6</td> <td>-5,901</td> <td>1,105</td> <td>13</td> <td>-6,972</td> <td>GGTCACCATCGACCCGCTACCC</td> <td>GTG</td> <td>6651</td> <td>87</td> </tr> <tr> <td>7</td> <td>-5,814</td> <td>1,147</td> <td>17</td> <td>-7,115</td> <td>TACCGAGATTGCACAGTCTGAC</td> <td>GTG</td> <td>6687</td> <td>51</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	-1,883	3,066	12	-2,929	AGACCAGTAAGGATCAITTTATC	ATG	6339	399	2	-5,312	1,393	16	-6,534	CGGTGTGATTTCTCCGCACAG	GTG	6417	321	3	-4,141	1,964	16	-5,362	AGTAGCGGGTGATGTTATTACC	GTG	6462	276	4	-5,608	1,248	17	-6,909	TACAACGGTCCGACTTTGAGTTT	GTG	6591	147	5	-4,694	1,694	13	-5,765	GCCAAATACGGCGCTAGGTGCC	ATG	6615	123	6	-5,901	1,105	13	-6,972	GGTCACCATCGACCCGCTACCC	GTG	6651	87	7	-5,814	1,147	17	-7,115	TACCGAGATTGCACAGTCTGAC	GTG	6687	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	-1,883	3,066	12	-2,929	AGACCAGTAAGGATCAITTTATC	ATG	6339	399																																																																										
2	-5,312	1,393	16	-6,534	CGGTGTGATTTCTCCGCACAG	GTG	6417	321																																																																										
3	-4,141	1,964	16	-5,362	AGTAGCGGGTGATGTTATTACC	GTG	6462	276																																																																										
4	-5,608	1,248	17	-6,909	TACAACGGTCCGACTTTGAGTTT	GTG	6591	147																																																																										
5	-4,694	1,694	13	-5,765	GCCAAATACGGCGCTAGGTGCC	ATG	6615	123																																																																										
6	-5,901	1,105	13	-6,972	GGTCACCATCGACCCGCTACCC	GTG	6651	87																																																																										
7	-5,814	1,147	17	-7,115	TACCGAGATTGCACAGTCTGAC	GTG	6687	51																																																																										
<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>	<p><i>ORF le plus long</i></p> <p><i>Score de -2</i></p> <p><i>ATG</i></p>																																																																																	
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>–</p>																																																																																	

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

Range 1: 1 to 136 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
122 bits(305)	3e-32	Compositional matrix adjust.	71/136(52%)	85/136(62%)	8/136(5%)

```

Query 1  MAPKYFRGSGTLTGGSSG---VDFSAQVTSCTVPEAG--DVIITVLSGDTIGN---YT 52
           MA K  G+GTLTL  SG  DFS+Q+TS  PEA  D I  VLSG+TI  +  +
Sbjct 1  MAIKTVTVGAGTLLAVKSGDGGAKDFSSQITSAARFEPEAKSDPILVLSGETIASADSF 60

Query 53  ANLSVNFQDLSTAGITAYSFDNAGTVDVDFEVPNTALGAMIAGKVTIDPLVGGATTEI 112
           LSV  FLQDLST  GI  YSF  NAG  DF  +  PNTA  A  +G  V  I+PLPVG  +  +I
Sbjct 61  GRLSVEFLQDLSTNGIDVYSFTNAGKEADFYVTPNTANKAKLAGTVIEIPLVGDVSDI 120

Query 113 AQSDVSWTCPELPTFT 128
           A++  VSW  P  LP  FT
Sbjct 121 AKASVSWQVPSLPRFT 136

```

[Download](#) ▼ [GenPept](#) [Graphics](#) ▼ Next ▲ Previous ◀ Descriptions

hypothetical protein [Corynebacterium diphtheriae]
 Sequence ID: [WP_014318097.1](#) Length: 139 Number of Matches: 1
[See 5 more title\(s\)](#) ▼ [See all Identical Proteins \(IPG\)](#)

Range 1: 1 to 136 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
121 bits(303)	5e-32	Compositional matrix adjust.	71/136(52%)	84/136(61%)	8/136(5%)

```

Query 1  MAPKYFRGSGTLTGGSSG---VDFSAQVTSCTVPEAG--DVIITVLSGDTIGN---YT 52
           MA K  G+GTLTL  SG  DFS+Q+TS  PEA  D I  VLSG+TI  +  +
Sbjct 1  MAIKTVTVGAGTLLAVKSGDGGAKDFSSQITSAARFEPEAKSDPILVLSGETIASADSF 60

Query 53  ANLSVNFQDLSTAGITAYSFDNAGTVDVDFEVPNTALGAMIAGKVTIDPLVGGATTEI 112
           LSV  FLQDLST  GI  YSF  NAG  DF  +  PNTA  A  + G  V  I+PLPVG  +  +I
Sbjct 61  GRLSVEFLQDLSTNGIDVYSFRNAGKEADFYVTPNTANKAKLGTGTVEIPLVGDVSDI 120

Query 113 AQSDVSWTCPELPTFT 128
           A++  VSW  P  LP  FT
Sbjct 121 AKASVSWQVPSLPRFT 136

```

DECISION:

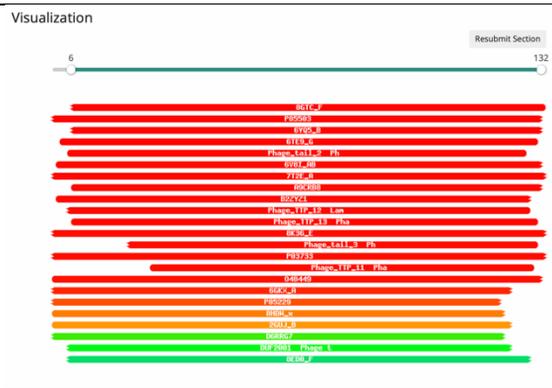
6339

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 style="text-align: center; color: purple;"><u>Distribution of 100 Blast Hits on the Query Sequence</u></p> <div style="font-size: x-small; margin-bottom: 5px;"> Mouse-over to show define and scores. Click to show alignments </div> <p style="text-align: center; font-weight: bold; font-size: small;">Color Key for Alignment Scores</p> <div style="display: flex; justify-content: center; align-items: center; gap: 10px; margin-bottom: 10px;"> <40 40-50 50-80 80-200 >=200 </div> <p>*** nr :</p> <p>(q#: s#) : [alignment] ; e-value :</p> <table border="0" style="width: 100%; font-size: x-small; margin-top: 10px;"> <tr> <td><input checked="" type="checkbox"/> major tail protein [Gordonia phage SallySpecial]</td> <td>Gordonia phage SallySpecial</td> <td>80.1</td> <td>80.1</td> <td>90%</td> <td>9e-16</td> <td>39.20%</td> <td>148</td> <td>YP_010674632.1</td> </tr> <tr> <td><input checked="" type="checkbox"/> TPA tail tube protein [Caudoviricetes.sp.]</td> <td>Caudoviricetes.sp.</td> <td>69.7</td> <td>69.7</td> <td>99%</td> <td>1e-11</td> <td>33.33%</td> <td>142</td> <td>DAP67703.1</td> </tr> <tr> <td><input checked="" type="checkbox"/> TPA tail tube protein [Caudoviricetes.sp.]</td> <td>Caudoviricetes.sp.</td> <td>79.7</td> <td>79.7</td> <td>94%</td> <td>2e-15</td> <td>38.93%</td> <td>159</td> <td>DAI73884.1</td> </tr> <tr> <td><input checked="" type="checkbox"/> major tail protein [Gordonia phage EricDab]</td> <td>Gordonia phage EricDab</td> <td>79.3</td> <td>79.3</td> <td>96%</td> <td>2e-15</td> <td>37.50%</td> <td>145</td> <td>YP_010674654.1</td> </tr> </table>	<input checked="" type="checkbox"/> major tail protein [Gordonia phage SallySpecial]	Gordonia phage SallySpecial	80.1	80.1	90%	9e-16	39.20%	148	YP_010674632.1	<input checked="" type="checkbox"/> TPA tail tube protein [Caudoviricetes.sp.]	Caudoviricetes.sp.	69.7	69.7	99%	1e-11	33.33%	142	DAP67703.1	<input checked="" type="checkbox"/> TPA tail tube protein [Caudoviricetes.sp.]	Caudoviricetes.sp.	79.7	79.7	94%	2e-15	38.93%	159	DAI73884.1	<input checked="" type="checkbox"/> major tail protein [Gordonia phage EricDab]	Gordonia phage EricDab	79.3	79.3	96%	2e-15	37.50%	145	YP_010674654.1
<input checked="" type="checkbox"/> major tail protein [Gordonia phage SallySpecial]	Gordonia phage SallySpecial	80.1	80.1	90%	9e-16	39.20%	148	YP_010674632.1																													
<input checked="" type="checkbox"/> TPA tail tube protein [Caudoviricetes.sp.]	Caudoviricetes.sp.	69.7	69.7	99%	1e-11	33.33%	142	DAP67703.1																													
<input checked="" type="checkbox"/> TPA tail tube protein [Caudoviricetes.sp.]	Caudoviricetes.sp.	79.7	79.7	94%	2e-15	38.93%	159	DAI73884.1																													
<input checked="" type="checkbox"/> major tail protein [Gordonia phage EricDab]	Gordonia phage EricDab	79.3	79.3	96%	2e-15	37.50%	145	YP_010674654.1																													

Sequences producing significant alignments									
Download Select columns Show 100									
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	
hypothetical protein [Corynebacterium belfantii]	Corynebacterium belfantii	122	122	96%	3e-32	52.21%	139	WP_197892065.1	
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	121	121	96%	5e-32	52.21%	139	WP_014318087.1	
hypothetical protein FRC0086_01805 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	120	120	96%	1e-31	52.21%	139	CAB0734176.1	
hypothetical protein CIP107559_01882 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	119	119	96%	2e-31	51.47%	139	CAB0612706.1	
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	119	119	96%	3e-31	51.47%	139	WP_088298565.1	
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	119	119	96%	4e-31	50.74%	139	WP_134884349.1	
hypothetical protein FRC0087_01816 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	119	119	96%	4e-31	51.47%	139	CAB0733813.1	
hypothetical protein CIP107524_01829 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	119	119	96%	4e-31	51.47%	139	CAB0565201.1	
hypothetical protein FRC0405_01824 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	117	117	96%	3e-30	50.74%	139	CAB0884021.1	
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	115	115	93%	5e-30	53.12%	135	HCT5225724.1	
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	115	115	96%	9e-30	49.24%	135	HCD1563133.1	
hypothetical protein [Corynebacterium sp. HMSC065A05]	Corynebacterium sp. HMSC065A05	114	114	93%	2e-29	52.34%	136	WP_070778423.1	
putative tail protein [Corynebacterium striatum]	Corynebacterium striatum	114	114	93%	3e-29	52.34%	135	QDD13876.1	
hypothetical protein [Corynebacterium]	Corynebacterium	114	114	96%	3e-29	48.48%	135	WP_070420700.1	
hypothetical protein [Corynebacterium sp. HMSC077D10]	Corynebacterium sp. HMSC077D10	114	114	93%	4e-29	52.34%	136	WP_070736118.1	
TPA: tail tube protein [Caudoviricetes sp.]	Caudoviricetes sp.	112	112	100%	2e-28	45.99%	137	DAX64369.1	
hypothetical protein [Corynebacterium]	Corynebacterium	110	110	100%	1e-27	45.96%	137	WP_040167460.1	

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 ?

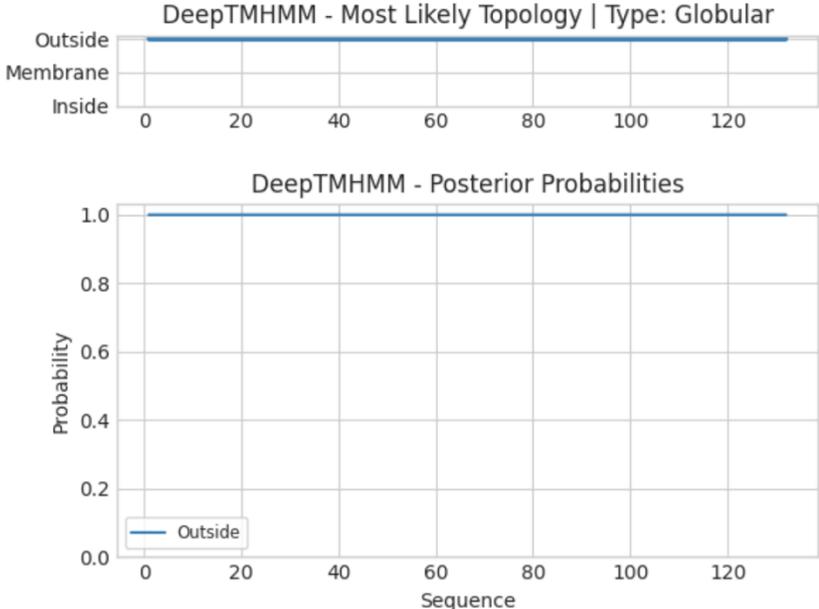


Pfam : Phage_tail_2 ; Phage tail tube protein
 PDB : Major tail protein; Marine bacteriophage, Cryo-EM, Siphophage, Baseplate, Megatron protein, Tail fibre protein, Distal tail protein, VIRAL; 4.5A {Dinoroseobacter phage vB_DshS-R4C}
 Uniprot : STRU3_BPPAJ Structural protein 3 OS=Pseudomonas phage PAJU2 OX=504346 PE=1 SV=2

Accession	Score	Expect	Start	End	Ident	Positives	Negatives	Gaps	Open	Closed
PF05983	122.0	1.7e-34	102	131	100	100	0	0	0	0
PF05983	121.0	5.6e-32	102	131	100	100	0	0	0	0
PF05983	120.0	1.6e-31	102	131	100	100	0	0	0	0
PF05983	119.0	2.1e-31	102	131	100	100	0	0	0	0
PF05983	119.0	3.1e-31	102	131	100	100	0	0	0	0
PF05983	119.0	4.1e-31	102	131	100	100	0	0	0	0
PF05983	119.0	4.1e-31	102	131	100	100	0	0	0	0
PF05983	117.0	3.0e-30	102	131	100	100	0	0	0	0
PF05983	115.0	5.5e-30	102	131	100	100	0	0	0	0
PF05983	115.0	9.0e-30	102	131	100	100	0	0	0	0
PF05983	114.0	2.0e-29	102	131	100	100	0	0	0	0
PF05983	114.0	3.0e-29	102	131	100	100	0	0	0	0
PF05983	114.0	4.0e-29	102	131	100	100	0	0	0	0
PF05983	112.0	2.0e-28	102	131	100	100	0	0	0	0
PF05983	110.0	1.0e-27	102	131	100	100	0	0	0	0

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 ?

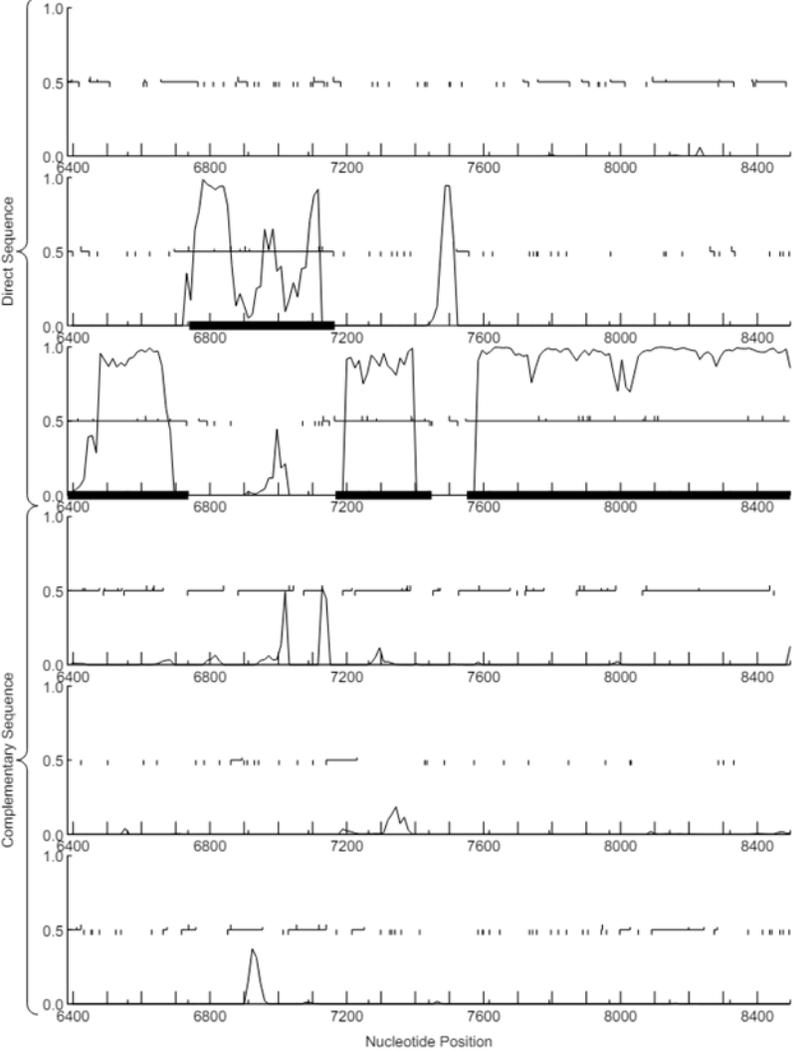
Gene avant : minor tail protein

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

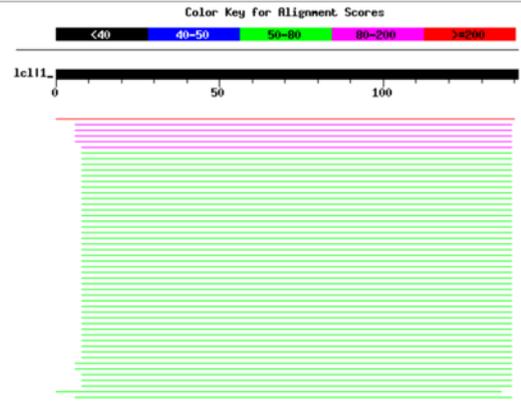
Student Gene Annotation Worksheet

<p>Basic Phage Information</p>	
<p>Nom du Phage</p>	<p>CyranoPS</p>
<p>Gène #</p>	<p>11</p>
<p>Coordonnées du Stop</p>	<p>7165</p>
<p>Direction (For/Rev)</p>	<p>For</p>
<p>Gap/chevauchement avec un autre gène</p>	<p>Gap 2</p>
<p>Coordonnées du Start retenu</p>	<p>6740</p>
<p>Fonction prédite</p>	<p>NKF</p>

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 data-bbox="560 367 1404 399">noPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 5/9</p>  <p data-bbox="609 409 1404 1459">The figure displays a GeneMark hmm prediction plot for a DNA sequence. The x-axis represents the Nucleotide Position, ranging from 6400 to 8400. The y-axis represents the signal strength, ranging from 0.0 to 1.0. The plot is divided into two main sections: Direct Sequence (top) and Complementary Sequence (bottom). Each section contains two signal traces and a GeneMark hmm prediction bar. The prediction bar shows a solid black line indicating a predicted coding region between approximately 6800 and 7200 nucleotide positions. The signal traces show peaks corresponding to the predicted coding region, with the highest peak reaching a signal strength of 1.0.</p>

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



Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyanoP5_Draft_11, function unknown, 141	284	6e-77
Schiebs_12, function unknown, 124	188	4e-24
Emperor_14, function unknown, 129	187	1e-23
SallySpecial_12, function unknown, 130	186	2e-23
EpicDab_13, function unknown, 127	181	6e-22
Swenson_11, minor tail protein, 125	88	1e-15
Toulouse_10, minor tail protein, 125	79	4e-15
Seume_11, minor tail protein, 125	79	4e-15
Taj14_11, minor tail protein, 125	78	7e-15
StewieGriff_11, minor tail protein, 125	78	7e-15
Saphira_11, minor tail protein, 125	78	7e-15
Lore_11, minor tail protein, 125	78	7e-15

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

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 [Connebacterium striatum]	Connebacterium striatum	188	188	97%	2e-58	64.49%	144	WP_201806824.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium]	Connebacterium	183	183	97%	2e-56	62.77%	140	WP_064833785.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium pseudodichtherium]	Connebacterium pseudodichth...	183	183	97%	2e-56	62.77%	140	WP_284849035.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium eroinoum]	Connebacterium eroinoum	182	182	97%	8e-56	62.77%	140	WP_284504216.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium hadale]	Connebacterium hadale	180	180	95%	3e-55	64.18%	142	WP_095275385.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium sp. HMSC06C06]	Connebacterium sp. HMSC06C06	180	180	87%	5e-55	69.92%	142	WP_246817189.1
<input checked="" type="checkbox"/> outative shape protein [Connebacterium striatum]	Connebacterium striatum	177	177	95%	9e-54	60.74%	147	C0013979.1
<input checked="" type="checkbox"/> hypothetical protein [unclassified Connebacterium]	unclassified Connebacterium	175	175	95%	4e-53	60.74%	147	WP_246817216.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Connebacterium striatum]	Connebacterium striatum	172	172	92%	4e-52	61.07%	138	HCT5225725.1
<input checked="" type="checkbox"/> hypothetical protein HMPREF2998_07920 [Connebacterium sp. HMSC06SA05]	Connebacterium sp. HMSC06S...	171	171	92%	1e-51	61.07%	138	QEP20320.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium dichtheriae]	Connebacterium dichtheriae	162	162	99%	5e-48	57.75%	137	WP_010935362.1
<input checked="" type="checkbox"/> hypothetical protein CIP197541_01712 [Connebacterium dichtheriae]	Connebacterium dichtheriae	162	162	99%	7e-48	57.75%	152	CAB0610097.1
<input checked="" type="checkbox"/> hypothetical protein FRC0431_01842 [Connebacterium dichtheriae]	Connebacterium dichtheriae	160	160	99%	3e-47	57.04%	152	CAB0923859.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium eroinoum]	Connebacterium eroinoum	159	159	96%	6e-47	53.68%	144	WP_239210984.1
<input checked="" type="checkbox"/> hypothetical protein [Connebacterium]	Connebacterium	159	159	94%	1e-46	54.89%	144	WP_065420615.1
<input checked="" type="checkbox"/> hypothetical protein WSM_08714 [Connebacterium dichtheriae bv. intermedium str. NCTC 5011] Connebacterium dichtheriae bv. ...	Connebacterium dichtheriae bv. ...	156	156	95%	5e-46	57.46%	129	EIK55670.1

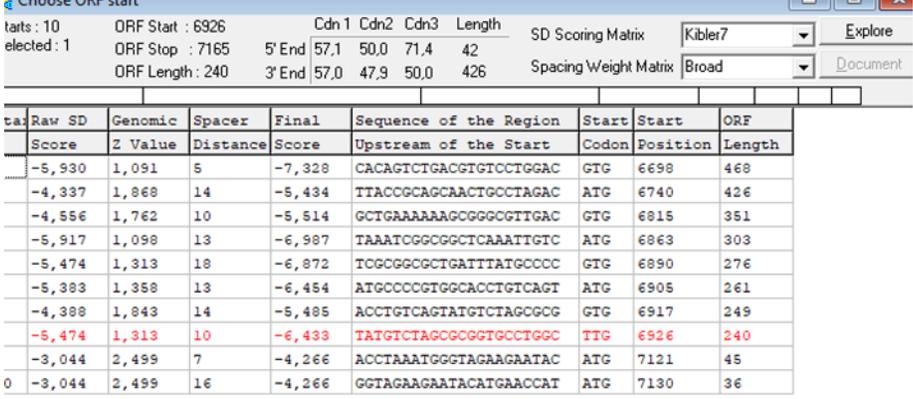
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																									
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 6740 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 6740</p>																																																																																									
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <table border="1" data-bbox="561 598 1356 892"> <thead> <tr> <th>Raw SD</th> <th>Genomic Z Value</th> <th>Spacer 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>-5,930</td> <td>1,091</td> <td>5</td> <td>-7,328</td> <td>CACAGTCTGACGTTCTCTGGAC</td> <td>GTG</td> <td>6698</td> <td>468</td> </tr> <tr> <td>-4,337</td> <td>1,868</td> <td>14</td> <td>-5,434</td> <td>TTACCGCAGCAACTGCCTAGAC</td> <td>ATG</td> <td>6740</td> <td>426</td> </tr> <tr> <td>-4,556</td> <td>1,762</td> <td>10</td> <td>-5,514</td> <td>GCTGAAAAAAGCGGGCGTTGAC</td> <td>GTG</td> <td>6815</td> <td>351</td> </tr> <tr> <td>-5,917</td> <td>1,098</td> <td>13</td> <td>-6,987</td> <td>TAAATCGGGGCTCAAATTGTC</td> <td>ATG</td> <td>6863</td> <td>303</td> </tr> <tr> <td>-5,474</td> <td>1,313</td> <td>18</td> <td>-6,872</td> <td>TCGGGGGCTGATTTATGCCCC</td> <td>GTG</td> <td>6890</td> <td>276</td> </tr> <tr> <td>-5,383</td> <td>1,358</td> <td>13</td> <td>-6,454</td> <td>ATGCCCCGTGGCACCTGTGAGT</td> <td>ATG</td> <td>6905</td> <td>261</td> </tr> <tr> <td>-4,388</td> <td>1,843</td> <td>14</td> <td>-5,485</td> <td>ACCTGTCAGTATGCTAGCGCG</td> <td>GTG</td> <td>6917</td> <td>249</td> </tr> <tr> <td>-5,474</td> <td>1,313</td> <td>10</td> <td>-6,433</td> <td>TATGCTAGCGCGGTGCTGGC</td> <td>TTG</td> <td>6926</td> <td>240</td> </tr> <tr> <td>-3,044</td> <td>2,499</td> <td>7</td> <td>-4,266</td> <td>ACCTAAATGGGTAGAAGAATAC</td> <td>ATG</td> <td>7121</td> <td>45</td> </tr> <tr> <td>0</td> <td>-3,044</td> <td>2,499</td> <td>16</td> <td>-4,266</td> <td>GGTAGAAGAATACATGAACCAT</td> <td>ATG</td> <td>7130</td> <td>36</td> </tr> </tbody> </table>	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length	-5,930	1,091	5	-7,328	CACAGTCTGACGTTCTCTGGAC	GTG	6698	468	-4,337	1,868	14	-5,434	TTACCGCAGCAACTGCCTAGAC	ATG	6740	426	-4,556	1,762	10	-5,514	GCTGAAAAAAGCGGGCGTTGAC	GTG	6815	351	-5,917	1,098	13	-6,987	TAAATCGGGGCTCAAATTGTC	ATG	6863	303	-5,474	1,313	18	-6,872	TCGGGGGCTGATTTATGCCCC	GTG	6890	276	-5,383	1,358	13	-6,454	ATGCCCCGTGGCACCTGTGAGT	ATG	6905	261	-4,388	1,843	14	-5,485	ACCTGTCAGTATGCTAGCGCG	GTG	6917	249	-5,474	1,313	10	-6,433	TATGCTAGCGCGGTGCTGGC	TTG	6926	240	-3,044	2,499	7	-4,266	ACCTAAATGGGTAGAAGAATAC	ATG	7121	45	0	-3,044	2,499	16	-4,266	GGTAGAAGAATACATGAACCAT	ATG	7130	36
Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length																																																																																			
-5,930	1,091	5	-7,328	CACAGTCTGACGTTCTCTGGAC	GTG	6698	468																																																																																			
-4,337	1,868	14	-5,434	TTACCGCAGCAACTGCCTAGAC	ATG	6740	426																																																																																			
-4,556	1,762	10	-5,514	GCTGAAAAAAGCGGGCGTTGAC	GTG	6815	351																																																																																			
-5,917	1,098	13	-6,987	TAAATCGGGGCTCAAATTGTC	ATG	6863	303																																																																																			
-5,474	1,313	18	-6,872	TCGGGGGCTGATTTATGCCCC	GTG	6890	276																																																																																			
-5,383	1,358	13	-6,454	ATGCCCCGTGGCACCTGTGAGT	ATG	6905	261																																																																																			
-4,388	1,843	14	-5,485	ACCTGTCAGTATGCTAGCGCG	GTG	6917	249																																																																																			
-5,474	1,313	10	-6,433	TATGCTAGCGCGGTGCTGGC	TTG	6926	240																																																																																			
-3,044	2,499	7	-4,266	ACCTAAATGGGTAGAAGAATAC	ATG	7121	45																																																																																			
0	-3,044	2,499	16	-4,266	GGTAGAAGAATACATGAACCAT	ATG	7130	36																																																																																		
<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>	<p><i>Plus long ORF avec un ATG</i></p>																																																																																									
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>—</p>																																																																																									

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	
<p>DECISION:</p>	<p>6740</p>

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

<p>Collection des éléments de réponse</p>	<p>Rationnelle</p>																								
<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>*** nr :</p> <table border="1"> <thead> <tr> <th>Sequences producing significant alignments:</th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>CyranoPS_Draft_11, function unknown, 141</td> <td>284</td> <td>6e-77</td> </tr> <tr> <td>Schiebs_12, function unknown, 124</td> <td>108</td> <td>4e-24</td> </tr> <tr> <td>Emperor_14, function unknown, 129</td> <td>107</td> <td>1e-23</td> </tr> <tr> <td>SallySpecial_12, function unknown, 130</td> <td>106</td> <td>2e-23</td> </tr> <tr> <td>EpicDab_13, function unknown, 127</td> <td>101</td> <td>6e-22</td> </tr> <tr> <td>Swenson_11, minor tail protein, 125</td> <td>80</td> <td>1e-15</td> </tr> <tr> <td>Toulouse_10, minor tail protein, 125</td> <td>79</td> <td>4e-15</td> </tr> </tbody> </table>	Sequences producing significant alignments:	Score (bits)	E Value	CyranoPS_Draft_11, function unknown, 141	284	6e-77	Schiebs_12, function unknown, 124	108	4e-24	Emperor_14, function unknown, 129	107	1e-23	SallySpecial_12, function unknown, 130	106	2e-23	EpicDab_13, function unknown, 127	101	6e-22	Swenson_11, minor tail protein, 125	80	1e-15	Toulouse_10, minor tail protein, 125	79	4e-15
Sequences producing significant alignments:	Score (bits)	E Value																							
CyranoPS_Draft_11, function unknown, 141	284	6e-77																							
Schiebs_12, function unknown, 124	108	4e-24																							
Emperor_14, function unknown, 129	107	1e-23																							
SallySpecial_12, function unknown, 130	106	2e-23																							
EpicDab_13, function unknown, 127	101	6e-22																							
Swenson_11, minor tail protein, 125	80	1e-15																							
Toulouse_10, minor tail protein, 125	79	4e-15																							

Compare these results against the new Clustered nr database [BLAST](#)

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 100 sequences selected GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

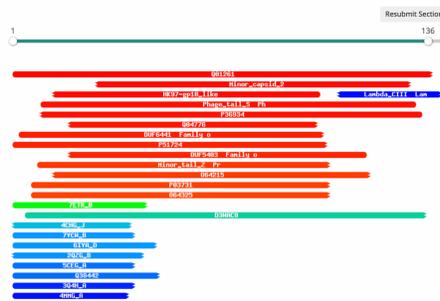
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium striatum]	<i>Corynebacterium striatum</i>	188	188	97%	2e-58	64.49%	144	WP_201808824.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium]	<i>Corynebacterium</i>	183	183	97%	2e-56	62.77%	140	WP_064833785.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodiphtheriticum]	<i>Corynebacterium pseud...</i>	183	183	97%	2e-56	62.77%	140	WP_284849035.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium propinquum]	<i>Corynebacterium propin...</i>	182	182	97%	8e-56	62.77%	140	WP_284594216.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hadale]	<i>Corynebacterium hadale</i>	180	180	95%	3e-55	64.16%	142	WP_095275385.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium sp. HMSC06C06]	<i>Corynebacterium sp. H...</i>	180	180	87%	5e-55	69.92%	142	WP_246817189.1
<input checked="" type="checkbox"/> putative phage protein [Corynebacterium striatum]	<i>Corynebacterium striatum</i>	177	177	95%	9e-54	60.74%	147	QQD13979.1
<input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium]	<i>unclassified Corynebac...</i>	175	175	95%	4e-53	60.74%	147	WP_246817216.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Corynebacterium striatum]	<i>Corynebacterium striatum</i>	172	172	92%	4e-52	61.07%	138	HCT15225725.1
<input checked="" type="checkbox"/> hypothetical protein HMPREF298_07820 [Corynebacterium sp. HMSC05A05]	<i>Corynebacterium sp. H...</i>	171	171	92%	1e-51	61.07%	138	QFP20320.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diphtheriae]	<i>Corynebacterium diphth...</i>	162	162	99%	5e-48	57.75%	137	WP_010935362.1
<input checked="" type="checkbox"/> hypothetical protein CIP107541_01712 [Corynebacterium diphtheriae]	<i>Corynebacterium diphth...</i>	162	162	99%	7e-48	57.75%	152	CAB0610097.1
<input checked="" type="checkbox"/> hypothetical protein FRC0431_01842 [Corynebacterium diphtheriae]	<i>Corynebacterium diphth...</i>	160	160	99%	3e-47	57.04%	152	CAB0923859.1
<input checked="" type="checkbox"/> TPA: HK97_gp10 family phage protein [Corynebacterium striatum]	<i>Corynebacterium striatum</i>	129	129	95%	4e-35	48.89%	131	HAT1153068.1
<input checked="" type="checkbox"/> TPA: HK97_gp10 family phage protein [Corynebacterium striatum]	<i>Corynebacterium striatum</i>	127	127	95%	1e-34	48.15%	131	HAT1137143.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Gulosibacter bifidus]	<i>Gulosibacter bifidus</i>	127	127	95%	2e-34	52.99%	125	WP_083524510.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium mastitidis]	<i>Corynebacterium mastitidis</i>	125	125	96%	9e-34	47.79%	131	WP_082422082.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium accolens]	<i>Corynebacterium accolens</i>	125	125	95%	1e-33	44.44%	131	WP_302526990.1
<input checked="" type="checkbox"/> TPA: putative tail component [Caudoviricetes sp.]	<i>Caudoviricetes sp.</i>	124	124	95%	3e-33	47.41%	131	DAL07359.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Corynebacterium mastitidis]	<i>Corynebacterium mastitidis</i>	124	124	96%	6e-33	46.32%	131	WP_337890914.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Gulosibacter hominis]	<i>Gulosibacter hominis</i>	123	123	95%	9e-33	51.49%	129	WP_201518892.1
<input checked="" type="checkbox"/> hypothetical protein [Schaalia hyovaginalis]	<i>Schaalia hyovaginalis</i>	118	118	97%	6e-31	48.18%	130	WP_303773435.1
<input checked="" type="checkbox"/> hypothetical protein [Schaalia georgiae]	<i>Schaalia georgiae</i>	115	115	99%	6e-30	43.36%	130	WP_005867136.1
<input checked="" type="checkbox"/> hypothetical protein [Actinomyces urogenitalis]	<i>Actinomyces urogenitalis</i>	115	115	98%	6e-30	46.76%	131	WP_278787886.1
<input checked="" type="checkbox"/> hypothetical protein [Microterricola viridarii]	<i>Microterricola viridarii</i>	114	114	95%	3e-29	45.19%	130	WP_067226933.1
<input checked="" type="checkbox"/> TPA: putative tail component [Caudoviricetes sp.]	<i>Caudoviricetes sp.</i>	114	114	97%	6e-29	46.38%	138	DAT06839.1
<input checked="" type="checkbox"/> TPA: putative tail component [Caudoviricetes sp.]	<i>Caudoviricetes sp.</i>	113	113	99%	7e-29	43.36%	130	DAP67704.1
<input checked="" type="checkbox"/> hypothetical protein [Microbacterium dauci]	<i>Microbacterium dauci</i>	112	112	100%	1e-28	46.81%	126	WP_283717022.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Bifidobacterium]	<i>Bifidobacterium</i>	111	111	97%	6e-28	46.72%	135	WP_278711690.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Bifidobacterium sp. ESL0690]	<i>Bifidobacterium sp. ESL0690</i>	109	109	99%	2e-27	44.29%	127	WP_277176301.1
<input checked="" type="checkbox"/> TPA: putative tail component [Caudoviricetes sp.]	<i>Caudoviricetes sp.</i>	109	109	95%	3e-27	41.79%	135	DAQ68535.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium freiburgense]	<i>Corynebacterium freiburgense</i>	109	109	95%	3e-27	43.28%	137	WP_027011474.1
<input checked="" type="checkbox"/> TPA: putative tail component [Caudoviricetes sp.]	<i>Caudoviricetes sp.</i>	109	109	95%	3e-27	41.79%	135	DAD60079.1
<input checked="" type="checkbox"/> TPA: putative tail component [Caudoviricetes sp.]	<i>Caudoviricetes sp.</i>	108	108	95%	7e-27	41.79%	135	DAK55028.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Corynebacterium matruchotii]	<i>Corynebacterium matruchotii</i>	108	108	95%	9e-27	41.79%	135	WP_126299898.1
<input checked="" type="checkbox"/> hypothetical protein [Schaalia vaccinae]	<i>Schaalia vaccinae</i>	108	108	95%	1e-26	44.67%	146	WP_022867147.1
<input checked="" type="checkbox"/> HK97_gp10 family putative phage morphogenesis protein [Canibacter zhoujuianae]	<i>Canibacter zhoujuianae</i>	107	107	99%	1e-26	41.43%	127	WP_166985206.1
<input checked="" type="checkbox"/> TPA: putative tail component [Siphoviridae sp. cCCy12]	<i>Siphoviridae sp. cCCy12</i>	107	107	95%	1e-26	41.79%	135	DAD8989.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Gemmatimonadaceae bacterium]	<i>Gemmatimonadaceae bacterium</i>	107	107	98%	2e-26	40.00%	129	HXG71289.1
<input checked="" type="checkbox"/> HK97_gp10 family phage protein [Microbacterium sorbilolivorans]	<i>Microbacterium sorbilolivorans</i>	107	107	98%	2e-26	43.17%	134	WP_114116776.1
<input checked="" type="checkbox"/> hypothetical protein [Glutamicobacter sp. BW80]	<i>Glutamicobacter sp. BW80</i>	107	107	95%	2e-26	43.28%	128	WP_096284140.1

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 $\geq 90\%$ et une couverture acceptable ?

Pfam : Minor_capsid_2 ; Minor capsid protein
 Uniport : GPG_BPMU Putative capsid assembly protein G
 OS=Escherichia phage Mu OX=10677 GN=G PE=2 SV=1

<input type="checkbox"/>	1	Q01261	GPG_BPMU Putative capsid assembly protein G OS=Escherichia phage Mu OX=10677 GN=G PE=2 SV=1	99.51	1.7e-13	85.1	7.1	124	156
<input type="checkbox"/>	2	PF11114.12	; Minor_capsid_2 ; Minor capsid protein	99.42	2.8e-12	74.8	7.1	83	105
<input type="checkbox"/>	3	PF04883.16	; HK97-gp10_like ; Bacteriophage HK97-gp10, putative tail-component	99.13	5.5e-10	60.68	5.7	72	74
<input type="checkbox"/>	4	PF05069.17	; Phage_tail_S ; Phage virion morphogenesis family	98.64	1.4e-7	58.5	4.7	113	145
<input type="checkbox"/>	5	P36934	VPS_BPP2 Tail completion protein S OS=Escherichia phage P2 OX=10679 GN=S PE=4 SV=1	98.6	7.6e-7	55.51	7	115	150
<input type="checkbox"/>	6	Q04776	YG36_BPMV4 Uncharacterized protein ORF6 OS=Lactococcus phage mv4 OX=12392 PE=4 SV=1	98.01	0.000024	44.44	3.6	61	85
<input type="checkbox"/>	7	PF20039.3	; DUF6441 ; Family of unknown function (DUF6441)	97.48	0.0015	43.84	6.3	87	215
<input type="checkbox"/>	8	P51724	YO22_BPHC1 Uncharacterized 26.0 kDa protein in rep-hol intergenic region OS=Haemophilus phage HP1 (strain HP1c1) OX=1289	97.15	0.0011	44.83	3.1	90	227
<input type="checkbox"/>	9	PF17395.6	; DUF5403 ; Family of unknown function (DUF5403)	97.02	0.0026	36.93	3.4	80	93
<input type="checkbox"/>	10	PF06763.15	; Minor_tail_Z ; Prophage minor tail protein Z (GPZ)	95.33	0.44	31.31	7.2	85	182
<input type="checkbox"/>	11	O64215	VG21_BPMD2 Gene 21 protein OS=Mycobacterium phage D29 OX=28369 GN=21 PE=4 SV=1	95.18	0.025	34.21	1.1	88	111

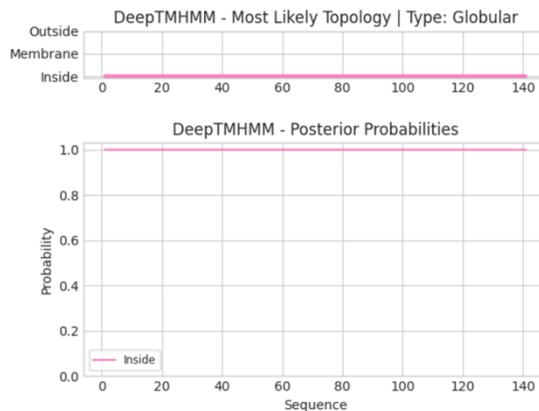
Visualization



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 synthèse

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 ?	Non
DECISION:	NKF

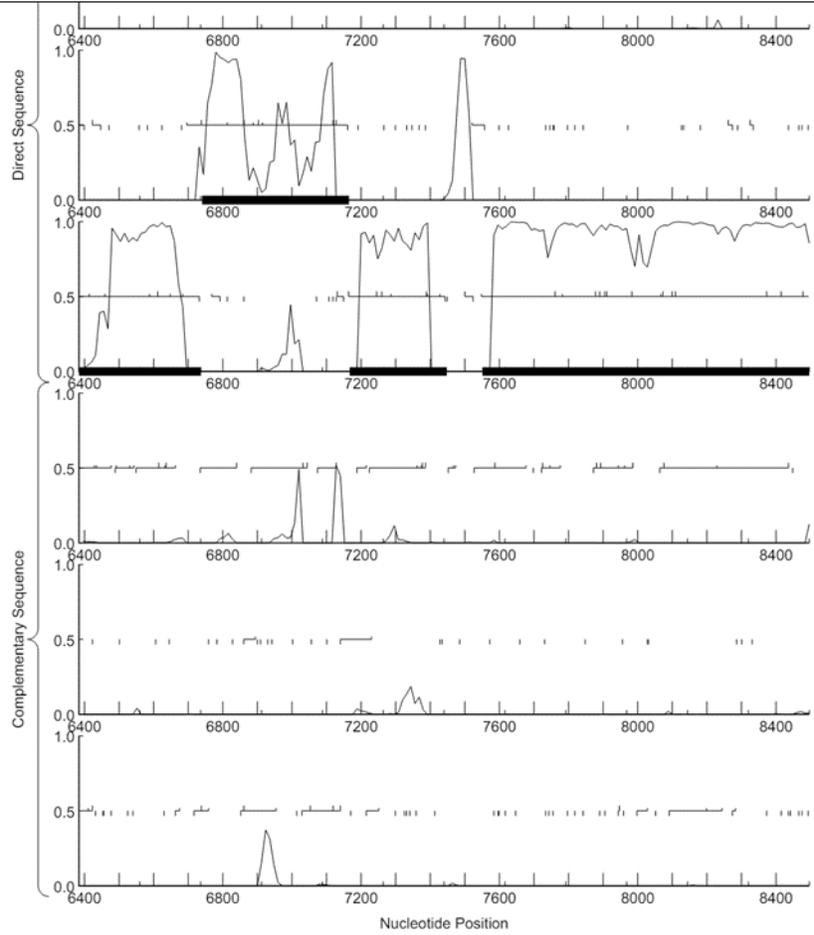
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	12
Coordonnées du Stop	7448
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 1
Coordonnées du Start retenu	7167
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 ?



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

Other reports

No significant similarity found. For reasons why [click here](#)

Query: CyranoP5_12 (93 letters)

Distribution of 3 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:

Sequence	Score (bits)	E Value
CyranoP5_Draft_12, function unknown, 93	186	2e-47
KeA11i_2, terminase large subunit, 565	28	7.9
DrSierra_2, terminase large subunit, 568	28	7.9

>CyranoP5_Draft_12, function unknown, 93
Length = 93

Score = 186 bits (472), Expect = 2e-47
Identities = 93/93 (100%), Positives = 93/93 (100%)

Query: 1 MAIQEVTVLDGKKETATISQGHRIAEHALSAMGRAVAQNAVTLNVAAFFAVVNPCKFN 60
Sbjct: 1 MAIQEVTVLDGKKETATISQGHRIAEHALSAMGRAVAQNAVTLNVAAFFAVVNPCKFN 60

Query: 61 FDEFLTMSDTIENIEMVEEQVNPQADLVLDPQ 93
FDEFLTMSDTIENIEMVEEQVNPQADLVLDPQ
Sbjct: 61 FDEFLTMSDTIENIEMVEEQVNPQADLVLDPQ 93

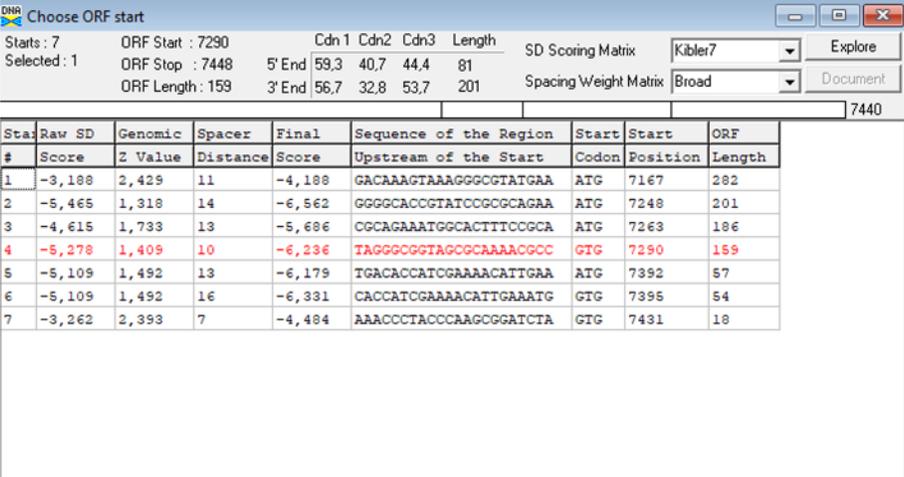
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

non

DECISION:

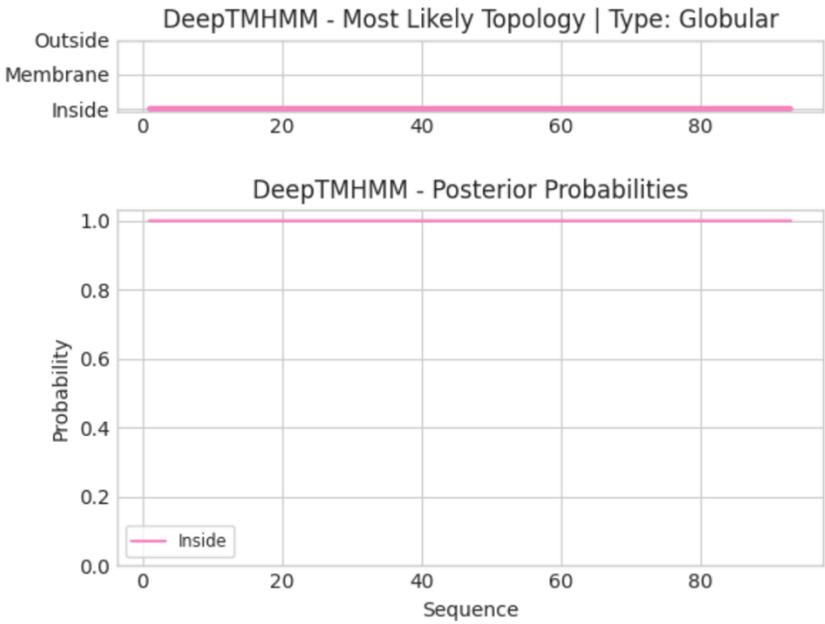
oui

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																	
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :7167 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :7167																																																																																	
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <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>-3,188</td> <td>2,429</td> <td>11</td> <td>-4,188</td> <td>GACAAAGTAAAGGGCGTATGAA</td> <td>ATG</td> <td>7167</td> <td>282</td> </tr> <tr> <td>2</td> <td>-5,465</td> <td>1,318</td> <td>14</td> <td>-6,562</td> <td>GGGGCACCGTATCGGGCAGAA</td> <td>ATG</td> <td>7248</td> <td>201</td> </tr> <tr> <td>3</td> <td>-4,615</td> <td>1,733</td> <td>13</td> <td>-5,686</td> <td>CGCAGAAATGGCACTTTCCGCA</td> <td>ATG</td> <td>7263</td> <td>186</td> </tr> <tr> <td>4</td> <td>-5,278</td> <td>1,409</td> <td>10</td> <td>-6,236</td> <td>TAGGGCGGTAGCGCAAACGCC</td> <td>GTG</td> <td>7290</td> <td>159</td> </tr> <tr> <td>5</td> <td>-5,109</td> <td>1,492</td> <td>13</td> <td>-6,179</td> <td>TGACCCATCGAAAACATTGAA</td> <td>ATG</td> <td>7392</td> <td>57</td> </tr> <tr> <td>6</td> <td>-5,109</td> <td>1,492</td> <td>16</td> <td>-6,331</td> <td>CACCATCGAAAACATTGAAATG</td> <td>GTG</td> <td>7395</td> <td>54</td> </tr> <tr> <td>7</td> <td>-3,262</td> <td>2,393</td> <td>7</td> <td>-4,484</td> <td>AAACCOCTACCCAAAGCGGATCTA</td> <td>GTG</td> <td>7431</td> <td>18</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	-3,188	2,429	11	-4,188	GACAAAGTAAAGGGCGTATGAA	ATG	7167	282	2	-5,465	1,318	14	-6,562	GGGGCACCGTATCGGGCAGAA	ATG	7248	201	3	-4,615	1,733	13	-5,686	CGCAGAAATGGCACTTTCCGCA	ATG	7263	186	4	-5,278	1,409	10	-6,236	TAGGGCGGTAGCGCAAACGCC	GTG	7290	159	5	-5,109	1,492	13	-6,179	TGACCCATCGAAAACATTGAA	ATG	7392	57	6	-5,109	1,492	16	-6,331	CACCATCGAAAACATTGAAATG	GTG	7395	54	7	-3,262	2,393	7	-4,484	AAACCOCTACCCAAAGCGGATCTA	GTG	7431	18
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	-3,188	2,429	11	-4,188	GACAAAGTAAAGGGCGTATGAA	ATG	7167	282																																																																										
2	-5,465	1,318	14	-6,562	GGGGCACCGTATCGGGCAGAA	ATG	7248	201																																																																										
3	-4,615	1,733	13	-5,686	CGCAGAAATGGCACTTTCCGCA	ATG	7263	186																																																																										
4	-5,278	1,409	10	-6,236	TAGGGCGGTAGCGCAAACGCC	GTG	7290	159																																																																										
5	-5,109	1,492	13	-6,179	TGACCCATCGAAAACATTGAA	ATG	7392	57																																																																										
6	-5,109	1,492	16	-6,331	CACCATCGAAAACATTGAAATG	GTG	7395	54																																																																										
7	-3,262	2,393	7	-4,484	AAACCOCTACCCAAAGCGGATCTA	GTG	7431	18																																																																										
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>Meilleur score,</i></p> <p><i>ORF le plus long</i></p> <p><i>ORF le plus long avec un ATG</i></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>Other reports</p> <p> No significant similarity found. For reasons why, click here</p>																																																																																	
DECISION:	7167																																																																																	

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>RIEN</p> <p>*** nr :</p> <p>(q#: s#) : [alignment] ; e-value :</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>																											
<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>"Pas de correspondance pertinente"</p>  <p>Hitlist</p> <p>Show 25 Entries Search:</p> <table border="1"> <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>cd11010</td> <td>S1-P1_nuclease; S1/P1 nucleases and related enzymes. This family summarizes both S1 and P1 nucleases (EC:3.</td> <td>54.53</td> <td>22</td> <td>24.11</td> <td>2.3</td> <td>44</td> <td>264</td> </tr> <tr> <td>2</td> <td>cd14487</td> <td>AlgX_C; C-terminal carbohydrate-binding domain of the alginate O-acetyltransferase AlgX. The alginate biosynthesis protei</td> <td>50.72</td> <td>19</td> <td>23.58</td> <td>1.5</td> <td>22</td> <td>125</td> </tr> </tbody> </table>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	1	cd11010	S1-P1_nuclease; S1/P1 nucleases and related enzymes. This family summarizes both S1 and P1 nucleases (EC:3.	54.53	22	24.11	2.3	44	264	2	cd14487	AlgX_C; C-terminal carbohydrate-binding domain of the alginate O-acetyltransferase AlgX. The alginate biosynthesis protei	50.72	19	23.58	1.5	22	125
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																				
1	cd11010	S1-P1_nuclease; S1/P1 nucleases and related enzymes. This family summarizes both S1 and P1 nucleases (EC:3.	54.53	22	24.11	2.3	44	264																				
2	cd14487	AlgX_C; C-terminal carbohydrate-binding domain of the alginate O-acetyltransferase AlgX. The alginate biosynthesis protei	50.72	19	23.58	1.5	22	125																				
<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>Gene precedent est inconnu mais le gene suivant est une tape mesure protein</p>																											

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	 <p>DeepTMHMM - Most Likely Topology Type: Globular</p> <p>DeepTMHMM - Posterior Probabilities</p>
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>Répondez Oui ou Non. Une fois que vous avez pris une décision quant-à-la fonction du gène, vérifiez la liste des 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.</p>
<p>DECISION:</p>	<p>NKF</p>

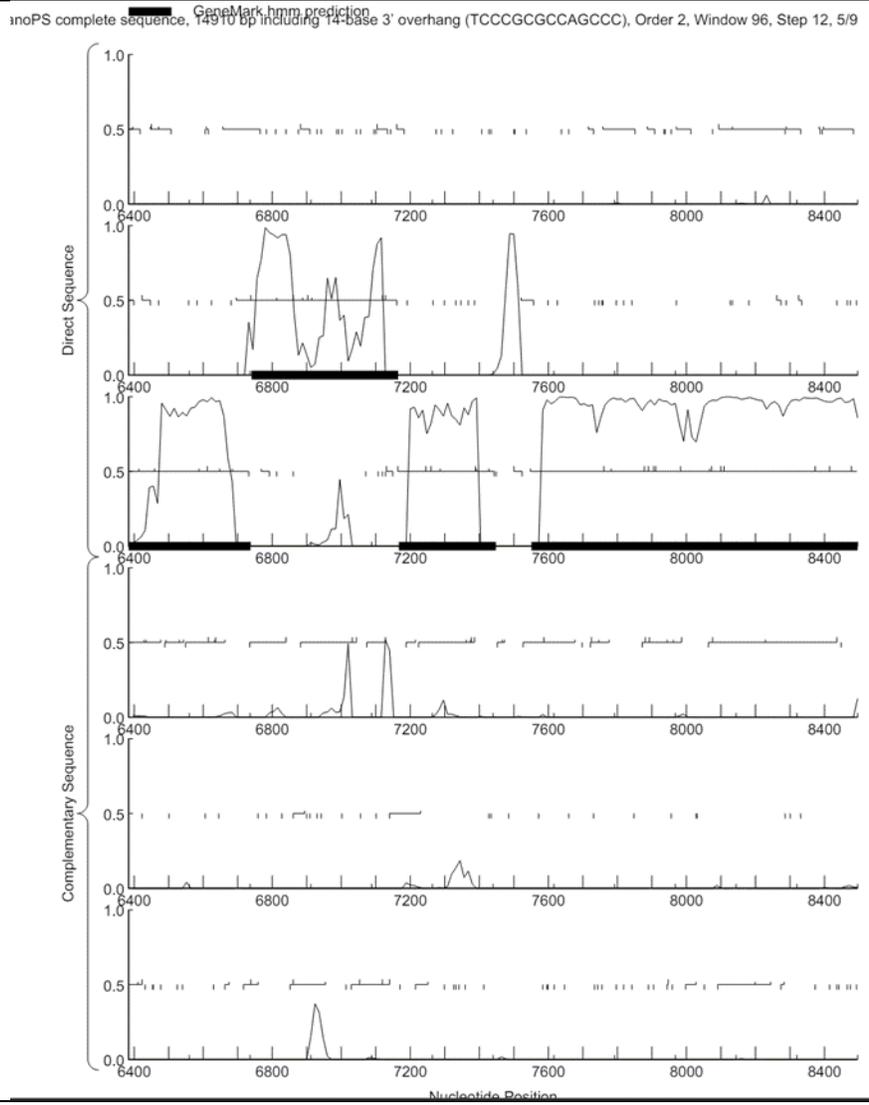
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	13
Coordonnées du Stop	9212
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 103
Coordonnées du Start retenu	7551
Fonction prédite	tape measure protein

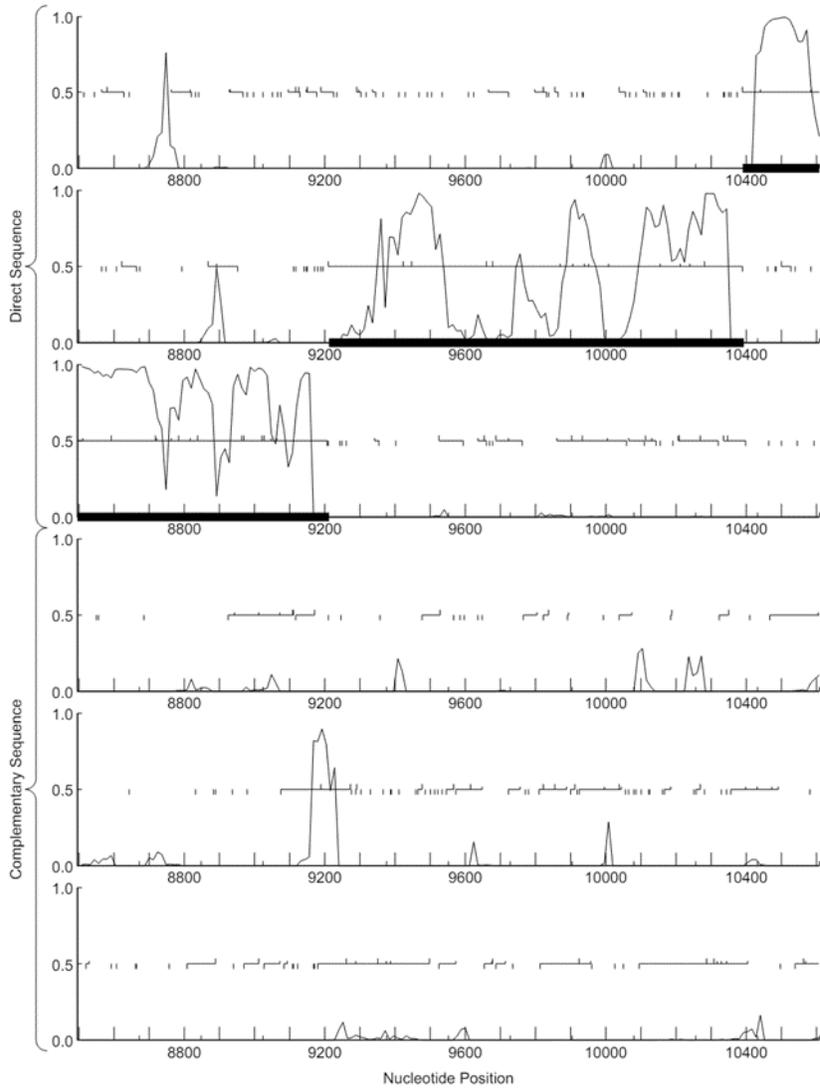
Décision #1 : Est-ce un gène ?

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 ?



GeneMark_hmm prediction
 anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCGCGCCAGCCC), Order 2, Window 96, Step 12, 6/9



Distribution of 102 Blast Hits on the Query Sequence

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

Color Key for Alignment Scores



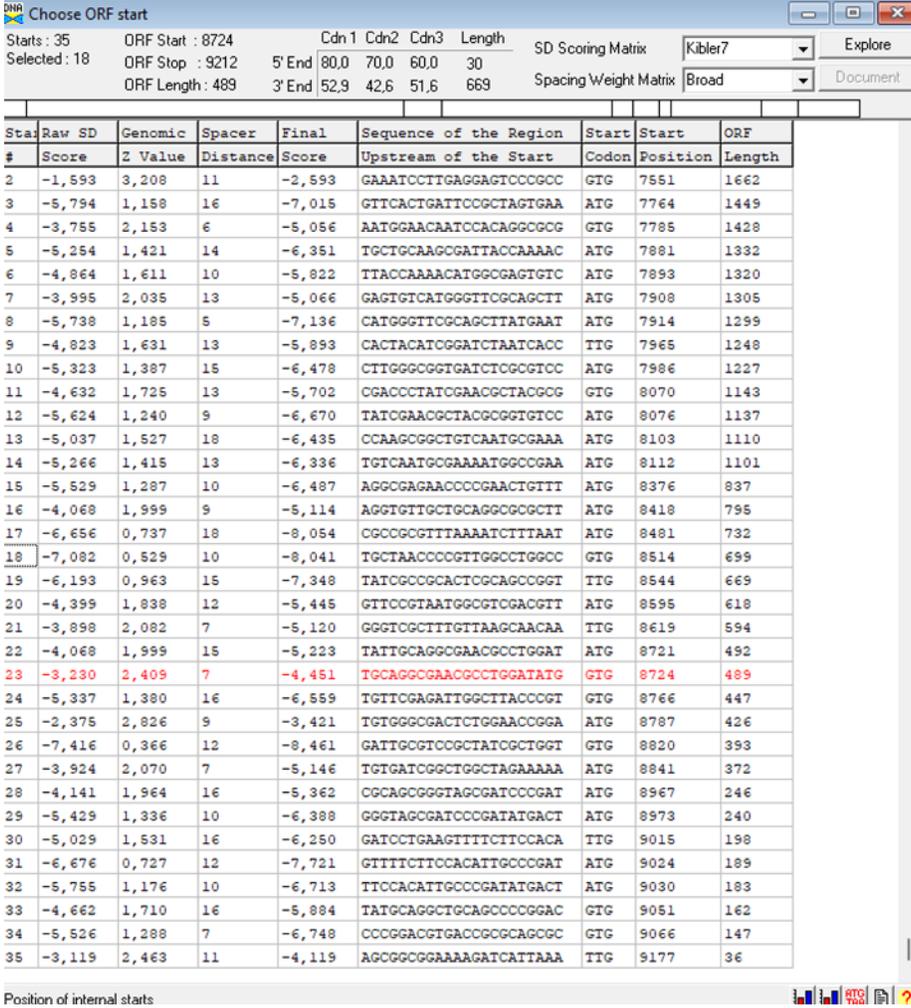
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?

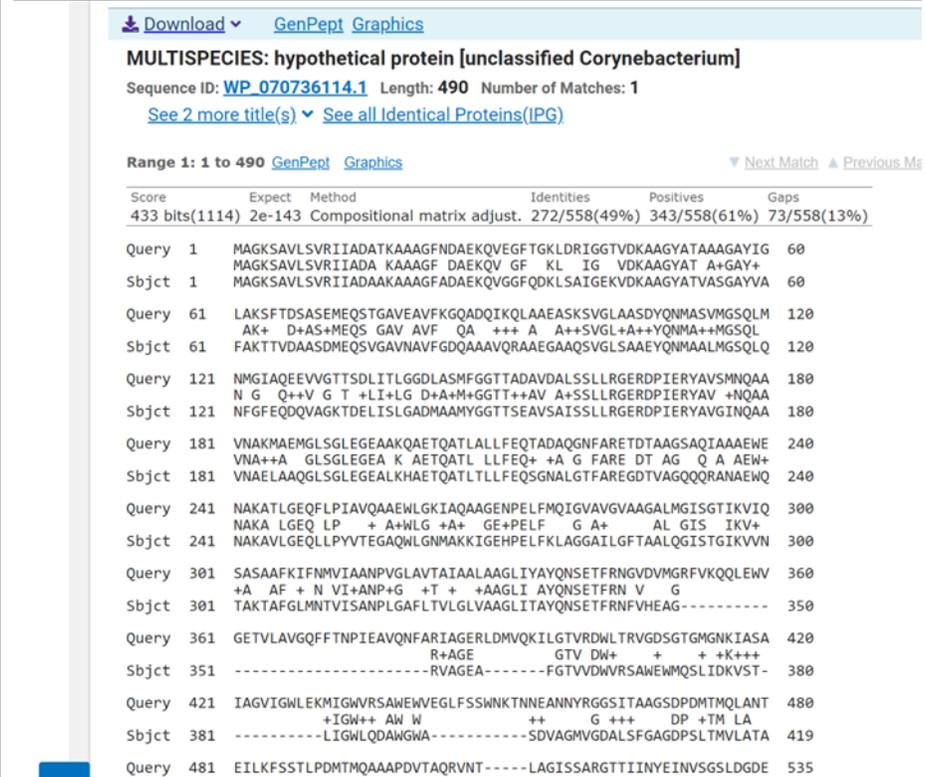
Sequences producing significant alignments:

Sequence	Score (bits)	E Value
CyranoPS_Draft_13, function unknown, 553	1059	0.0
SerialPhiller_11, tape measure protein, 607	323	9e-88
Kels_11, tape measure protein, 607	323	9e-88
Arielagos_11, tape measure protein, 607	323	9e-88
StewieGriff_13, tape measure protein, 665	322	1e-87
Lore_13, tape measure protein, 665	322	1e-87
Laila_15, tape measure protein, 665	322	1e-87
KylieMac 14, tape measure protein, 665	322	1e-87

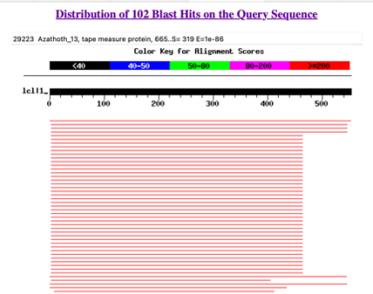
	<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium] Corynebacterium 444 444 100% 1e-147 49.10% 492 WP_070420702.1 <input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 436 436 100% 3e-144 48.92% 490 HGT5225727.1 <input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium] unclassified Corynebacterium 433 433 100% 2e-143 48.75% 490 WP_070736114.1 <input checked="" type="checkbox"/> putative chape protein [Corynebacterium striatum] Corynebacterium striatum 433 433 100% 3e-143 48.75% 490 COD13985.1 <input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 432 432 100% 7e-143 48.74% 492 HCD1553136.1 <input checked="" type="checkbox"/> hypothetical protein CIP107524_01832 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 391 391 80% 2e-126 53.63% 523 CAB0565259.1 <input checked="" type="checkbox"/> hypothetical protein FRC0370_01709 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 389 389 96% 1e-125 47.54% 523 CAB0870047.1 <input checked="" type="checkbox"/> hypothetical protein FRC0378_01882 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 388 388 96% 3e-125 48.09% 523 CAB0871289.1 <input checked="" type="checkbox"/> hypothetical protein FRC0474_02088 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 388 388 80% 4e-125 53.19% 523 CAB0972304.1 <input checked="" type="checkbox"/> hypothetical protein D9R17_02160 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 387 387 80% 5e-125 53.19% 503 BLP10245.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 387 387 80% 6e-125 53.19% 523 WP_182001102.1 <input checked="" type="checkbox"/> hypothetical protein FRC0430_01824 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 387 387 80% 6e-125 53.19% 523 CAB0913975.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hadate] Corynebacterium hadate 386 386 100% 7e-125 46.32% 488 WP_065275383.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 386 386 96% 2e-124 47.72% 523 WP_088288569.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium plaucum] Corynebacterium plaucum 385 385 100% 2e-124 42.58% 489 WP_301925793.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 385 385 80% 5e-124 52.75% 523 WP_209917723.1 <input checked="" type="checkbox"/> hypothetical protein CIP107521_01967 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 384 384 96% 1e-123 47.54% 523 CAB0567951.1 <input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 385 385 99% 6e-123 39.30% 625 HAT6525551.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium coccinuosum] Corynebacterium coccinuosum 382 382 98% 2e-122 43.49% 554 WP_284594213.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodichtherricum] Corynebacterium pseudodichtherricum 380 380 98% 1e-121 44.50% 551 WP_284596340.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 378 378 80% 2e-121 53.41% 523 WP_342351204.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium] Corynebacterium 379 379 99% 3e-121 43.09% 551 WP_064833148.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodichtherricum] Corynebacterium pseudodichtherricum 379 379 98% 3e-121 43.67% 554 WP_284849037.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium coccinuosum] Corynebacterium coccinuosum 379 379 98% 4e-121 43.65% 554 WP_302524538.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 377 377 80% 5e-121 53.41% 523 WP_106361639.1	
	<input checked="" type="checkbox"/> hypothetical protein [Winkia] Winkia 284 284 67% 9e-86 44.25% 462 WP_004805609.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 289 358 100% 1e-85 46.40% 627 WP_201816501.1 <input checked="" type="checkbox"/> hypothetical protein [Proionimicrobium hymebohilum] Proionimicrobium hymebohilum 283 283 79% 5e-84 40.81% 556 WP_285112190.1 <input checked="" type="checkbox"/> hypothetical protein HMPREF1550_01549 [Actinomyces sp. oral taxon 877 str. F0543] Actinomyces sp. oral taxon 877 str. F0543 280 280 68% 7e-84 47.30% 486 ERH31088.1 <input checked="" type="checkbox"/> hypothetical protein GCM10023233_27360 [Brevibacterium ottidis] Brevibacterium ottidis 280 280 79% 5e-83 39.87% 539 BEF07767.1 <input checked="" type="checkbox"/> hypothetical protein [Schaalia peoriae] Schaalia peoriae 277 277 68% 2e-82 47.45% 495 WP_005867126.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium covariense] Corynebacterium covariense 278 278 64% 2e-82 43.77% 529 WP_187974193.1 <input checked="" type="checkbox"/> chape tail lace measure protein [Corynebacterium oculi] Corynebacterium oculi 278 338 55% 7e-81 63.11% 680 WP_150114274.1 <input checked="" type="checkbox"/> hypothetical protein [uncultured Leifsonia sp.] uncultured Leifsonia sp. 273 273 53% 2e-80 52.88% 525 WP_314148240.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Swenson] Arthrobacter chape Swenson 273 273 62% 4e-79 49.14% 665 ASR83993.1 <input checked="" type="checkbox"/> hypothetical protein [Gordonia westfalica] Gordonia westfalica 269 269 98% 5e-79 37.01% 527 WP_074850009.1 <input checked="" type="checkbox"/> tail length lace measure protein [Gordonia chape EricDab] Gordonia chape EricDab 266 266 72% 3e-78 41.63% 513 YP_010674657.1 <input checked="" type="checkbox"/> hypothetical protein [Cellulosimicrobium so. J38E] Cellulosimicrobium so. J38E 269 269 67% 5e-78 49.62% 636 WP_064317077.1 <input checked="" type="checkbox"/> hypothetical protein [Gulosibacter bifidus] Gulosibacter bifidus 266 266 97% 1e-77 38.53% 536 WP_066057107.1 <input checked="" type="checkbox"/> hypothetical protein [Microbacterium so. A/2022] Microbacterium so. A/2022 263 263 57% 2e-77 49.54% 477 WP_311878682.1 <input checked="" type="checkbox"/> tail length lace measure protein [Arthrobacter chape Decurso] Arthrobacter chape Decurso 268 268 62% 3e-77 49.14% 665 YP_009191306.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Jessica] Arthrobacter chape Jessica 268 268 62% 3e-77 49.14% 665 ALF00769.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Mullie] Arthrobacter chape Mullie 268 268 62% 4e-77 49.14% 665 ALY09769.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Toulouse] Arthrobacter chape Toulouse 268 268 62% 4e-77 49.14% 665 ALY10652.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Moloch] Arthrobacter chape Moloch 268 268 62% 4e-77 49.14% 665 ALY09743.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Elkhorn] Arthrobacter chape Elkhorn 265 265 62% 3e-76 49.14% 665 ASR83520.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape KylieMac] Arthrobacter chape KylieMac 265 265 62% 3e-76 49.14% 665 ASR83547.1 <input checked="" type="checkbox"/> hypothetical protein [Brevibacterium so. UMB1308A] Brevibacterium so. UMB1308A 261 261 73% 1e-75 41.49% 555 WP_284926673.1 <input checked="" type="checkbox"/> hypothetical protein [Microbacterium so. cx-55] Microbacterium so. cx-55 259 259 71% 2e-75 42.45% 524 WP_223721111.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Sachita] Arthrobacter chape Sachita 263 263 62% 3e-75 48.56% 665 QGH75234.1	
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	NON	
DECISION:	OUI	

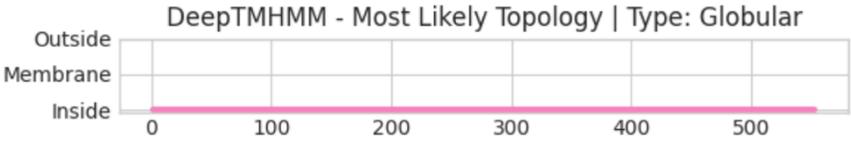
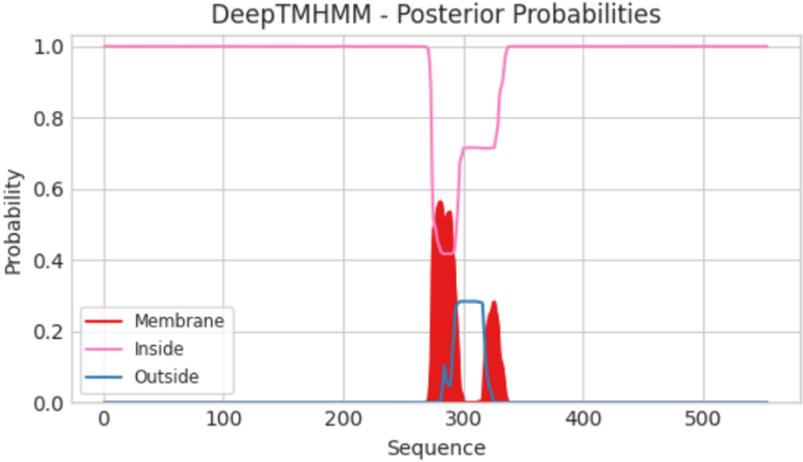
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																																																																																																																																																																																				
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :7551 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :7551</p>																																																																																																																																																																																																																																																																																																																																				
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <p>DNR Choose ORF start</p> <p>Starts : 35 ORF Start : 8724 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler7 Explore Selected : 18 ORF Stop : 9212 5' End 80,0 70,0 60,0 30 Spacing Weight Matrix Broad Document ORF Length : 489 3' End 52,9 42,6 51,6 669</p> <table border="1"> <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>2</td><td>-1,593</td><td>3,208</td><td>11</td><td>-2,593</td><td>GAAATCCITGAGGAGTCCCGCC</td><td>GTG</td><td>7551</td><td>1662</td></tr> <tr><td>3</td><td>-5,794</td><td>1,158</td><td>16</td><td>-7,015</td><td>GTTCACTGATTCOGCTAGTGAA</td><td>ATG</td><td>7764</td><td>1449</td></tr> <tr><td>4</td><td>-3,755</td><td>2,153</td><td>6</td><td>-5,056</td><td>AATGGAAACAATCCACAGGCGG</td><td>GTG</td><td>7785</td><td>1428</td></tr> <tr><td>5</td><td>-5,254</td><td>1,421</td><td>14</td><td>-6,351</td><td>TGCTGCAAGCGATTACCAAAAC</td><td>ATG</td><td>7881</td><td>1332</td></tr> <tr><td>6</td><td>-4,864</td><td>1,611</td><td>10</td><td>-5,822</td><td>TTACCAAAACATGGCGAGTGTC</td><td>ATG</td><td>7893</td><td>1320</td></tr> <tr><td>7</td><td>-3,995</td><td>2,035</td><td>13</td><td>-5,066</td><td>GAGTGTCAATGGGTTCCGAGCTT</td><td>ATG</td><td>7908</td><td>1305</td></tr> <tr><td>8</td><td>-5,738</td><td>1,185</td><td>5</td><td>-7,136</td><td>CATGGGTTCCGACGTTATGAAT</td><td>ATG</td><td>7914</td><td>1299</td></tr> <tr><td>9</td><td>-4,823</td><td>1,631</td><td>13</td><td>-5,893</td><td>CACTACATCGGATCTAATCACC</td><td>TTG</td><td>7965</td><td>1248</td></tr> <tr><td>10</td><td>-5,323</td><td>1,387</td><td>15</td><td>-6,478</td><td>CTTGGGCGGTGATCTCGCGTCC</td><td>ATG</td><td>7986</td><td>1227</td></tr> <tr><td>11</td><td>-4,632</td><td>1,725</td><td>13</td><td>-5,702</td><td>CGACCCATATCGAACGCTACGGG</td><td>GTG</td><td>8070</td><td>1143</td></tr> <tr><td>12</td><td>-5,624</td><td>1,240</td><td>9</td><td>-6,670</td><td>TATCGAACGCTACGCGGTGCC</td><td>ATG</td><td>8076</td><td>1137</td></tr> <tr><td>13</td><td>-5,037</td><td>1,527</td><td>18</td><td>-6,435</td><td>CCRAGCGGCTGCAATCGGAAA</td><td>ATG</td><td>8103</td><td>1110</td></tr> <tr><td>14</td><td>-5,266</td><td>1,415</td><td>13</td><td>-6,336</td><td>TGTCAATCGCAAAATGGCCGAA</td><td>ATG</td><td>8112</td><td>1101</td></tr> <tr><td>15</td><td>-5,529</td><td>1,287</td><td>10</td><td>-6,487</td><td>AGGCGAAGACCCGAACTGTTT</td><td>ATG</td><td>8376</td><td>837</td></tr> <tr><td>16</td><td>-4,068</td><td>1,999</td><td>9</td><td>-5,114</td><td>AGGTGTTGCTGCAGCGCGGCTT</td><td>ATG</td><td>8418</td><td>795</td></tr> <tr><td>17</td><td>-6,656</td><td>0,737</td><td>18</td><td>-8,054</td><td>CGCCGGGTTAAAATCTTTAAT</td><td>ATG</td><td>8481</td><td>732</td></tr> <tr><td>18</td><td>-7,082</td><td>0,529</td><td>10</td><td>-8,041</td><td>TGCTAACCCCGTTGGCCTGGCC</td><td>GTG</td><td>8514</td><td>699</td></tr> <tr><td>19</td><td>-6,193</td><td>0,963</td><td>15</td><td>-7,348</td><td>TATCGCCGCACTCGCAGCCGGT</td><td>TTG</td><td>8544</td><td>669</td></tr> <tr><td>20</td><td>-4,399</td><td>1,838</td><td>12</td><td>-5,445</td><td>GTTCCGTAATGGGCTCGACGTT</td><td>ATG</td><td>8595</td><td>618</td></tr> <tr><td>21</td><td>-3,898</td><td>2,082</td><td>7</td><td>-5,120</td><td>GGGTCCGTTTGTAAAGCAACAA</td><td>TTG</td><td>8619</td><td>594</td></tr> <tr><td>22</td><td>-4,068</td><td>1,999</td><td>15</td><td>-5,223</td><td>TATTCAGGCGAACGCCITGGAT</td><td>ATG</td><td>8721</td><td>492</td></tr> <tr><td>23</td><td>-3,230</td><td>2,409</td><td>7</td><td>-4,451</td><td>TGCAGGCGAACGCCITGGATG</td><td>GTG</td><td>8724</td><td>489</td></tr> <tr><td>24</td><td>-5,337</td><td>1,380</td><td>16</td><td>-6,559</td><td>TGTTGAGATTGGCTTACCCGT</td><td>GTG</td><td>8766</td><td>447</td></tr> <tr><td>25</td><td>-2,375</td><td>2,826</td><td>9</td><td>-3,421</td><td>TGTGGCGACTCTGGAACCGGA</td><td>ATG</td><td>8787</td><td>426</td></tr> <tr><td>26</td><td>-7,416</td><td>0,366</td><td>12</td><td>-8,461</td><td>GATTCGGTCCGCTATCGCTGGT</td><td>GTG</td><td>8820</td><td>393</td></tr> <tr><td>27</td><td>-3,924</td><td>2,070</td><td>7</td><td>-5,146</td><td>TGTGATCGGCTGGCTAGAAAAA</td><td>ATG</td><td>8841</td><td>372</td></tr> <tr><td>28</td><td>-4,141</td><td>1,964</td><td>16</td><td>-5,362</td><td>CGCAGCGGTTAGCGATCCCGAT</td><td>ATG</td><td>8967</td><td>246</td></tr> <tr><td>29</td><td>-5,429</td><td>1,336</td><td>10</td><td>-6,388</td><td>GGGTAGCGATCCCGATATGACT</td><td>ATG</td><td>8973</td><td>240</td></tr> <tr><td>30</td><td>-5,029</td><td>1,531</td><td>16</td><td>-6,250</td><td>GATCCTGAAGTTTTCTTCCACA</td><td>TTG</td><td>9015</td><td>198</td></tr> <tr><td>31</td><td>-6,676</td><td>0,727</td><td>12</td><td>-7,721</td><td>GTTTTCTCCACATTGCCCGAT</td><td>ATG</td><td>9024</td><td>189</td></tr> <tr><td>32</td><td>-5,755</td><td>1,176</td><td>10</td><td>-6,713</td><td>TTCCACATTGCCCGATATGACT</td><td>ATG</td><td>9030</td><td>183</td></tr> <tr><td>33</td><td>-4,662</td><td>1,710</td><td>16</td><td>-5,884</td><td>TATGCAGGCTGCAGCCCGGAC</td><td>GTG</td><td>9051</td><td>162</td></tr> <tr><td>34</td><td>-5,526</td><td>1,288</td><td>7</td><td>-6,748</td><td>CCCGACGTTGACCGCGCAGCGC</td><td>GTG</td><td>9066</td><td>147</td></tr> <tr><td>35</td><td>-3,119</td><td>2,463</td><td>11</td><td>-4,119</td><td>AGGGGCGGAAAAGATCATTAAA</td><td>TTG</td><td>9177</td><td>36</td></tr> </tbody> </table> <p>Position of internal starts</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	2	-1,593	3,208	11	-2,593	GAAATCCITGAGGAGTCCCGCC	GTG	7551	1662	3	-5,794	1,158	16	-7,015	GTTCACTGATTCOGCTAGTGAA	ATG	7764	1449	4	-3,755	2,153	6	-5,056	AATGGAAACAATCCACAGGCGG	GTG	7785	1428	5	-5,254	1,421	14	-6,351	TGCTGCAAGCGATTACCAAAAC	ATG	7881	1332	6	-4,864	1,611	10	-5,822	TTACCAAAACATGGCGAGTGTC	ATG	7893	1320	7	-3,995	2,035	13	-5,066	GAGTGTCAATGGGTTCCGAGCTT	ATG	7908	1305	8	-5,738	1,185	5	-7,136	CATGGGTTCCGACGTTATGAAT	ATG	7914	1299	9	-4,823	1,631	13	-5,893	CACTACATCGGATCTAATCACC	TTG	7965	1248	10	-5,323	1,387	15	-6,478	CTTGGGCGGTGATCTCGCGTCC	ATG	7986	1227	11	-4,632	1,725	13	-5,702	CGACCCATATCGAACGCTACGGG	GTG	8070	1143	12	-5,624	1,240	9	-6,670	TATCGAACGCTACGCGGTGCC	ATG	8076	1137	13	-5,037	1,527	18	-6,435	CCRAGCGGCTGCAATCGGAAA	ATG	8103	1110	14	-5,266	1,415	13	-6,336	TGTCAATCGCAAAATGGCCGAA	ATG	8112	1101	15	-5,529	1,287	10	-6,487	AGGCGAAGACCCGAACTGTTT	ATG	8376	837	16	-4,068	1,999	9	-5,114	AGGTGTTGCTGCAGCGCGGCTT	ATG	8418	795	17	-6,656	0,737	18	-8,054	CGCCGGGTTAAAATCTTTAAT	ATG	8481	732	18	-7,082	0,529	10	-8,041	TGCTAACCCCGTTGGCCTGGCC	GTG	8514	699	19	-6,193	0,963	15	-7,348	TATCGCCGCACTCGCAGCCGGT	TTG	8544	669	20	-4,399	1,838	12	-5,445	GTTCCGTAATGGGCTCGACGTT	ATG	8595	618	21	-3,898	2,082	7	-5,120	GGGTCCGTTTGTAAAGCAACAA	TTG	8619	594	22	-4,068	1,999	15	-5,223	TATTCAGGCGAACGCCITGGAT	ATG	8721	492	23	-3,230	2,409	7	-4,451	TGCAGGCGAACGCCITGGATG	GTG	8724	489	24	-5,337	1,380	16	-6,559	TGTTGAGATTGGCTTACCCGT	GTG	8766	447	25	-2,375	2,826	9	-3,421	TGTGGCGACTCTGGAACCGGA	ATG	8787	426	26	-7,416	0,366	12	-8,461	GATTCGGTCCGCTATCGCTGGT	GTG	8820	393	27	-3,924	2,070	7	-5,146	TGTGATCGGCTGGCTAGAAAAA	ATG	8841	372	28	-4,141	1,964	16	-5,362	CGCAGCGGTTAGCGATCCCGAT	ATG	8967	246	29	-5,429	1,336	10	-6,388	GGGTAGCGATCCCGATATGACT	ATG	8973	240	30	-5,029	1,531	16	-6,250	GATCCTGAAGTTTTCTTCCACA	TTG	9015	198	31	-6,676	0,727	12	-7,721	GTTTTCTCCACATTGCCCGAT	ATG	9024	189	32	-5,755	1,176	10	-6,713	TTCCACATTGCCCGATATGACT	ATG	9030	183	33	-4,662	1,710	16	-5,884	TATGCAGGCTGCAGCCCGGAC	GTG	9051	162	34	-5,526	1,288	7	-6,748	CCCGACGTTGACCGCGCAGCGC	GTG	9066	147	35	-3,119	2,463	11	-4,119	AGGGGCGGAAAAGATCATTAAA	TTG	9177	36
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																																																																																																																																																																																																																																																																																																																													
2	-1,593	3,208	11	-2,593	GAAATCCITGAGGAGTCCCGCC	GTG	7551	1662																																																																																																																																																																																																																																																																																																																													
3	-5,794	1,158	16	-7,015	GTTCACTGATTCOGCTAGTGAA	ATG	7764	1449																																																																																																																																																																																																																																																																																																																													
4	-3,755	2,153	6	-5,056	AATGGAAACAATCCACAGGCGG	GTG	7785	1428																																																																																																																																																																																																																																																																																																																													
5	-5,254	1,421	14	-6,351	TGCTGCAAGCGATTACCAAAAC	ATG	7881	1332																																																																																																																																																																																																																																																																																																																													
6	-4,864	1,611	10	-5,822	TTACCAAAACATGGCGAGTGTC	ATG	7893	1320																																																																																																																																																																																																																																																																																																																													
7	-3,995	2,035	13	-5,066	GAGTGTCAATGGGTTCCGAGCTT	ATG	7908	1305																																																																																																																																																																																																																																																																																																																													
8	-5,738	1,185	5	-7,136	CATGGGTTCCGACGTTATGAAT	ATG	7914	1299																																																																																																																																																																																																																																																																																																																													
9	-4,823	1,631	13	-5,893	CACTACATCGGATCTAATCACC	TTG	7965	1248																																																																																																																																																																																																																																																																																																																													
10	-5,323	1,387	15	-6,478	CTTGGGCGGTGATCTCGCGTCC	ATG	7986	1227																																																																																																																																																																																																																																																																																																																													
11	-4,632	1,725	13	-5,702	CGACCCATATCGAACGCTACGGG	GTG	8070	1143																																																																																																																																																																																																																																																																																																																													
12	-5,624	1,240	9	-6,670	TATCGAACGCTACGCGGTGCC	ATG	8076	1137																																																																																																																																																																																																																																																																																																																													
13	-5,037	1,527	18	-6,435	CCRAGCGGCTGCAATCGGAAA	ATG	8103	1110																																																																																																																																																																																																																																																																																																																													
14	-5,266	1,415	13	-6,336	TGTCAATCGCAAAATGGCCGAA	ATG	8112	1101																																																																																																																																																																																																																																																																																																																													
15	-5,529	1,287	10	-6,487	AGGCGAAGACCCGAACTGTTT	ATG	8376	837																																																																																																																																																																																																																																																																																																																													
16	-4,068	1,999	9	-5,114	AGGTGTTGCTGCAGCGCGGCTT	ATG	8418	795																																																																																																																																																																																																																																																																																																																													
17	-6,656	0,737	18	-8,054	CGCCGGGTTAAAATCTTTAAT	ATG	8481	732																																																																																																																																																																																																																																																																																																																													
18	-7,082	0,529	10	-8,041	TGCTAACCCCGTTGGCCTGGCC	GTG	8514	699																																																																																																																																																																																																																																																																																																																													
19	-6,193	0,963	15	-7,348	TATCGCCGCACTCGCAGCCGGT	TTG	8544	669																																																																																																																																																																																																																																																																																																																													
20	-4,399	1,838	12	-5,445	GTTCCGTAATGGGCTCGACGTT	ATG	8595	618																																																																																																																																																																																																																																																																																																																													
21	-3,898	2,082	7	-5,120	GGGTCCGTTTGTAAAGCAACAA	TTG	8619	594																																																																																																																																																																																																																																																																																																																													
22	-4,068	1,999	15	-5,223	TATTCAGGCGAACGCCITGGAT	ATG	8721	492																																																																																																																																																																																																																																																																																																																													
23	-3,230	2,409	7	-4,451	TGCAGGCGAACGCCITGGATG	GTG	8724	489																																																																																																																																																																																																																																																																																																																													
24	-5,337	1,380	16	-6,559	TGTTGAGATTGGCTTACCCGT	GTG	8766	447																																																																																																																																																																																																																																																																																																																													
25	-2,375	2,826	9	-3,421	TGTGGCGACTCTGGAACCGGA	ATG	8787	426																																																																																																																																																																																																																																																																																																																													
26	-7,416	0,366	12	-8,461	GATTCGGTCCGCTATCGCTGGT	GTG	8820	393																																																																																																																																																																																																																																																																																																																													
27	-3,924	2,070	7	-5,146	TGTGATCGGCTGGCTAGAAAAA	ATG	8841	372																																																																																																																																																																																																																																																																																																																													
28	-4,141	1,964	16	-5,362	CGCAGCGGTTAGCGATCCCGAT	ATG	8967	246																																																																																																																																																																																																																																																																																																																													
29	-5,429	1,336	10	-6,388	GGGTAGCGATCCCGATATGACT	ATG	8973	240																																																																																																																																																																																																																																																																																																																													
30	-5,029	1,531	16	-6,250	GATCCTGAAGTTTTCTTCCACA	TTG	9015	198																																																																																																																																																																																																																																																																																																																													
31	-6,676	0,727	12	-7,721	GTTTTCTCCACATTGCCCGAT	ATG	9024	189																																																																																																																																																																																																																																																																																																																													
32	-5,755	1,176	10	-6,713	TTCCACATTGCCCGATATGACT	ATG	9030	183																																																																																																																																																																																																																																																																																																																													
33	-4,662	1,710	16	-5,884	TATGCAGGCTGCAGCCCGGAC	GTG	9051	162																																																																																																																																																																																																																																																																																																																													
34	-5,526	1,288	7	-6,748	CCCGACGTTGACCGCGCAGCGC	GTG	9066	147																																																																																																																																																																																																																																																																																																																													
35	-3,119	2,463	11	-4,119	AGGGGCGGAAAAGATCATTAAA	TTG	9177	36																																																																																																																																																																																																																																																																																																																													
<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>	<p><i>Plus long orf mais pas avec un ATG</i></p> <p><i>Meilleur score</i></p>																																																																																																																																																																																																																																																																																																																																				
<p>Est-ce que le start est conservé chez les</p>	<p>—</p>																																																																																																																																																																																																																																																																																																																																				

homologues voir Starterator ?	
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	
DECISION:	7551

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 : (q#: s#) : [alignment] ; e-value :</p>  <p>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th>Sequence</th> <th>Score (bits)</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>Cyranop_Draft_13, function unknown, 553</td> <td>1035</td> <td>0.0</td> </tr> <tr> <td>SocialPhiler_11, tape measure protein, 607</td> <td>322</td> <td>9e-88</td> </tr> <tr> <td>Kela_11, tape measure protein, 607</td> <td>322</td> <td>9e-88</td> </tr> <tr> <td>Ariehaps_11, tape measure protein, 607</td> <td>322</td> <td>9e-88</td> </tr> <tr> <td>StewieDriff_13, tape measure protein, 665</td> <td>322</td> <td>1e-87</td> </tr> <tr> <td>Lore_13, tape measure protein, 665</td> <td>322</td> <td>1e-87</td> </tr> <tr> <td>Isalia_15, tape measure protein, 665</td> <td>322</td> <td>1e-87</td> </tr> </tbody> </table>	Sequence	Score (bits)	E-value	Cyranop_Draft_13, function unknown, 553	1035	0.0	SocialPhiler_11, tape measure protein, 607	322	9e-88	Kela_11, tape measure protein, 607	322	9e-88	Ariehaps_11, tape measure protein, 607	322	9e-88	StewieDriff_13, tape measure protein, 665	322	1e-87	Lore_13, tape measure protein, 665	322	1e-87	Isalia_15, tape measure protein, 665	322	1e-87
Sequence	Score (bits)	E-value																							
Cyranop_Draft_13, function unknown, 553	1035	0.0																							
SocialPhiler_11, tape measure protein, 607	322	9e-88																							
Kela_11, tape measure protein, 607	322	9e-88																							
Ariehaps_11, tape measure protein, 607	322	9e-88																							
StewieDriff_13, tape measure protein, 665	322	1e-87																							
Lore_13, tape measure protein, 665	322	1e-87																							
Isalia_15, tape measure protein, 665	322	1e-87																							

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>NON ?</p>  
<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>tape mesure protein</p>

Student Gene Annotation Worksheet

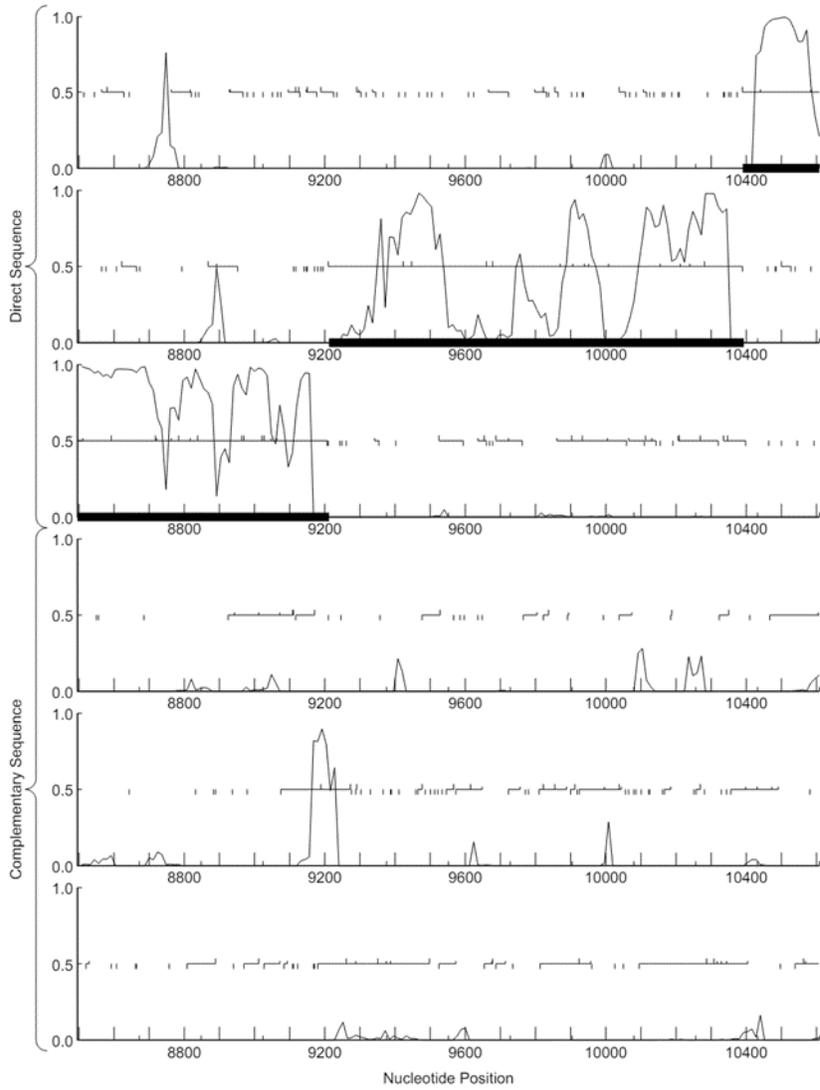
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	13
Coordonnées du Stop	9212
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 2

Coordonnées du Start retenu	7551
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>inoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGGCCAGCCC), Order 2, Window 96, Step 12, 5/9</p> <p>The figure displays a GeneMark hmm prediction plot for a DNA sequence. The x-axis represents the Nucleotide Position, ranging from 6400 to 8400. The y-axis represents the probability of a region being a coding sequence, ranging from 0.0 to 1.0. The plot is divided into two main sections: 'Direct Sequence' (top) and 'Complementary Sequence' (bottom). Each section contains two tracks: a line graph showing the predicted probability and a bar chart showing the predicted exons. In the Direct Sequence section, there are several prominent peaks, with the highest peak reaching a probability of 1.0 around position 7200. In the Complementary Sequence section, there are also several peaks, with the highest peak reaching a probability of approximately 0.8 around position 7200. The bar charts in both sections show several exons of varying lengths, with the longest exon in the Direct Sequence section spanning from approximately 6800 to 7200. The overall prediction indicates a strong potential for a coding sequence in this region.</p>

GeneMark_hmm prediction
 anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGGCCAGCCC), Order 2, Window 96, Step 12, 6/9



Distribution of 102 Blast Hits on the Query Sequence

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

Color Key for Alignment Scores



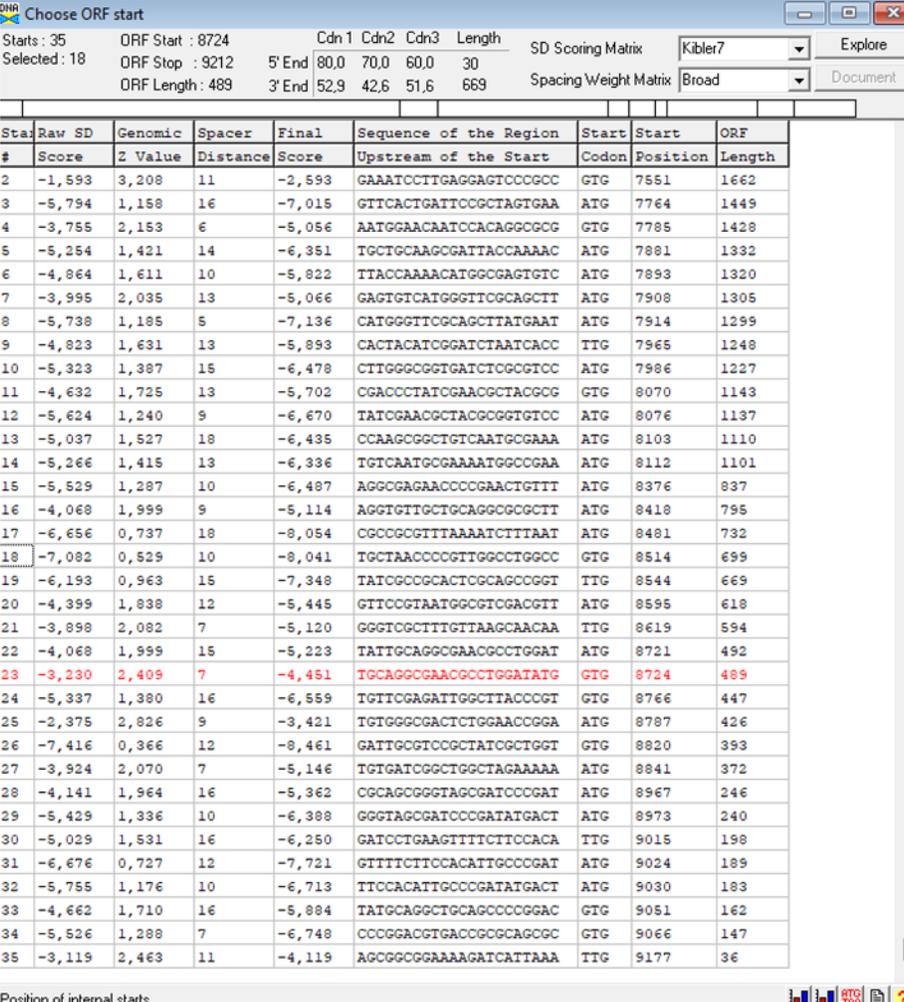
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?

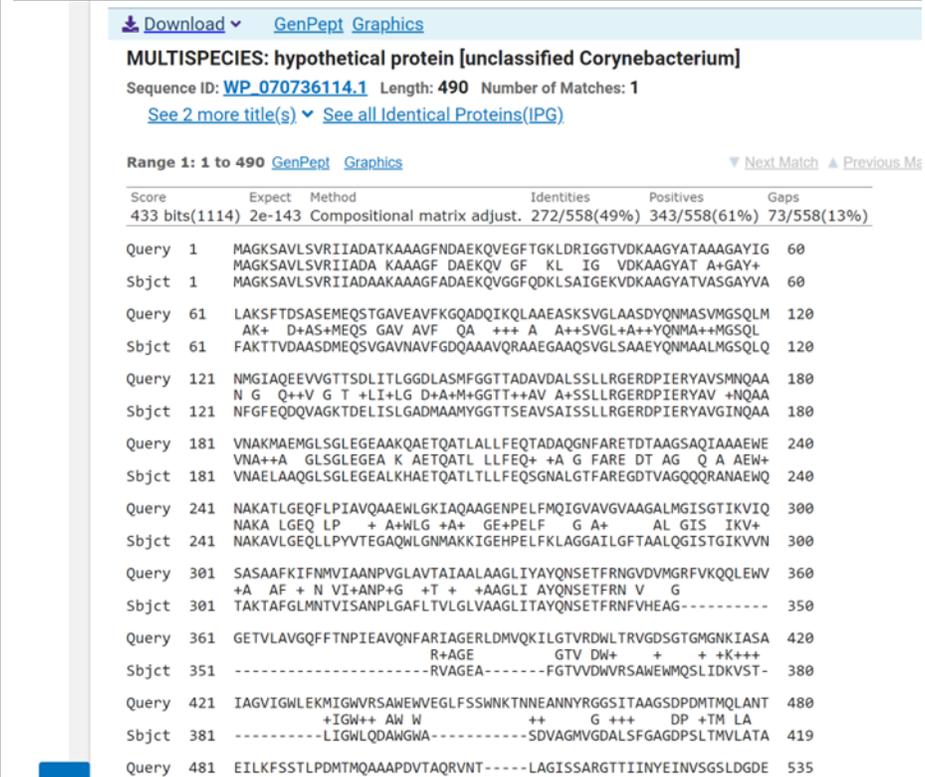
Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyranoPS_Draft_13, function unknown, 553	1059	0.0
SerialPhiller_11, tape measure protein, 607	323	9e-88
Kels_11, tape measure protein, 607	323	9e-88
Arielagos_11, tape measure protein, 607	323	9e-88
StewieGriff_13, tape measure protein, 665	322	1e-87
Lore_13, tape measure protein, 665	322	1e-87
Laila_15, tape measure protein, 665	322	1e-87
KylieMac 14, tape measure protein, 665	322	1e-87

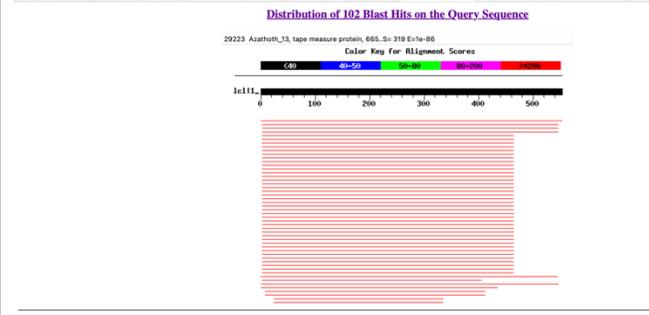
	<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium] Corynebacterium 444 444 100% 1e-147 49.10% 492 WP_070420702.1 <input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 436 436 100% 3e-144 48.92% 490 HGT5225727.1 <input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium] unclassified Corynebacterium 433 433 100% 2e-143 48.75% 490 WP_070736114.1 <input checked="" type="checkbox"/> putative chape protein [Corynebacterium striatum] Corynebacterium striatum 433 433 100% 3e-143 48.75% 490 COD13985.1 <input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 432 432 100% 7e-143 48.74% 492 HCD1553136.1 <input checked="" type="checkbox"/> hypothetical protein CIP107524_01832 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 391 391 80% 2e-126 53.63% 523 CAB0565259.1 <input checked="" type="checkbox"/> hypothetical protein FRC0370_01709 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 389 389 96% 1e-125 47.54% 523 CAB0870047.1 <input checked="" type="checkbox"/> hypothetical protein FRC0378_01882 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 388 388 96% 3e-125 48.09% 523 CAB0871289.1 <input checked="" type="checkbox"/> hypothetical protein FRC0474_02088 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 388 388 80% 4e-125 53.19% 523 CAB0972304.1 <input checked="" type="checkbox"/> hypothetical protein D9R17_02160 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 387 387 80% 5e-125 53.19% 503 BLP10245.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 387 387 80% 6e-125 53.19% 523 WP_182001102.1 <input checked="" type="checkbox"/> hypothetical protein FRC0430_01824 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 387 387 80% 6e-125 53.19% 523 CAB0913975.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hadate] Corynebacterium hadate 386 386 100% 7e-125 46.32% 488 WP_065275383.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 386 386 96% 2e-124 47.72% 523 WP_088288569.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium plaucum] Corynebacterium plaucum 385 385 100% 2e-124 42.58% 489 WP_301925793.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 385 385 80% 5e-124 52.75% 523 WP_209917723.1 <input checked="" type="checkbox"/> hypothetical protein CIP107521_01967 [Corynebacterium diobtheriae] Corynebacterium diobtheriae 384 384 96% 1e-123 47.54% 523 CAB0567951.1 <input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 385 385 99% 6e-123 39.35% 625 HAT652555.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium coccinuosum] Corynebacterium coccinuosum 382 382 98% 2e-122 43.49% 554 WP_284594213.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodichtherricum] Corynebacterium pseudodichtherricum 380 380 98% 1e-121 44.50% 551 WP_284596340.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 378 378 80% 2e-121 53.41% 523 WP_342351204.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium] Corynebacterium 379 379 99% 3e-121 43.09% 551 WP_064833148.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium pseudodichtherricum] Corynebacterium pseudodichtherricum 379 379 98% 3e-121 43.67% 554 WP_284849037.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium coccinuosum] Corynebacterium coccinuosum 379 379 98% 4e-121 43.65% 554 WP_302524538.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium diobtheriae] Corynebacterium diobtheriae 377 377 80% 5e-121 53.41% 523 WP_106361639.1	
	<input checked="" type="checkbox"/> hypothetical protein [Winkia] Winkia 284 284 67% 9e-86 44.25% 462 WP_004805609.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 289 358 100% 1e-85 46.40% 627 WP_201816501.1 <input checked="" type="checkbox"/> hypothetical protein [Proionimicrobium hymebohilum] Proionimicrobium hymebohilum 283 283 79% 5e-84 40.81% 556 WP_285112190.1 <input checked="" type="checkbox"/> hypothetical protein HMPREF1550_01549 [Actinomyces sp. oral taxon 877 str. F0543] Actinomyces sp. oral taxon 877 str. F0543 280 280 68% 7e-84 47.30% 486 ERH31088.1 <input checked="" type="checkbox"/> hypothetical protein GCM10023233_27360 [Brevibacterium otidids] Brevibacterium otidids 280 280 79% 5e-83 39.87% 539 BEF07767.1 <input checked="" type="checkbox"/> hypothetical protein [Schaalia peoriae] Schaalia peoriae 277 277 68% 2e-82 47.45% 495 WP_005867126.1 <input checked="" type="checkbox"/> hypothetical protein [Corynebacterium covariense] Corynebacterium covariense 278 278 64% 2e-82 43.77% 529 WP_187974193.1 <input checked="" type="checkbox"/> chape tail lace measure protein [Corynebacterium oculi] Corynebacterium oculi 278 338 55% 7e-81 63.11% 680 WP_150114274.1 <input checked="" type="checkbox"/> hypothetical protein [uncultured Leifsonia sp.] uncultured Leifsonia sp. 273 273 53% 2e-80 52.88% 525 WP_314148240.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Swenson] Arthrobacter chape Swenson 273 273 62% 4e-79 49.14% 665 ASR83993.1 <input checked="" type="checkbox"/> hypothetical protein [Gordonia westfalica] Gordonia westfalica 269 269 98% 5e-79 37.01% 527 WP_074850009.1 <input checked="" type="checkbox"/> tail length lace measure protein [Gordonia chape EricDab] Gordonia chape EricDab 266 266 72% 3e-78 41.63% 513 YP_010674657.1 <input checked="" type="checkbox"/> hypothetical protein [Cellulosimicrobium so. J38E] Cellulosimicrobium so. J38E 269 269 67% 5e-78 49.62% 636 WP_064317077.1 <input checked="" type="checkbox"/> hypothetical protein [Gutsoibacter bifidus] Gutsoibacter bifidus 266 266 97% 1e-77 38.53% 536 WP_066057107.1 <input checked="" type="checkbox"/> hypothetical protein [Microbacterium so. A/2022] Microbacterium so. A/2022 263 263 57% 2e-77 49.54% 477 WP_311878682.1 <input checked="" type="checkbox"/> tail length lace measure protein [Arthrobacter chape Decurso] Arthrobacter chape Decurso 268 268 62% 3e-77 49.14% 665 YP_009191306.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Jessica] Arthrobacter chape Jessica 268 268 62% 3e-77 49.14% 665 ALF00769.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Mullie] Arthrobacter chape Mullie 268 268 62% 4e-77 49.14% 665 ALY09769.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Toulouse] Arthrobacter chape Toulouse 268 268 62% 4e-77 49.14% 665 ALY10652.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Moloch] Arthrobacter chape Moloch 268 268 62% 4e-77 49.14% 665 ALY09743.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Elkhorn] Arthrobacter chape Elkhorn 265 265 62% 3e-76 49.14% 665 ASR83520.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape KylieMac] Arthrobacter chape KylieMac 265 265 62% 3e-76 49.14% 665 ASR83547.1 <input checked="" type="checkbox"/> hypothetical protein [Brevibacterium so. UMB1308A] Brevibacterium so. UMB1308A 261 261 73% 1e-75 41.49% 555 WP_284926673.1 <input checked="" type="checkbox"/> hypothetical protein [Microbacterium so. cx-55] Microbacterium so. cx-55 259 259 71% 2e-75 42.45% 524 WP_223721111.1 <input checked="" type="checkbox"/> lace measure protein [Arthrobacter chape Sachita] Arthrobacter chape Sachita 263 263 62% 3e-75 48.56% 665 QGH75234.1	
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																																																																																																																																																																																																																																																																																																																																				
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :7551 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :7551</p>																																																																																																																																																																																																																																																																																																																																				
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <p>DNR Choose ORF start</p> <p>Starts : 35 ORF Start : 8724 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler7 Explore Selected : 18 ORF Stop : 9212 5' End 80,0 70,0 60,0 30 Spacing Weight Matrix Broad Document ORF Length : 489 3' End 52,9 42,6 51,6 669</p> <table border="1"> <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>2</td><td>-1,593</td><td>3,208</td><td>11</td><td>-2,593</td><td>GAAATCCITGAGGAGTCCCGCC</td><td>GTG</td><td>7551</td><td>1662</td></tr> <tr><td>3</td><td>-5,794</td><td>1,158</td><td>16</td><td>-7,015</td><td>GTTCACTGATTCOGCTAGTGAA</td><td>ATG</td><td>7764</td><td>1449</td></tr> <tr><td>4</td><td>-3,755</td><td>2,153</td><td>6</td><td>-5,056</td><td>AATGGAAACAATCCACAGGGCG</td><td>GTG</td><td>7785</td><td>1428</td></tr> <tr><td>5</td><td>-5,254</td><td>1,421</td><td>14</td><td>-6,351</td><td>TGCTGCAAGCGATTACCAAAAC</td><td>ATG</td><td>7881</td><td>1332</td></tr> <tr><td>6</td><td>-4,864</td><td>1,611</td><td>10</td><td>-5,822</td><td>TTACCAAAACATGGCGAGTGTC</td><td>ATG</td><td>7893</td><td>1320</td></tr> <tr><td>7</td><td>-3,995</td><td>2,035</td><td>13</td><td>-5,066</td><td>GAGTGTATGGGTTCCGAGCTT</td><td>ATG</td><td>7908</td><td>1305</td></tr> <tr><td>8</td><td>-5,738</td><td>1,185</td><td>5</td><td>-7,136</td><td>CATGGGTTCCGACGTTATGAAT</td><td>ATG</td><td>7914</td><td>1299</td></tr> <tr><td>9</td><td>-4,823</td><td>1,631</td><td>13</td><td>-5,893</td><td>CACTACATCGGATCTAATCACC</td><td>TTG</td><td>7965</td><td>1248</td></tr> <tr><td>10</td><td>-5,323</td><td>1,387</td><td>15</td><td>-6,478</td><td>CTTGGGCGGTGATCTCGCGTCC</td><td>ATG</td><td>7986</td><td>1227</td></tr> <tr><td>11</td><td>-4,632</td><td>1,725</td><td>13</td><td>-5,702</td><td>CGACCCATATCGAACGCTACGGG</td><td>GTG</td><td>8070</td><td>1143</td></tr> <tr><td>12</td><td>-5,624</td><td>1,240</td><td>9</td><td>-6,670</td><td>TATCGAACGCTACGCGGTGCC</td><td>ATG</td><td>8076</td><td>1137</td></tr> <tr><td>13</td><td>-5,037</td><td>1,527</td><td>18</td><td>-6,435</td><td>CCRAGCGGCTGCAATCGGAAA</td><td>ATG</td><td>8103</td><td>1110</td></tr> <tr><td>14</td><td>-5,266</td><td>1,415</td><td>13</td><td>-6,336</td><td>TGTCATATCGGAAATGGCCGAA</td><td>ATG</td><td>8112</td><td>1101</td></tr> <tr><td>15</td><td>-5,529</td><td>1,287</td><td>10</td><td>-6,487</td><td>AGGCGAAGACCCGAACTGTTT</td><td>ATG</td><td>8376</td><td>837</td></tr> <tr><td>16</td><td>-4,068</td><td>1,999</td><td>9</td><td>-5,114</td><td>AGGTGTTGCTGCAGCGCGGCTT</td><td>ATG</td><td>8418</td><td>795</td></tr> <tr><td>17</td><td>-6,656</td><td>0,737</td><td>18</td><td>-8,054</td><td>CGCCGGGTTAAAATCTTTAAT</td><td>ATG</td><td>8481</td><td>732</td></tr> <tr><td>18</td><td>-7,082</td><td>0,529</td><td>10</td><td>-8,041</td><td>TGCTAACCCCGTTGGCCTGGCC</td><td>GTG</td><td>8514</td><td>699</td></tr> <tr><td>19</td><td>-6,193</td><td>0,963</td><td>15</td><td>-7,348</td><td>TATCGCCGCACTCGCAGCCGGT</td><td>TTG</td><td>8544</td><td>669</td></tr> <tr><td>20</td><td>-4,399</td><td>1,838</td><td>12</td><td>-5,445</td><td>GTTCCGTAATGGGCTCGACGTT</td><td>ATG</td><td>8595</td><td>618</td></tr> <tr><td>21</td><td>-3,898</td><td>2,082</td><td>7</td><td>-5,120</td><td>GGGTCCGTTTGTAAAGCAACAA</td><td>TTG</td><td>8619</td><td>594</td></tr> <tr><td>22</td><td>-4,068</td><td>1,999</td><td>15</td><td>-5,223</td><td>TATTCAGGCGAAGCCTGGAT</td><td>ATG</td><td>8721</td><td>492</td></tr> <tr><td>23</td><td>-3,230</td><td>2,409</td><td>7</td><td>-4,451</td><td>TGCAGGCGAAGCCTGGATATG</td><td>GTG</td><td>8724</td><td>489</td></tr> <tr><td>24</td><td>-5,337</td><td>1,380</td><td>16</td><td>-6,559</td><td>TGTTGAGATTGGCTTACCCGT</td><td>GTG</td><td>8766</td><td>447</td></tr> <tr><td>25</td><td>-2,375</td><td>2,826</td><td>9</td><td>-3,421</td><td>TGTGGCGACTCTGGAACCGGA</td><td>ATG</td><td>8787</td><td>426</td></tr> <tr><td>26</td><td>-7,416</td><td>0,366</td><td>12</td><td>-8,461</td><td>GATTCGGTCCGCTATCGCTGGT</td><td>GTG</td><td>8820</td><td>393</td></tr> <tr><td>27</td><td>-3,924</td><td>2,070</td><td>7</td><td>-5,146</td><td>TGTGATCGGCTGGCTAGAAAAA</td><td>ATG</td><td>8841</td><td>372</td></tr> <tr><td>28</td><td>-4,141</td><td>1,964</td><td>16</td><td>-5,362</td><td>CGCAGCGGTTAGCGATCCCGAT</td><td>ATG</td><td>8967</td><td>246</td></tr> <tr><td>29</td><td>-5,429</td><td>1,336</td><td>10</td><td>-6,388</td><td>GGGTAGCGATCCCGATATGACT</td><td>ATG</td><td>8973</td><td>240</td></tr> <tr><td>30</td><td>-5,029</td><td>1,531</td><td>16</td><td>-6,250</td><td>GATCCTGAAGTTTTCTTCCACA</td><td>TTG</td><td>9015</td><td>198</td></tr> <tr><td>31</td><td>-6,676</td><td>0,727</td><td>12</td><td>-7,721</td><td>GTTTTCTCCACATTGCCCGAT</td><td>ATG</td><td>9024</td><td>189</td></tr> <tr><td>32</td><td>-5,755</td><td>1,176</td><td>10</td><td>-6,713</td><td>TTCCACATTGCCCGATATGACT</td><td>ATG</td><td>9030</td><td>183</td></tr> <tr><td>33</td><td>-4,662</td><td>1,710</td><td>16</td><td>-5,884</td><td>TATGCAGGCTGCAGCCCGGAC</td><td>GTG</td><td>9051</td><td>162</td></tr> <tr><td>34</td><td>-5,526</td><td>1,288</td><td>7</td><td>-6,748</td><td>CCCGACGTTGACCGCGCAGCGC</td><td>GTG</td><td>9066</td><td>147</td></tr> <tr><td>35</td><td>-3,119</td><td>2,463</td><td>11</td><td>-4,119</td><td>AGGGGCGGAAAAGATCATTTAA</td><td>TTG</td><td>9177</td><td>36</td></tr> </tbody> </table> <p>Position of internal starts</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	2	-1,593	3,208	11	-2,593	GAAATCCITGAGGAGTCCCGCC	GTG	7551	1662	3	-5,794	1,158	16	-7,015	GTTCACTGATTCOGCTAGTGAA	ATG	7764	1449	4	-3,755	2,153	6	-5,056	AATGGAAACAATCCACAGGGCG	GTG	7785	1428	5	-5,254	1,421	14	-6,351	TGCTGCAAGCGATTACCAAAAC	ATG	7881	1332	6	-4,864	1,611	10	-5,822	TTACCAAAACATGGCGAGTGTC	ATG	7893	1320	7	-3,995	2,035	13	-5,066	GAGTGTATGGGTTCCGAGCTT	ATG	7908	1305	8	-5,738	1,185	5	-7,136	CATGGGTTCCGACGTTATGAAT	ATG	7914	1299	9	-4,823	1,631	13	-5,893	CACTACATCGGATCTAATCACC	TTG	7965	1248	10	-5,323	1,387	15	-6,478	CTTGGGCGGTGATCTCGCGTCC	ATG	7986	1227	11	-4,632	1,725	13	-5,702	CGACCCATATCGAACGCTACGGG	GTG	8070	1143	12	-5,624	1,240	9	-6,670	TATCGAACGCTACGCGGTGCC	ATG	8076	1137	13	-5,037	1,527	18	-6,435	CCRAGCGGCTGCAATCGGAAA	ATG	8103	1110	14	-5,266	1,415	13	-6,336	TGTCATATCGGAAATGGCCGAA	ATG	8112	1101	15	-5,529	1,287	10	-6,487	AGGCGAAGACCCGAACTGTTT	ATG	8376	837	16	-4,068	1,999	9	-5,114	AGGTGTTGCTGCAGCGCGGCTT	ATG	8418	795	17	-6,656	0,737	18	-8,054	CGCCGGGTTAAAATCTTTAAT	ATG	8481	732	18	-7,082	0,529	10	-8,041	TGCTAACCCCGTTGGCCTGGCC	GTG	8514	699	19	-6,193	0,963	15	-7,348	TATCGCCGCACTCGCAGCCGGT	TTG	8544	669	20	-4,399	1,838	12	-5,445	GTTCCGTAATGGGCTCGACGTT	ATG	8595	618	21	-3,898	2,082	7	-5,120	GGGTCCGTTTGTAAAGCAACAA	TTG	8619	594	22	-4,068	1,999	15	-5,223	TATTCAGGCGAAGCCTGGAT	ATG	8721	492	23	-3,230	2,409	7	-4,451	TGCAGGCGAAGCCTGGATATG	GTG	8724	489	24	-5,337	1,380	16	-6,559	TGTTGAGATTGGCTTACCCGT	GTG	8766	447	25	-2,375	2,826	9	-3,421	TGTGGCGACTCTGGAACCGGA	ATG	8787	426	26	-7,416	0,366	12	-8,461	GATTCGGTCCGCTATCGCTGGT	GTG	8820	393	27	-3,924	2,070	7	-5,146	TGTGATCGGCTGGCTAGAAAAA	ATG	8841	372	28	-4,141	1,964	16	-5,362	CGCAGCGGTTAGCGATCCCGAT	ATG	8967	246	29	-5,429	1,336	10	-6,388	GGGTAGCGATCCCGATATGACT	ATG	8973	240	30	-5,029	1,531	16	-6,250	GATCCTGAAGTTTTCTTCCACA	TTG	9015	198	31	-6,676	0,727	12	-7,721	GTTTTCTCCACATTGCCCGAT	ATG	9024	189	32	-5,755	1,176	10	-6,713	TTCCACATTGCCCGATATGACT	ATG	9030	183	33	-4,662	1,710	16	-5,884	TATGCAGGCTGCAGCCCGGAC	GTG	9051	162	34	-5,526	1,288	7	-6,748	CCCGACGTTGACCGCGCAGCGC	GTG	9066	147	35	-3,119	2,463	11	-4,119	AGGGGCGGAAAAGATCATTTAA	TTG	9177	36
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																																																																																																																																																																																																																																																																																																																													
2	-1,593	3,208	11	-2,593	GAAATCCITGAGGAGTCCCGCC	GTG	7551	1662																																																																																																																																																																																																																																																																																																																													
3	-5,794	1,158	16	-7,015	GTTCACTGATTCOGCTAGTGAA	ATG	7764	1449																																																																																																																																																																																																																																																																																																																													
4	-3,755	2,153	6	-5,056	AATGGAAACAATCCACAGGGCG	GTG	7785	1428																																																																																																																																																																																																																																																																																																																													
5	-5,254	1,421	14	-6,351	TGCTGCAAGCGATTACCAAAAC	ATG	7881	1332																																																																																																																																																																																																																																																																																																																													
6	-4,864	1,611	10	-5,822	TTACCAAAACATGGCGAGTGTC	ATG	7893	1320																																																																																																																																																																																																																																																																																																																													
7	-3,995	2,035	13	-5,066	GAGTGTATGGGTTCCGAGCTT	ATG	7908	1305																																																																																																																																																																																																																																																																																																																													
8	-5,738	1,185	5	-7,136	CATGGGTTCCGACGTTATGAAT	ATG	7914	1299																																																																																																																																																																																																																																																																																																																													
9	-4,823	1,631	13	-5,893	CACTACATCGGATCTAATCACC	TTG	7965	1248																																																																																																																																																																																																																																																																																																																													
10	-5,323	1,387	15	-6,478	CTTGGGCGGTGATCTCGCGTCC	ATG	7986	1227																																																																																																																																																																																																																																																																																																																													
11	-4,632	1,725	13	-5,702	CGACCCATATCGAACGCTACGGG	GTG	8070	1143																																																																																																																																																																																																																																																																																																																													
12	-5,624	1,240	9	-6,670	TATCGAACGCTACGCGGTGCC	ATG	8076	1137																																																																																																																																																																																																																																																																																																																													
13	-5,037	1,527	18	-6,435	CCRAGCGGCTGCAATCGGAAA	ATG	8103	1110																																																																																																																																																																																																																																																																																																																													
14	-5,266	1,415	13	-6,336	TGTCATATCGGAAATGGCCGAA	ATG	8112	1101																																																																																																																																																																																																																																																																																																																													
15	-5,529	1,287	10	-6,487	AGGCGAAGACCCGAACTGTTT	ATG	8376	837																																																																																																																																																																																																																																																																																																																													
16	-4,068	1,999	9	-5,114	AGGTGTTGCTGCAGCGCGGCTT	ATG	8418	795																																																																																																																																																																																																																																																																																																																													
17	-6,656	0,737	18	-8,054	CGCCGGGTTAAAATCTTTAAT	ATG	8481	732																																																																																																																																																																																																																																																																																																																													
18	-7,082	0,529	10	-8,041	TGCTAACCCCGTTGGCCTGGCC	GTG	8514	699																																																																																																																																																																																																																																																																																																																													
19	-6,193	0,963	15	-7,348	TATCGCCGCACTCGCAGCCGGT	TTG	8544	669																																																																																																																																																																																																																																																																																																																													
20	-4,399	1,838	12	-5,445	GTTCCGTAATGGGCTCGACGTT	ATG	8595	618																																																																																																																																																																																																																																																																																																																													
21	-3,898	2,082	7	-5,120	GGGTCCGTTTGTAAAGCAACAA	TTG	8619	594																																																																																																																																																																																																																																																																																																																													
22	-4,068	1,999	15	-5,223	TATTCAGGCGAAGCCTGGAT	ATG	8721	492																																																																																																																																																																																																																																																																																																																													
23	-3,230	2,409	7	-4,451	TGCAGGCGAAGCCTGGATATG	GTG	8724	489																																																																																																																																																																																																																																																																																																																													
24	-5,337	1,380	16	-6,559	TGTTGAGATTGGCTTACCCGT	GTG	8766	447																																																																																																																																																																																																																																																																																																																													
25	-2,375	2,826	9	-3,421	TGTGGCGACTCTGGAACCGGA	ATG	8787	426																																																																																																																																																																																																																																																																																																																													
26	-7,416	0,366	12	-8,461	GATTCGGTCCGCTATCGCTGGT	GTG	8820	393																																																																																																																																																																																																																																																																																																																													
27	-3,924	2,070	7	-5,146	TGTGATCGGCTGGCTAGAAAAA	ATG	8841	372																																																																																																																																																																																																																																																																																																																													
28	-4,141	1,964	16	-5,362	CGCAGCGGTTAGCGATCCCGAT	ATG	8967	246																																																																																																																																																																																																																																																																																																																													
29	-5,429	1,336	10	-6,388	GGGTAGCGATCCCGATATGACT	ATG	8973	240																																																																																																																																																																																																																																																																																																																													
30	-5,029	1,531	16	-6,250	GATCCTGAAGTTTTCTTCCACA	TTG	9015	198																																																																																																																																																																																																																																																																																																																													
31	-6,676	0,727	12	-7,721	GTTTTCTCCACATTGCCCGAT	ATG	9024	189																																																																																																																																																																																																																																																																																																																													
32	-5,755	1,176	10	-6,713	TTCCACATTGCCCGATATGACT	ATG	9030	183																																																																																																																																																																																																																																																																																																																													
33	-4,662	1,710	16	-5,884	TATGCAGGCTGCAGCCCGGAC	GTG	9051	162																																																																																																																																																																																																																																																																																																																													
34	-5,526	1,288	7	-6,748	CCCGACGTTGACCGCGCAGCGC	GTG	9066	147																																																																																																																																																																																																																																																																																																																													
35	-3,119	2,463	11	-4,119	AGGGGCGGAAAAGATCATTTAA	TTG	9177	36																																																																																																																																																																																																																																																																																																																													
<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>	<p><i>Plus long orf mais pas avec un ATG</i></p> <p><i>Meilleur score</i></p>																																																																																																																																																																																																																																																																																																																																				
<p>Est-ce que le start est conservé chez les</p>	<p>—</p>																																																																																																																																																																																																																																																																																																																																				

homologues voir Starterator ?	
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	
DECISION:	7551

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 : (q#: s#) : [alignment] ; e-value :</p>  <p>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th>Sequence</th> <th>Score (bits)</th> <th>E-value</th> </tr> </thead> <tbody> <tr> <td>Cyranop_Draft_13, function unknown, 553</td> <td>1035</td> <td>0.0</td> </tr> <tr> <td>SocialPhiler_11, tape measure protein, 607</td> <td>322</td> <td>9e-88</td> </tr> <tr> <td>Kela_11, tape measure protein, 607</td> <td>322</td> <td>9e-88</td> </tr> <tr> <td>Ariehaps_11, tape measure protein, 607</td> <td>322</td> <td>9e-88</td> </tr> <tr> <td>StewieDriff_13, tape measure protein, 665</td> <td>322</td> <td>1e-87</td> </tr> <tr> <td>Lore_13, tape measure protein, 665</td> <td>322</td> <td>1e-87</td> </tr> <tr> <td>Ialia_15, tape measure protein, 665</td> <td>322</td> <td>1e-87</td> </tr> </tbody> </table>	Sequence	Score (bits)	E-value	Cyranop_Draft_13, function unknown, 553	1035	0.0	SocialPhiler_11, tape measure protein, 607	322	9e-88	Kela_11, tape measure protein, 607	322	9e-88	Ariehaps_11, tape measure protein, 607	322	9e-88	StewieDriff_13, tape measure protein, 665	322	1e-87	Lore_13, tape measure protein, 665	322	1e-87	Ialia_15, tape measure protein, 665	322	1e-87
Sequence	Score (bits)	E-value																							
Cyranop_Draft_13, function unknown, 553	1035	0.0																							
SocialPhiler_11, tape measure protein, 607	322	9e-88																							
Kela_11, tape measure protein, 607	322	9e-88																							
Ariehaps_11, tape measure protein, 607	322	9e-88																							
StewieDriff_13, tape measure protein, 665	322	1e-87																							
Lore_13, tape measure protein, 665	322	1e-87																							
Ialia_15, tape measure protein, 665	322	1e-87																							

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 ?

Uniprot : TMP_BPPHC Probable tape measure protein
 OS=Streptomyces phage phiC31 OX=10719 GN=43 PE=3 SV=1
 Pdb : Tape Measure Protein, gp57; phage tail, tail tip, tape measure protein, VIRAL PROTEIN; 3.7A {Staphylococcus virus 80alpha}

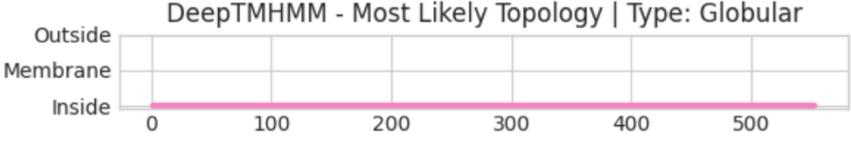
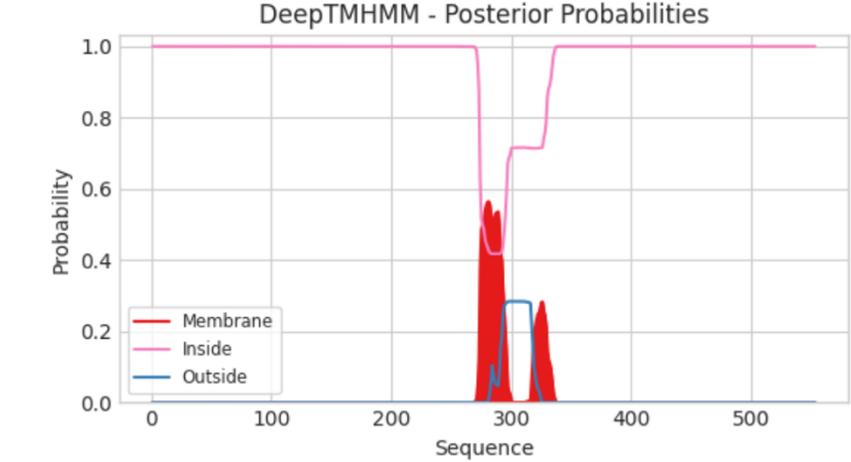
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	Q9ZXAS	TMP_BPPHC Probable tape measure protein OS=Streptomyces phage phiC31 OX=10719 GN=43 PE=3 SV=1	99.91	6.8e-19	188.89	53.9	37	729
<input type="checkbox"/>	P51731	TMP_BPHC1 Probable tape measure protein OS=Haemophilus phage HP1 (strain HP1c1) OX=1289570 PE=3 SV=1	99.91	6.8e-18	180.05	57.5	229	689
<input type="checkbox"/>	O64314	TMP_BPP2 Probable tape measure protein OS=Escherichia phage P2 OX=10679 GN=T PE=3 SV=1	99.84	9.2e-16	167.21	49	139	815
<input type="checkbox"/>	Q6QIAS	TMP_BPBUM Probable tape measure protein OS=Burkholderia phage BcepMu (isolate - /United States/Summer/2002) OX=1283335 GN	99.75	3e-12	140.88	53.4	502	846
<input type="checkbox"/>	E7DNB6	TMP_BPDP1 Tape measure protein OS=Pneumococcus phage Dp-1 OX=59241 GN=TMP PE=4 SV=1	99.7	5e-12	144.02	48.5	431	1149
<input type="checkbox"/>	Q24L11	TMP_BPPCD Probable tape measure protein OS=Clostridium phage phiCD119 (strain Clostridium difficile/United States/Govind	99.64	9.2e-10	123.25	56.1	438	944
<input type="checkbox"/>	Q6KGH8	TMP_BPFO1 Probable tape measure protein OS=Salmonella phage Felix O1 (isolate Felix	99.56	2.8e-10	123.67	41.3	398	742

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 ?

Gene precedent est inconnu

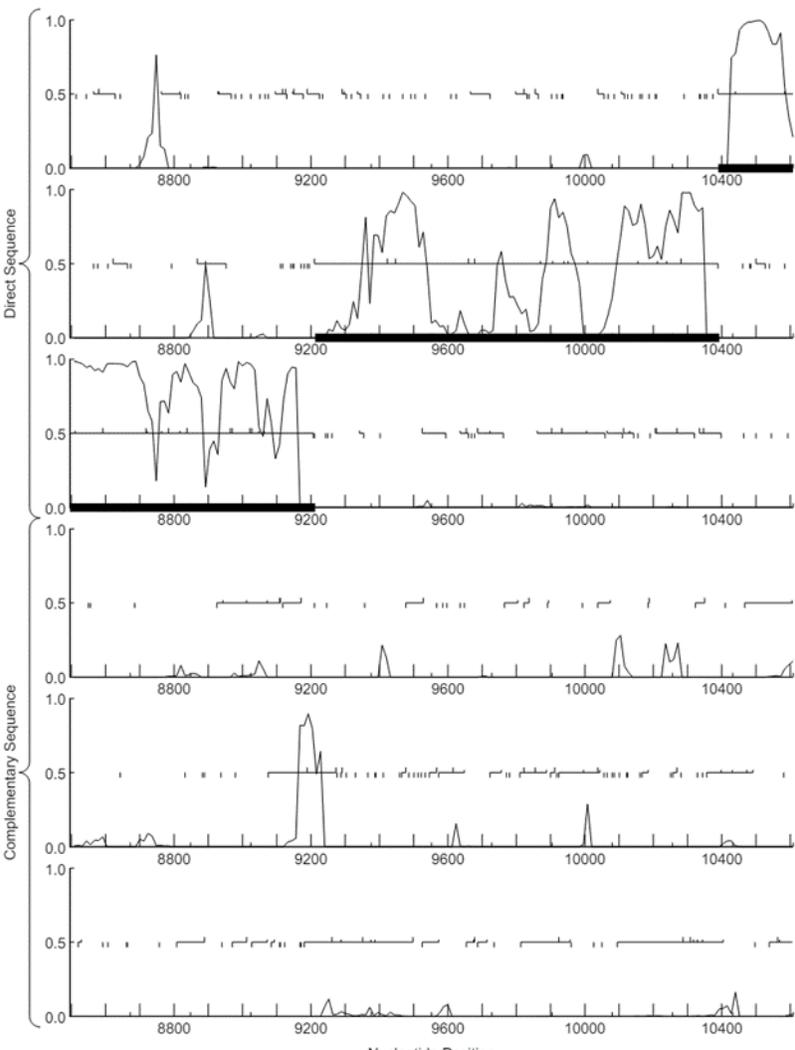
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>NON ?</p>  
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>Oui.</p>
<p>DECISION:</p>	<p>tape mesure protein</p>

Student Gene Annotation Worksheet

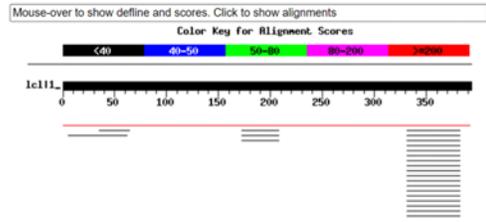
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	14
Coordonnées du Stop	10393
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 1
Coordonnées du Start retenu	9212

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 ?	<p>anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 6/9</p>  <p>The figure displays six tracks of genomic data. The top three tracks are labeled 'Direct Sequence' and the bottom three are 'Complementary Sequence'. Each track shows signal intensity (0.0 to 1.0) across a nucleotide position range from 8800 to 10400. A thick black bar at the bottom of each track indicates the predicted coding region. The top track also includes GeneMark hmm prediction annotations (vertical lines) above the signal. The signal intensity is notably high in the predicted coding region, supporting the potential for a coding sequence.</p>

Distribution of 24 Blast Hits on the Query Sequence



Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyranopS_Draft_14, function unknown, 393	887	0.0
EpicDab_16, minor tail protein, 478	33	1.6
Emperor_17, minor tail protein, 483	32	2.1
Tiff81_Draft_23, function unknown, 433	32	3.5
YoungHarleeyz_Draft_21, function unknown, 433	31	4.6
Seahorse_22, minor tail protein, 433	31	4.6
Sakai_21, minor tail protein, 433	31	4.6
Raphaella_Draft_22, function unknown, 433	31	4.6
RadFad_21, minor tail protein, 433	31	4.6
Phrank15_Draft_21, function unknown, 433	31	4.6
Isolde_21, minor tail protein, 433	31	4.6
Hillester_Draft_21, function unknown, 433	31	4.6
Hestia_21, minor tail protein, 433	31	4.6
Gorpy_21, minor tail protein, 433	31	4.6
CookieBear_Draft_22, function unknown, 433	31	4.6
BillyTP_Draft_21, function unknown, 433	31	4.6
Auxilium_21, minor tail protein, 433	31	4.6
Aikyam_Draft_21, function unknown, 433	31	4.6
Richie_21, minor tail protein, 433	30	7.8
Pikmin_58, function unknown, 48	30	7.8

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

Checkmark	TPA hypothetical protein [Siphoviridae so. cF3212]	Siphoviridae so. cF3212	98.7	98.7	90%	1e-17	29.81%	396	U625833.1
✓	hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	92.8	92.8	98%	2e-16	28.29%	378	WP_201816500.1
✓	TPA hypothetical protein [Siphoviridae so. cF3212]	Siphoviridae so. cF3212	90.9	90.9	90%	1e-15	27.20%	411	DAD68080.1
✓	hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	88.6	88.6	89%	5e-15	26.68%	375	WP_070736112.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	88.6	88.6	90%	8e-15	27.13%	411	DAV46101.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	87.0	87.0	95%	2e-14	23.49%	396	DAD60086.1
✓	conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	85.5	85.5	89%	5e-14	25.27%	375	CCQ13987.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	85.5	85.5	91%	7e-14	25.98%	408	DAW70282.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	85.5	85.5	91%	7e-14	25.71%	385	DAU54398.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	85.5	85.5	91%	8e-14	25.46%	409	DAP19134.1
✓	hypothetical protein [Corynebacterium matruchoti]	Corynebacterium matruchoti	84.7	84.7	94%	1e-13	23.64%	395	WP_126299861.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.7	84.7	91%	1e-13	25.97%	385	DA25515.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.3	84.3	91%	2e-13	26.23%	408	DAS66265.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.0	84.0	91%	2e-13	25.97%	408	DAS95148.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.0	84.0	91%	2e-13	25.65%	408	DAW81670.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	3e-13	25.97%	409	DAK21463.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	4e-13	25.71%	408	DAJ35125.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	4e-13	25.72%	408	DAW44110.1
✓	TPA actin-like protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	4e-13	25.71%	411	DAW69952.1
✓	TPA actin-like protein [Caudoviricetes so.]	Caudoviricetes so.	83.2	83.2	91%	4e-13	25.65%	412	DAP14260.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.2	83.2	91%	5e-13	25.52%	407	DAX60235.1
✓	TPA hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	82.4	82.4	89%	6e-13	27.22%	367	HAT6525552.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	82.0	82.0	91%	1e-12	25.97%	378	DAQ14881.1
✓	hypothetical protein [Actinomyces griseoviridis]	Actinomyces griseoviridis	81.6	81.6	91%	1e-12	26.49%	378	MBE0930126.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.3	84.3	91%	2e-13	26.23%	408	DAS66265.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.0	84.0	91%	2e-13	25.97%	408	DAS95148.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	84.0	84.0	91%	2e-13	25.65%	408	DAW81670.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	3e-13	25.97%	409	DAK21463.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	4e-13	25.71%	408	DAJ35125.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.6	83.6	91%	4e-13	25.72%	408	DAW44110.1
✓	TPA actin-like protein [Caudoviricetes so.]	Caudoviricetes so.	83.2	83.2	91%	4e-13	25.65%	412	DAP14260.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	83.2	83.2	91%	5e-13	25.52%	407	DAX60235.1
✓	TPA hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	82.4	82.4	89%	6e-13	27.22%	367	HAT6525552.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	82.0	82.0	91%	1e-12	25.97%	378	DAQ14881.1
✓	hypothetical protein [Actinomyces griseoviridis]	Actinomyces griseoviridis	81.6	81.6	91%	1e-12	26.49%	378	MBE0930126.1
✓	TPA hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	80.9	80.9	89%	2e-12	26.94%	367	HCG52976170.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	79.3	79.3	91%	9e-12	25.97%	408	DAR82652.1
✓	TPA hypothetical protein [Caudoviricetes so.]	Caudoviricetes so.	79.3	79.3	90%	9e-12	25.40%	414	DAU31952.1
✓	TPA hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	66.2	66.2	78%	1e-07	25.38%	312	HCT5225728.1
✓	hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	63.9	63.9	91%	9e-07	23.58%	395	WP_03852566.1
✓	hypothetical protein FRQ0431_01846 [Corynebacterium dichtherae]	Corynebacterium dichtherae	62.8	62.8	91%	2e-06	23.31%	395	CAB0923875.1
✓	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	62.8	62.8	91%	2e-06	28.90%	382	WP_027011473.1
✓	hypothetical protein [Corynebacterium dichtherae]	Corynebacterium dichtherae	62.8	62.8	91%	2e-06	23.31%	395	WP_196075731.1
✓	hypothetical protein [Corynebacterium pseudodichtherricum]	Corynebacterium pseudodichtherricum	60.5	60.5	91%	1e-05	22.19%	383	WP_284567142.1
✓	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	50.8	50.8	18%	8e-04	44.16%	99	WP_205618476.1
✓	hypothetical protein [Corynebacterium glaucum]	Corynebacterium glaucum	52.8	52.8	91%	0.003	23.37%	385	WP_301925792.1

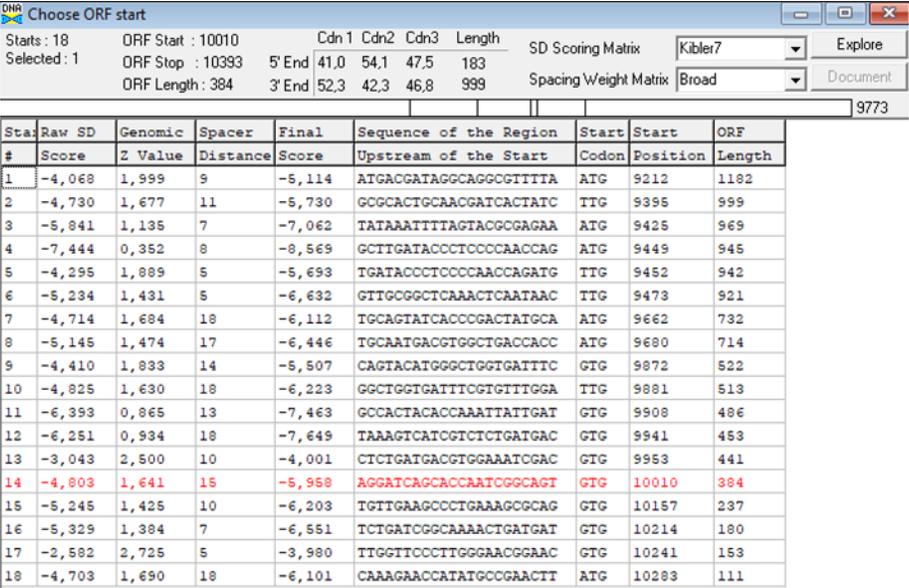
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																											
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :9212 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :9212																																																																																																																																																																											
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <p>Choose ORF start</p> <p>Starts: 18 ORF Start: 10010 Cdn1 Cdn2 Cdn3 Length SD Scoring Matrix Kibler7 Explore Selected: 1 ORF Stop: 10393 5' End 41,0 54,1 47,5 183 Spacing/Weight Matrix Broad Document ORF Length: 384 3' End 52,3 42,3 46,8 999 9773</p> <table border="1"> <thead> <tr> <th>Sta</th> <th>Raw 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> </thead> <tbody> <tr><td>1</td><td>-4,068</td><td>1,999</td><td>9</td><td>-5,114</td><td>ATGACGATAGGCAGGCGTTTA</td><td>ATG</td><td>9212</td><td>1182</td></tr> <tr><td>2</td><td>-4,730</td><td>1,677</td><td>11</td><td>-5,730</td><td>GCGCACTGCAACGATCACTATC</td><td>TTG</td><td>9395</td><td>999</td></tr> <tr><td>3</td><td>-5,841</td><td>1,135</td><td>7</td><td>-7,062</td><td>TATAAATTTTAGTACGGAGAA</td><td>ATG</td><td>9425</td><td>969</td></tr> <tr><td>4</td><td>-7,444</td><td>0,352</td><td>8</td><td>-8,569</td><td>GCTTGATACCCCTCCCCAACCAG</td><td>ATG</td><td>9449</td><td>945</td></tr> <tr><td>5</td><td>-4,295</td><td>1,889</td><td>5</td><td>-5,693</td><td>TGATACCCCTCCCCAACCCAGATG</td><td>TTG</td><td>9452</td><td>942</td></tr> <tr><td>6</td><td>-5,234</td><td>1,431</td><td>5</td><td>-6,632</td><td>GTTGCGGCTCAAACTCAATAAC</td><td>TTG</td><td>9473</td><td>921</td></tr> <tr><td>7</td><td>-4,714</td><td>1,684</td><td>18</td><td>-6,112</td><td>TGCAGTATCACCCGACTATGCA</td><td>ATG</td><td>9662</td><td>732</td></tr> <tr><td>8</td><td>-5,145</td><td>1,474</td><td>17</td><td>-6,446</td><td>TGCAATGACGTGGCTGACCACC</td><td>ATG</td><td>9680</td><td>714</td></tr> <tr><td>9</td><td>-4,410</td><td>1,833</td><td>14</td><td>-5,507</td><td>CAGTACATGGGCTGGTGATTC</td><td>GTG</td><td>9872</td><td>522</td></tr> <tr><td>10</td><td>-4,825</td><td>1,630</td><td>18</td><td>-6,223</td><td>GGCTGGTGATTTCTGTITGGA</td><td>TTG</td><td>9881</td><td>513</td></tr> <tr><td>11</td><td>-6,393</td><td>0,865</td><td>13</td><td>-7,463</td><td>GCCACTACACCAAAATTATGAT</td><td>GTG</td><td>9908</td><td>486</td></tr> <tr><td>12</td><td>-6,251</td><td>0,934</td><td>18</td><td>-7,649</td><td>TAAAGTCATCGTCTCTGATGAC</td><td>GTG</td><td>9941</td><td>453</td></tr> <tr><td>13</td><td>-3,043</td><td>2,500</td><td>10</td><td>-4,001</td><td>CTCTGATGACGTGGAAATCGAC</td><td>GTG</td><td>9953</td><td>441</td></tr> <tr><td>14</td><td>-4,803</td><td>1,641</td><td>15</td><td>-5,958</td><td>AGGATCAGCACCAAATCGGCAGT</td><td>GTG</td><td>10010</td><td>384</td></tr> <tr><td>15</td><td>-5,245</td><td>1,425</td><td>10</td><td>-6,203</td><td>TGTTGAAGCCCTGAAAGCGCAG</td><td>GTG</td><td>10157</td><td>237</td></tr> <tr><td>16</td><td>-5,329</td><td>1,384</td><td>7</td><td>-6,551</td><td>TCTGATCGGCAAAACCTGATGAT</td><td>GTG</td><td>10214</td><td>180</td></tr> <tr><td>17</td><td>-2,582</td><td>2,725</td><td>5</td><td>-3,980</td><td>TTGGTTCCCTTGGGAACGGAAAC</td><td>GTG</td><td>10241</td><td>153</td></tr> <tr><td>18</td><td>-4,703</td><td>1,690</td><td>18</td><td>-6,101</td><td>CARAGAACCATATGCCGAACTT</td><td>ATG</td><td>10283</td><td>111</td></tr> </tbody> </table>	Sta	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length	1	-4,068	1,999	9	-5,114	ATGACGATAGGCAGGCGTTTA	ATG	9212	1182	2	-4,730	1,677	11	-5,730	GCGCACTGCAACGATCACTATC	TTG	9395	999	3	-5,841	1,135	7	-7,062	TATAAATTTTAGTACGGAGAA	ATG	9425	969	4	-7,444	0,352	8	-8,569	GCTTGATACCCCTCCCCAACCAG	ATG	9449	945	5	-4,295	1,889	5	-5,693	TGATACCCCTCCCCAACCCAGATG	TTG	9452	942	6	-5,234	1,431	5	-6,632	GTTGCGGCTCAAACTCAATAAC	TTG	9473	921	7	-4,714	1,684	18	-6,112	TGCAGTATCACCCGACTATGCA	ATG	9662	732	8	-5,145	1,474	17	-6,446	TGCAATGACGTGGCTGACCACC	ATG	9680	714	9	-4,410	1,833	14	-5,507	CAGTACATGGGCTGGTGATTC	GTG	9872	522	10	-4,825	1,630	18	-6,223	GGCTGGTGATTTCTGTITGGA	TTG	9881	513	11	-6,393	0,865	13	-7,463	GCCACTACACCAAAATTATGAT	GTG	9908	486	12	-6,251	0,934	18	-7,649	TAAAGTCATCGTCTCTGATGAC	GTG	9941	453	13	-3,043	2,500	10	-4,001	CTCTGATGACGTGGAAATCGAC	GTG	9953	441	14	-4,803	1,641	15	-5,958	AGGATCAGCACCAAATCGGCAGT	GTG	10010	384	15	-5,245	1,425	10	-6,203	TGTTGAAGCCCTGAAAGCGCAG	GTG	10157	237	16	-5,329	1,384	7	-6,551	TCTGATCGGCAAAACCTGATGAT	GTG	10214	180	17	-2,582	2,725	5	-3,980	TTGGTTCCCTTGGGAACGGAAAC	GTG	10241	153	18	-4,703	1,690	18	-6,101	CARAGAACCATATGCCGAACTT	ATG	10283	111
Sta	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length																																																																																																																																																																				
1	-4,068	1,999	9	-5,114	ATGACGATAGGCAGGCGTTTA	ATG	9212	1182																																																																																																																																																																				
2	-4,730	1,677	11	-5,730	GCGCACTGCAACGATCACTATC	TTG	9395	999																																																																																																																																																																				
3	-5,841	1,135	7	-7,062	TATAAATTTTAGTACGGAGAA	ATG	9425	969																																																																																																																																																																				
4	-7,444	0,352	8	-8,569	GCTTGATACCCCTCCCCAACCAG	ATG	9449	945																																																																																																																																																																				
5	-4,295	1,889	5	-5,693	TGATACCCCTCCCCAACCCAGATG	TTG	9452	942																																																																																																																																																																				
6	-5,234	1,431	5	-6,632	GTTGCGGCTCAAACTCAATAAC	TTG	9473	921																																																																																																																																																																				
7	-4,714	1,684	18	-6,112	TGCAGTATCACCCGACTATGCA	ATG	9662	732																																																																																																																																																																				
8	-5,145	1,474	17	-6,446	TGCAATGACGTGGCTGACCACC	ATG	9680	714																																																																																																																																																																				
9	-4,410	1,833	14	-5,507	CAGTACATGGGCTGGTGATTC	GTG	9872	522																																																																																																																																																																				
10	-4,825	1,630	18	-6,223	GGCTGGTGATTTCTGTITGGA	TTG	9881	513																																																																																																																																																																				
11	-6,393	0,865	13	-7,463	GCCACTACACCAAAATTATGAT	GTG	9908	486																																																																																																																																																																				
12	-6,251	0,934	18	-7,649	TAAAGTCATCGTCTCTGATGAC	GTG	9941	453																																																																																																																																																																				
13	-3,043	2,500	10	-4,001	CTCTGATGACGTGGAAATCGAC	GTG	9953	441																																																																																																																																																																				
14	-4,803	1,641	15	-5,958	AGGATCAGCACCAAATCGGCAGT	GTG	10010	384																																																																																																																																																																				
15	-5,245	1,425	10	-6,203	TGTTGAAGCCCTGAAAGCGCAG	GTG	10157	237																																																																																																																																																																				
16	-5,329	1,384	7	-6,551	TCTGATCGGCAAAACCTGATGAT	GTG	10214	180																																																																																																																																																																				
17	-2,582	2,725	5	-3,980	TTGGTTCCCTTGGGAACGGAAAC	GTG	10241	153																																																																																																																																																																				
18	-4,703	1,690	18	-6,101	CARAGAACCATATGCCGAACTT	ATG	10283	111																																																																																																																																																																				
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>ORF le plus long</i> <i>ATG</i> <i>Score : -5</i>																																																																																																																																																																											
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 ?

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

[Next](#) | [Previous](#) | [Descriptions](#)

hypothetical protein [Corynebacterium striatum]
 Sequence ID: [WP_201816500.1](#) Length: 378 Number of Matches: 1
[See 2 more title\(s\)](#) | [See all Identical Proteins \(IPG\)](#)

Range 1: 2 to 370 [GenPept](#) | [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
92.8 bits(229)	2e-16	Compositional matrix adjust.	114/403(28%)	182/403(45%)	51/403(12%)
Query 1	MSTLADSFELQLGIRVITTSIGITITWVRRQAVEGVSRMGROSLDQPKQRTATIT				60
Sbjct 2	VSRLAASTEIKIQSDQNGITYRPGNI-----LTGISITWGRQSLMDAPVKRSCKFT				52
Query 61	I-LEDINFSREMLDILPNQMLRLKLNLSYPVIFEGIIDIKPFHGSKQGITVSATES				119
Sbjct 53	ISYEDIQ-AARDQWFTFLRANVTVLDGT--VVFVGHISMEFEIIGDQWIAVSVEA				108
Query 120	FFTSFDRLGTIAEPEGNFLRAVSPDYAMTLTMRVYV--PASTFLRTMPSGASVF				177
Sbjct 109	--TQWNLMSRTF-ETRANPALMASMEIQSG--GIRPNIIEPATTEVFRFP-----				157
Query 178	AQPEAQLTKVQIIEVAANRPLTYPSPDKRQIRST-WAGDFVFLQLHQIIDVPSA				236
Sbjct 158	AQPVT--LTVEQNGAVAAPWPLSHPNQSPDYQCSATTNRLDHAWGAN----ITV-FA				209
Query 237	KVIVSDVDIVRTPYKSWTYQQSGAPISGVSEFYPKQSRPQGNVAVSDHLNFNPNY				296
Sbjct 210	NEVAGKPRMDAGEMTRITYWESGGTYGESNNGTQLSSYNAPGSEGGANSGRKSTNPYA				269
Query 297	VTGINDVIKPGVEALKAQVTSRPRKFRSHNL-----IGKDDVHFHFNERNVQIQRI				349
Sbjct 270	VKSYPLEHKAASNAQAQITAPREFTFYDGLMEIQIGQWPLF-YTMEQR-RRFKI				327
Query 350	T-----KEPYAE-LMGIDPVFVPIGGTLKITADVDTHVNCVM 386				
Sbjct 328	NPDGSGNDPYAATLFDLQPLVYVIGGTLTRHQQTTHVNTALY 370				

DECISION:

9212

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 : RIEN</p> <p>*** nr :</p> <table border="1" style="width: 100%; border-collapse: collapse; font-size: 0.7em;"> <tbody> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>85.5</td><td>85.5</td><td>91%</td><td>7e-14</td><td>25.71%</td><td>385</td><td>DAU54398.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>85.5</td><td>85.5</td><td>91%</td><td>8e-14</td><td>25.46%</td><td>409</td><td>DAP19134.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium matruchotii]</td><td>Corynebacterium matruchotii</td><td>84.7</td><td>84.7</td><td>94%</td><td>1e-13</td><td>23.64%</td><td>395</td><td>WP_126299861.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>84.7</td><td>84.7</td><td>91%</td><td>1e-13</td><td>25.97%</td><td>385</td><td>DAJ2551.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>84.3</td><td>84.3</td><td>91%</td><td>2e-13</td><td>26.23%</td><td>408</td><td>DAS66265.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>84.0</td><td>84.0</td><td>91%</td><td>2e-13</td><td>25.97%</td><td>408</td><td>DAS95148.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>84.0</td><td>84.0</td><td>91%</td><td>2e-13</td><td>25.65%</td><td>408</td><td>DAW81670.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>83.6</td><td>83.6</td><td>91%</td><td>3e-13</td><td>25.97%</td><td>409</td><td>DAK21463.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>83.6</td><td>83.6</td><td>91%</td><td>4e-13</td><td>25.71%</td><td>408</td><td>DAJ35125.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>83.6</td><td>83.6</td><td>91%</td><td>4e-13</td><td>25.72%</td><td>408</td><td>DAV44110.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: actin-like protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>83.6</td><td>83.6</td><td>91%</td><td>4e-13</td><td>25.71%</td><td>411</td><td>DAW89952.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: actin-like protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>83.2</td><td>83.2</td><td>91%</td><td>4e-13</td><td>25.65%</td><td>412</td><td>DAP14260.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>83.2</td><td>83.2</td><td>91%</td><td>5e-13</td><td>25.52%</td><td>407</td><td>DAX60235.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Corynebacterium striatum]</td><td>Corynebacterium striatum</td><td>82.4</td><td>82.4</td><td>89%</td><td>6e-13</td><td>27.22%</td><td>367</td><td>HAT6625552.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>82.0</td><td>82.0</td><td>91%</td><td>1e-12</td><td>25.97%</td><td>378</td><td>DAQ14881.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Actinomyces graevenitzi]</td><td>Actinomyces graevenitzi</td><td>81.6</td><td>81.6</td><td>91%</td><td>1e-12</td><td>26.49%</td><td>378</td><td>MBF0930126.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Corynebacterium striatum]</td><td>Corynebacterium striatum</td><td>80.9</td><td>80.9</td><td>89%</td><td>2e-12</td><td>26.94%</td><td>367</td><td>HCG2878170.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>79.3</td><td>79.3</td><td>91%</td><td>9e-12</td><td>25.97%</td><td>408</td><td>DAR82652.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Caudoviricetes sp.]</td><td>Caudoviricetes sp.</td><td>79.3</td><td>79.3</td><td>90%</td><td>9e-12</td><td>25.40%</td><td>414</td><td>DAU31952.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>TPA: hypothetical protein [Corynebacterium striatum]</td><td>Corynebacterium striatum</td><td>66.2</td><td>66.2</td><td>78%</td><td>1e-07</td><td>25.38%</td><td>312</td><td>HCT5225728.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>63.9</td><td>63.9</td><td>91%</td><td>9e-07</td><td>23.58%</td><td>395</td><td>WP_003852566.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [FRC0431_01846 [Corynebacterium diphtheriae]]</td><td>Corynebacterium diphtheriae</td><td>62.8</td><td>62.8</td><td>91%</td><td>2e-06</td><td>23.31%</td><td>395</td><td>CAB0923875.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium freiburgense]</td><td>Corynebacterium freiburgense</td><td>62.8</td><td>62.8</td><td>53%</td><td>2e-06</td><td>28.90%</td><td>382</td><td>WP_027011473.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>62.8</td><td>62.8</td><td>91%</td><td>2e-06</td><td>23.31%</td><td>395</td><td>WP_196975731.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium pseudodiphthericum]</td><td>Corynebacterium pseudodiphthericum</td><td>60.5</td><td>60.5</td><td>91%</td><td>1e-05</td><td>22.19%</td><td>383</td><td>WP_284587142.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium freiburgense]</td><td>Corynebacterium freiburgense</td><td>50.8</td><td>50.8</td><td>18%</td><td>8e-04</td><td>44.16%</td><td>99</td><td>WP_205618476.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>hypothetical protein [Corynebacterium glaucum]</td><td>Corynebacterium glaucum</td><td>52.8</td><td>52.8</td><td>91%</td><td>0.003</td><td>23.37%</td><td>385</td><td>WP_301925792.1</td></tr> </tbody> </table>	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	7e-14	25.71%	385	DAU54398.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	8e-14	25.46%	409	DAP19134.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium matruchotii]	Corynebacterium matruchotii	84.7	84.7	94%	1e-13	23.64%	395	WP_126299861.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.7	84.7	91%	1e-13	25.97%	385	DAJ2551.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.3	84.3	91%	2e-13	26.23%	408	DAS66265.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.0	84.0	91%	2e-13	25.97%	408	DAS95148.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.0	84.0	91%	2e-13	25.65%	408	DAW81670.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	3e-13	25.97%	409	DAK21463.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	4e-13	25.71%	408	DAJ35125.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	4e-13	25.72%	408	DAV44110.1	<input checked="" type="checkbox"/>	TPA: actin-like protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	4e-13	25.71%	411	DAW89952.1	<input checked="" type="checkbox"/>	TPA: actin-like protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.2	83.2	91%	4e-13	25.65%	412	DAP14260.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.2	83.2	91%	5e-13	25.52%	407	DAX60235.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	82.4	82.4	89%	6e-13	27.22%	367	HAT6625552.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	82.0	82.0	91%	1e-12	25.97%	378	DAQ14881.1	<input checked="" type="checkbox"/>	hypothetical protein [Actinomyces graevenitzi]	Actinomyces graevenitzi	81.6	81.6	91%	1e-12	26.49%	378	MBF0930126.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	80.9	80.9	89%	2e-12	26.94%	367	HCG2878170.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	79.3	79.3	91%	9e-12	25.97%	408	DAR82652.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	79.3	79.3	90%	9e-12	25.40%	414	DAU31952.1	<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	66.2	66.2	78%	1e-07	25.38%	312	HCT5225728.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.9	63.9	91%	9e-07	23.58%	395	WP_003852566.1	<input checked="" type="checkbox"/>	hypothetical protein [FRC0431_01846 [Corynebacterium diphtheriae]]	Corynebacterium diphtheriae	62.8	62.8	91%	2e-06	23.31%	395	CAB0923875.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	62.8	62.8	53%	2e-06	28.90%	382	WP_027011473.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	62.8	62.8	91%	2e-06	23.31%	395	WP_196975731.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	60.5	60.5	91%	1e-05	22.19%	383	WP_284587142.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	50.8	50.8	18%	8e-04	44.16%	99	WP_205618476.1	<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium glaucum]	Corynebacterium glaucum	52.8	52.8	91%	0.003	23.37%	385	WP_301925792.1
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	7e-14	25.71%	385	DAU54398.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	8e-14	25.46%	409	DAP19134.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium matruchotii]	Corynebacterium matruchotii	84.7	84.7	94%	1e-13	23.64%	395	WP_126299861.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.7	84.7	91%	1e-13	25.97%	385	DAJ2551.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.3	84.3	91%	2e-13	26.23%	408	DAS66265.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.0	84.0	91%	2e-13	25.97%	408	DAS95148.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.0	84.0	91%	2e-13	25.65%	408	DAW81670.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	3e-13	25.97%	409	DAK21463.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	4e-13	25.71%	408	DAJ35125.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	4e-13	25.72%	408	DAV44110.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: actin-like protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.6	83.6	91%	4e-13	25.71%	411	DAW89952.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: actin-like protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.2	83.2	91%	4e-13	25.65%	412	DAP14260.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	83.2	83.2	91%	5e-13	25.52%	407	DAX60235.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	82.4	82.4	89%	6e-13	27.22%	367	HAT6625552.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	82.0	82.0	91%	1e-12	25.97%	378	DAQ14881.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Actinomyces graevenitzi]	Actinomyces graevenitzi	81.6	81.6	91%	1e-12	26.49%	378	MBF0930126.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	80.9	80.9	89%	2e-12	26.94%	367	HCG2878170.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	79.3	79.3	91%	9e-12	25.97%	408	DAR82652.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	79.3	79.3	90%	9e-12	25.40%	414	DAU31952.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	66.2	66.2	78%	1e-07	25.38%	312	HCT5225728.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.9	63.9	91%	9e-07	23.58%	395	WP_003852566.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [FRC0431_01846 [Corynebacterium diphtheriae]]	Corynebacterium diphtheriae	62.8	62.8	91%	2e-06	23.31%	395	CAB0923875.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	62.8	62.8	53%	2e-06	28.90%	382	WP_027011473.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	62.8	62.8	91%	2e-06	23.31%	395	WP_196975731.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodiphthericum	60.5	60.5	91%	1e-05	22.19%	383	WP_284587142.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	50.8	50.8	18%	8e-04	44.16%	99	WP_205618476.1																																																																																																																																																																																																																																																																						
<input checked="" type="checkbox"/>	hypothetical protein [Corynebacterium glaucum]	Corynebacterium glaucum	52.8	52.8	91%	0.003	23.37%	385	WP_301925792.1																																																																																																																																																																																																																																																																						

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 35 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"/> TPA: hypothetical protein [Siphoviridae sp. c CCv12]	Siphoviridae sp. c CCv12	98.7	96.7	95%	1e-17	25.81%	396	DAD90005.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	92.8	92.8	98%	2e-16	28.29%	378	WP_201818500.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Siphoviridae sp. c Rz6]	Siphoviridae sp. c Rz6	90.9	90.9	90%	1e-15	27.20%	411	DAD68080.1
<input checked="" type="checkbox"/> hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	88.6	88.6	89%	5e-15	26.68%	375	WP_070736112.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	88.6	88.6	90%	8e-15	27.13%	411	DAV46101.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	87.0	87.0	95%	2e-14	23.49%	396	DAD60086.1
<input checked="" type="checkbox"/> conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	85.5	85.5	89%	5e-14	25.27%	375	QQD13987.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	7e-14	25.98%	408	DAW70282.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	7e-14	25.71%	385	DAU54398.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	85.5	85.5	91%	8e-14	25.46%	409	DAP19134.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium matrucholii]	Corynebacterium matrucholii	84.7	84.7	94%	1e-13	23.64%	395	WP_126299861.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.7	84.7	91%	1e-13	25.97%	385	DAJ25515.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.3	84.3	91%	2e-13	26.23%	408	DAS66265.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.0	84.0	91%	2e-13	25.97%	408	DAS95148.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	84.0	84.0	91%	2e-13	25.65%	408	DAW81670.1

Alignment Scores: Show Conserved Domains Alignment Scores < 40 40 - 50 50 - 80 80 - 200 >= 200

35 sequences selected No putative conserved domains have been detected

Distribution of the top 35 Blast Hits on 35 subject sequences

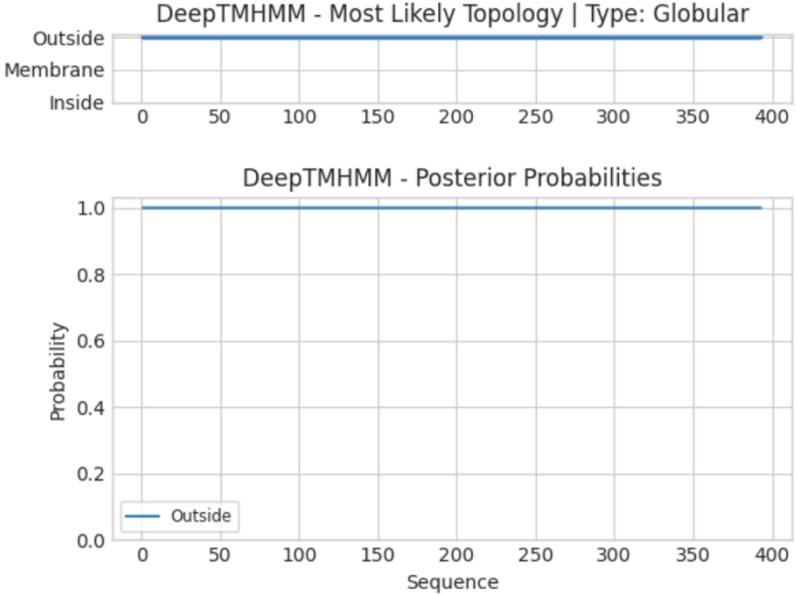
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Pfam : Phage_GPD ; Phage tail baseplate hub (GPD)
 Uniprot : 43 kDa tail protein; BACTERIOPHAGE MU, BASEPLATE, GENE PRODUCT 44, STRUCTURAL PROTEIN; 2.1A {Enterobacteria phage Mu} SC
 Uniprot : BP44_BPMU Baseplate hub protein gp44 OS=Escherichia phage Mu OX=10677 GN=P PE=1 SV=1

<input type="checkbox"/>	1	1WRJ_U	43 kDa tail protein; BACTERIOPHAGE MU, BASEPLATE, GENE PRODUCT 44, STRUCTURAL PROTEIN; 2.1A {Enterobacteria phage Mu} SC	95.68	1.3	45.28	15.9	167	379
<input type="checkbox"/>	2	3CDD_B	Prophage Mu502, 43 kDa tail protein; Mu502, Shewanella oneidensis MR-1, Structural Genomics, PSI-2, Protein Structure In	94.93	1.8	43.68	14	155	361
<input type="checkbox"/>	3	P08558	BP44_BPMU Baseplate hub protein gp44 OS=Escherichia phage Mu OX=10677 GN=P PE=1 SV=1	94.84	2.7	43.4	15.2	165	379
<input type="checkbox"/>	4	8EON_E	Baseplate hub gp41; Pseudomonas, phage, baseplate, VIRUS{Pseudomonas phage vB_PaeM_E217}	94.82	3	41.91	15.2	168	287
<input type="checkbox"/>	5	3D37_B	Tail protein, 43 kDa; tail protein, structural genomics, PSI, MCS6, Protein Structure Initiative, Midwest Center for Str	94.78	2.5	43.49	14.8	152	381
<input type="checkbox"/>	6	8GRA_G	Type VI secretion system spike protein VgrG, Type VI Secretion System, VgrG, HgsG, PAAR, TRANSPORT PROTEIN, {Bacteroides}	94.58	4.6	44.92	17.2	150	616
<input type="checkbox"/>	7	6J0M_B	PvcR; assembly, Photorhabdus asymbiotica, PVC, contractile injection system, bacteriophage-like, PROTEIN TRANSPORT; 3.9A	94.4	3.3	45.01	15.2	143	538
<input type="checkbox"/>	8	6RBK_C	Alp8; Anti-feeding prophage, secretion system, AFP, contractile, VIRUS LIKE PARTICLE, baseplate; 3.4A {Serratia entomoph}	93.58	7.1	42.41	15.8	152	529
<input type="checkbox"/>	9	6USH_B	Probable bacteriophage protein Pycocin R2;	91.88	15	36.36	14.3	145	329

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
 Gene avant : tape measure protein

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<h3>DeepTMHMM - Predictions</h3> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p> 
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>non</p>
<p>DECISION:</p>	<p>Minor tail protein</p>

Student Gene Annotation Worksheet

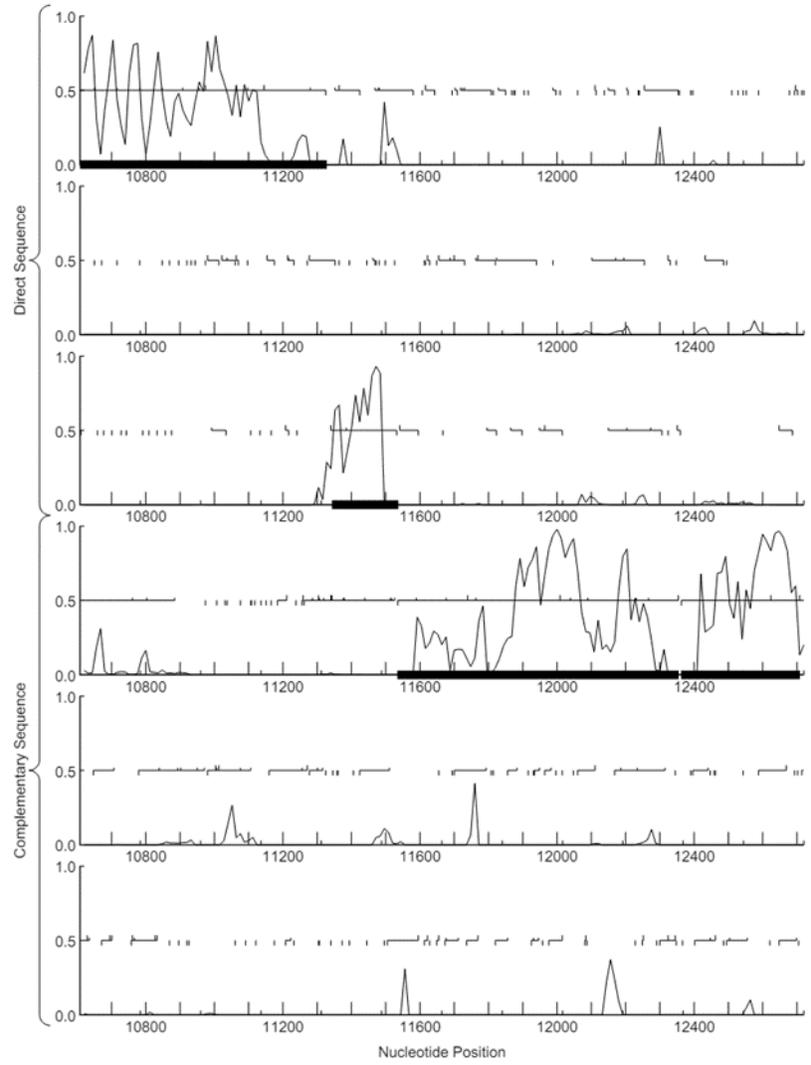
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	15
Coordonnées du Stop	11328
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 4

Coordonnées du Start retenu	10390
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 ?	<p>GeneMark hm prediction</p> <p>YanoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 6/9</p> <p>Direct Sequence</p> <p>Complementary Sequence</p> <p>Nucleotide Position</p>

GeneMark.hmm prediction
ranoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 7/9



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

Sequences producing significant alignments Download Select columns Show 100

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 [Corynebacterium aoulae]	Corynebacterium aoulae	150	394	47%	1e-36	55.77%	666	WP_075725710.1
<input checked="" type="checkbox"/> hypothetical protein [Clostridia bacterium]	Clostridia bacterium	132	298	50%	1e-31	51.23%	327	MBR5768827.1
<input checked="" type="checkbox"/> hypothetical protein [Clostridia bacterium]	Clostridia bacterium	132	359	50%	1e-31	51.23%	328	MBO4583800.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Pseudomonas aeruginosa]	Pseudomonas aeruginosa	124	197	37%	1e-27	56.90%	481	HCF5853879.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Allosethinoviricella so.]	Allosethinoviricella so.	123	501	48%	3e-27	53.72%	537	HZF95778.1
<input checked="" type="checkbox"/> rhage tail repeat domain-containing protein [Paludibacteraceae bacterium]	Paludibacteraceae bacterium	123	675	45%	7e-27	53.72%	1040	MCQ2313286.1
<input checked="" type="checkbox"/> hypothetical protein [Prescottella equi]	Prescottella equi	119	321	39%	2e-26	57.72%	394	WP_286460814.1
<input checked="" type="checkbox"/> hypothetical protein [Paratrunkia so. EUN1f]	Paratrunkia so. EUN1f	119	119	62%	4e-26	45.05%	423	WP_006545789.1
<input checked="" type="checkbox"/> rhage tail protein [Mannheimia oemjora]	Mannheimia oemjora	118	223	38%	4e-25	52.89%	790	WP_176808477.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium glutamicum]	Corynebacterium glutamicum	115	115	29%	9e-25	62.37%	415	WP_211438912.1
<input checked="" type="checkbox"/> rhage tail fiber protein [Stutzerimonas stutzeri]	Stutzerimonas stutzeri	114	502	39%	7e-24	55.74%	623	WP_020699854.1
<input checked="" type="checkbox"/> TPA_minor tail protein [Caudoviricoetes so.]	Caudoviricoetes so.	114	271	47%	7e-24	55.96%	804	DAV45625.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Allosethinoviricella so.]	Allosethinoviricella so.	113	289	39%	9e-24	49.30%	523	HEV2865932.1
<input checked="" type="checkbox"/> hypothetical protein [Thalassovita litoralis]	Thalassovita litoralis	108	180	50%	3e-22	45.28%	495	WP_142493021.1
<input checked="" type="checkbox"/> TPA_minor tail protein [Caudoviricoetes so.]	Caudoviricoetes so.	109	282	42%	3e-22	52.80%	601	DAT66149.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium ciconiae]	Corynebacterium ciconiae	109	109	36%	4e-22	50.42%	664	WP_301354965.1
<input checked="" type="checkbox"/> rhage tail fiber protein [Pseudomonas outida]	Pseudomonas outida	104	378	39%	6e-21	60.44%	459	WP_212628419.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium so. TAE3-ERU2]	Corynebacterium so. TAE3-ERU2	105	276	38%	7e-21	52.63%	663	WP_218366510.1
<input checked="" type="checkbox"/> hypothetical protein BC39_27945 [Pseudomonas montelli]	Pseudomonas montelli	104	285	37%	1e-20	60.64%	459	KXK68009.1

Distribution of 202 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	E Value
622	e-178
69	1e-11
69	1e-11
67	7e-11
64	6e-10
63	1e-09
62	1e-09
62	2e-09

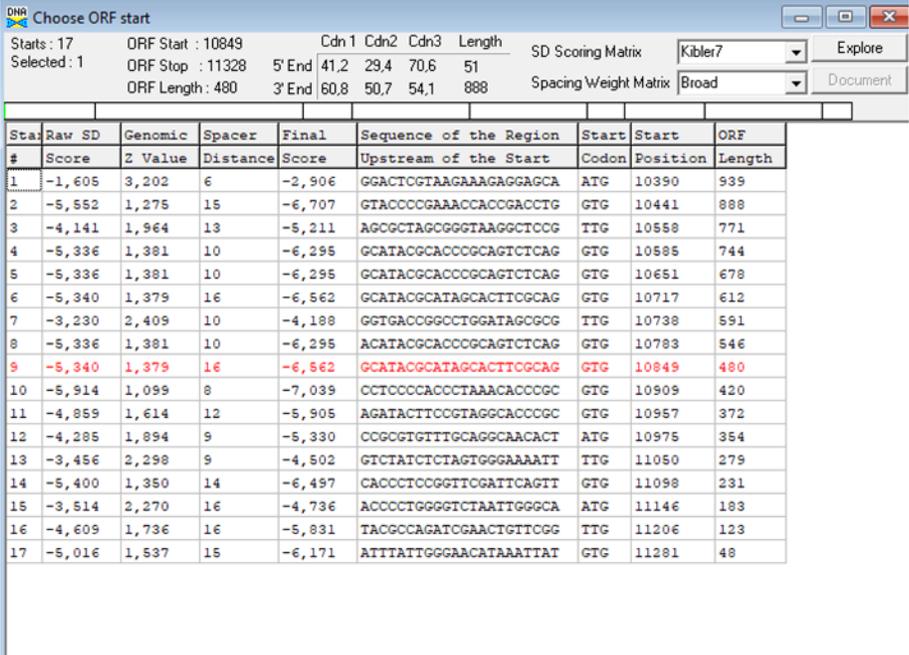
<input checked="" type="checkbox"/> hypothetical protein TVAG_128080 [Trichomonas vaginalis G3]	Trichomonas vaginalis G3	87.4	490	40%	8e-15	33.86%	486	EAY09999.1
<input checked="" type="checkbox"/> hypothetical protein [Physcisphaerales bacterium]	Physcisphaerales bacterium	85.9	318	35%	8e-15	69.49%	305	MBY0263617.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	151	35%	9e-15	42.25%	560	HAT1184756.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	87.4	36%	9e-15	41.78%	509	HCG2979289.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	87.4	38%	9e-15	41.78%	528	HCG2973818.1
<input checked="" type="checkbox"/> rhage tail repeat-like family [Trichomonas vaginalis G3]	Trichomonas vaginalis G3	85.5	162	40%	9e-15	33.07%	286	XP_051082150.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	87.4	36%	9e-15	41.78%	520	HCG2992671.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	87.4	36%	1e-14	41.78%	523	HCH2243518.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	87.4	36%	1e-14	41.78%	513	HCG2995325.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	151	35%	1e-14	42.25%	560	HAT1426081.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.4	151	35%	1e-14	42.25%	560	HAT6623892.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	87.0	87.0	36%	1e-14	43.84%	468	HCG3003954.1
<input checked="" type="checkbox"/> hypothetical protein [Planctomycetaceae bacterium]	Planctomycetaceae bacterium	87.4	169	50%	1e-14	40.74%	796	MCD8138439.1
<input checked="" type="checkbox"/> TPA_tail fiber protein [Escherichia coli]	Escherichia coli	86.3	86.3	48%	1e-14	29.72%	383	HDP5136300.1
<input checked="" type="checkbox"/> short-chain fatty acid transporter [Escherichia coli]	Escherichia coli	85.9	85.9	48%	1e-14	30.42%	352	MZW84979.1
<input checked="" type="checkbox"/> TPA_minor tail protein [Caudoviricoetes so.]	Caudoviricoetes so.	81.6	233	41%	2e-14	49.04%	148	DAI53149.1
<input checked="" type="checkbox"/> hypothetical protein TVAG_583930 [Trichomonas vaginalis G3]	Trichomonas vaginalis G3	84.7	236	40%	2e-14	32.28%	291	EAX68211.1
<input checked="" type="checkbox"/> rhage tail fiber protein [Pseudomonas aeruginosa]	Pseudomonas aeruginosa	86.7	231	38%	2e-14	64.38%	528	WP_311056509.1
<input checked="" type="checkbox"/> rhage tail protein [Escherichia coli]	Escherichia coli	86.7	86.7	48%	2e-14	29.72%	611	WP_119174799.1
<input checked="" type="checkbox"/> TPA_minor tail protein [Caudoviricoetes so.]	Caudoviricoetes so.	86.3	440	53%	2e-14	35.60%	543	DAM58653.1
<input checked="" type="checkbox"/> hypothetical protein [Nitrate reductase]	Nitrate reductase	86.3	150	68%	2e-14	34.33%	471	MCB1423177.1
<input checked="" type="checkbox"/> hypothetical protein TVAG_399810 [Trichomonas vaginalis G3]	Trichomonas vaginalis G3	85.9	299	45%	2e-14	35.54%	459	EAX79942.1
<input checked="" type="checkbox"/> rhage tail protein [Escherichia coli]	Escherichia coli	86.3	86.3	48%	2e-14	30.42%	594	WP_265934434.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Alphaproteobacteria bacterium]	Alphaproteobacteria bacterium	86.3	233	39%	2e-14	47.93%	615	HRY77510.1
<input checked="" type="checkbox"/> TPA_hypothetical protein [Alphaproteobacteria bacterium]	Alphaproteobacteria bacterium	85.9	158	51%	3e-14	40.24%	588	HEX2753421.1

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

NON

DECISION:	OUI
------------------	------------

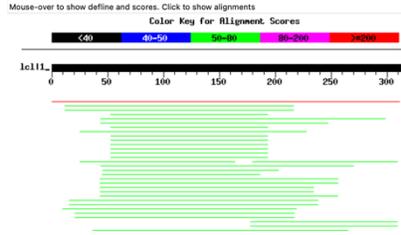
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																																		
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :10390 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :10390																																																																																																																																																																		
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <thead> <tr> <th>Start</th> <th>Raw SD</th> <th>Genomic Z Value</th> <th>Spacer 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>-1,605</td><td>3,202</td><td>6</td><td>-2,906</td><td>GGACTCGTAAGAAAGAGGAGCA</td><td>ATG</td><td>10390</td><td>939</td></tr> <tr><td>2</td><td>-5,552</td><td>1,275</td><td>15</td><td>-6,707</td><td>GTACCCCGAAACCACCGACCTG</td><td>GTG</td><td>10441</td><td>888</td></tr> <tr><td>3</td><td>-4,141</td><td>1,964</td><td>13</td><td>-5,211</td><td>AGCGCTAGCGGGTAAGGCTCCG</td><td>TTG</td><td>10558</td><td>771</td></tr> <tr><td>4</td><td>-5,336</td><td>1,381</td><td>10</td><td>-6,295</td><td>GCATACGCACCCGCGAGTCTCAG</td><td>GTG</td><td>10585</td><td>744</td></tr> <tr><td>5</td><td>-5,336</td><td>1,381</td><td>10</td><td>-6,295</td><td>GCATACGCACCCGCGAGTCTCAG</td><td>GTG</td><td>10651</td><td>678</td></tr> <tr><td>6</td><td>-5,340</td><td>1,379</td><td>16</td><td>-6,562</td><td>GCATACGCATAGCACTTCGCAG</td><td>GTG</td><td>10717</td><td>612</td></tr> <tr><td>7</td><td>-3,230</td><td>2,409</td><td>10</td><td>-4,188</td><td>GGTGACCGGCTGGATAGCGCG</td><td>TTG</td><td>10738</td><td>591</td></tr> <tr><td>8</td><td>-5,336</td><td>1,381</td><td>10</td><td>-6,295</td><td>ACATACGCACCCGCGAGTCTCAG</td><td>GTG</td><td>10783</td><td>546</td></tr> <tr><td>9</td><td>-5,340</td><td>1,379</td><td>16</td><td>-6,562</td><td>GCATACGCATAGCACTTCGCAG</td><td>GTG</td><td>10849</td><td>480</td></tr> <tr><td>10</td><td>-5,914</td><td>1,099</td><td>8</td><td>-7,039</td><td>CCTCCCCACCCTAAACACCCGC</td><td>GTG</td><td>10909</td><td>420</td></tr> <tr><td>11</td><td>-4,859</td><td>1,614</td><td>12</td><td>-5,905</td><td>AGATACTTCGGTAGGCACCCGC</td><td>GTG</td><td>10957</td><td>372</td></tr> <tr><td>12</td><td>-4,285</td><td>1,894</td><td>9</td><td>-5,330</td><td>COGOGITTTGCAGGCCAACACT</td><td>ATG</td><td>10975</td><td>354</td></tr> <tr><td>13</td><td>-3,456</td><td>2,298</td><td>9</td><td>-4,502</td><td>GTCTATCTCTAGTGGGAAAATT</td><td>TTG</td><td>11050</td><td>279</td></tr> <tr><td>14</td><td>-5,400</td><td>1,350</td><td>14</td><td>-6,497</td><td>CACCTCCGGTTCGATTGAGTT</td><td>GTG</td><td>11098</td><td>231</td></tr> <tr><td>15</td><td>-3,514</td><td>2,270</td><td>16</td><td>-4,736</td><td>ACCCCTGGGGTCTAATTGGGCA</td><td>ATG</td><td>11146</td><td>183</td></tr> <tr><td>16</td><td>-4,609</td><td>1,736</td><td>16</td><td>-5,831</td><td>TACGCCAGATCGAAGTGTCCG</td><td>TTG</td><td>11206</td><td>123</td></tr> <tr><td>17</td><td>-5,016</td><td>1,537</td><td>15</td><td>-6,171</td><td>ATTATTGGGACATAAATTAT</td><td>GTG</td><td>11281</td><td>48</td></tr> </tbody> </table>	Start	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length	1	-1,605	3,202	6	-2,906	GGACTCGTAAGAAAGAGGAGCA	ATG	10390	939	2	-5,552	1,275	15	-6,707	GTACCCCGAAACCACCGACCTG	GTG	10441	888	3	-4,141	1,964	13	-5,211	AGCGCTAGCGGGTAAGGCTCCG	TTG	10558	771	4	-5,336	1,381	10	-6,295	GCATACGCACCCGCGAGTCTCAG	GTG	10585	744	5	-5,336	1,381	10	-6,295	GCATACGCACCCGCGAGTCTCAG	GTG	10651	678	6	-5,340	1,379	16	-6,562	GCATACGCATAGCACTTCGCAG	GTG	10717	612	7	-3,230	2,409	10	-4,188	GGTGACCGGCTGGATAGCGCG	TTG	10738	591	8	-5,336	1,381	10	-6,295	ACATACGCACCCGCGAGTCTCAG	GTG	10783	546	9	-5,340	1,379	16	-6,562	GCATACGCATAGCACTTCGCAG	GTG	10849	480	10	-5,914	1,099	8	-7,039	CCTCCCCACCCTAAACACCCGC	GTG	10909	420	11	-4,859	1,614	12	-5,905	AGATACTTCGGTAGGCACCCGC	GTG	10957	372	12	-4,285	1,894	9	-5,330	COGOGITTTGCAGGCCAACACT	ATG	10975	354	13	-3,456	2,298	9	-4,502	GTCTATCTCTAGTGGGAAAATT	TTG	11050	279	14	-5,400	1,350	14	-6,497	CACCTCCGGTTCGATTGAGTT	GTG	11098	231	15	-3,514	2,270	16	-4,736	ACCCCTGGGGTCTAATTGGGCA	ATG	11146	183	16	-4,609	1,736	16	-5,831	TACGCCAGATCGAAGTGTCCG	TTG	11206	123	17	-5,016	1,537	15	-6,171	ATTATTGGGACATAAATTAT	GTG	11281	48
Start	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length																																																																																																																																																											
1	-1,605	3,202	6	-2,906	GGACTCGTAAGAAAGAGGAGCA	ATG	10390	939																																																																																																																																																											
2	-5,552	1,275	15	-6,707	GTACCCCGAAACCACCGACCTG	GTG	10441	888																																																																																																																																																											
3	-4,141	1,964	13	-5,211	AGCGCTAGCGGGTAAGGCTCCG	TTG	10558	771																																																																																																																																																											
4	-5,336	1,381	10	-6,295	GCATACGCACCCGCGAGTCTCAG	GTG	10585	744																																																																																																																																																											
5	-5,336	1,381	10	-6,295	GCATACGCACCCGCGAGTCTCAG	GTG	10651	678																																																																																																																																																											
6	-5,340	1,379	16	-6,562	GCATACGCATAGCACTTCGCAG	GTG	10717	612																																																																																																																																																											
7	-3,230	2,409	10	-4,188	GGTGACCGGCTGGATAGCGCG	TTG	10738	591																																																																																																																																																											
8	-5,336	1,381	10	-6,295	ACATACGCACCCGCGAGTCTCAG	GTG	10783	546																																																																																																																																																											
9	-5,340	1,379	16	-6,562	GCATACGCATAGCACTTCGCAG	GTG	10849	480																																																																																																																																																											
10	-5,914	1,099	8	-7,039	CCTCCCCACCCTAAACACCCGC	GTG	10909	420																																																																																																																																																											
11	-4,859	1,614	12	-5,905	AGATACTTCGGTAGGCACCCGC	GTG	10957	372																																																																																																																																																											
12	-4,285	1,894	9	-5,330	COGOGITTTGCAGGCCAACACT	ATG	10975	354																																																																																																																																																											
13	-3,456	2,298	9	-4,502	GTCTATCTCTAGTGGGAAAATT	TTG	11050	279																																																																																																																																																											
14	-5,400	1,350	14	-6,497	CACCTCCGGTTCGATTGAGTT	GTG	11098	231																																																																																																																																																											
15	-3,514	2,270	16	-4,736	ACCCCTGGGGTCTAATTGGGCA	ATG	11146	183																																																																																																																																																											
16	-4,609	1,736	16	-5,831	TACGCCAGATCGAAGTGTCCG	TTG	11206	123																																																																																																																																																											
17	-5,016	1,537	15	-6,171	ATTATTGGGACATAAATTAT	GTG	11281	48																																																																																																																																																											
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>ORF le plus long (avec un ATG)</i> <i>Score de -2</i>																																																																																																																																																																		
Est-ce que le start est conservé chez les homologues voir Starterator ?	—																																																																																																																																																																		

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: right;"> Download GenPept Graphics Sort by: E value </p> <p>hypothetical protein [Corynebacterium aquilae]</p> <p>Sequence ID: WP_075725710.1 Length: 666 Number of Matches: 3</p> <p style="text-align: center;"> See 1 more title(s) See all Identical Proteins(IPG) </p> <hr/> <p> Range 1: 141 to 296 GenPept Graphics ▼ Next Match ▲ Previous Match </p> <table border="1" style="width: 100%; border-collapse: collapse; font-size: small;"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>150 bits(380)</td> <td>1e-36</td> <td>Compositional matrix adjust.</td> <td>87/156(56%)</td> <td>99/156(63%)</td> <td>13/156(8%)</td> </tr> </tbody> </table> <pre style="font-family: monospace; font-size: x-small; margin-top: 5px;"> Query 31 ELIDTFLYEQEK---QTAIDAALAGKAPLSHTHPQSQVTGLDAALAGKAPLSHTHPQS 86 LID L + + + + + A + GKA H H VTGLDAAL GKA SHTHP S Sbjct 141 RLIDEGLDQGQVRVKHLNSEVKALIDGKAGKKQHMGDVTGLDAALGKADKSHTHPMS 200 Query 87 QVTGLDAALAGKAPLSHTSTSQVTGLDSALQ-----GKAPLSHTHPQSQVTGLDA 137 QV GLD ALAGKAP +HTHST+QVTGLD+AL+ GKA SHTH SQVTGLDA Sbjct 201 QVVGLDDALAGKAPKTHHSTAQVTGLDAALKAKADVSALGGKAAKSHTHAMSQVTGLDA 260 Query 138 ALAGKAPLSHTSTSQVTGLDAALAGKAPLPTLNTR 173 AL GKA SHTHS V+GLD AL GKA L + Sbjct 261 ALGGKAAKSHTHSMGDVSGLDLALRGKADTADLQRK 296 Range 2: 181 to 325 GenPept Graphics ▼ Next Match ▲ Previous Match ⬆ First Match </pre> <p>Remarque : si vous proposez plusieurs starts, indiquez pourquoi et fournissez l'information pour chaque start.</p> </div>	Score	Expect	Method	Identities	Positives	Gaps	150 bits(380)	1e-36	Compositional matrix adjust.	87/156(56%)	99/156(63%)	13/156(8%)
Score	Expect	Method	Identities	Positives	Gaps								
150 bits(380)	1e-36	Compositional matrix adjust.	87/156(56%)	99/156(63%)	13/156(8%)								
<p>DECISION:</p>	<p>10390</p>												

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

<p>Collection des éléments de réponse</p>	<p>Rationnelle</p>
<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>



Sequences producing significant alignments:

Sequence	Score	E Value
Cyran08_Draft_15, function unknown, 312	522	e-178
Stomaxpedion_24, minor tail protein, 802	59	1e-11
RedWattleHog_29, minor tail protein, 802	69	1e-11
ChiannaKitsuine_25, minor tail protein, 517	67	7e-11
Nannuculus_Draft_43, function unknown, 450	64	6e-10
BrubMoment_46, function unknown, 450	63	1e-09
Hanen_29, minor tail protein, 551	62	1e-09
DrYang_27, minor tail protein, 770	62	2e-09
Schomber_28, minor tail protein, 551	62	2e-09
Fakusa_Draft_28, function unknown, 551	62	2e-09
Kabocho_30, minor tail protein, 551	62	2e-09
Gray_29, minor tail protein, 551	62	2e-09
Chidiebere_29, minor tail protein, 551	62	2e-09
Aloki_Draft_28, function unknown, 551	62	2e-09
Whack_28, minor tail protein, 475	60	5e-09
Abba_23, minor tail protein, 591	60	5e-09
Altadena_23, minor tail protein, 600	60	7e-09

*** nr :

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

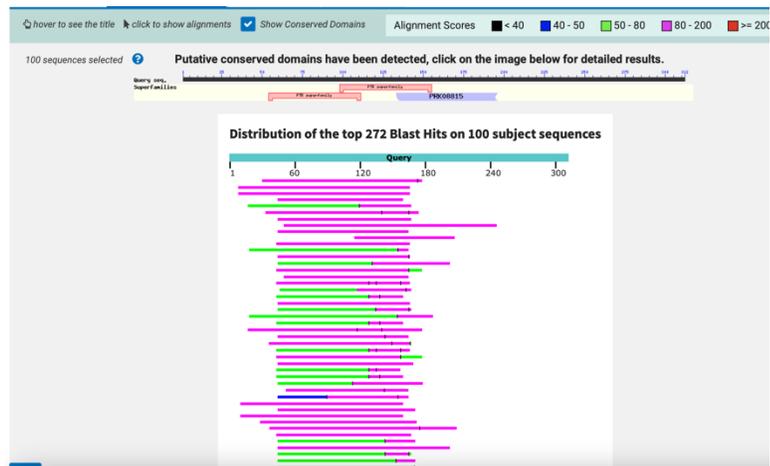
Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments

Download | Select columns | Show 100

select all 100 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium aquilae]	Corynebacterium aquilae	150	394	47%	1e-36	55.77%	666	WP_075725710.1
<input checked="" type="checkbox"/> hypothetical protein [Clostridia bacterium]	Clostridia bacterium	132	298	50%	1e-31	51.23%	327	MBR5768827.1
<input checked="" type="checkbox"/> hypothetical protein [Clostridia bacterium]	Clostridia bacterium	132	359	50%	1e-31	51.23%	328	MBQ4583800.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Pseudomonas aeruginosa]	Pseudomonas aeruginosa	124	197	37%	1e-27	56.90%	481	HCF5853879.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Alloshingonocella sp.]	Alloshingonocella sp.	123	501	48%	3e-27	53.72%	537	HZF95778.1
<input checked="" type="checkbox"/> phage tail repeat domain-containing protein [Paludibacteraceae bacterium]	Paludibacteraceae bacterium	123	675	45%	7e-27	53.72%	1040	MCQ2313286.1
<input checked="" type="checkbox"/> hypothetical protein [Prescottella equi]	Prescottella equi	119	321	39%	2e-26	57.72%	394	WP_286460814.1
<input checked="" type="checkbox"/> hypothetical protein [Parafrankia sp. EUN1f]	Parafrankia sp. EUN1f	119	119	62%	4e-26	45.05%	423	WP_006545789.1
<input checked="" type="checkbox"/> phage tail protein [Mannheimia pemigra]	Mannheimia pemigra	118	223	38%	4e-25	52.89%	790	WP_176808477.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium glutamicum]	Corynebacterium glutamicum	115	115	29%	9e-25	62.37%	415	WP_211438912.1
<input checked="" type="checkbox"/> phage tail fiber protein [Stutzerimonas stutzeri]	Stutzerimonas stutzeri	114	502	39%	7e-24	55.74%	623	WP_020699854.1
<input checked="" type="checkbox"/> TPA, minor tail protein [Caudoviricetes sp.]	Caudoviricetes sp.	114	271	47%	7e-24	55.96%	804	DAW45625.1
<input checked="" type="checkbox"/> TPA, hypothetical protein [Alloshingonocella sp.]	Alloshingonocella sp.	113	289	39%	9e-24	49.30%	523	HEV2865932.1
<input checked="" type="checkbox"/> hypothetical protein [Thalassovita libralis]	Thalassovita libralis	108	180	50%	3e-22	45.28%	495	WP_142493021.1



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 $\geq 90\%$ et une couverture acceptable ?

Pfam : PTR ; Phage tail repeat like
 Cd proba de 75 % : RBP-H; Head domain of virus receptor-binding proteins (RBP). Virus receptor-binding proteins (RBPs) are found in lactoco
 PDB : Major structural protein 1; bacteriophages, Lactococcus lactis, Siphoviridae, receptor binding protein, single-chain nan
 Uniport : RBP_BPLSK Receptor binding protein OS=Lactococcus phage SK1 OX=31532 PE=3 SV=1

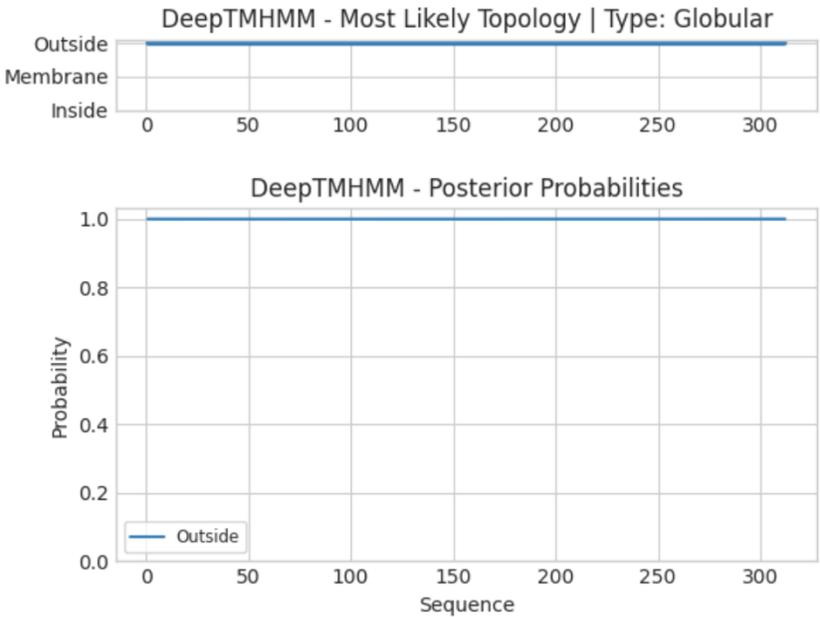
<input type="checkbox"/>	1	PF12789.11	; PTR ; Phage tail repeat like	97.95	0.00013	55.49	2.3	57	60
<input type="checkbox"/>	2	PF12789.11	; PTR ; Phage tail repeat like	97.6	0.00013	50.43	2.9	58	60
<input type="checkbox"/>	3	5E7F_H	Major structural protein 1; bacteriophages, Lactococcus lactis, Siphoviridae, receptor binding protein, single-chain nan	97.38	0.0034	54.48	9.3	84	174
<input type="checkbox"/>	4	2FSD_A	putative baseplate protein; Phage, Lactococcus lactis, receptor binding protein, head domain, VIRAL PROTEIN; 2.3A {unide	96.32	0.072	44.95	8.6	94	142
<input type="checkbox"/>	5	2FOC_C	Phage tp901-1 ORF49 (BPP); Beta-barrel, beta prism, 3 helix parallel bundle, VIRAL PROTEIN; HET: GOL; 1.65A {Lactococcus	96.32	0.085	46.62	9.4	81	191
<input type="checkbox"/>	6	3DAQ_C	Cleaved chimeric receptor binding protein from bacteriophages TP901-1 and p2; lactococcal phage p2, lactococcal phage TP	96.31	0.11	42.45	9.7	82	136
<input type="checkbox"/>	7	4HEM_A	BPP; Alpha-beta, Phage Receptor Binding Protein, Llma VHH, viral protein, CELL ADHESION-IMMUNE SYSTEM complex; 1.65A {L	96.2	0.1	44.52	9.1	80	163
<input type="checkbox"/>	8	1ZRU_A	lactophage p2 receptor binding protein; 3 domains: BETA BARREL, BETA PRISM, BETA BARREL, Structural Genomics, Structural	96.04	0.13	48.07	9.7	80	264
<input type="checkbox"/>	9	PF08931.14	; Caudo_bapla_RBP ; Lactophage receptor-binding protein C-terminal head domain	95.73	0.26	37.96	8.8	79	101
<input type="checkbox"/>	10	4IOS_B	BPP; all beta, jelly roll motif, receptor binding protein, neutralizing llama antibody domain, CELL ADHESION-IMMUNE SYST	95.49	0.49	35.89	9.4	80	100

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".

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<h3>DeepTMHMM - Predictions</h3> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p> 
<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>Minor tail protein</p>

Student Gene Annotation Worksheet

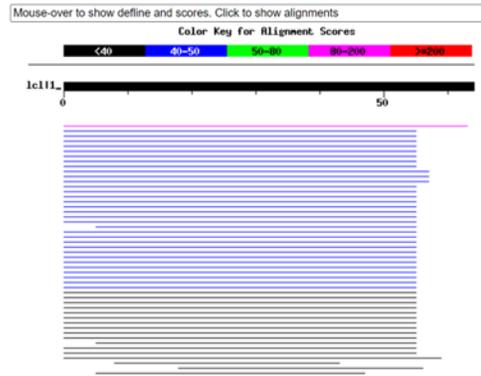
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	16
Coordonnées du Stop	11537
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 16
Coordonnées du Start retenu	11343

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 style="font-size: small;">anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCC GCGCCAGCCC), Order 2, Window 96, Step 12, 7/9</p> <p>The figure displays a GeneMark hmh prediction plot for the anoPS complete sequence. It consists of two main sections: 'Direct Sequence' and 'Complementary Sequence', each with three sub-plots. The x-axis represents the Nucleotide Position from 10800 to 12400. The y-axis represents the prediction score from 0.0 to 1.0. The top plot in each section shows a line graph with peaks indicating predicted coding regions. The middle plot shows a series of horizontal bars representing predicted exons and introns. The bottom plot shows a series of vertical bars representing predicted exons. A thick black bar highlights a specific region around nucleotide 11500, which is identified as a GeneMark hmh prediction. The plot also shows a 14-base 3' overhang (TCCC GCGCCAGCCC) at the end of the sequence.</p>

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



Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyranopS_Draft_16, function unknown, 64	128	4e-30
Litotes_3, function unknown, 58	49	6e-06
Wheelbite_29, function unknown, 59	48	7e-06
Wayne_3, function unknown, 58	48	7e-06
Laroye_30, function unknown, 59	48	7e-06
Partycup_Draft_3, function unknown, 58	47	1e-05
Canowicakte_3, function unknown, 58	47	1e-05
AppleCider_3, function unknown, 58	47	1e-05
CallieOMalley_3, function unknown, 58	47	2e-05
Zion_28, function unknown, 62	46	3e-05
PotatoChin_28, function unknown, 62	46	3e-05

select all 15 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 [Corynebacterium glutamicum]	Corynebacterium glutamicum	79.7	79.7	93%	1e-17	63.33%	67	WP_211438911.1
<input checked="" type="checkbox"/> hypothetical protein KDJ02_pe03 [Arthrobacter cbae Litotes]	Arthrobacter cbae Litotes	47.0	47.0	87%	8e-05	37.50%	58	YP_010050172.1
<input checked="" type="checkbox"/> hypothetical protein SEA_APPLECIDER_3 [Arthrobacter cbae AppleCider]	Arthrobacter cbae AppleCider	45.8	45.8	87%	3e-04	37.50%	58	QHR847173.1
<input checked="" type="checkbox"/> hypothetical protein FDH63_pe03 [Arthrobacter cbae Wayne]	Arthrobacter cbae Wayne	45.8	45.8	87%	3e-04	37.50%	58	YP_009602930.1
<input checked="" type="checkbox"/> hypothetical protein SEA_CANOWICAKTE_3 [Arthrobacter cbae Canowicakte]	Arthrobacter cbae Canowicakte	45.4	45.4	87%	4e-04	37.50%	58	ASR83239.1
<input checked="" type="checkbox"/> hypothetical protein SEA_CALLIEOMALLEY_3 [Arthrobacter cbae CallieOMalley]	Arthrobacter cbae CallieOMalley	45.1	45.1	87%	6e-04	37.50%	58	AZF97630.1
<input checked="" type="checkbox"/> hypothetical protein FDH59_pe03 [Arthrobacter cbae Joann]	Arthrobacter cbae Joann	44.7	44.7	87%	7e-04	37.50%	58	YP_009602683.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hylobafts]	Corynebacterium hylobafts	45.1	45.1	84%	8e-04	40.74%	79	WP_126121245.1
<input checked="" type="checkbox"/> hypothetical protein EDJ12_pe28 [Corynebacterium cbae Zion]	Corynebacterium cbae Zion	44.7	44.7	90%	8e-04	37.93%	62	YP_009620373.1
<input checked="" type="checkbox"/> hypothetical protein [Brachybacterium sp. NBEC-018]	Brachybacterium sp. NBEC-018	44.3	44.3	82%	0.001	43.40%	78	WP_259360684.1
<input checked="" type="checkbox"/> hypothetical protein SEA_LAKSHMI_3 [Arthrobacter cbae Lakshmi]	Arthrobacter cbae Lakshmi	43.5	43.5	87%	0.002	37.50%	58	WV085560.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium lemuis]	Corynebacterium lemuis	43.5	43.5	84%	0.003	40.74%	74	WP_269427917.1
<input checked="" type="checkbox"/> hypothetical protein KD97_pe03 [Arthrobacter cbae GreenHearts]	Arthrobacter cbae GreenHearts	42.0	42.0	87%	0.009	37.50%	58	YP_010049923.1
<input checked="" type="checkbox"/> hypothetical protein [Pseudoclavibacter sp. CFCC 14310]	Pseudoclavibacter sp. CFCC 14310	42.0	42.0	87%	0.009	35.71%	59	WP_158037309.1
<input checked="" type="checkbox"/> hypothetical protein SEA_VALLEJO_3 [Arthrobacter cbae Vallejo]	Arthrobacter cbae Vallejo	40.0	40.0	87%	0.045	33.93%	58	AOT24093.1

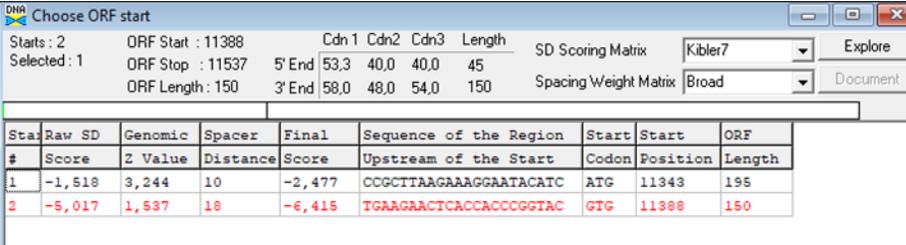
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

DECISION:

OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																											
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :11343 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :11343																											
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <thead> <tr> <th>Start</th> <th>Raw SD</th> <th>Genomic Z Value</th> <th>Spacer 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>-1,518</td> <td>3,244</td> <td>10</td> <td>-2,477</td> <td>CGCTTAAGAAAGGAATACATC</td> <td>ATG</td> <td>11343</td> <td>195</td> </tr> <tr> <td>2</td> <td>-5,017</td> <td>1,537</td> <td>18</td> <td>-6,415</td> <td>TGAAGAACTCACCACCCGGTAC</td> <td>GTG</td> <td>11388</td> <td>150</td> </tr> </tbody> </table>	Start	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length	1	-1,518	3,244	10	-2,477	CGCTTAAGAAAGGAATACATC	ATG	11343	195	2	-5,017	1,537	18	-6,415	TGAAGAACTCACCACCCGGTAC	GTG	11388	150
Start	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length																				
1	-1,518	3,244	10	-2,477	CGCTTAAGAAAGGAATACATC	ATG	11343	195																				
2	-5,017	1,537	18	-6,415	TGAAGAACTCACCACCCGGTAC	GTG	11388	150																				
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>Plus long ORF</i></p> <p>ATG</p> <p><i>Score faible : -2.4</i></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>hypothetical protein [Corynebacterium glutamicum] Sequence ID: WP_211438911.1 Length: 67 Number of Matches: 1</p> <p>Range 1: 1 to 60 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>79.7 bits(195)</td> <td>1e-17</td> <td>Compositional matrix adjust.</td> <td>38/60(63%)</td> <td>48/60(80%)</td> <td>0/60(0%)</td> </tr> </tbody> </table> <p>Query 1 MSIRDLSDEELTRYVEANEELSRRRLSEIPTI EIRLAQQWRDNGGDSAIIDAIHTTP 60 MSIRDLSDEEL RY+EAN+E+RRRL+ IP EI ELAQQ R+ G E++ +IDA+ P Sbjct 1 MSIRDLSDEELNARYIEANDEIARRRLAAI PAETSELAQQREIGVEEATMIDAVTRAP 60</p> <p>Download GenPept Graphics Next Previous Descriptions</p> <p>hypothetical protein KDJ02_gp03 [Arthrobacter phage Litotes] Sequence ID: YP_010050172.1 Length: 58 Number of Matches: 1</p> <p>See 1 more title(s) See all Identical Proteins(IPG)</p> <p>Range 1: 1 to 56 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>47.0 bits(110)</td> <td>8e-05</td> <td>Compositional matrix adjust.</td> <td>21/56(38%)</td> <td>34/56(60%)</td> <td>0/56(0%)</td> </tr> </tbody> </table> <p>Query 1 MSIRDLSDEELTRYVEANEELSRRRLSEIPTI EIRLAQQWRDNGGDSAIIDAI 56 M RD+SDE+L E E RR R+ IP+++ LA Q+R+ GG+ + ++ A+ Sbjct 1 MDRFMSDEQLDNLHNEVNEKERERIKNIPSQVAMLATQFREGGGQTELVAAV 56</p> <p>Related Information Gene - associated gene details Identical Proteins - Identical proteins to YP_010050172.1</p>	Score	Expect	Method	Identities	Positives	Gaps	79.7 bits(195)	1e-17	Compositional matrix adjust.	38/60(63%)	48/60(80%)	0/60(0%)	Score	Expect	Method	Identities	Positives	Gaps	47.0 bits(110)	8e-05	Compositional matrix adjust.	21/56(38%)	34/56(60%)	0/56(0%)			
Score	Expect	Method	Identities	Positives	Gaps																							
79.7 bits(195)	1e-17	Compositional matrix adjust.	38/60(63%)	48/60(80%)	0/60(0%)																							
Score	Expect	Method	Identities	Positives	Gaps																							
47.0 bits(110)	8e-05	Compositional matrix adjust.	21/56(38%)	34/56(60%)	0/56(0%)																							
DECISION:	11343																											

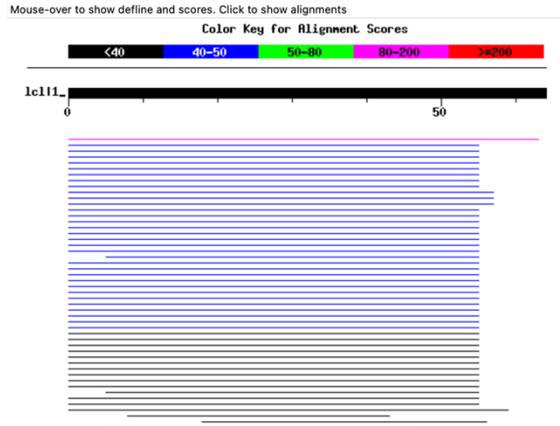
Décision #3 : Quelle est la fonction de la protéine putative ?

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

***** PhagesDB :**



Sequences producing significant alignments:

	Score (bits)	E Value
Cyranops_Draft_16, function unknown, 64	128	4e-30
Litotes_3, function unknown, 58	42	6e-06
Wheelbite_29, function unknown, 59	48	7e-06
Wayne_3, function unknown, 58	48	7e-06
Laroye_30, function unknown, 59	48	7e-06
PartyCup_Draft_3, function unknown, 58	47	1e-05
Canowicakte_3, function unknown, 58	47	1e-05
AppleCider_3, function unknown, 58	47	1e-05
CallieOMalley_3, function unknown, 58	47	2e-05
Zion_28, function unknown, 62	46	3e-05

***** nr :**

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

15 sequences selected No putative conserved domains have been detected

Distribution of the top 15 Blast Hits on 15 subject sequences

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 15 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 [Corynebacterium glutamicum]	Corynebacterium glutamicum	79.7	79.7	93%	1e-17	63.33%	67	WP_2114388911.1
<input checked="" type="checkbox"/> hypothetical protein KDJ02_gp03 [Arthrobacter phage Litotes]	Arthrobacter phage Litotes	47.0	47.0	87%	8e-05	37.50%	58	YP_010050172.1
<input checked="" type="checkbox"/> hypothetical protein SEA_APPLECIDER_3 [Arthrobacter phage AppleCider]	Arthrobacter phage AppleCider	45.8	45.8	87%	3e-04	37.50%	58	QHR847173.1
<input checked="" type="checkbox"/> hypothetical protein FDH63_gp03 [Arthrobacter phage Wayne]	Arthrobacter phage Wayne	45.8	45.8	87%	3e-04	37.50%	58	YP_009602830.1
<input checked="" type="checkbox"/> hypothetical protein SEA_CANOWICAKTE_3 [Arthrobacter phage Canowicakte]	Arthrobacter phage Canowicakte	45.4	45.4	87%	4e-04	37.50%	58	ASR83239.1
<input checked="" type="checkbox"/> hypothetical protein SEA_CALLIEOMALLEY_3 [Arthrobacter phage CallieOMalley]	Arthrobacter phage CallieOMalley	45.1	45.1	87%	6e-04	37.50%	58	AZF97639.1
<input checked="" type="checkbox"/> hypothetical protein FDH59_gp03 [Arthrobacter phage Joann]	Arthrobacter phage Joann	44.7	44.7	87%	7e-04	37.50%	58	YP_009602683.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium hylobatis]	Corynebacterium hylobatis	45.1	45.1	84%	8e-04	40.74%	79	WP_128121745.1
<input checked="" type="checkbox"/> hypothetical protein FDJ12_gp28 [Corynebacterium phage Zion]	Corynebacterium phage Zion	44.7	44.7	90%	8e-04	37.93%	62	YP_009620373.1
<input checked="" type="checkbox"/> hypothetical protein [Brachybaculum sp. NBEC-018]	Brachybaculum sp. NBEC-018	44.3	44.3	82%	0.001	43.40%	78	WP_259360684.1
<input checked="" type="checkbox"/> hypothetical protein SEA_LAKSHMI_3 [Arthrobacter phage Lakshmi]	Arthrobacter phage Lakshmi	43.5	43.5	87%	0.002	37.50%	58	WKW85560.1
<input checked="" type="checkbox"/> hypothetical protein [Corynebacterium lemuris]	Corynebacterium lemuris	43.5	43.5	84%	0.003	40.74%	74	WP_259427917.1
<input checked="" type="checkbox"/> hypothetical protein KD197_gp03 [Arthrobacter phage GreenHearts]	Arthrobacter phage GreenHearts	42.0	42.0	87%	0.009	37.50%	58	YP_010049923.1
<input checked="" type="checkbox"/> hypothetical protein [Pseudoclavibacter sp. CFCC 14310]	Pseudoclavibacter sp. CFCC 14310	42.0	42.0	87%	0.009	35.71%	59	WP_158037309.1
<input checked="" type="checkbox"/> hypothetical protein SEA_VALLEJO_3 [Arthrobacter phage Vallejo]	Arthrobacter phage Vallejo	40.0	40.0	87%	0.045	33.93%	58	AOT24093.1

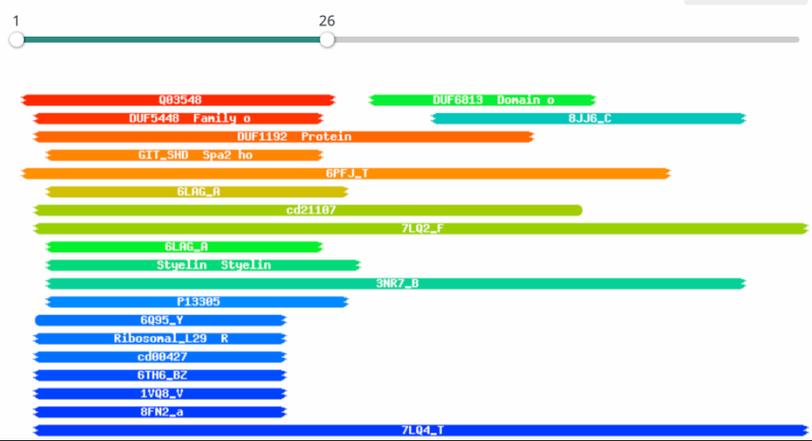
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors

**PFAM : DUF5448 ; Family of unknown function (DUF5448)
Cd proba de 79.61% : RsiG; anti-sigma factor RsiG (AmfC). RsiG is an anti-sigma factor that binds and sequesters the sporulation-specific sig**

d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?

Uniprot : VEFB_BPP22 Eaf protein OS=Salmonella phage P22
 OX=10754 GN=eaf PE=4 SV=3
 PDB : AmfC protein; sigma, anti-sigma, c-di-GMP, developmental switch, TRANSCRIPTION; HET: C2E; 2.08A {Streptomyces venezuelae}

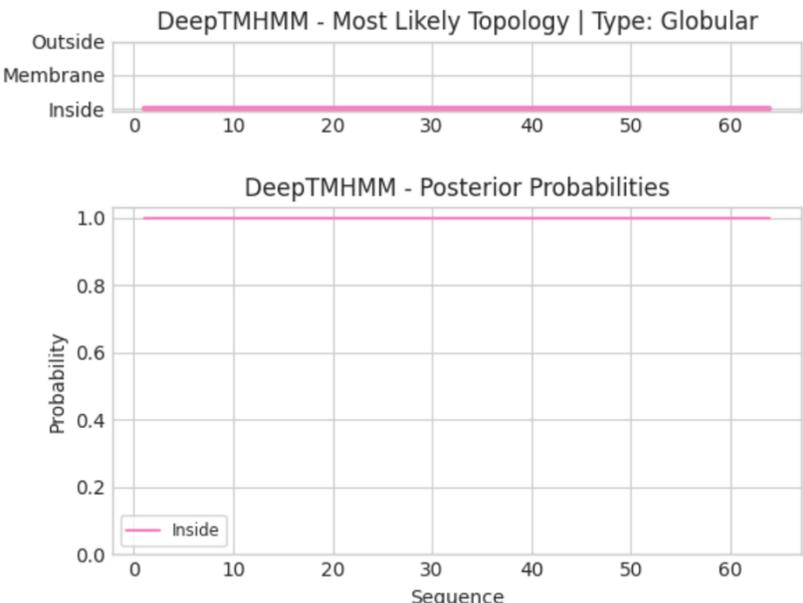
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	1	Q03548 VEFB_BPP22 Eaf protein OS=Salmonella phage P22 OX=10754 GN=eaf PE=4 SV=3	96.72	0.0055	41.69	3.5	26	119
<input type="checkbox"/>	2	PF17526.6 ; DUF5448 ; Family of unknown function (DUF5448)	95.64	0.038	37.78	3.2	24	118
<input type="checkbox"/>	3	PF06698.15 ; DUF1192 ; Protein of unknown function (DUF1192)	91.29	4.2	24.78	6.1	41	61
<input type="checkbox"/>	4	PF08518.15 ; GIT_SHD ; Spa2 homology domain (SHD) of GIT	88.78	2.4	19.72	3.1	23	29
<input type="checkbox"/>	5	6PFJ_T AmfC protein; sigma, anti-sigma, c-di-GMP, developmental switch, TRANSCRIPTION; HET: C2E; 2.08A (Streptomyces venezuelae)	87.8	11	27.56	6.9	53	176
<input type="checkbox"/>	6	6LAG_A Spa2-like protein; SHD, SPA-2, GIT-PIX, polarity, SIGNALING PROTEIN; NMR (Neurospora crassa)	82.48	4.9	28.24	3.3	25	155
<input type="checkbox"/>	7	cd21107 RsiG; anti-sigma factor RsiG (AmfC). RsiG is an anti-sigma factor that binds and sequesters the sporulation-specific sig	79.61	31	24.53	6.3	45	143



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 ?

Gene apres = integrase ?

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>Répondez Oui ou Non. Une fois que vous avez pris une décision quant-à-la fonction du gène, vérifiez la liste des 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.</p>
<p>DECISION:</p>	<p>NKF</p>

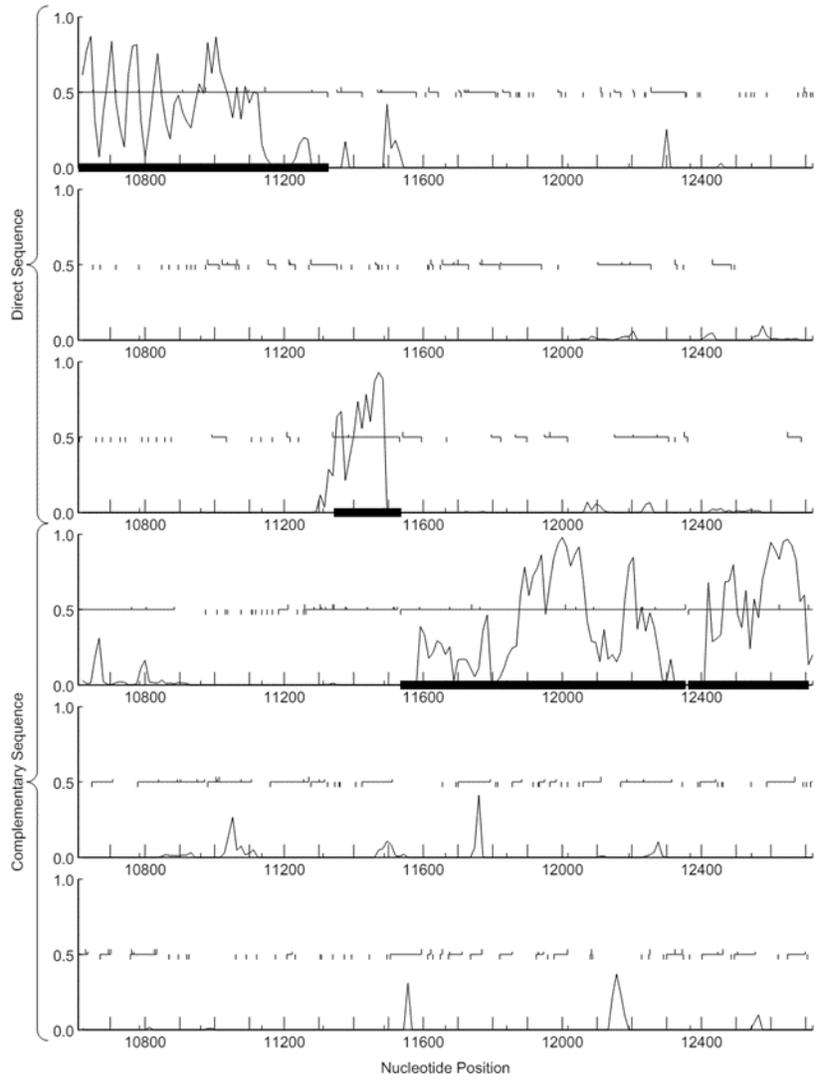
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	17
Coordonnées du Stop	11534
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	Gap 6
Coordonnées du Start retenu	12355
Fonction prédite	Tyrosine integrase

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

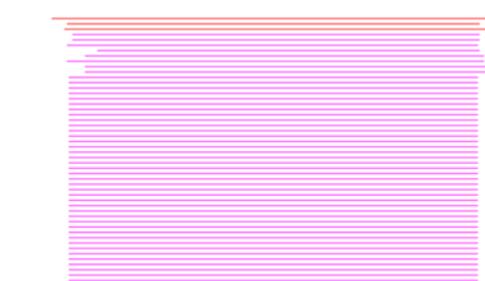
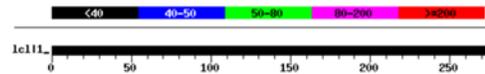
Y-a-t-il des éléments supportant un potentiel codant ?



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

172205 GRU3_17, putative integrase, 297..S= 171 E=2e-42

Color Key for Alignment Scores



Sequences producing significant alignments:

Sequence Name	Score (bits)	E Value
CyranoPS_Draft_17, function unknown, 273	553	e-157
Emperor_19, integrase, 275	211	2e-54
EpicDab_18, integrase, 276	202	6e-54
SallySpecial_17, integrase, 266	192	8e-51
Rahul_19, tyrosine integrase, 275	192	6e-49
Weirdo19_Draft_45, function unknown, 411	191	2e-48
Coeur_19, integrase, 273	187	2e-47
GRU3_17, putative integrase, 297	171	2e-42
Cracklewink_50, tyrosine integrase, 410	170	4e-42
McGonagall_18, integrase, 298	162	9e-42

Sequences producing significant alignments		Download	Select columns	Show				
<input checked="" type="checkbox"/> 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"/> lysine-tyrosine recombinase/integrase [Corynebacterium freiburgense]	<i>Corynebacterium freiburgense</i>	284	284	98%	5e-92	50.00%	272	WP_027011485.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	280	280	95%	2e-90	51.33%	272	WP_205917729.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	280	280	95%	2e-90	51.33%	272	CAB0883992.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	279	279	95%	3e-90	51.71%	272	CAB11005300.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	279	279	95%	3e-90	51.33%	272	CAB0972377.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium hadale]	<i>Corynebacterium hadale</i>	278	278	96%	7e-90	53.79%	272	WP_095275381.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	278	278	95%	8e-90	51.33%	272	WP_088298575.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	278	278	95%	1e-89	50.95%	272	WP_342351201.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	278	278	95%	1e-89	50.95%	272	CAB0565317.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	278	278	95%	1e-89	51.33%	272	CAB0918901.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	277	277	95%	2e-89	51.33%	272	CAB0565583.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	277	277	95%	2e-89	50.95%	272	WP_106361636.1
<input checked="" type="checkbox"/> integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	276	276	95%	5e-89	50.57%	272	CAB0612923.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium oculi]	<i>Corynebacterium oculi</i>	275	275	98%	2e-88	50.56%	270	WP_169786392.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium mastitidis]	<i>Corynebacterium mastitidis</i>	274	274	98%	3e-88	51.30%	269	WP_337891380.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium saquilis]	<i>Corynebacterium saquilis</i>	266	266	96%	3e-85	53.99%	272	WP_259927055.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium hindlerae]	<i>Corynebacterium hindlerae</i>	264	264	99%	2e-84	48.71%	268	WP_182385532.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	260	260	94%	8e-83	48.65%	271	WP_071574084.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium striatum]	<i>Corynebacterium striatum</i>	259	259	96%	3e-82	51.33%	268	WP_284790704.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium dichtheriae]	<i>Corynebacterium dichtheriae</i>	256	256	95%	3e-81	48.85%	271	WP_014317013.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Rhodococcus fascians]	<i>Rhodococcus fascians</i>	189	189	95%	1e-54	42.59%	279	WP_196249166.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	188	188	93%	1e-54	44.14%	273	DA90862.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	188	188	93%	1e-54	44.14%	273	DAK36682.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	187	187	93%	2e-54	44.40%	271	DAQ32442.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	187	187	93%	4e-54	43.97%	271	DAN74143.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	187	187	93%	4e-54	43.75%	273	DAX34272.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Rhodococcus so. 14-2496-1d]	Rhodococcus so. 14-2496-1d	187	187	95%	4e-54	42.21%	282	WP_179277775.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium mastitidis]	<i>Corynebacterium mastitidis</i>	183	183	54%	4e-54	59.46%	149	WP_337890933.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Mycobacteroides abscessus]	<i>Mycobacteroides abscessus</i>	188	188	94%	5e-54	37.31%	303	WP_079630806.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Mycobacteroides franklinii]	<i>Mycobacteroides franklinii</i>	188	188	94%	5e-54	37.69%	302	WP_078336654.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Dermatophilaceae bacterium]	Dermatophilaceae bacterium	187	187	94%	5e-54	42.86%	268	MBP888079.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Rhodococcus so. Leaf278]	Rhodococcus so. Leaf278	187	187	93%	7e-54	42.02%	282	WP_162248372.1
<input checked="" type="checkbox"/> hypotheical protein CH296_18965 [Rhodococcus so. 14-2496-1d]	Rhodococcus so. 14-2496-1d	187	187	95%	7e-54	42.21%	299	OZF28235.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Schaalia hyvoainalis]	<i>Schaalia hyvoainalis</i>	185	185	91%	1e-53	42.46%	254	WP_320772312.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Mycobacteroides abscessus]	<i>Mycobacteroides abscessus</i>	186	186	94%	2e-53	36.19%	283	WP_005121816.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	185	185	97%	2e-53	41.57%	259	DAL07361.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Mycobacteroides abscessus]	<i>Mycobacteroides abscessus</i>	185	185	94%	2e-53	36.19%	283	WP_005130835.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Dietzia alimentaria]	<i>Dietzia alimentaria</i>	185	185	93%	3e-53	43.36%	279	WP_158217841.1
<input checked="" type="checkbox"/> lysine integrase [Gordonia chajae Rahul]	<i>Gordonia chajae Rahul</i>	184	184	93%	4e-53	42.02%	275	WNO26681.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Corynebacterium procioum]	<i>Corynebacterium procioum</i>	181	181	63%	5e-53	56.07%	175	WP_284584211.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Mycobacteroides abscessus]	<i>Mycobacteroides abscessus</i>	185	185	94%	6e-53	36.19%	302	WP_016894272.1
<input checked="" type="checkbox"/> lysine-tyrosine recombinase/integrase [Mycobacteroides abscessus]	<i>Mycobacteroides abscessus</i>	184	184	94%	1e-52	36.19%	303	WP_079676398.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	183	183	94%	1e-52	43.13%	276	DAP71614.1
<input checked="" type="checkbox"/> TPA: SITE SPECIFIC RECOMBINASE XERD [Caudoviricetes so.]	Caudoviricetes so.	183	183	94%	1e-52	40.15%	270	DAR25295.1

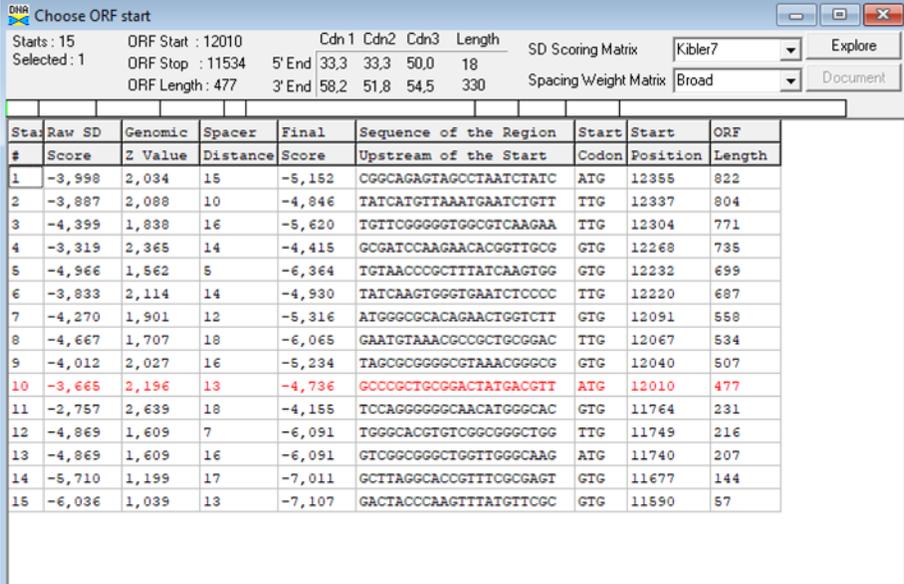
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

NON

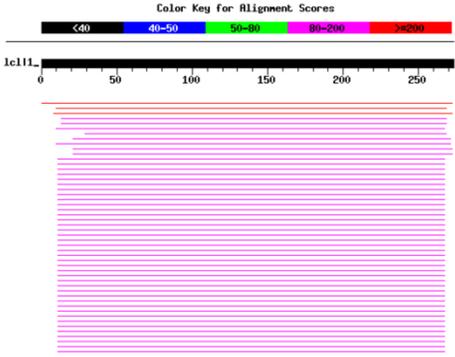
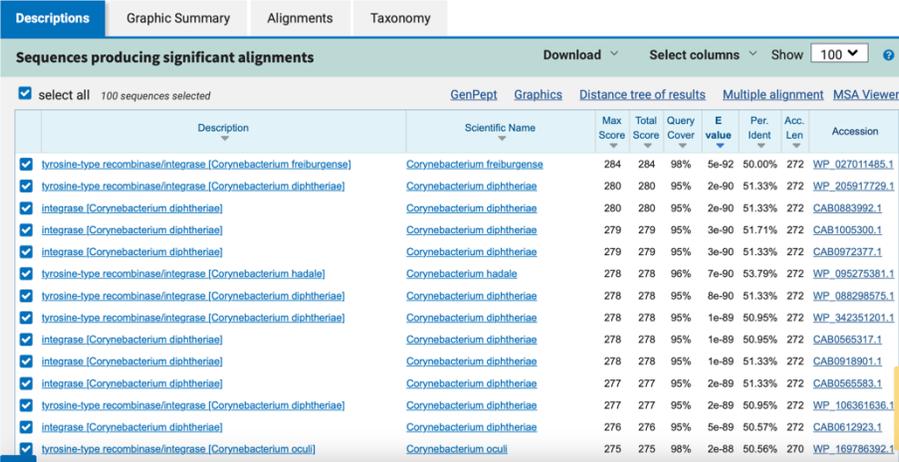
DECISION:

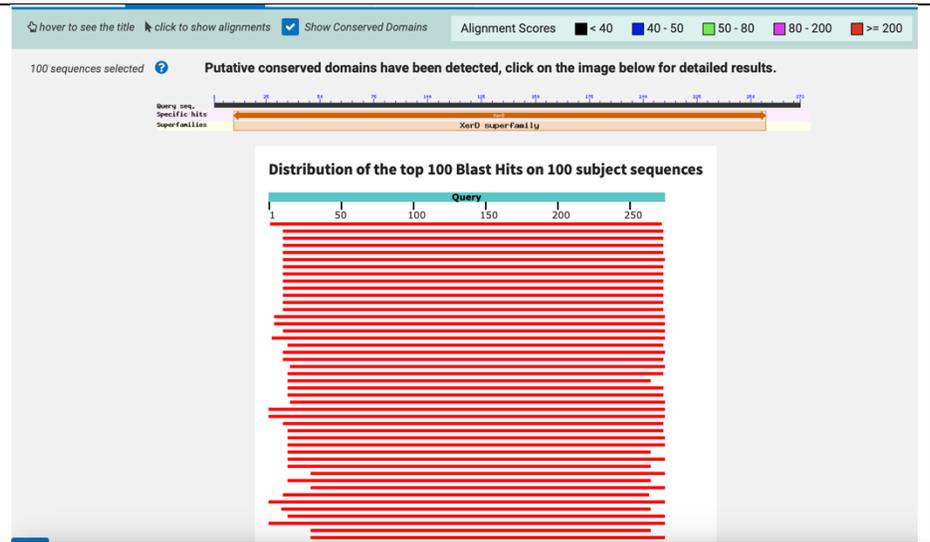
OUI

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																																																																																																																																
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 12355 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 12355																																																																																																																																																
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <thead> <tr> <th>#</th> <th>Raw SD</th> <th>Genomic Z Value</th> <th>Spacer 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>-3,998</td><td>2,034</td><td>15</td><td>-5,152</td><td>CGGCAGAGTAGCCTAATCTATC</td><td>ATG</td><td>12355</td><td>822</td></tr> <tr><td>2</td><td>-3,887</td><td>2,088</td><td>10</td><td>-4,846</td><td>TATCATGTTAAATGAATCTGTT</td><td>TTG</td><td>12337</td><td>804</td></tr> <tr><td>3</td><td>-4,399</td><td>1,838</td><td>16</td><td>-5,620</td><td>TGTTGGGGGTGGCGTCAAGAA</td><td>TTG</td><td>12304</td><td>771</td></tr> <tr><td>4</td><td>-3,319</td><td>2,365</td><td>14</td><td>-4,415</td><td>GCGATCCAAGAACACGGTTCGG</td><td>GTG</td><td>12268</td><td>735</td></tr> <tr><td>5</td><td>-4,966</td><td>1,562</td><td>5</td><td>-6,364</td><td>TGTAAACCCGCTTTATCAAGTGG</td><td>GTG</td><td>12232</td><td>699</td></tr> <tr><td>6</td><td>-3,833</td><td>2,114</td><td>14</td><td>-4,930</td><td>TATCAAGTGGGTGAATCTCCCC</td><td>TTG</td><td>12220</td><td>687</td></tr> <tr><td>7</td><td>-4,270</td><td>1,901</td><td>12</td><td>-5,316</td><td>ATGGGGCCACAGAACTGGTCTT</td><td>GTG</td><td>12091</td><td>558</td></tr> <tr><td>8</td><td>-4,667</td><td>1,707</td><td>18</td><td>-6,065</td><td>GAATGTAAACCCGCTGCGGAC</td><td>TTG</td><td>12067</td><td>534</td></tr> <tr><td>9</td><td>-4,012</td><td>2,027</td><td>16</td><td>-5,234</td><td>TAGCGGGGGCGTAAACGGGGG</td><td>GTG</td><td>12040</td><td>507</td></tr> <tr><td>10</td><td>-3,665</td><td>2,196</td><td>13</td><td>-4,736</td><td>GCCCGCTGCCGACTATGACGTT</td><td>ATG</td><td>12010</td><td>477</td></tr> <tr><td>11</td><td>-2,757</td><td>2,639</td><td>18</td><td>-4,155</td><td>TCCAGGGGGCCACATGGGCAC</td><td>GTG</td><td>11764</td><td>231</td></tr> <tr><td>12</td><td>-4,869</td><td>1,609</td><td>7</td><td>-6,091</td><td>TGGGCACGTGTGCGCGGGCTGG</td><td>TTG</td><td>11749</td><td>216</td></tr> <tr><td>13</td><td>-4,869</td><td>1,609</td><td>16</td><td>-6,091</td><td>GTCGGGGGGCTGGTTGGGCAAG</td><td>ATG</td><td>11740</td><td>207</td></tr> <tr><td>14</td><td>-5,710</td><td>1,199</td><td>17</td><td>-7,011</td><td>GCTTAGGCACCGTTTCGCGAGT</td><td>GTG</td><td>11677</td><td>144</td></tr> <tr><td>15</td><td>-6,036</td><td>1,039</td><td>13</td><td>-7,107</td><td>GACTACCCAAAGTTTATGTTCCG</td><td>GTG</td><td>11590</td><td>57</td></tr> </tbody> </table>	#	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length	1	-3,998	2,034	15	-5,152	CGGCAGAGTAGCCTAATCTATC	ATG	12355	822	2	-3,887	2,088	10	-4,846	TATCATGTTAAATGAATCTGTT	TTG	12337	804	3	-4,399	1,838	16	-5,620	TGTTGGGGGTGGCGTCAAGAA	TTG	12304	771	4	-3,319	2,365	14	-4,415	GCGATCCAAGAACACGGTTCGG	GTG	12268	735	5	-4,966	1,562	5	-6,364	TGTAAACCCGCTTTATCAAGTGG	GTG	12232	699	6	-3,833	2,114	14	-4,930	TATCAAGTGGGTGAATCTCCCC	TTG	12220	687	7	-4,270	1,901	12	-5,316	ATGGGGCCACAGAACTGGTCTT	GTG	12091	558	8	-4,667	1,707	18	-6,065	GAATGTAAACCCGCTGCGGAC	TTG	12067	534	9	-4,012	2,027	16	-5,234	TAGCGGGGGCGTAAACGGGGG	GTG	12040	507	10	-3,665	2,196	13	-4,736	GCCCGCTGCCGACTATGACGTT	ATG	12010	477	11	-2,757	2,639	18	-4,155	TCCAGGGGGCCACATGGGCAC	GTG	11764	231	12	-4,869	1,609	7	-6,091	TGGGCACGTGTGCGCGGGCTGG	TTG	11749	216	13	-4,869	1,609	16	-6,091	GTCGGGGGGCTGGTTGGGCAAG	ATG	11740	207	14	-5,710	1,199	17	-7,011	GCTTAGGCACCGTTTCGCGAGT	GTG	11677	144	15	-6,036	1,039	13	-7,107	GACTACCCAAAGTTTATGTTCCG	GTG	11590	57
#	Raw SD	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region	Start Codon	Start Position	ORF Length																																																																																																																																									
1	-3,998	2,034	15	-5,152	CGGCAGAGTAGCCTAATCTATC	ATG	12355	822																																																																																																																																									
2	-3,887	2,088	10	-4,846	TATCATGTTAAATGAATCTGTT	TTG	12337	804																																																																																																																																									
3	-4,399	1,838	16	-5,620	TGTTGGGGGTGGCGTCAAGAA	TTG	12304	771																																																																																																																																									
4	-3,319	2,365	14	-4,415	GCGATCCAAGAACACGGTTCGG	GTG	12268	735																																																																																																																																									
5	-4,966	1,562	5	-6,364	TGTAAACCCGCTTTATCAAGTGG	GTG	12232	699																																																																																																																																									
6	-3,833	2,114	14	-4,930	TATCAAGTGGGTGAATCTCCCC	TTG	12220	687																																																																																																																																									
7	-4,270	1,901	12	-5,316	ATGGGGCCACAGAACTGGTCTT	GTG	12091	558																																																																																																																																									
8	-4,667	1,707	18	-6,065	GAATGTAAACCCGCTGCGGAC	TTG	12067	534																																																																																																																																									
9	-4,012	2,027	16	-5,234	TAGCGGGGGCGTAAACGGGGG	GTG	12040	507																																																																																																																																									
10	-3,665	2,196	13	-4,736	GCCCGCTGCCGACTATGACGTT	ATG	12010	477																																																																																																																																									
11	-2,757	2,639	18	-4,155	TCCAGGGGGCCACATGGGCAC	GTG	11764	231																																																																																																																																									
12	-4,869	1,609	7	-6,091	TGGGCACGTGTGCGCGGGCTGG	TTG	11749	216																																																																																																																																									
13	-4,869	1,609	16	-6,091	GTCGGGGGGCTGGTTGGGCAAG	ATG	11740	207																																																																																																																																									
14	-5,710	1,199	17	-7,011	GCTTAGGCACCGTTTCGCGAGT	GTG	11677	144																																																																																																																																									
15	-6,036	1,039	13	-7,107	GACTACCCAAAGTTTATGTTCCG	GTG	11590	57																																																																																																																																									
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 = 822 pb ATG 12 355, gap de 6pb																																																																																																																																																
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 ?	Le START 12 355 n'est pas conservé																																																																																																																																																
DECISION:	ATG 12 355																																																																																																																																																

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

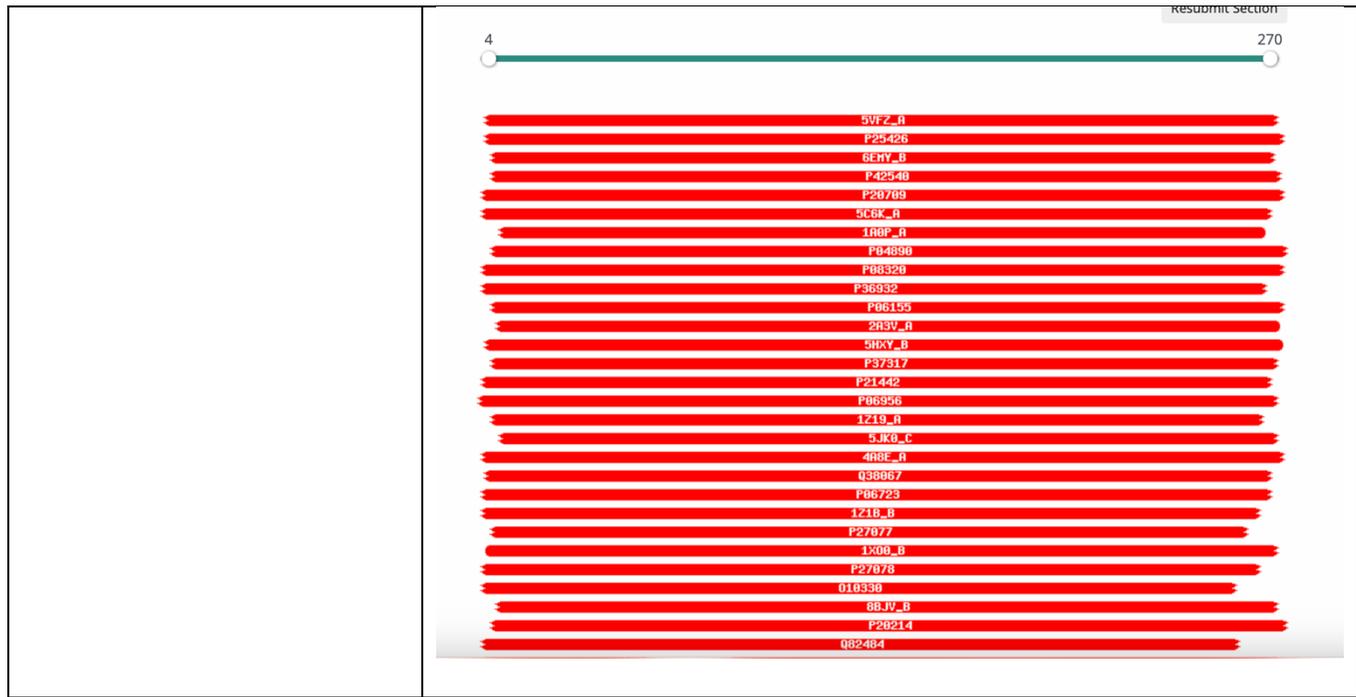
Collection des éléments de réponse	Rationnelle																																																																																																																																																																														
<p>Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée 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>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th></th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr><td>CyranoPS_Draft_17, function unknown, 273</td><td>553</td><td>e-157</td></tr> <tr><td>Emperor_19, integrase, 275</td><td>211</td><td>2e-54</td></tr> <tr><td>EpicDab_18, integrase, 276</td><td>209</td><td>6e-54</td></tr> <tr><td>SallySpecial_17, integrase, 266</td><td>199</td><td>8e-51</td></tr> <tr><td>Rahul_19, tyrosine integrase, 275</td><td>192</td><td>6e-49</td></tr> <tr><td>Weirdo19_Draft_45, function unknown, 411</td><td>191</td><td>2e-48</td></tr> <tr><td>Coeur_19, integrase, 273</td><td>187</td><td>2e-47</td></tr> <tr><td>GRU3_17, putative integrase, 297</td><td>171</td><td>2e-42</td></tr> <tr><td>Cracklewink_50, tyrosine integrase, 410</td><td>170</td><td>4e-42</td></tr> <tr><td>McGonagall_18, integrase, 298</td><td>169</td><td>9e-42</td></tr> <tr><td>Jeanie_19, integrase, 298</td><td>169</td><td>9e-42</td></tr> <tr><td>zombie_32, integrase, 398</td><td>167</td><td>2e-41</td></tr> </tbody> </table> <p>*** nr :</p>  <p>Sequences producing significant alignments:</p> <table border="1"> <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>tyrosine-type recombinase/integrase [Corynebacterium freiburgense]</td><td>Corynebacterium freiburgense</td><td>284</td><td>284</td><td>98%</td><td>5e-92</td><td>50.00%</td><td>272</td><td>WP_027011485.1</td></tr> <tr><td>tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>280</td><td>280</td><td>95%</td><td>2e-90</td><td>51.33%</td><td>272</td><td>WP_205917729.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>280</td><td>280</td><td>95%</td><td>2e-90</td><td>51.33%</td><td>272</td><td>CAB0883992.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>279</td><td>279</td><td>95%</td><td>3e-90</td><td>51.71%</td><td>272</td><td>CAB1005300.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>279</td><td>279</td><td>95%</td><td>3e-90</td><td>51.33%</td><td>272</td><td>CAB0972377.1</td></tr> <tr><td>tyrosine-type recombinase/integrase [Corynebacterium hadate]</td><td>Corynebacterium hadate</td><td>278</td><td>278</td><td>96%</td><td>7e-90</td><td>53.79%</td><td>272</td><td>WP_095275381.1</td></tr> <tr><td>tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>278</td><td>278</td><td>95%</td><td>8e-90</td><td>51.33%</td><td>272</td><td>WP_088298575.1</td></tr> <tr><td>tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>278</td><td>278</td><td>95%</td><td>1e-89</td><td>50.95%</td><td>272</td><td>WP_342351201.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>278</td><td>278</td><td>95%</td><td>1e-89</td><td>50.95%</td><td>272</td><td>CAB0565317.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>278</td><td>278</td><td>95%</td><td>1e-89</td><td>51.33%</td><td>272</td><td>CAB0918901.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>277</td><td>277</td><td>95%</td><td>2e-89</td><td>51.33%</td><td>272</td><td>CAB0565583.1</td></tr> <tr><td>tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>277</td><td>277</td><td>95%</td><td>2e-89</td><td>50.95%</td><td>272</td><td>WP_106361636.1</td></tr> <tr><td>integrase [Corynebacterium diphtheriae]</td><td>Corynebacterium diphtheriae</td><td>276</td><td>276</td><td>95%</td><td>5e-89</td><td>50.57%</td><td>272</td><td>CAB0612923.1</td></tr> <tr><td>tyrosine-type recombinase/integrase [Corynebacterium oculi]</td><td>Corynebacterium oculi</td><td>275</td><td>275</td><td>98%</td><td>2e-88</td><td>50.56%</td><td>270</td><td>WP_169786392.1</td></tr> </tbody> </table>		Score (bits)	E Value	CyranoPS_Draft_17, function unknown, 273	553	e-157	Emperor_19, integrase, 275	211	2e-54	EpicDab_18, integrase, 276	209	6e-54	SallySpecial_17, integrase, 266	199	8e-51	Rahul_19, tyrosine integrase, 275	192	6e-49	Weirdo19_Draft_45, function unknown, 411	191	2e-48	Coeur_19, integrase, 273	187	2e-47	GRU3_17, putative integrase, 297	171	2e-42	Cracklewink_50, tyrosine integrase, 410	170	4e-42	McGonagall_18, integrase, 298	169	9e-42	Jeanie_19, integrase, 298	169	9e-42	zombie_32, integrase, 398	167	2e-41	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	tyrosine-type recombinase/integrase [Corynebacterium freiburgense]	Corynebacterium freiburgense	284	284	98%	5e-92	50.00%	272	WP_027011485.1	tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	280	280	95%	2e-90	51.33%	272	WP_205917729.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	280	280	95%	2e-90	51.33%	272	CAB0883992.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	279	279	95%	3e-90	51.71%	272	CAB1005300.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	279	279	95%	3e-90	51.33%	272	CAB0972377.1	tyrosine-type recombinase/integrase [Corynebacterium hadate]	Corynebacterium hadate	278	278	96%	7e-90	53.79%	272	WP_095275381.1	tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	8e-90	51.33%	272	WP_088298575.1	tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	1e-89	50.95%	272	WP_342351201.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	1e-89	50.95%	272	CAB0565317.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	1e-89	51.33%	272	CAB0918901.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	277	277	95%	2e-89	51.33%	272	CAB0565583.1	tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	277	277	95%	2e-89	50.95%	272	WP_106361636.1	integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	276	276	95%	5e-89	50.57%	272	CAB0612923.1	tyrosine-type recombinase/integrase [Corynebacterium oculi]	Corynebacterium oculi	275	275	98%	2e-88	50.56%	270	WP_169786392.1
	Score (bits)	E Value																																																																																																																																																																													
CyranoPS_Draft_17, function unknown, 273	553	e-157																																																																																																																																																																													
Emperor_19, integrase, 275	211	2e-54																																																																																																																																																																													
EpicDab_18, integrase, 276	209	6e-54																																																																																																																																																																													
SallySpecial_17, integrase, 266	199	8e-51																																																																																																																																																																													
Rahul_19, tyrosine integrase, 275	192	6e-49																																																																																																																																																																													
Weirdo19_Draft_45, function unknown, 411	191	2e-48																																																																																																																																																																													
Coeur_19, integrase, 273	187	2e-47																																																																																																																																																																													
GRU3_17, putative integrase, 297	171	2e-42																																																																																																																																																																													
Cracklewink_50, tyrosine integrase, 410	170	4e-42																																																																																																																																																																													
McGonagall_18, integrase, 298	169	9e-42																																																																																																																																																																													
Jeanie_19, integrase, 298	169	9e-42																																																																																																																																																																													
zombie_32, integrase, 398	167	2e-41																																																																																																																																																																													
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium freiburgense]	Corynebacterium freiburgense	284	284	98%	5e-92	50.00%	272	WP_027011485.1																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	280	280	95%	2e-90	51.33%	272	WP_205917729.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	280	280	95%	2e-90	51.33%	272	CAB0883992.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	279	279	95%	3e-90	51.71%	272	CAB1005300.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	279	279	95%	3e-90	51.33%	272	CAB0972377.1																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium hadate]	Corynebacterium hadate	278	278	96%	7e-90	53.79%	272	WP_095275381.1																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	8e-90	51.33%	272	WP_088298575.1																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	1e-89	50.95%	272	WP_342351201.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	1e-89	50.95%	272	CAB0565317.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	278	278	95%	1e-89	51.33%	272	CAB0918901.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	277	277	95%	2e-89	51.33%	272	CAB0565583.1																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	277	277	95%	2e-89	50.95%	272	WP_106361636.1																																																																																																																																																																							
integrase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	276	276	95%	5e-89	50.57%	272	CAB0612923.1																																																																																																																																																																							
tyrosine-type recombinase/integrase [Corynebacterium oculi]	Corynebacterium oculi	275	275	98%	2e-88	50.56%	270	WP_169786392.1																																																																																																																																																																							



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 $\geq 90\%$ et une couverture acceptable ?

**Cd : INT_C_like_2; Uncharacterized site-specific tyrosine recombinase, C-terminal catalytic domain.
 PFAM : Phage_integrase ; Phage integrase family
 PDB : Gp33; Bacteriophage, Brujita, DNA-binding, Integrase, DNA BINDING PROTEIN; HET: ACT, GOL; 1.847A {Mycobacterium phage Br
 Uniprot : VINT_BPMFR Integrase OS=Mycobacterium phage FRAT1 OX=12388 GN=int PE=3 SV=1**

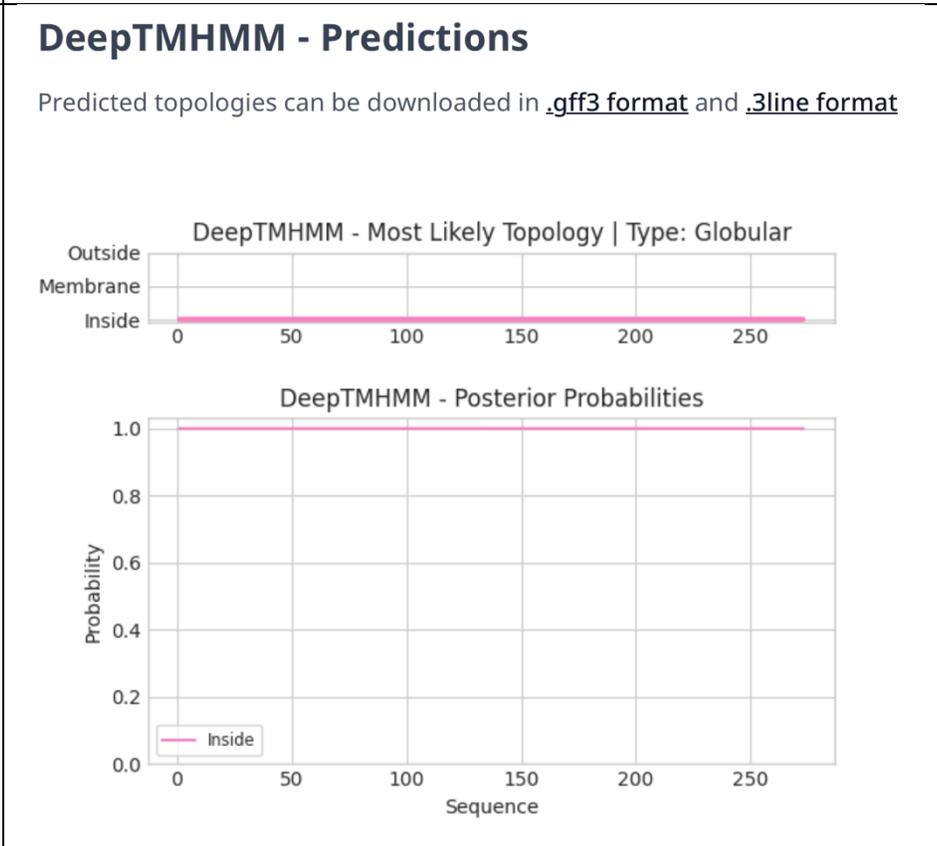
<input type="checkbox"/>	11	P06155	VINT_BPPH8 Integrase OS=Enterobacteria phage phi80 OX=10713 GN=int PE=3 SV=1	100	3.7e-30	198.68	27.1	264	402
<input type="checkbox"/>	12	2A3V_A	site-specific recombinase IntI4; Protein-DNA complex, RECOMBINATION; 2.8A (Vibrio cholerae O1 biovar eltor str. N16961)	100	5.3e-30	191.05	26	260	320
<input type="checkbox"/>	13	5HXY_B	Tyrosine recombinase XerA; recombinase, XerA, RECOMBINATION; HET: MSE, PO4; 2.5A (Thermoplasma acidophilum DSM 1728)	100	1.1e-29	189.91	27.5	262	317
<input type="checkbox"/>	14	P37317	VINT_BPSFV Integrase OS=Shigella phage Sf6 OX=10761 GN=int PE=3 SV=1	100	1e-29	195.11	26.2	264	385
<input type="checkbox"/>	15	P21442	VINT_BPHC1 Integrase OS=Haemophilus phage HP1 (strain HP1c1) OX=1289570 GN=int PE=1 SV=1	100	2.3e-29	189.82	26.4	254	337
<input type="checkbox"/>	16	P06956	RECR_BPP1 Recombinase cre OS=Escherichia phage P1 OX=10678 GN=cre PE=1 SV=1	99.98	1.3e-29	191.19	22.3	264	343
<input type="checkbox"/>	17	1Z19_A	Integrase; PROTEIN-DNA COMPLEX, DNA BINDING PROTEIN-DNA COMPLEX; HET: PTR, MSE; 2.8A (Enterobacteria phage lambda) SCOP:	99.97	9.8e-29	181.53	25.3	255	283
<input type="checkbox"/>	18	5JK0_C	Tyrosine recombinase XerH; Xer, tyrosine recombinase, site-specific recombinase, chromosome dimer resolution, cell cycle	99.97	5.2e-28	184.06	29.3	260	363
<input type="checkbox"/>	19	4A8E_A	PROBABLE TYROSINE RECOMBINASE XERC-LIKE; CELL CYCLE, CHROMOSOME DIMER RESOLUTION, PAB0255; HET: SO4; 2.99A (PYROCOCCUS A	99.97	6.6e-28	178.31	28.7	260	292
<input type="checkbox"/>	20	Q38067	INTG_BPPF1 Putative integrase OS=Pseudomonas phage Pf1 OX=2011081 PE=3 SV=2	99.97	3.3e-28	183.27	26	257	333
<input type="checkbox"/>	1	5VFZ_A	Gp33; Bacteriophage, Brujita, DNA-binding, Integrase, DNA BINDING PROTEIN; HET: ACT, GOL; 1.847A (Mycobacterium phage Br	100	2.3e-33	209.44	31.2	265	318
<input type="checkbox"/>	2	P25426	VINT_BPMFR Integrase OS=Mycobacterium phage FRAT1 OX=12388 GN=int PE=3 SV=1	100	5.3e-32	203.6	28.6	265	333
<input type="checkbox"/>	3	6EMY_B	Int protein; transposase protein-DNA complex, tyrosine recombinase, Y-transposase, Tn916-like conjugative transposon, an	100	4.9e-31	196.31	29.3	262	317
<input type="checkbox"/>	4	P42540	VINT_BPL2 Probable integrase/recombinase OS=Acholeplasma phage L2 OX=46014 PE=3 SV=1	100	2e-30	191.04	28.1	259	289
<input type="checkbox"/>	5	P20709	VINT_BPL54 Integrase OS=Staphylococcus phage L54a OX=10727 GN=int PE=3 SV=1	100	8.9e-31	198.05	26.9	269	354
<input type="checkbox"/>	6	5C6K_A	Integrase; Integrase, tyrosine recombinase, integration, site-specific recombination, hydrolase; 1.9A (Enterobacteria ph	100	5.3e-30	189.26	28.9	255	292
<input type="checkbox"/>	7	1A0P_A	SITE-SPECIFIC RECOMBINASE XERD; XERD, RECOMBINASE, DNA BINDING, DNA RECOMBINATION; 2.5A (Escherichia coli) SCOP: d.163.1	100	6.3e-30	188.18	29	256	290
<input type="checkbox"/>	8	P04890	VINT_BPP22 Integrase OS=Salmonella phage P22 OX=10754 GN=int PE=3 SV=1	100	3.5e-30	197.72	28.2	262	387
<input type="checkbox"/>	9	P08320	VINT_BPP4 Integrase OS=Enterobacteria phage P4 OX=10680 GN=int PE=3 SV=2	100	2.9e-30	201.28	27.8	268	439
<input type="checkbox"/>	10	P36932	VINT_BPP2 Integrase OS=Escherichia phage P2 OX=10679 GN=int PE=1 SV=2	100	2.8e-30	194.7	26.8	257	337



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 ?

Le gene apres est un immunity repressor ou helix-turn-helix-transcription factor

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



Est-ce que la fonction proposée fait partie de liste

SEA PHAGE on a soit « serine integrase » soit « tyrosine integrase »

de fonctions approuvées par SEA-PHAGES ?	
DECISION:	Tyrosine Integrase

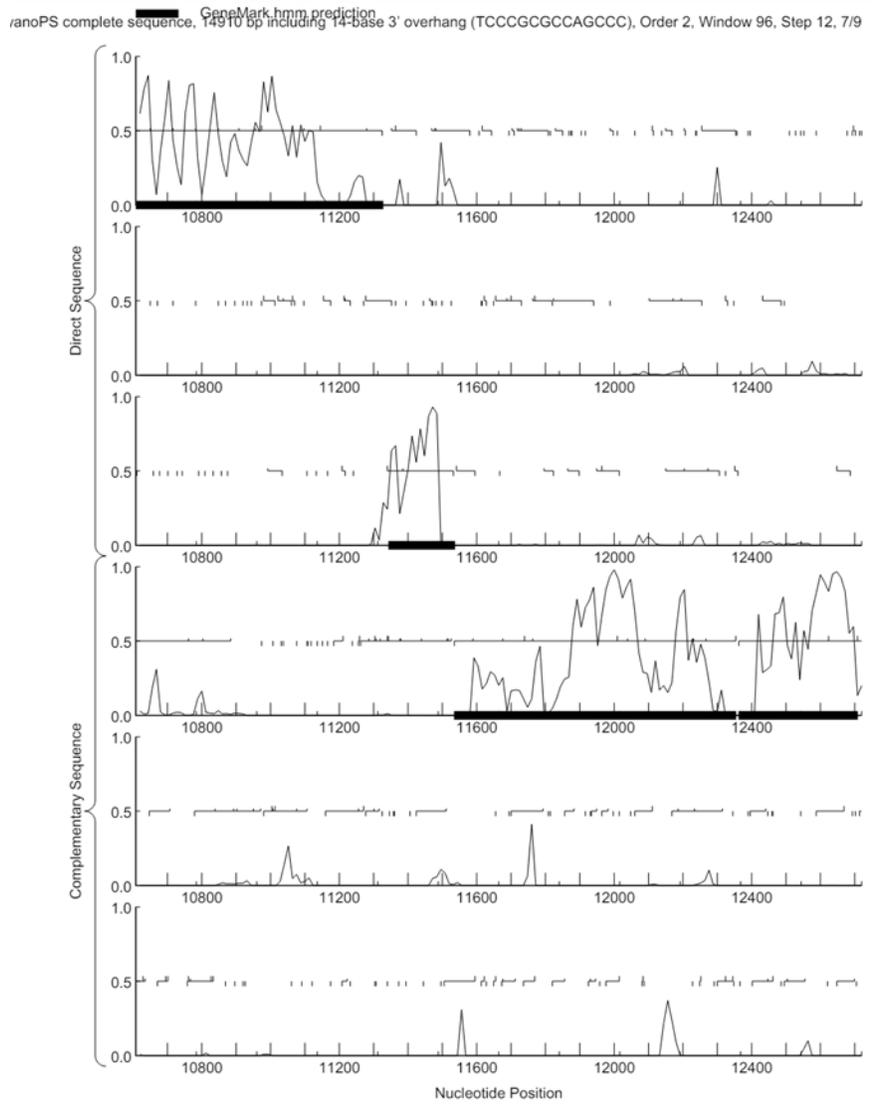
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	18
Coordonnées du Stop	12 362
Direction (For/Rev)	Rev
Gap/chevauchement avec un autre gène	Gap 103
Coordonnées du Start retenu	12 709
Fonction prédite	Immunity repressor (ou helix-turn-helix transcriptional regulator?)

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

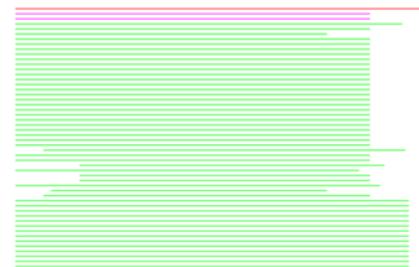
Y-a-t-il des éléments supportant un potentiel codant ?



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

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

Color Key for Alignment Scores



Sequences producing significant alignments:

	Score (bits)	E Value
Cyranos_Draft_18, function unknown, 115	236	1e-62
McGonagall_19, immunity repressor, 116	82	6e-16
Jeanie_19, immunity repressor, 116	82	6e-16
Bakery_54, immunity repressor, 137	74	1e-13
Doggs_37, immunity repressor, 126	74	2e-13
Arri_51, immunity repressor, 137	74	2e-13
VanDeWege_51, immunity repressor, 137	73	2e-13
Valary_50, immunity repressor, 137	73	2e-13
Twister6_50, immunity repressor, 137	73	2e-13
Togo_47, immunity repressor, 137	73	2e-13

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	123	123	85%	1e-33	67.68%	101	WP_235696498.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	121	121	85%	6e-33	66.67%	101	WP_226813673.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	121	121	85%	7e-33	66.67%	107	WP_310795193.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	119	119	81%	3e-32	68.42%	97	UJL54986.1
<input checked="" type="checkbox"/> HTH family transcriptional regulator [Corynebacterium diphtheriae by Intermedia str. NCTC 5011]	Corynebacterium diphtheri...	118	118	81%	1e-31	67.37%	97	EIK55653.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	117	117	86%	3e-31	59.00%	102	QQJ78348.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	116	116	76%	9e-31	70.79%	101	WP_339326148.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Corynebacterium diphtheriae by oravis]	Corynebacterium diphtheri...	114	114	73%	3e-30	70.24%	88	UWF02311.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	113	113	86%	1e-29	57.28%	104	WP_205917829.1
<input checked="" type="checkbox"/> transcriptional regulator [Corynebacterium diphtheriae by oravis]	Corynebacterium diphtheri...	112	112	73%	1e-29	71.76%	91	QWN57024.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	112	112	86%	2e-29	56.31%	104	WP_106361736.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium heidelve...	112	112	85%	2e-29	58.59%	101	WP_112769907.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium accolens	112	112	85%	3e-29	59.60%	101	WP_284900651.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium heidelve...	111	111	85%	8e-29	58.59%	112	WQZ37841.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium freiburgense]	Corynebacterium freiburg...	110	110	85%	2e-28	59.60%	101	WP_051256139.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	109	109	85%	3e-28	56.57%	101	WP_306592755.1
<input checked="" type="checkbox"/> XRE family transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	108	108	71%	5e-28	71.08%	88	AR887912.2
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium erocinoum]	Corynebacterium erocinou...	108	108	86%	6e-28	57.00%	102	WP_284571910.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	108	108	86%	1e-27	53.40%	104	WP_088296544.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	85%	2e-27	54.55%	101	WP_284790726.1
<input checked="" type="checkbox"/> TPA_helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	86%	3e-27	55.00%	102	HAT6625290.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium erocinoum]	Corynebacterium erocinou...	106	106	86%	8e-27	53.00%	102	WP_284594249.1
<input checked="" type="checkbox"/> helix-turn-helix transcriptional regulator [Corynebacterium pseudodiphtherioides]	Corynebacterium pseudod...	106	106	85%	8e-27	55.56%	101	WP_284587248.1

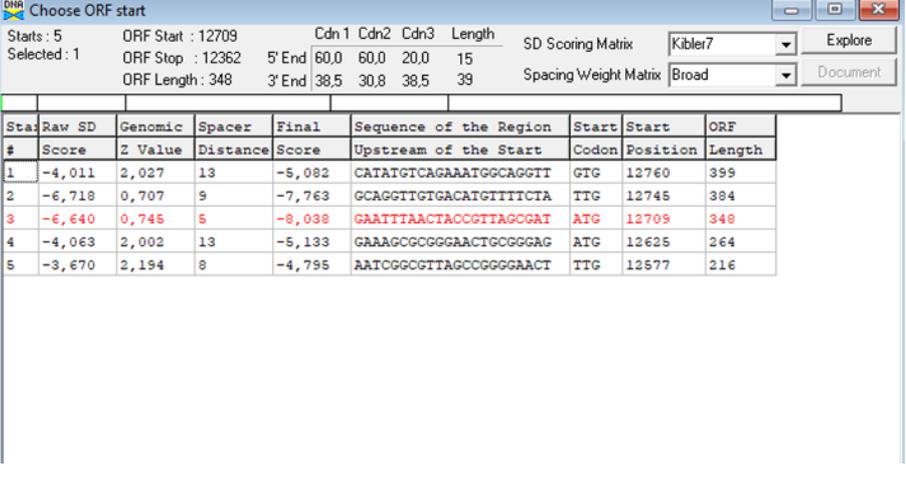
Est-ce que le candidat est en contradiction avec les principes d'annotation ?

non

DECISION:

oui

Décision #2 : Quel est le start du gène ?

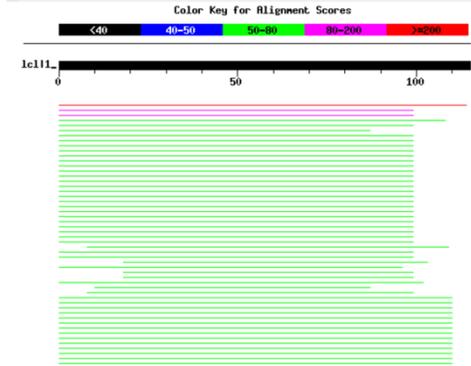
Collection des éléments de réponse	Rationnelle																																																							
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :12709 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :12709</p>																																																							
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <table border="1" data-bbox="565 615 1341 779"> <thead> <tr> <th>Sta#</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> </thead> <tbody> <tr> <td>1</td> <td>-4,011</td> <td>2,027</td> <td>13</td> <td>-5,082</td> <td>CATAATGTCAGAAATGGCAGGTT</td> <td>GTG</td> <td>12760</td> <td>399</td> </tr> <tr> <td>2</td> <td>-6,718</td> <td>0,707</td> <td>9</td> <td>-7,763</td> <td>GCAGGTTGTGACATGTTTCTA</td> <td>TTG</td> <td>12745</td> <td>384</td> </tr> <tr> <td>3</td> <td>-8,640</td> <td>0,745</td> <td>5</td> <td>-8,039</td> <td>GAATTTAACTACCGTTAGCGAT</td> <td>ATG</td> <td>12709</td> <td>348</td> </tr> <tr> <td>4</td> <td>-4,063</td> <td>2,002</td> <td>13</td> <td>-5,133</td> <td>GAAAGCGCGGGAACCTGCGGGAG</td> <td>ATG</td> <td>12625</td> <td>264</td> </tr> <tr> <td>5</td> <td>-3,670</td> <td>2,194</td> <td>8</td> <td>-4,795</td> <td>AAICGGCGTTAGCCGGGGAAC</td> <td>TTG</td> <td>12577</td> <td>216</td> </tr> </tbody> </table>	Sta#	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	-4,011	2,027	13	-5,082	CATAATGTCAGAAATGGCAGGTT	GTG	12760	399	2	-6,718	0,707	9	-7,763	GCAGGTTGTGACATGTTTCTA	TTG	12745	384	3	-8,640	0,745	5	-8,039	GAATTTAACTACCGTTAGCGAT	ATG	12709	348	4	-4,063	2,002	13	-5,133	GAAAGCGCGGGAACCTGCGGGAG	ATG	12625	264	5	-3,670	2,194	8	-4,795	AAICGGCGTTAGCCGGGGAAC	TTG	12577	216
Sta#	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	-4,011	2,027	13	-5,082	CATAATGTCAGAAATGGCAGGTT	GTG	12760	399																																																
2	-6,718	0,707	9	-7,763	GCAGGTTGTGACATGTTTCTA	TTG	12745	384																																																
3	-8,640	0,745	5	-8,039	GAATTTAACTACCGTTAGCGAT	ATG	12709	348																																																
4	-4,063	2,002	13	-5,133	GAAAGCGCGGGAACCTGCGGGAG	ATG	12625	264																																																
5	-3,670	2,194	8	-4,795	AAICGGCGTTAGCCGGGGAAC	TTG	12577	216																																																
<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>	<p><i>Troisième plus longue ORF avec un ATG score (-8)</i></p> <p><i>Mais plus long ORF avec un ATG</i></p>																																																							
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>—</p>																																																							

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<pre> >Arri_51, immunity repressor, 137 Length = 137 Score = 73.6 bits (179), Expect = 2e-13 Identities = 40/100 (40%), Positives = 62/100 (62%), Gaps = 1/100 (1%) Query: 1 MTSAFNQAGIIPWTKQDRIRKARELREMKQSDLATAIGVSRGTLASIEQGVREPRRGEV 60 MT+A+ ++G +PE + + R+R ARE ++Q LA +GVSR T+++ E+G PR+ + Sbjct: 1 MTTAY-ESGRVPEISLRHRLRIAREEAGLEQQSLAERMGVSRRNTVSAAEKGNAPRKVVV 59 Query: 61 IAISFATGVSLDWLETGKTPADDNGGGEKWAHWSNRPRA 100 A + ATG + WLETG P G + A DSNP+P+ Sbjct: 60 NAWALATGFDVKWLETGVAPQPGPEGDDGCARRDSNPKPS 99 >VanDeWege_51, immunity repressor, 137 Length = 137 Score = 73.2 bits (178), Expect = 2e-13 Identities = 40/100 (40%), Positives = 62/100 (62%), Gaps = 1/100 (1%) Query: 1 MTSAFNQAGIIPWTKQDRIRKARELREMKQSDLATAIGVSRGTLASIEQGVREPRRGEV 60 MT+A+ ++G +PE + + R+R ARE ++Q LA +GVSR T+++ E+G PR+ + Sbjct: 1 MTTAY-ESGRVPEISLRHRLRIAREEAGLEQQSLAERMGVSRRNTVSAAEKGNAPRKVVV 59 Query: 61 IAISFATGVSLDWLETGKTPADDNGGGEKWAHWSNRPRA 100 A + ATG + WLETG P G + A DSNP+P+ Sbjct: 60 NAWALATGFDVKWLETGVIAPQPGPEGDDGCARRDSNPKPS 99 >Valary_50, immunity repressor, 137 Length = 137 Score = 73.2 bits (178), Expect = 2e-13 Identities = 40/100 (40%), Positives = 62/100 (62%), Gaps = 1/100 (1%) Query: 1 MTSAFNQAGIIPWTKQDRIRKARELREMKQSDLATAIGVSRGTLASIEQGVREPRRGEV 60 MT+A+ ++G +PE + + R+R ARE ++Q LA +GVSR T+++ E+G PR+ + Sbjct: 1 MTTAY-ESGRVPEISLRHRLRIAREEAGLEQQSLAERMGVSRRNTVSAAEKGNAPRKVVV 59 </pre>
<p>DECISION:</p>	<p>12 709 ATG</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>

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



Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_18, function unknown, 115	236	1e-62
McGonagall_19, immunity repressor, 116	82	6e-16
Jeanie_19, immunity repressor, 116	82	6e-16
Bakery_54, immunity repressor, 137	74	1e-13
Doggs_37, immunity repressor, 126	74	2e-13
Arri_51, immunity repressor, 137	74	2e-13
VanDeWege_51, immunity repressor, 137	73	2e-13
Valary_50, immunity repressor, 137	73	2e-13
Twister6_50, immunity repressor, 137	73	2e-13

*** nr :

hover to see the title click to show alignments Show Conserved Domains Alignment Scores < 40 40 - 50 50 - 80 80 - 200 >= 200

100 sequences selected

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. HTSAPRQNDIIPENFKQDRISKARLDELKENSQSLAFAICVSRQLASLEGQVREPRQEVIAISPTQSLDLLETGKTPADQNGQDENWNNHDSNPPFACICXPSDQGGKVA
Specific hits XRE
Superfamilies XRE superFamily

Distribution of the top 100 Blast Hits on 100 subject sequences



<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium procioum]	Corynebacterium procioum	108	108	86%	6e-28	57.00%	102	WP_284571910.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	108	108	86%	1e-27	53.40%	104	WP_088298544.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	85%	2e-27	54.55%	101	WP_284790726.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	86%	3e-27	55.00%	102	HAT6825290.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium procioum]	Corynebacterium procioum	106	106	86%	8e-27	53.00%	102	WP_284594249.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodi...	106	106	85%	8e-27	55.56%	101	WP_284587248.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodi...	105	105	85%	1e-26	54.55%	101	WP_284594587.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [uncultured Corynebacterium sp.]	uncultured Corynebacteriu...	105	105	88%	1e-26	50.49%	108	WP_296216267.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	105	105	85%	1e-26	54.55%	101	HAT1503366.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Corynebacterium diphtheriae bv. gravis]	Corynebacterium diphtheri...	103	103	80%	2e-26	77.46%	73	UWF08262.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium accolens	104	104	85%	4e-26	53.54%	101	MDK4334013.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium]	Corynebacterium	103	103	74%	4e-26	60.92%	90	WP_141740943.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [uncultured Corynebacterium sp.]	uncultured Corynebacteriu...	103	103	86%	1e-25	52.48%	102	WP_296111086.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	102	102	85%	2e-25	53.54%	101	HAT1137157.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	100	100	85%	2e-24	52.53%	101	WP_201818247.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	99.4	99.4	85%	4e-24	51.52%	101	HAT1153054.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	99.0	99.0	86%	5e-24	56.00%	101	HCT5225811.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	99.0	99.0	80%	6e-24	58.51%	102	WP_134884368.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [uncultured Corynebacterium sp.]	uncultured Corynebacteriu...	98.6	98.6	85%	8e-24	51.52%	101	WP_298894171.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium accolens	98.2	98.2	79%	9e-24	55.43%	95	WP_284641497.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium pseudodiphthericum]	Corynebacterium pseudodi...	98.6	98.6	86%	9e-24	50.00%	102	WP_284494094.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Corynebacterium diphtheriae bv. gravis]	Corynebacterium diphtheri...	97.1	97.1	48%	1e-23	85.71%	60	UWFE8692.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Corynebacterium poyangense]	Corynebacterium poyange...	97.1	97.1	86%	3e-23	51.00%	102	QNG91501.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium poyangense]	Corynebacterium poyange...	97.1	97.1	86%	4e-23	51.00%	108	WP_316932489.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium sanguinis]	Corynebacterium sanguinis	96.7	96.7	85%	5e-23	50.51%	101	WP_259927076.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium sp. EPI-003-04-2554_SCH2473822]	Corynebacterium sp. EPI-0...	96.3	96.3	85%	7e-23	49.49%	101	WP_064834154.1
<input checked="" type="checkbox"/>	transcriptional regulator [Corynebacterium hadale]	Corynebacterium hadale	95.9	95.9	85%	9e-23	54.55%	100	PAJ70834.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium accolens	95.9	95.9	85%	9e-23	53.54%	101	WP_302526980.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium procioum]	Corynebacterium procioum	95.9	95.9	85%	1e-22	50.51%	101	WP_284589724.1

Sequences producing significant alignments Download Select columns Show 100

<input checked="" type="checkbox"/> 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"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	123	123	85%	1e-33	67.68%	101	WP_235696498.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	121	121	85%	6e-33	66.67%	101	WP_226813673.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	121	121	85%	7e-33	66.67%	107	WP_310795193.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	119	119	81%	3e-32	68.42%	97	UJL54886.1
<input checked="" type="checkbox"/>	HTH family transcriptional regulator [Corynebacterium diphtheriae bv. intermedius str. NCTC 5011]	Corynebacterium...	118	118	81%	1e-31	67.37%	97	EIK55653.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium s...	117	117	86%	3e-31	59.00%	102	QQU78348.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	116	116	76%	9e-31	70.79%	101	WP_339326148.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Corynebacterium diphtheriae bv. gravis]	Corynebacterium...	114	114	73%	3e-30	70.24%	88	UWF02311.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	113	113	86%	1e-29	57.28%	104	WP_205917829.1
<input checked="" type="checkbox"/>	transcriptional regulator [Corynebacterium diphtheriae bv. gravis]	Corynebacterium...	112	112	73%	1e-29	71.76%	91	OVNS7024.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	112	112	86%	2e-29	56.31%	104	WP_106361736.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium...	112	112	85%	2e-29	58.59%	101	WP_112768907.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium...	112	112	85%	3e-29	59.60%	101	WP_284900651.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium...	111	111	85%	8e-29	58.59%	112	WCZ37641.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium freiburgense]	Corynebacterium f...	110	110	85%	2e-28	59.60%	101	WP_051256139.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium s...	109	109	85%	3e-28	56.57%	101	WP_306592755.1
<input checked="" type="checkbox"/>	XRE family transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	108	108	71%	5e-28	71.08%	88	ARB67912.2
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium procioum]	Corynebacterium...	108	108	86%	6e-28	57.00%	102	WP_284571910.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium...	108	108	86%	1e-27	53.40%	104	WP_088298544.1
<input checked="" type="checkbox"/>	helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium s...	107	107	85%	2e-27	54.55%	101	WP_284790726.1

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 ?

PFAM : BetR ; BetR domain
 CDD : DUF1870 ; Domain of unknown function (DUF1870)
 Cd : HTH_XRE; Helix-turn-helix XRE-family like proteins. Prokaryotic DNA binding proteins belonging to the xenobiotic response element family of transcriptional regulators.
 PDB : ComR; Streptococcus, Competence, Quorum sensing, ComR, TRANSCRIPTION REGULATOR; 2.9A {Streptococcus suis (strain 05ZYH33
 Uniprot : RPC_BP163 Repressor protein C OS=Rhizobium phage 16-3 OX=10704 GN=C PE=1 SV=4

<input type="checkbox"/>	1	5FD4_B	ComR; Streptococcus, Competence, Quorum sensing, ComR, TRANSCRIPTION REGULATORY; 2.9A (Streptococcus suis (strain 05ZYH33	98.46	0.000022	45.11	9.5	69	324
<input type="checkbox"/>	2	P15238	RPC_BP163 Repressor protein C OS=Rhizobium phage 16-3 OX=10704 GN=C PE=1 SV=4	98.39	0.000034	43.09	8.9	65	263
<input type="checkbox"/>	3	6B9T_D	Methylphosphonate synthase; Phosphonate, Methylphosphonate, Iron, OXIDOREDUCTASE; HET: FMT, 2HE; 2.35A (Nitrosopumilus m	98.22	0.00011	44.41	9	65	457
<input type="checkbox"/>	4	6WPZ_B	Pf4r; DNA BINDING PROTEIN; 1.993A (Pseudomonas aeruginosa)	97.94	0.00057	32.12	9	90	95
<input type="checkbox"/>	5	P18680	RPC1_BPHK0 26 kDa repressor protein OS=Escherichia phage HK022 OX=10742 GN=CI-HTT PE=4 SV=2	97.93	0.00051	37.76	7.5	68	235
<input type="checkbox"/>	6	6H49_A	Orf20; SaPI, Repressor, STRUCTURAL PROTEIN; HET: SO4; 1.8A (Staphylococcus aureus)	97.92	0.001	34.13	9.2	78	157
<input type="checkbox"/>	7	6XGT_B	Cyanate hydratase; cyanase, HYDROLASE; HET: MLI, FMT; 2.2A (Thermomyces lanuginosus)	97.78	0.0024	34.06	9.3	74	181
<input type="checkbox"/>	8	7N1N_B	ComR; Paratox, bacteriophage, quorum sensing, Streptococcus, ComR, natural competence, VIRAL PROTEIN-TRANSCRIPTION compl	97.7	0.0012	28.41	8.9	67	68
<input type="checkbox"/>	9	P04132	RPC_BPP2 Repressor protein C OS=Escherichia phage P2 OX=10679 GN=C PE=1 SV=1	97.66	0.0021	30.23	9.4	97	99
<input type="checkbox"/>	19	8EZT_A	HipB(Lp); toxin-antitoxin complex, legionella pneumophila, structural genomics, Center for Structural Genomics of Infect	97.25	0.007	26.66	9.3	76	76
<input type="checkbox"/>	20	7JVT_D	Repressor protein CI; DNA binding transcriptional regulator, GENE REGULATION, GENE REGULATION-DNA complex; 3.16A (Escher	97.24	0.012	32.15	6.7	75	214
<input type="checkbox"/>	21	7P4A_A	StI; Mobile Genetic Element, Repressor, SaPI, HTH domain, DNA BINDING PROTEIN; HET: MSE; 2.901A (Staphylococcus aureus)	97.21	0.024	32.05	9.4	69	247
<input type="checkbox"/>	22	2EF8_A	Putative transcription factor; helix-turn-helix, DNA binding protein, TRANSCRIPTION REGULATORY; HET: CME; 1.95A (Enteroba	97.19	0.0088	26.72	6.8	80	84
<input type="checkbox"/>	23	PF08667.14	; BetR ; BetR domain	97.16	0.022	30.59	8.9	68	174
<input type="checkbox"/>	24	P03034	RPC1_LAMBD Repressor protein cI OS=Escherichia phage lambda OX=10710 GN=cI PE=1 SV=2	97.15	0.013	32.53	6.4	77	237
<input type="checkbox"/>	25	P14819	RPC1_BPPH8 Repressor protein CI OS=Enterobacteria phage phi80 OX=10713 GN=CI PE=1 SV=2	97.13	0.027	31.16	9.4	68	236

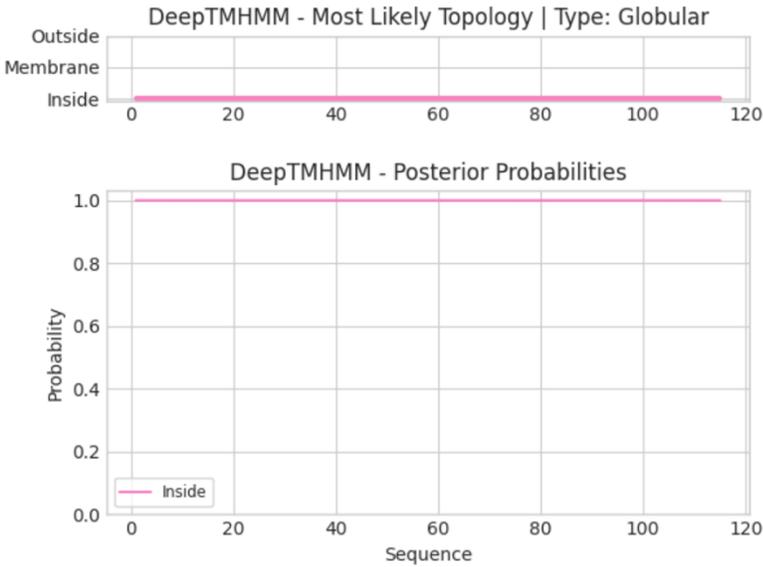
<input type="checkbox"/>	10	2WIU_D	HTH-TYPE TRANSCRIPTIONAL REGULATOR HIPB; TRANSFERASE TRANSCRIPTION COMPLEX, SERINE KINASE, DNA-BINDING, MERCURY DERIVATI	97.6	0.0021	28.52	9.2	85	88
<input type="checkbox"/>	11	1X57_A	Endothelial differentiation-related factor 1; EDF1, HMBF1alpha, helix-turn-helix, Structural Genomics, NPP5FA, National	97.4	0.0045	27.57	9.1	90	91
<input type="checkbox"/>	12	3U3W_B	Transcriptional activator PlcR protein; ternary complex, PlcR-PAPR7-DNA, HTH DNA-binding domain, Quorum Sensing, HTH_3 (97.37	0.0095	33.28	7.5	64	293
<input type="checkbox"/>	13	6B9R_D	Hydroxyethylphosphonate dioxygenase; Phosphonate, Hydroxymethylphosphonate, Iron, OXIDOREDUCTASE; HET: GOL, 2HE; 1.802A	97.37	0.013	35.89	8.3	83	450
<input type="checkbox"/>	14	3TYS_A	Predicted transcriptional regulator; Structural Genomics, Center for Structural Genomics of Infectious Diseases, CSGID,	97.35	0.0059	27.91	9.4	78	88
<input type="checkbox"/>	15	3F6W_C	XRE-family like protein; helix-turn-helix DNA binding protein xenobiotic response element family of transcriptional regu	97.33	0.0051	26.78	8.5	74	83
<input type="checkbox"/>	16	7XI1_A	anti-CRISPR protein AcrIF24; Anti-CRISPR, Acr; Inhibitor of type I-F Cascade, IMMUNE SYSTEM; 2.53A (Pseudomonas aerugin	97.32	0.017	32.59	8.3	59	236
<input type="checkbox"/>	17	8JFQ_C	AcrIIA15; IIA type anti-crispr protein, VIRAL PROTEIN; 2.3A (Staphylococcus delphini)	97.26	0.0072	31.82	5.9	61	171
<input type="checkbox"/>	18	6B9T_D	Methylphosphonate synthase; Phosphonate, Methylphosphonate, Iron, OXIDOREDUCTASE; HET: FMT, 2HE; 2.35A (Nitrosopumilus m	97.26	0.033	34.51	9.5	75	457

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 ?

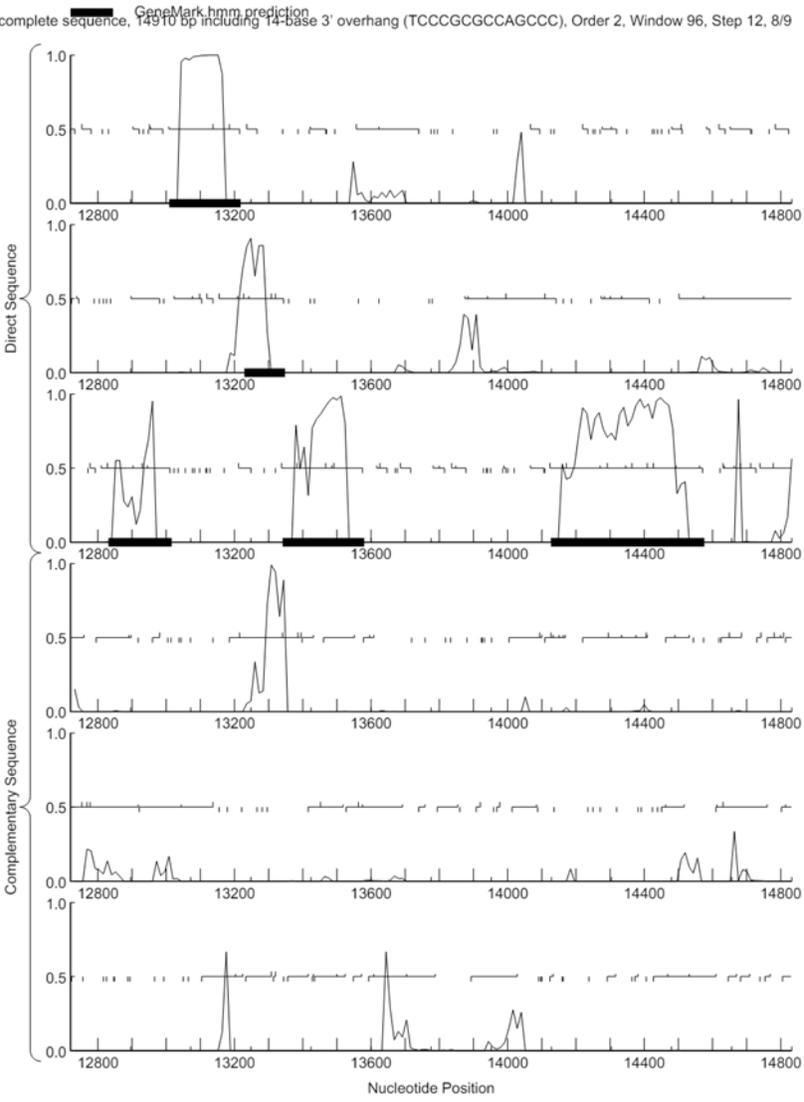
Oui le gene avant = integrase et le gene apres helix turn helix

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>DeepTMHMM - Predictions</p> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p> 
<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>Immunity repressor</p>

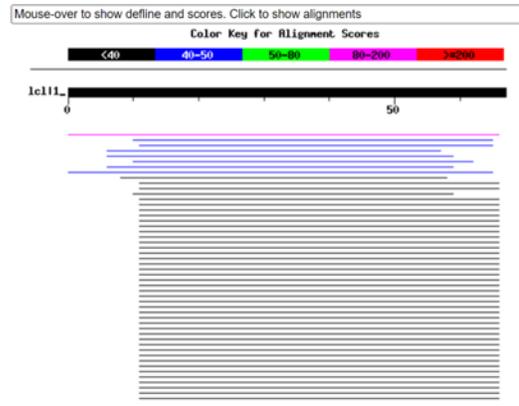
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	19
Coordonnées du Stop	13016
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Gap 103
Coordonnées du Start retenu	12813
Fonction prédite	excise

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>anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCGCGCCAGCCC), Order 2, Window 96, Step 12, 8/9</p>  <p>The figure displays a GeneMark.hmm prediction plot for a DNA sequence. The x-axis represents the Nucleotide Position, ranging from 12800 to 14800. The y-axis represents signal intensity, ranging from 0.0 to 1.0. The plot is divided into two main sections: Direct Sequence (top) and Complementary Sequence (bottom). The Direct Sequence section shows a prominent peak around 13200 bp, indicating a potential coding region. The Complementary Sequence section shows a corresponding peak around 13200 bp, suggesting a potential coding region on the opposite strand. The GeneMark.hmm prediction bar at the top indicates a predicted coding region from approximately 13100 to 13300 bp.</p>

Distribution of 100 Blast Hits on the Query Sequence



Sequences producing significant alignments:

	Score	E
	(bits)	Value
CyranoPS_Draft_19, function unknown, 67	136	2e-32
Gusanita_41, helix-turn-helix DNA binding domain protein, 69	48	8e-06
Emperor_22, helix-turn-helix DNA binding protein, 65	48	8e-06
SallySpecial_19, helix-turn-helix DNA binding protein, 63	48	1e-05
GMA5_20, function unknown, 59	47	2e-05
Idaho_17, DNA binding protein, 85	41	0.001
GRU3_20, function unknown, 62	40	0.002
Elesar_38, helix-turn-helix DNA-binding protein, 84	40	0.002

Sequences producing significant alignments

Download Select columns Show 100

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> hypothetical protein VHI5_05880 [Corynebacterium ulcerans]	Corynebacterium ulcerans	78.2	78.2	91%	5e-17	60.66%	62	KKOR7258.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Brachybacterium caraeonoplomeratum]	Brachybacterium caraeonoplomeratum	70.9	70.9	83%	4e-14	58.93%	62	WP_269361840.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Dermabacter linivensis]	Dermabacter linivensis	61.2	61.2	92%	2e-10	50.00%	66	WP_096882434.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Cellulomonas so_A375-1]	Cellulomonas so_A375-1	60.5	60.5	82%	5e-10	54.55%	61	WP_048343072.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Agrococcus casei]	Agrococcus casei	60.1	60.1	85%	6e-10	50.88%	60	WP_086992702.1
<input checked="" type="checkbox"/> hypothetical protein CWC39_02120 [Corynebacterium heidelbergense]	Corynebacterium heidelbergense	60.1	60.1	76%	8e-10	58.82%	60	RAV34623.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [bacterium]	bacterium	60.1	60.1	89%	1e-09	56.67%	72	MBK8168118.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Devriesea agamanum]	Devriesea agamanum	59.3	59.3	86%	2e-09	51.72%	71	WP_058234152.1
<input checked="" type="checkbox"/> hypothetical protein [Mycobacterium so.]	Mycobacterium so.	58.9	58.9	83%	2e-09	50.00%	62	MDZ2826881.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Microcella so.]	Microcella so.	57.8	57.8	77%	5e-09	53.85%	56	MDO9590038.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	83%	5e-09	53.57%	59	MBK8168434.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	80%	7e-09	59.26%	66	MBK8168292.1
<input checked="" type="checkbox"/> DNA-binding protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	57.0	57.0	94%	1e-08	49.21%	62	CAB0923827.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Glutamicibacter soil]	Glutamicibacter soil	57.0	57.0	85%	1e-08	45.61%	63	WP_161450204.1
<input checked="" type="checkbox"/> TPA: helix-turn-helix domain-containing protein [Acidothermaceae bacterium]	Acidothermaceae bacterium	56.2	56.2	74%	3e-08	54.00%	60	HXR41760.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	56.2	56.2	85%	3e-08	45.61%	62	WP_157759411.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Phytoactinoochyspora limicola]	Phytoactinoochyspora limicola	56.6	56.6	77%	4e-08	51.92%	101	WP_166345202.1
<input checked="" type="checkbox"/> hypothetical protein HMPREF2946_07410 [Actinomyces so_HMSC062G12]	Actinomyces so_HMSC062G12	56.6	56.6	83%	4e-08	48.21%	101	QFQ23805.1
<input checked="" type="checkbox"/> TPA: helix-turn-helix domain-containing protein [Nitrospirales bacterium]	Nitrospirales bacterium	55.8	55.8	97%	5e-08	44.62%	70	HEX9285267.1
<input checked="" type="checkbox"/> excisionase and transcriptional regulator [Gordonia chape_GMA5]	Gordonia chape_GMA5	47.0	47.0	80%	1e-04	37.04%	59	YP_009273617.1
<input checked="" type="checkbox"/> helix-turn-helix DNA binding domain protein [Arthrobacter chape_Gusanita]	Arthrobacter chape_Gusanita	47.0	47.0	83%	1e-04	44.64%	69	WNM72261.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Arthrobacter woluwensis]	Arthrobacter woluwensis	46.6	46.6	79%	2e-04	43.40%	61	WP_208189087.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [uncultured Caudovirales chape]	uncultured Caudovirales chape	46.2	46.2	70%	2e-04	51.06%	47	CAB4136963.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Clavibacter michiganensis]	Clavibacter michiganensis	46.6	46.6	92%	2e-04	37.10%	69	WP_220015090.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Mycobacterium aquaticum]	Mycobacterium aquaticum	46.6	46.6	76%	2e-04	45.10%	68	WP_158087256.1
<input checked="" type="checkbox"/> hypothetical protein BST13_35350 [Mycobacterium aquaticum]	Mycobacterium aquaticum	46.6	46.6	76%	2e-04	45.10%	72	ORA23429.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Arthrobacter so_LUM1]	Arthrobacter so_LUM1	47.0	47.0	92%	2e-04	37.50%	82	WP_226339756.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Ornithinimicrobium lamniarum]	Ornithinimicrobium lamniarum	46.6	46.6	79%	2e-04	45.28%	72	MBW8171822.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Rhodococcus so_EPR-147]	Rhodococcus so_EPR-147	46.2	46.2	77%	3e-04	46.15%	66	WP_068047953.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Rhodococcus]	Rhodococcus	45.8	45.8	77%	3e-04	46.15%	58	WP_080732798.1
<input checked="" type="checkbox"/> DNA-binding protein [Clavibacter michiganensis subsp. michiganensis]	Clavibacter michiganensis subsp. michiganensis	46.2	46.2	92%	3e-04	37.10%	79	MBW8025326.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Rothia nasimurum]	Rothia nasimurum	46.2	46.2	86%	3e-04	40.00%	69	WP_237296078.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Jamia majanobamensis]	Jamia majanobamensis	45.4	45.4	76%	4e-04	47.06%	62	WP_272737436.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Glutamicibacter halohydrocola]	Glutamicibacter halohydrocola	45.4	45.4	76%	4e-04	39.22%	62	WP_171919166.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Rhodococcus fascians]	Rhodococcus fascians	45.4	45.4	88%	6e-04	33.90%	65	WP_037165103.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Rhodococcus fascians]	Rhodococcus fascians	45.1	45.1	85%	6e-04	40.35%	58	WP_179277582.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Streptomyces so_BK022]	Streptomyces so_BK022	47.8	47.8	71%	6e-04	40.42%	235	WP_130335191.1
<input checked="" type="checkbox"/> hypothetical protein [uncultured Microbacterium so.]	uncultured Microbacterium so.	44.7	44.7	71%	7e-04	43.75%	51	WP_298038354.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Rhodococcus fascians]	Rhodococcus fascians	45.1	45.1	85%	8e-04	40.35%	65	WP_094650624.1
<input checked="" type="checkbox"/> TPA: helix-turn-helix domain-containing protein [Acidothermaceae bacterium]	Acidothermaceae bacterium	45.1	45.1	83%	8e-04	39.29%	64	HEY5201640.1
<input checked="" type="checkbox"/> TPA: hypothetical protein [Acidimicrobiales bacterium]	Acidimicrobiales bacterium	44.7	44.7	73%	0.001	46.94%	65	HET8917187.1
<input checked="" type="checkbox"/> TPA: Pyropin activator protein PTH [Caudovirales so.]	Caudovirales so.	45.4	45.4	70%	0.001	42.55%	90	DA086406.1
<input checked="" type="checkbox"/> helix-turn-helix domain-containing protein [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	44.7	44.7	76%	0.001	41.18%	66	WP_257418416.1
<input checked="" type="checkbox"/> TPA: helix-turn-helix domain-containing protein [Streptosporangiales bacterium]	Streptosporangiales bacterium	44.7	44.7	86%	0.001	36.21%	66	HXJ24621.1

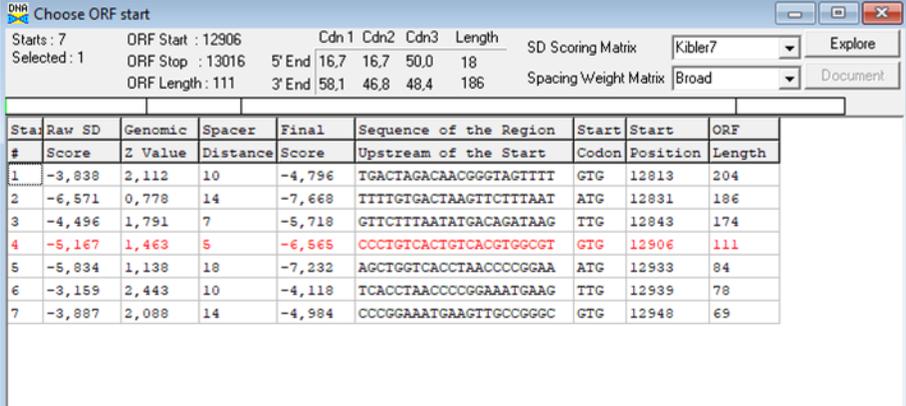
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?

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

NON

DECISION:	OUI
------------------	------------

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :12813 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :12831</p>
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	
<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>	<p><i>Le start de GeneMark (12831) est un ATG. Le start de glimmer (12813) est un GTG mais il conduit à une ORF plus longue, un chevauchement moindre et a un meilleur score.</i></p> <p>(mais hit identiques avec l'autre start) :</p> 
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>—</p>

<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	<p>Download GenPept Graphics Next Previous Descriptions</p> <p>hypothetical protein VH15_05880 [Corynebacterium ulcerans] Sequence ID: KK087258.1 Length: 62 Number of Matches: 1</p> <p>Range 1: 1 to 61 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>78.2 bits(191)</td> <td>5e-17</td> <td>Compositional matrix adjust.</td> <td>37/61(61%)</td> <td>45/61(73%)</td> <td>0/61(0%)</td> </tr> </tbody> </table> <pre> Query 7 MTDKLLTAEVAERLGI SPVTVTRVRAGHLTPMKLPGVNGAYLFPASIDQLTKETQS 66 M++KLLTTA VA+ LG+ TVTWR R G LTP MKLPG+NGAYLF+P I+QL + Sbjct 1 MSKLLTTAAVAKILGLPHSTVTRWRARTGRLLTPMKLPGINGAYLFDPREIEQLASKKGP 60 Query 67 A 67 A Sbjct 61 A 61 </pre>	Score	Expect	Method	Identities	Positives	Gaps	78.2 bits(191)	5e-17	Compositional matrix adjust.	37/61(61%)	45/61(73%)	0/61(0%)
	Score	Expect	Method	Identities	Positives	Gaps							
78.2 bits(191)	5e-17	Compositional matrix adjust.	37/61(61%)	45/61(73%)	0/61(0%)								
<p>Download GenPept Graphics Next Previous Descriptions</p> <p>helix-turn-helix domain-containing protein [Brachy bacterium paraconglomeratum] Sequence ID: WP_269361840.1 Length: 62 Number of Matches: 1 See 1 more title(s) See all Identical Proteins (IPG)</p> <p>Range 1: 3 to 58 GenPept Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>70.9 bits(172)</td> <td>4e-14</td> <td>Compositional matrix adjust.</td> <td>33/56(59%)</td> <td>39/56(69%)</td> <td>0/56(0%)</td> </tr> </tbody> </table> <pre> Query 8 TDKLLTAEVAERLGI SPVTVTRVRAGHLTPMKLPGVNGAYLFPASIDQLTKE 63 + KLLT EVAERLGI+ TV+HRV G L P KLPNGAYLF+ +D+L E Sbjct 3 SSKLLTWEVAERLGIARQTVSRVARGQLPEVQKLPGRNGAYLFDEEQDRLAAE 58 </pre> <p style="text-align: right;">Related Information Identical Proteins - Identical proteins to WP_269361840.1</p>	Score	Expect	Method	Identities	Positives	Gaps	70.9 bits(172)	4e-14	Compositional matrix adjust.	33/56(59%)	39/56(69%)	0/56(0%)	
Score	Expect	Method	Identities	Positives	Gaps								
70.9 bits(172)	4e-14	Compositional matrix adjust.	33/56(59%)	39/56(69%)	0/56(0%)								
DECISION:	12813												

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

Rationnelle

Listez le meilleur hit Blastp pour chaque source :

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

[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	80-200	≥200
-----	-------	-------	--------	------

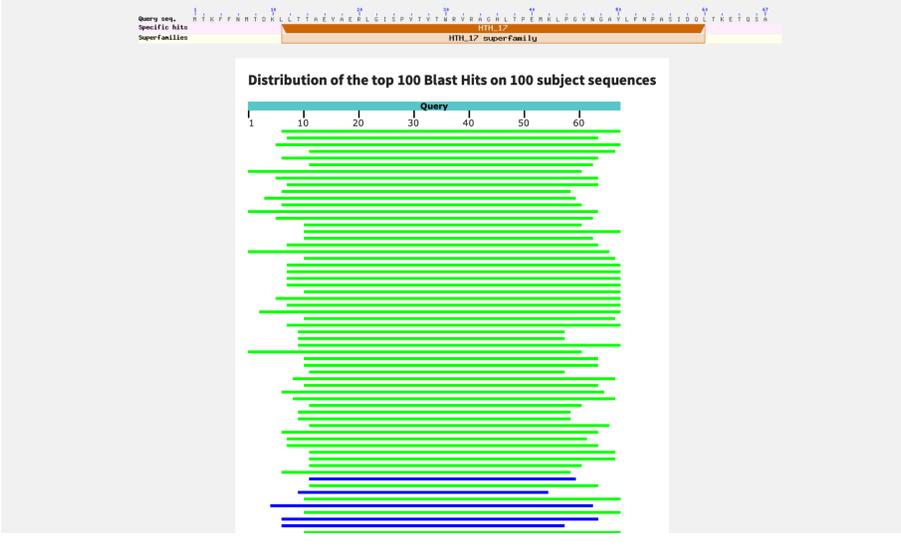
Sequences producing significant alignments:	Score	E
	(bits)	Value
Cyranops_Draft_19, function unknown, 67	136	2e-32
Gusanita_41, helix-turn-helix DNA binding domain protein, 69	48	8e-06
Emperor_22, helix-turn-helix DNA binding protein, 65	48	8e-06
SallySpecial_19, helix-turn-helix DNA binding protein, 63	48	1e-05
GMA5_20, function unknown, 59	47	2e-05

***** nr :**

<input checked="" type="checkbox"/>	hypothetical protein VHI15_05880 [Corynebacterium ulcerans]	Corynebacterium ulcerans	78.2	78.2	91%	5e-17	60.66%	62	KKO87258.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Brachybacterium paraconglomeratum]	Brachybacterium paraconglomeratum	70.9	70.9	83%	4e-14	58.93%	62	WP_269361840.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Dermabacter linjuensis]	Dermabacter linjuensis	61.2	61.2	92%	2e-10	50.00%	66	WP_086882434.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Cellulomonas sp. A375-1]	Cellulomonas sp. A375-1	60.5	60.5	82%	6e-10	54.55%	61	WP_048343072.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Agrococcus casei]	Agrococcus casei	60.1	60.1	85%	6e-10	50.88%	60	WP_086927202.1
<input checked="" type="checkbox"/>	hypothetical protein CWC39_02120 [Corynebacterium heidelbergense]	Corynebacterium heidelbergense	60.1	60.1	76%	8e-10	58.82%	60	RAV34823.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [bacterium]	bacterium	60.1	60.1	89%	1e-09	56.67%	72	MBK8168118.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Devriesea agamarum]	Devriesea agamarum	59.3	59.3	86%	2e-09	51.72%	71	WP_058234152.1
<input checked="" type="checkbox"/>	hypothetical protein [Mycobacterium sp.]	Mycobacterium sp.	58.9	58.9	83%	2e-09	50.00%	62	MOF2828881.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Microcella sp.]	Microcella sp.	57.8	57.8	77%	5e-09	53.85%	56	MDO9590038.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	83%	5e-09	53.57%	59	MBK8168434.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	80%	7e-09	59.26%	66	MBK8168292.1
<input checked="" type="checkbox"/>	DNA-binding protein [Corynebacterium dioltheriae]	Corynebacterium dioltheriae	57.0	57.0	94%	1e-08	49.21%	62	CAB0923827.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Glutamicibacter soli]	Glutamicibacter soli	57.0	57.0	85%	1e-08	45.61%	63	WP_161450204.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain-containing protein [Acidothermaceae bacterium]	Acidothermaceae bacterium	56.2	56.2	74%	3e-08	54.00%	60	HXR41760.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	56.2	56.2	85%	3e-08	45.61%	62	WP_157759411.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Phytoaetionoplyspora limicola]	Phytoaetionoplyspora limicola	56.6	56.6	77%	4e-08	51.92%	101	WP_166345202.1
<input checked="" type="checkbox"/>	hypothetical protein HMPREF2946_07410 [Actinomyces sp. HMSC062G12]	Actinomyces sp. HMSC062G12	56.6	56.6	83%	4e-08	48.21%	101	OFQ238085.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain-containing protein [Nitrospirales bacterium]	Nitrospirales bacterium	55.8	55.8	97%	5e-08	44.62%	70	HEX9285267.1
<input checked="" type="checkbox"/>	hypothetical protein [Leifsonia sp.]	Leifsonia sp.	55.5	55.5	83%	5e-08	50.00%	59	MAT117174.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [uncultured Actinomyces sp.]	uncultured Actinomyces sp.	56.2	56.2	89%	6e-08	43.33%	104	WP_297911466.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	56.2	56.2	89%	7e-08	43.33%	106	DAX62379.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Isotriecola variabilis]	Isotriecola variabilis	56.2	56.2	89%	7e-08	43.33%	104	MBF1252321.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	56.2	56.2	89%	8e-08	43.33%	106	DAR87009.1
<input checked="" type="checkbox"/>	hypothetical protein CBR64_00030 [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	55.8	55.8	85%	1e-07	45.61%	106	ARU50140.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	54.3	54.3	92%	2e-07	40.32%	67	WP_156532974.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	55.1	55.1	89%	2e-07	41.67%	106	DAN08877.1
<input checked="" type="checkbox"/>	hypothetical protein A0J59_14960 [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	54.3	54.3	97%	2e-07	38.46%	73	KZM77960.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	55.1	55.1	89%	2e-07	41.67%	106	DAN08877.1
<input checked="" type="checkbox"/>	hypothetical protein A0J59_14960 [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	54.3	54.3	97%	2e-07	38.46%	73	KZM77960.1
<input checked="" type="checkbox"/>	hypothetical protein [Microbacterium lacticum]	Microbacterium lacticum	53.5	53.5	83%	3e-07	46.43%	61	WP_314429462.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	54.7	54.7	89%	3e-07	41.67%	106	DAO74759.1
<input checked="" type="checkbox"/>	hypothetical protein GCM10025863_06820 [Microbacterium suwonense]	Microbacterium suwonense	53.5	53.5	71%	4e-07	58.33%	72	BZ38088.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Microbacterium suwonense]	Microbacterium suwonense	53.5	53.5	71%	4e-07	58.33%	70	WP_286303025.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain-containing protein [Rhodogobius sp.]	Rhodogobius sp.	53.1	53.1	86%	5e-07	49.18%	66	HOW00973.1
<input checked="" type="checkbox"/>	Helix-turn-helix domain-containing protein [Microbacterium sp. ru370.1]	Microbacterium sp. ru370.1	53.1	53.1	89%	5e-07	45.00%	71	SDO77819.1
<input checked="" type="checkbox"/>	DNA-binding protein [Actinomycetota bacterium]	Actinomycetota bacterium	52.4	52.4	79%	1e-06	52.83%	69	RPI11310.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Rhodococcus sp. FXJ9.536]	Rhodococcus sp. FXJ9.536	52.0	52.0	79%	1e-06	47.17%	60	WP_255969728.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Curtobacterium sp. VKM Ac-1376]	Curtobacterium sp. VKM Ac-1376	52.0	52.0	68%	1e-06	52.17%	62	WP_194626840.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	51.6	51.6	86%	2e-06	43.10%	64	DAI53416.1
<input checked="" type="checkbox"/>	hypothetical protein [Streptoporangiaceae bacterium]	Streptoporangiaceae bacterium	51.2	51.2	79%	2e-06	49.06%	61	MCW2900027.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Corynebacterium auriscanis]	Corynebacterium auriscanis	52.0	52.0	86%	2e-06	46.55%	80	WP_282939148.1
<input checked="" type="checkbox"/>	hypothetical protein N579_08860 [Corynebacterium pseudodiphthericum 090104]	Corynebacterium pseudodiphthericum 0...	51.2	51.2	86%	2e-06	43.10%	64	ERJ43622.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [unclassified Microbacterium]	unclassified Microbacterium	51.2	51.2	73%	3e-06	53.06%	59	WP_143016360.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Dermabacter sp. HSID17554]	Dermabacter sp. HSID17554	51.2	51.2	73%	3e-06	51.02%	64	RUP86599.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Dermabacter sp. HSID17554]	Dermabacter sp. HSID17554	50.8	50.8	73%	3e-06	51.02%	62	WP_161970202.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Betaproteobacteria bacterium]	Betaproteobacteria bacterium	50.8	50.8	80%	3e-06	49.09%	59	MBK6602370.1
<input checked="" type="checkbox"/>	hypothetical protein [Microbacterium sp. GCS4]	Microbacterium sp. GCS4	50.4	50.4	85%	4e-06	45.61%	59	WP_152926074.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Gulosibacter bifidus]	Gulosibacter bifidus	50.4	50.4	80%	4e-06	40.74%	66	WP_110477041.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Acidipropionibacterium timonense]	Acidipropionibacterium timonense	50.8	50.8	83%	5e-06	46.43%	70	WP_130864622.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Gordonia westfalica]	Gordonia westfalica	50.4	50.4	82%	5e-06	47.27%	65	WP_159441558.1
<input checked="" type="checkbox"/>	DNA binding domain-containing protein, excisionase family [Gordonia westfalica]	Gordonia westfalica	50.4	50.4	82%	5e-06	47.27%	74	SDU50629.1
<input checked="" type="checkbox"/>	hypothetical protein D1609_07700 [Corynebacterium urealyticum]	Corynebacterium urealyticum	50.4	50.4	73%	6e-06	51.02%	77	PZO99721.1
<input checked="" type="checkbox"/>	excisionase and transcriptional regulator [Gordonia phage SallySpecial]	Gordonia phage SallySpecial	50.1	50.1	77%	7e-06	50.00%	63	YP_010674640.1
<input checked="" type="checkbox"/>	TPA: helix-turn-helix domain-containing protein [Propionibacteriaceae bacterium]	Propionibacteriaceae bacterium	49.7	49.7	71%	8e-06	47.92%	63	HEY5482789.1
<input checked="" type="checkbox"/>	helix-turn-helix domain-containing protein [Metallococcus cam]	Metallococcus cam	50.1	50.1	77%	9e-06	44.23%	73	NHN55809.1

hover to see the title click to show alignments Show Conserved Domains Alignment Scores < 40 40 - 50 50 - 80 80 - 200 >= 200

100 sequences selected Putative conserved domains have been detected, click on the image below for detailed results.



PFAM : DUF1233 ; Putative excisionase (DUF1233)

Cd avec proba : HTH_MerR-like; Helix-Turn-Helix DNA binding domain of MerR-like transcription regulators. Helix-turn-helix (HTH) MerR-like transcription regulator, N-terminal domain.

PDB : AlpA family phage regulatory protein; excisionase, mobile genetic elements, recombination, DNA BINDING PROTEIN; 2.11A {A

Uniprot : VXIS_BPSF5 Excisionase OS=Shigella phage SfV OX=55884 GN=xis PE=3 SV=1



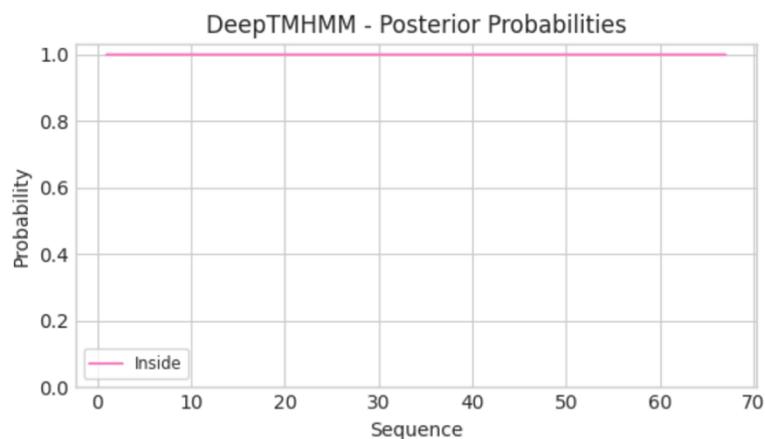
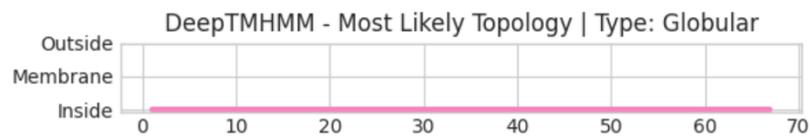
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	5FD4_B	ComR; Streptococcus, Competence, Quorum sensing, ComR, TRANSCRIPTION REGULATORY; 2.9A (Streptococcus suis (strain 05ZYH33	98.46	0.000022	45.11	9.5	69	324
<input type="checkbox"/>	P15238	RPC_BP163 Repressor protein C OS=Rhizobium phage 16-3 OX=10704 GN=C PE=1 SV=4	98.39	0.000034	43.09	8.9	65	263
<input type="checkbox"/>	6B9T_D	Methylphosphonate synthase; Phosphonate, Methylphosphonate, Iron, OXIDOREDUCTASE; HET: FMT, 2HE; 2.35A (Nitrosopumilus m	98.22	0.00011	44.41	9	65	457
<input type="checkbox"/>	6WPZ_B	Pf4r; DNA BINDING PROTEIN; 1.993A (Pseudomonas aeruginosa)	97.94	0.00057	32.12	9	90	95
<input type="checkbox"/>	P18680	RPC1_BPHK0 26 kDa repressor protein OS=Escherichia phage HK022 OX=10742 GN=CI-HTT PE=4 SV=2	97.93	0.00051	37.76	7.5	68	235
<input type="checkbox"/>	6H49_A	Orf20; SaPI, Repressor, STRUCTURAL PROTEIN; HET: SO4; 1.8A (Staphylococcus aureus)	97.92	0.001	34.13	9.2	78	157
<input type="checkbox"/>	6XGT_B	Cyanate hydratase; cyanase, HYDROLASE; HET: MLI, FMT; 2.2A (Thermomyces lanuginosus)	97.78	0.0024	34.06	9.3	74	181

<input type="checkbox"/>	8	6G1T_A	AM32; Repressor, Protein-DNA complex, Type IV secretion system, DNA BINDING PROTEIN; HET: EPE; 1.93A (Enterococcus faeca	98.94	1.9e-8	50	6.1	62	122
<input type="checkbox"/>	9	PF06806.16	; DUF1233 ; Putative excisionase (DUF1233)	98.93	3.1e-8	43.81	6.1	61	70
<input type="checkbox"/>	10	2KWV_A	Putative excisionase; excisionase, DNA binding, wHTH motif, Structural Genomics, NORTHEAST STRUCTURAL GENOMICS CONSORTIU	98.91	4.7e-8	45.1	6.4	66	78
<input type="checkbox"/>	11	P51705	VCOX_BPHC1 Regulatory protein cox OS=Haemophilus phage HP1 (strain HP1c1) OX=1289570 GN=cox PE=4 SV=1	98.9	1.1e-7	42.49	7.3	67	79
<input type="checkbox"/>	12	PF04936.16	; DUF658 ; Protein of unknown function (DUF658)	98.88	5.8e-8	51.11	6.9	56	186
<input type="checkbox"/>	13	6HLK_A	Redirecting phage packaging protein C (RppC); Redirecting packaging protein, DNA Binding protein, Homo-dimer, phage inte	98.86	8e-8	48.65	6.8	59	153
<input type="checkbox"/>	14	7LWR_A	Terminase, small subunit; DNA Packaging, Terminase, VIRAL PROTEIN; 2.35A (Enterobacteria phage P21}	98.84	1.8e-7	38.48	6.6	54	54
<input type="checkbox"/>	15	PF11112.12	; PyocinActivator ; Pyocin activator protein PrtN	98.81	1.4e-7	43	6.2	59	73
<input type="checkbox"/>	16	P25135	VG090_BPPF1 10.1 kDa protein OS=Pseudomonas phage Pf1 OX=2011081 PE=4 SV=1	98.8	1.7e-7	43.87	6.4	63	90
<input type="checkbox"/>	17	P21681	VAPL_BP186 Protein apl OS=Escherichia phage 186 OX=29252 GN=apl PE=4 SV=1	98.79	1.1e-7	44.62	5.6	67	87

Oui le gene avant = immunity repressor ou helix-turn-helix transcriptional regulator

DeepTMHMM - Predictions

Predicted topologies can be downloaded in [.gff3 format](#) and [.3line format](#)



Oui mais à voir comment le nommer

excise

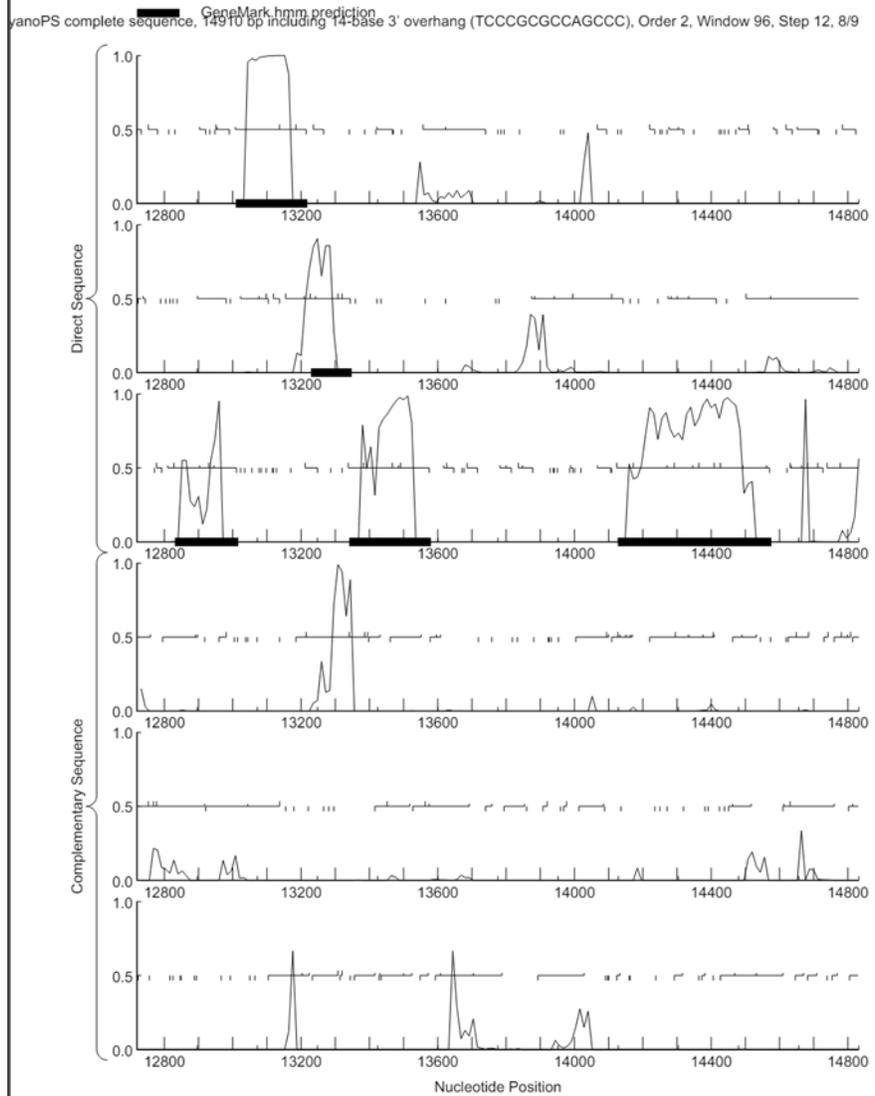
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	20
Coordonnées du Stop	13218
Direction (For/Rev)	for
Gap/chevauchement avec un autre gène	Ove 8
Coordonnées du Start retenu	13009
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

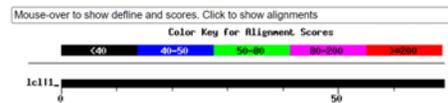
Y-a-t-il des éléments supportant un potentiel codant ?



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

Query: CyranoPS_20
(69 letters)

Distribution of 1 Blast Hits on the Query Sequence



Sequences producing significant alignments:

Score	E
(bits)	Value
148	9e-34

CyranoPS_Draft_20, function unknown, 69
Length = 69

>CyranoPS_Draft_20, function unknown, 69
Length = 69
Score = 148 bits (354), Expect = 9e-34
Identities = 69/69 (100%), Positives = 69/69 (100%)

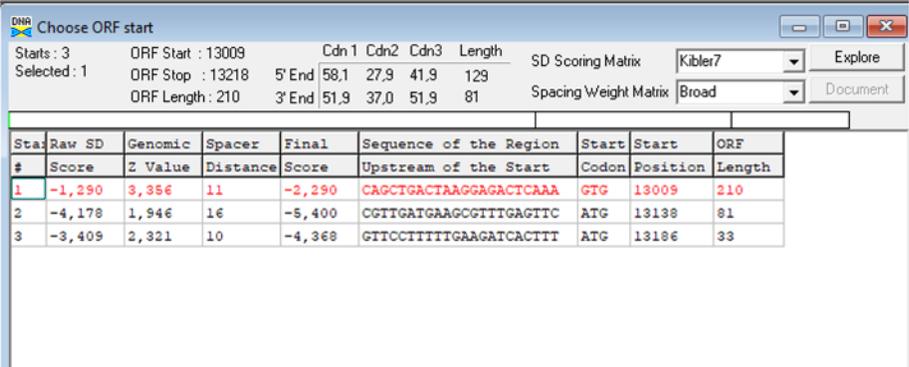
Query: 1 MLKEDRDMAIKLTASILEGSIHQREIENSKYDLEAIVDEAFEPFKETGTGVLFLFEDHFM 60
MLKEDRDMAIKLTASILEGSIHQREIENSKYDLEAIVDEAFEPFKETGTGVLFLFEDHFM 60
Sbjct: 1 MLKEDRDMAIKLTASILEGSIHQREIENSKYDLEAIVDEAFEPFKETGTGVLFLFEDHFM 60

Query: 61 EIARKHRR 69
EIARKHRR 69
Sbjct: 61 EIARKHRR 69

No significant similarity found. For reasons why [click here](#)

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 ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 13009 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 13009																																													
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <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>-1,290</td> <td>3,356</td> <td>11</td> <td>-2,290</td> <td>CAGCTGACTAAGGAGACTCAA</td> <td>GTG</td> <td>13009</td> <td>210</td> </tr> <tr> <td>2</td> <td>-4,178</td> <td>1,946</td> <td>16</td> <td>-5,400</td> <td>CGTTGATGAAGCGTTTGGAGTTC</td> <td>ATG</td> <td>13138</td> <td>81</td> </tr> <tr> <td>3</td> <td>-3,409</td> <td>2,321</td> <td>10</td> <td>-4,368</td> <td>GTTCCTTTTTGAAGATCACTTT</td> <td>ATG</td> <td>13186</td> <td>33</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	-1,290	3,356	11	-2,290	CAGCTGACTAAGGAGACTCAA	GTG	13009	210	2	-4,178	1,946	16	-5,400	CGTTGATGAAGCGTTTGGAGTTC	ATG	13138	81	3	-3,409	2,321	10	-4,368	GTTCCTTTTTGAAGATCACTTT	ATG	13186	33
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	-1,290	3,356	11	-2,290	CAGCTGACTAAGGAGACTCAA	GTG	13009	210																																						
2	-4,178	1,946	16	-5,400	CGTTGATGAAGCGTTTGGAGTTC	ATG	13138	81																																						
3	-3,409	2,321	10	-4,368	GTTCCTTTTTGAAGATCACTTT	ATG	13186	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 ?	ORF la plus longue avec le GTG 13009 ☐ 210pb , chevauchement de 8																																													
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 ?	Pas de hit																																													
DECISION:	START GTG ☐ meilleur score et ORF la plus longue et chevauchement ok																																													

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 : Rien

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

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 ?

<input type="checkbox"/>	1	2L1L_B	Exportin-1; Nuclear Export, PKI NES, CRM1, RangTP, NUCLEAR PROTEIN; NMR (Homo sapiens)	66.57	21	23.92	2.9	37	127
<input type="checkbox"/>	2	PF02108.20	; FliH; Flagellar assembly protein FliH	63.48	50	20.06	3.9	65	144
<input type="checkbox"/>	3	PF17657.5	; DNA_pol3_finger; Bacterial DNA polymerase III alpha subunit finger domain	62.74	17	22.61	1.9	17	164
<input type="checkbox"/>	4	PF16925.9	; TetR_C_13; Tetracyclin repressor-like, C-terminal domain	60.77	36	16.23	3.6	40	115
<input type="checkbox"/>	5	1SGM_A	Putative HTH-type transcriptional regulator yxaF; TRANSCRIPTION, Structural Genomics, PSI, Protein Structure Initiative,	57.99	60	17.94	3.9	44	191
<input type="checkbox"/>	6	PF06570.15	; DUF1129; Protein of unknown function (DUF1129)	54.17	66	21.34	3.6	42	201
<input type="checkbox"/>	7	PF17934.5	; TetR_C_26; Tetracyclin repressor-like, C-terminal domain	51.55	57	15.73	4.3	42	108
<input type="checkbox"/>	8	6AZH_A	TetR family transcriptional regulator; Regulator, TRANSCRIPTION; HET: MSE; 1.75A (Clostridium perfringens)	48.78	87	17.05	4	42	192

Resubmit Section

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

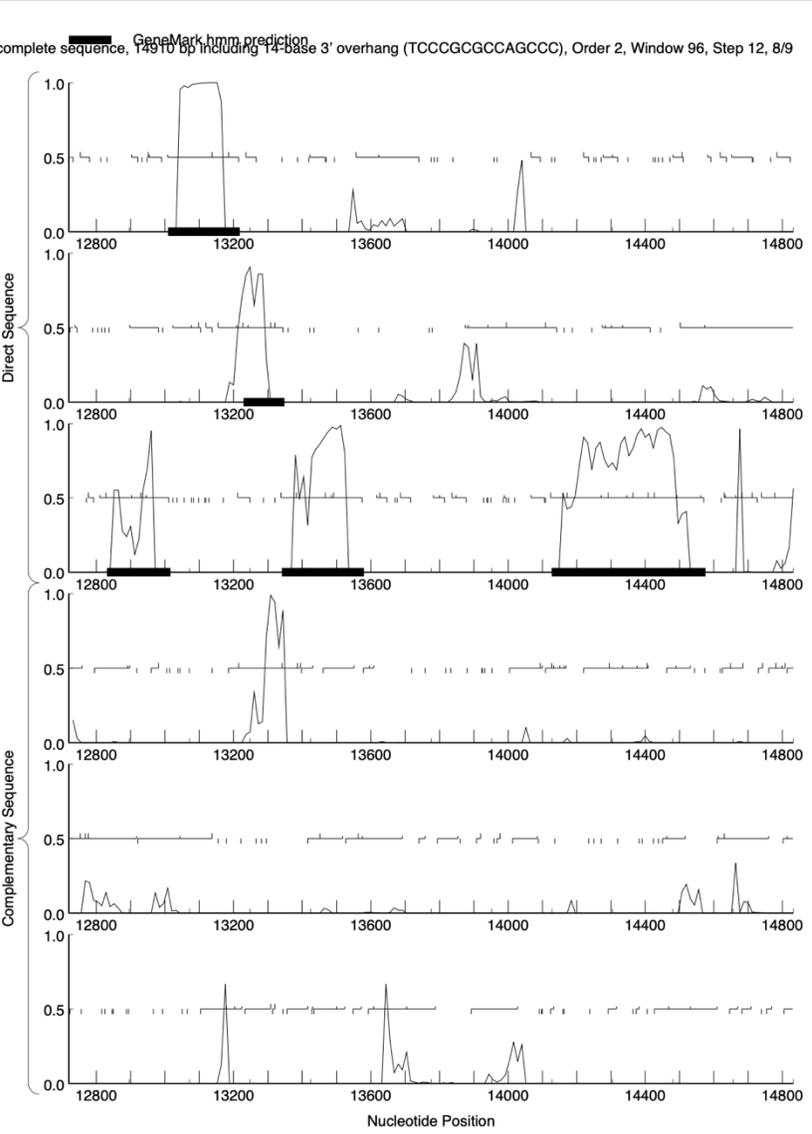
"Aucune synténie observée".

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 ?	non
DECISION:	NKF

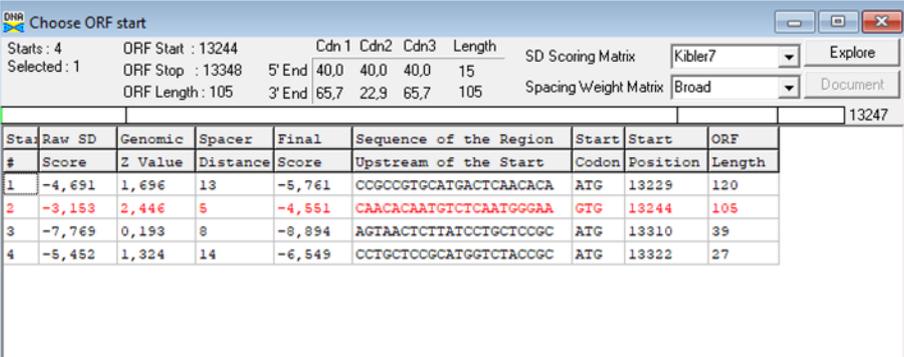
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	21
Coordonnées du Stop	13348
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 11
Coordonnées du Start retenu	13229
Fonction prédite	Membrane protein

Décision #1 : Est-ce un gène ?

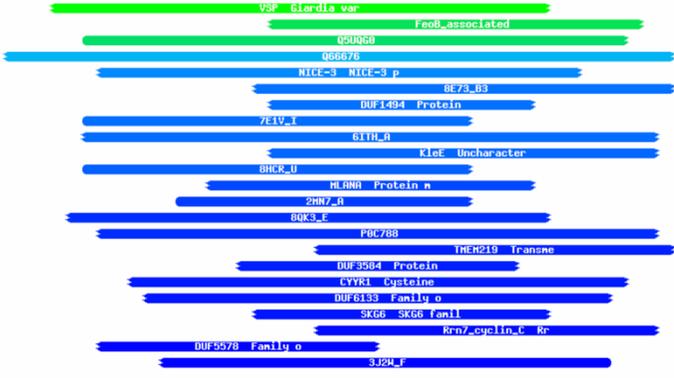
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	<p>GeneMarkS</p> <p>GenoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 8/9</p>  <p>The figure displays a GeneMarkS prediction plot for a DNA sequence. The x-axis represents the Nucleotide Position, ranging from 12800 to 14800. The y-axis represents the signal intensity, ranging from 0.0 to 1.0. The plot is divided into two main sections: Direct Sequence (top) and Complementary Sequence (bottom). Each section contains two sub-plots, one for each strand. The Direct Sequence plots show a significant peak between 13200 and 13600 bp, reaching a signal intensity of approximately 1.0. The Complementary Sequence plots show a corresponding peak in the same region, also reaching a signal intensity of approximately 1.0. The plot is annotated with GeneMarkS prediction results, showing a solid black bar indicating a predicted coding region. The plot is titled 'GenoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 8/9'.</p>
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	non
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	non
DECISION:	OUI

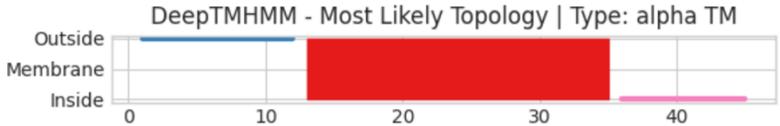
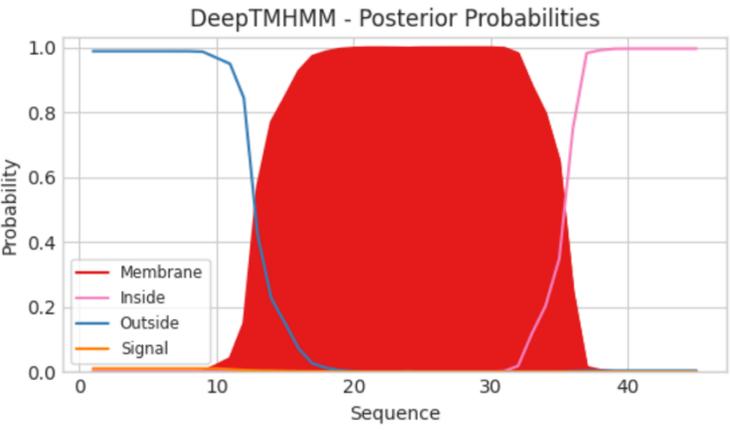
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle																																													
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 13211 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 13229																																													
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <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> </thead> <tbody> <tr> <td>1</td> <td>-4,691</td> <td>1,696</td> <td>13</td> <td>-5,761</td> <td>COGCOGTGCATGACTCAACACA</td> <td>ATG</td> <td>13229</td> <td>120</td> </tr> <tr> <td>2</td> <td>-3,153</td> <td>2,446</td> <td>5</td> <td>-4,551</td> <td>CAACACAATGTCTCAATGGGA</td> <td>GTG</td> <td>13244</td> <td>105</td> </tr> <tr> <td>3</td> <td>-7,769</td> <td>0,193</td> <td>8</td> <td>-8,894</td> <td>AGTAACTCTTATCCCTCCGCG</td> <td>ATG</td> <td>13310</td> <td>39</td> </tr> <tr> <td>4</td> <td>-5,452</td> <td>1,324</td> <td>14</td> <td>-6,549</td> <td>CCTGCTCCGCATGGTCTACCGC</td> <td>ATG</td> <td>13322</td> <td>27</td> </tr> </tbody> </table>	Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF	1	-4,691	1,696	13	-5,761	COGCOGTGCATGACTCAACACA	ATG	13229	120	2	-3,153	2,446	5	-4,551	CAACACAATGTCTCAATGGGA	GTG	13244	105	3	-7,769	0,193	8	-8,894	AGTAACTCTTATCCCTCCGCG	ATG	13310	39	4	-5,452	1,324	14	-6,549	CCTGCTCCGCATGGTCTACCGC	ATG	13322	27
Sta	Raw SD	Genomic	Spacer	Final	Sequence of the Region	Start	Start	ORF																																						
1	-4,691	1,696	13	-5,761	COGCOGTGCATGACTCAACACA	ATG	13229	120																																						
2	-3,153	2,446	5	-4,551	CAACACAATGTCTCAATGGGA	GTG	13244	105																																						
3	-7,769	0,193	8	-8,894	AGTAACTCTTATCCCTCCGCG	ATG	13310	39																																						
4	-5,452	1,324	14	-6,549	CCTGCTCCGCATGGTCTACCGC	ATG	13322	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 ?	ORF la plus longue : ATG 13 157 ☐ 192pb , chevauchement de 62pb ORf de 138pb GTG 13211 ☐ chevauchement 8pb ATG 13229 ☐ ORf de 120pb ☐ gap de 11pb																																													
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 ?	rien																																													
DECISION:	ATG 13229 , bon ORF et gap de 11 (a voir avec 13211 car meilleur score et ORF plus long)																																													

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 : rien</p> <p>*** nr : rien</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>PFAM proba de 70% : VSP ; Giardia variant-specific surface protein</p>  <table border="1"> <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"/></td> <td>1</td> <td>PF03302.17 ; VSP ; Giardia variant-specific surface protein</td> <td>70.05</td> <td>27</td> <td>21.55</td> <td>3.6</td> <td>33</td> <td>403</td> </tr> <tr> <td><input type="checkbox"/></td> <td>2</td> <td>PF12669.11 ; FeoB_associated ; FeoB-associated Cys-rich membrane protein</td> <td>64.92</td> <td>33</td> <td>17.18</td> <td>3.8</td> <td>25</td> <td>43</td> </tr> <tr> <td><input type="checkbox"/></td> <td>3</td> <td>Q5UQG0 YR486_MIMIV Putative PAN domain-containing protein R486 OS=Acanthamoeba polyphaga mimivirus OX=212035 GN=MIMI_R486 PE=1</td> <td>63.67</td> <td>46</td> <td>19.95</td> <td>3.7</td> <td>36</td> <td>223</td> </tr> <tr> <td><input type="checkbox"/></td> <td>4</td> <td>Q66676 VGE9_EHV2 Uncharacterized gene E9 protein OS=Equine herpesvirus 2 (strain 86/87) OX=82831 GN=E9 PE=3 SV=1</td> <td>55.57</td> <td>90</td> <td>21.11</td> <td>4.4</td> <td>44</td> <td>205</td> </tr> <tr> <td><input type="checkbox"/></td> <td>5</td> <td>PF07406.15 ; NICE-3 ; NICE-3 protein</td> <td>51.64</td> <td>90</td> <td>19.5</td> <td>3.7</td> <td>31</td> <td>178</td> </tr> <tr> <td><input type="checkbox"/></td> <td>6</td> <td>8E73_B3 NDUB3; respiratory supercomplex, nadh-cyt c oxidoreductase, membrane complex, ELECTRON TRANSPORT; HET: PC1, ZMP, FMN, CD</td> <td>49.73</td> <td>100</td> <td>18.18</td> <td>3.6</td> <td>28</td> <td>68</td> </tr> </tbody> </table>	Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	<input type="checkbox"/>	1	PF03302.17 ; VSP ; Giardia variant-specific surface protein	70.05	27	21.55	3.6	33	403	<input type="checkbox"/>	2	PF12669.11 ; FeoB_associated ; FeoB-associated Cys-rich membrane protein	64.92	33	17.18	3.8	25	43	<input type="checkbox"/>	3	Q5UQG0 YR486_MIMIV Putative PAN domain-containing protein R486 OS=Acanthamoeba polyphaga mimivirus OX=212035 GN=MIMI_R486 PE=1	63.67	46	19.95	3.7	36	223	<input type="checkbox"/>	4	Q66676 VGE9_EHV2 Uncharacterized gene E9 protein OS=Equine herpesvirus 2 (strain 86/87) OX=82831 GN=E9 PE=3 SV=1	55.57	90	21.11	4.4	44	205	<input type="checkbox"/>	5	PF07406.15 ; NICE-3 ; NICE-3 protein	51.64	90	19.5	3.7	31	178	<input type="checkbox"/>	6	8E73_B3 NDUB3; respiratory supercomplex, nadh-cyt c oxidoreductase, membrane complex, ELECTRON TRANSPORT; HET: PC1, ZMP, FMN, CD	49.73	100	18.18	3.6	28	68
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length																																																								
<input type="checkbox"/>	1	PF03302.17 ; VSP ; Giardia variant-specific surface protein	70.05	27	21.55	3.6	33	403																																																								
<input type="checkbox"/>	2	PF12669.11 ; FeoB_associated ; FeoB-associated Cys-rich membrane protein	64.92	33	17.18	3.8	25	43																																																								
<input type="checkbox"/>	3	Q5UQG0 YR486_MIMIV Putative PAN domain-containing protein R486 OS=Acanthamoeba polyphaga mimivirus OX=212035 GN=MIMI_R486 PE=1	63.67	46	19.95	3.7	36	223																																																								
<input type="checkbox"/>	4	Q66676 VGE9_EHV2 Uncharacterized gene E9 protein OS=Equine herpesvirus 2 (strain 86/87) OX=82831 GN=E9 PE=3 SV=1	55.57	90	21.11	4.4	44	205																																																								
<input type="checkbox"/>	5	PF07406.15 ; NICE-3 ; NICE-3 protein	51.64	90	19.5	3.7	31	178																																																								
<input type="checkbox"/>	6	8E73_B3 NDUB3; respiratory supercomplex, nadh-cyt c oxidoreductase, membrane complex, ELECTRON TRANSPORT; HET: PC1, ZMP, FMN, CD	49.73	100	18.18	3.6	28	68																																																								
<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>"Aucune synténie observée".</p>																																																															

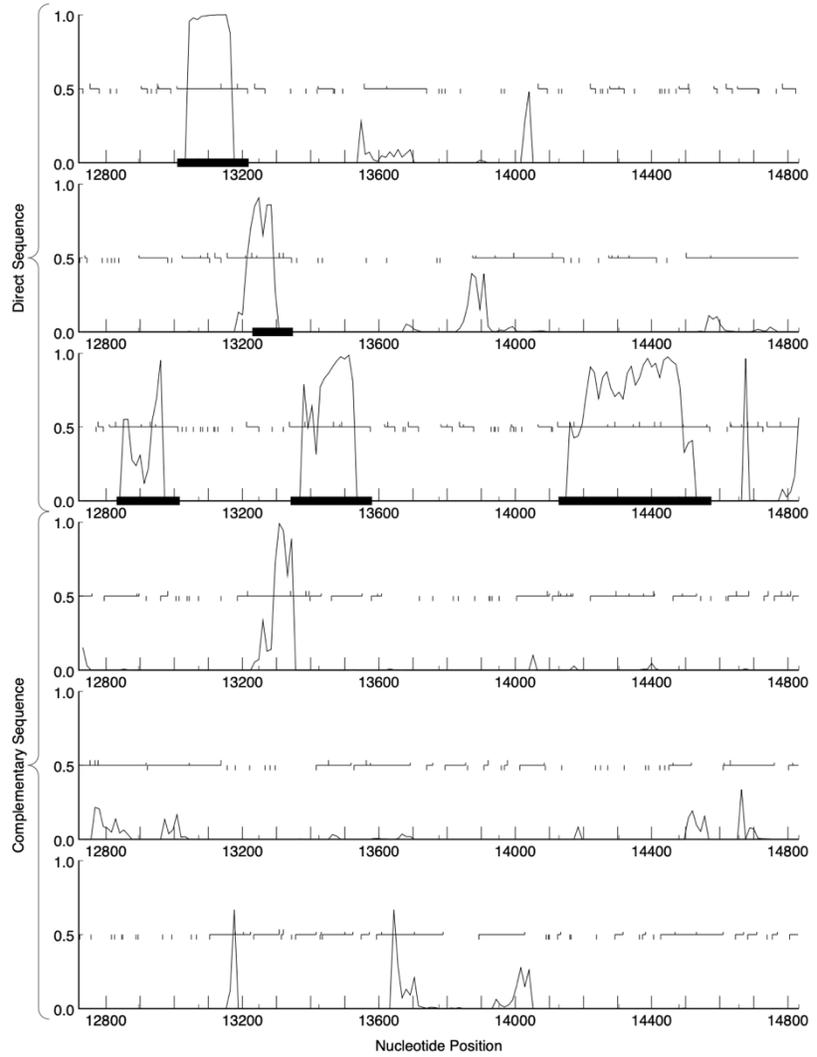
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<p>OUI.</p> <p>DeepTMHMM - Predictions</p> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p>  
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>non</p>
<p>DECISION:</p>	<p>Membrane protein</p>

<p>Basic Phage Information</p>	
<p>Nom du Phage</p>	<p>CyranoPS</p>
<p>Gène #</p>	<p>22</p>
<p>Coordonnées du Stop</p>	<p>13580</p>
<p>Direction (For/Rev)</p>	<p>for</p>
<p>Gap/chevauchement avec un autre gène</p>	<p>Ove 8</p>
<p>Coordonnées du Start retenu</p>	<p>13 341</p>
<p>Fonction prédite</p>	<p>NFK</p>

Décision #1 : Est-ce un gène ?

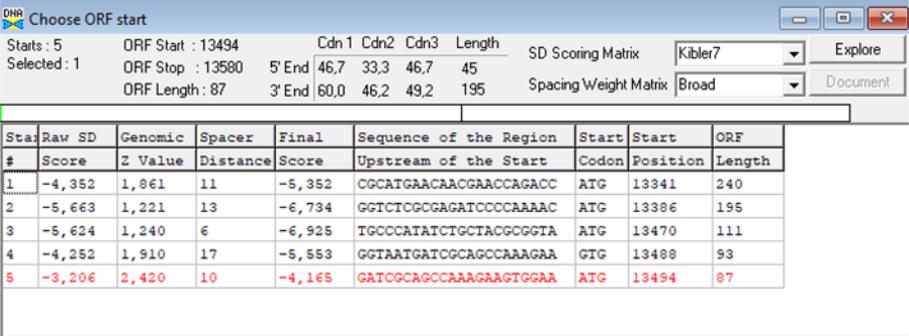
Collection des éléments de réponse	
	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	YES BOTH
Y-a-t-il des éléments supportant un potentiel codant ?	GeneMarkS :

GeneMark.hmm prediction
 yanoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 8/9



<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<p>rien</p>
<p>Est-ce que le candidat est en contradiction avec les principes d'annotation ?</p>	<p>Pas d'homologues</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 ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 13386 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 13341																																																						
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 <table border="1"> <thead> <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> </thead> <tbody> <tr> <td>1</td> <td>-4,352</td> <td>1,861</td> <td>11</td> <td>-5,352</td> <td>CGCATGAACAACGAACCAGACC</td> <td>ATG</td> <td>13341</td> <td>240</td> </tr> <tr> <td>2</td> <td>-5,663</td> <td>1,221</td> <td>13</td> <td>-6,734</td> <td>GGTCTGGGAGATCCCCAAAAC</td> <td>ATG</td> <td>13386</td> <td>195</td> </tr> <tr> <td>3</td> <td>-5,624</td> <td>1,240</td> <td>6</td> <td>-6,925</td> <td>TGCCCATATCTGCTACGGGGTA</td> <td>ATG</td> <td>13470</td> <td>111</td> </tr> <tr> <td>4</td> <td>-4,252</td> <td>1,910</td> <td>17</td> <td>-5,553</td> <td>GGTAATGATCGCAGCCAAAGAA</td> <td>GTG</td> <td>13488</td> <td>93</td> </tr> <tr> <td>5</td> <td>-3,206</td> <td>2,420</td> <td>10</td> <td>-4,165</td> <td>GATCGCAGCCAAAGAAGTGGAA</td> <td>ATG</td> <td>13494</td> <td>87</td> </tr> </tbody> </table>	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	-4,352	1,861	11	-5,352	CGCATGAACAACGAACCAGACC	ATG	13341	240	2	-5,663	1,221	13	-6,734	GGTCTGGGAGATCCCCAAAAC	ATG	13386	195	3	-5,624	1,240	6	-6,925	TGCCCATATCTGCTACGGGGTA	ATG	13470	111	4	-4,252	1,910	17	-5,553	GGTAATGATCGCAGCCAAAGAA	GTG	13488	93	5	-3,206	2,420	10	-4,165	GATCGCAGCCAAAGAAGTGGAA	ATG	13494	87
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	-4,352	1,861	11	-5,352	CGCATGAACAACGAACCAGACC	ATG	13341	240																																															
2	-5,663	1,221	13	-6,734	GGTCTGGGAGATCCCCAAAAC	ATG	13386	195																																															
3	-5,624	1,240	6	-6,925	TGCCCATATCTGCTACGGGGTA	ATG	13470	111																																															
4	-4,252	1,910	17	-5,553	GGTAATGATCGCAGCCAAAGAA	GTG	13488	93																																															
5	-3,206	2,420	10	-4,165	GATCGCAGCCAAAGAAGTGGAA	ATG	13494	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) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	ORF la plus longue = 240pb ☑ ATG 13341, chevauchement de 8 ORF 195pb ☑ ATG 13386, gap de 38pb																																																						
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 ?	NON 																																																						
DECISION:	13 341																																																						

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	Listez le meilleur hit Blastp pour chaque source :

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

*** PhagesDB :

rien

*** nr :

rien

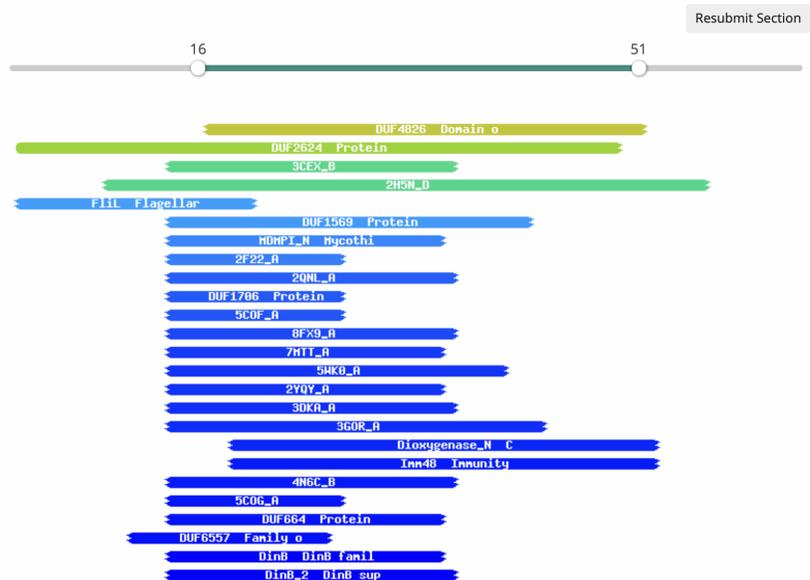
A No significant similarity found. For reasons why, [click here](#)

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 ?

PFAM proba de 81% : DUF4826 ; Domain of unknown function (DUF4826)

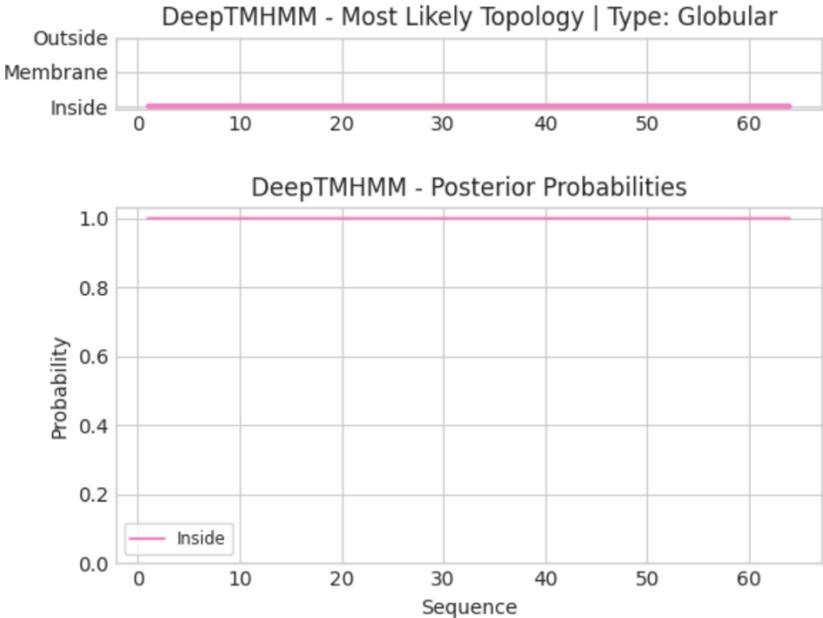
<input type="checkbox"/>	1	PF16108.9	; DUF4826 ; Domain of unknown function (DUF4826)	81.31	13	24.54	4.6	36	126
<input type="checkbox"/>	2	PF11116.12	; DUF2624 ; Protein of unknown function (DUF2624)	78.5	25	21.65	5	43	83
<input type="checkbox"/>	3	3CEX_B	Uncharacterized protein; Enterococcus faecalis, structural genomics, EF_3021, PSI-2, Protein Structure Initiative, Midwe	62.2	39	18.36	3	24	172
<input type="checkbox"/>	4	2H5N_D	Hypothetical protein PG_1108; hypothetical protein, SAD, MCSG, PSI, Structural Genomics, Protein Structure Initiative, M	61.91	50	18.27	5.7	48	133
<input type="checkbox"/>	5	PF03748.18	; FIL ; Flagellar basal body-associated protein FIL	53.35	31	19.41	1.8	20	92
<input type="checkbox"/>	6	PF07606.15	; DUF1569 ; Protein of unknown function (DUF1569)	53.06	80	18.36	3.4	30	151
<input type="checkbox"/>	7	PF11716.12	; MDMPI_N ; Mycothiol maleylpyruvate isomerase N-terminal domain	51.35	68	16.56	2.8	23	135
<input type="checkbox"/>	8	2F22_A	BH3987; PUTATIVE DNA DAMAGE-INDUCABLE (DINB) PROTEIN, STRUCTURAL GENOMICS, JOINT CENTER FOR STRUCTURAL GENOMICS, JCSG, P	50.77	43	18.5	2.1	15	144
<input type="checkbox"/>	9	2QNL_A	Uncharacterized protein; PUTATIVE DNA DAMAGE-INDUCIBLE PROTEIN, STRUCTURAL GENOMICS, JOINT CENTER FOR STRUCTURAL GENOMIC	48.07	89	16.92	3	24	162

Visualization



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

Gene apres = WhiB family transcripctor

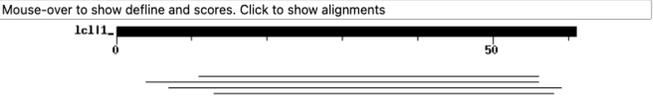
<p>dans une région du génome qui montre une forte conservation de l'ordre des gènes ?</p>	
<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	 <p>DeepTMHMM - Most Likely Topology Type: Globular</p> <p>DeepTMHMM - Posterior Probabilities</p>
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	
<p>DECISION:</p>	<p>NKF</p>

Student Gene Annotation Worksheet

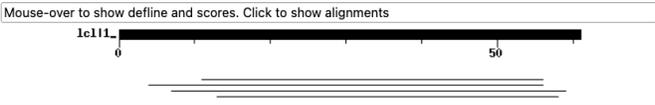
Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	23
Coordonnées du Stop	13743
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Ove 21
Coordonnées du Start retenu	13558
Fonction prédite	NKF

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)?	NO
Y-a-t-il des éléments supportant un potentiel codant ?	<p style="text-align: center;">Green frame below</p> <p style="font-size: small;">cyranoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGGCCAGCCCC), Order 2, Window 96, Step 12, 8/9</p> <p style="font-size: x-small;">GeneMark hmm prediction</p>

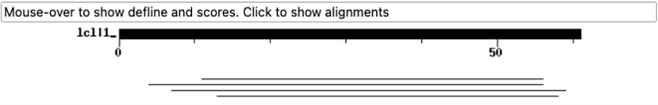
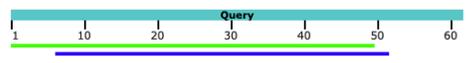
<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<p>Query= (61 letters)</p> <p style="text-align: center;">Distribution of 4 Blast Hits on the Query Sequence</p> <p style="text-align: center;">Mouse-over to show define and scores. Click to show alignments</p>  <p>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th></th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>Schiebs_21, function unknown, 62</td> <td>32</td> <td>0.74</td> </tr> <tr> <td>EpicDab_22, function unknown, 106</td> <td>30</td> <td>2.2</td> </tr> <tr> <td>Mbo4_11, minor capsid protein, 156</td> <td>29</td> <td>4.8</td> </tr> <tr> <td>Donovan_48, function unknown, 130</td> <td>29</td> <td>4.8</td> </tr> </tbody> </table> <p>>Schiebs_21, function unknown, 62 Length = 62</p> <p>Score = 31.6 bits (70), Expect = 0.74 Identities = 20/47 (42%), Positives = 24/47 (51%), Gaps = 6/47 (12%)</p> <p>Query: 12 PSQRDPHGRKVKCRRIFAADAVPAPDIDGNRICTDCK-HRPALNPP 57 P + H +C CRR FAA VP DG IC DC+ + A PP Sbjct: 11 PDRASAH--RCECCRKFAAGHVPT---DGLLICRDCRAEQEAAAPP 52</p> <p>>EpicDab_22, function unknown, 106 Length = 106</p> <p>Score = 30.0 bits (66), Expect = 2.2 Identities = 19/55 (34%), Positives = 25/55 (45%), Gaps = 6/55 (10%)</p> <p>Query: 5 PHTNEFRPSQRDPHGRKVKCRRIFAADAVPAPDIDGNRICTDCKHRPAL--NPP 57 P T E R + +CV C R+FA P D+ +C DC+ AL PP Sbjct: 48 PSTTEQRRALLAAKPRCVCERLFAPRVEPPRDL----LCRDCRTARALAQEPP 98</p>		Score (bits)	E Value	Schiebs_21, function unknown, 62	32	0.74	EpicDab_22, function unknown, 106	30	2.2	Mbo4_11, minor capsid protein, 156	29	4.8	Donovan_48, function unknown, 130	29	4.8
	Score (bits)	E Value														
Schiebs_21, function unknown, 62	32	0.74														
EpicDab_22, function unknown, 106	30	2.2														
Mbo4_11, minor capsid protein, 156	29	4.8														
Donovan_48, function unknown, 130	29	4.8														
<p>Est-ce que le candidat est en contradiction avec les principes d'annotation ?</p>	<p>This candidate gene was revealed when we analyzed large gaps in the annotation (gap 546 bp)</p>															
<p>DECISION:</p>	<p>OUI</p>															

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle															
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : NA Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : NA</p>															
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>																
<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>																
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	<p>NON</p> <p>Query= (61 letters)</p> <p style="text-align: center;">Distribution of 4 Blast Hits on the Query Sequence</p>  <p>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th></th> <th>Score (bits)</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>Schiebs_21, function unknown, 62</td> <td>32</td> <td>0.74</td> </tr> <tr> <td>EpicDab_22, function unknown, 106</td> <td>30</td> <td>2.2</td> </tr> <tr> <td>Mbo4_11, minor capsid protein, 156</td> <td>29</td> <td>4.8</td> </tr> <tr> <td>Donovan_48, function unknown, 130</td> <td>29</td> <td>4.8</td> </tr> </tbody> </table> <p>>Schiebs_21, function unknown, 62 Length = 62</p> <p>Score = 31.6 bits (70), Expect = 0.74 Identities = 20/47 (42%), Positives = 24/47 (51%), Gaps = 6/47 (12%)</p> <p>Query: 12 PSQRDPHGRKVKCRRIFAADAVPAPDIDGNRICTDCK-HRPALNPP 57 P + H +C CRR FAA VP DG IC DC+ + A PP Sbjct: 11 PDRASAH---RCECCRKFAAGHVPT---DGLLICRDCRAEQEAAAPP 52</p> <p>>EpicDab_22, function unknown, 106 Length = 106</p> <p>Score = 30.0 bits (66), Expect = 2.2 Identities = 19/55 (34%), Positives = 25/55 (45%), Gaps = 6/55 (10%)</p> <p>Query: 5 PHTNEFRPSQRDPHGRKVKCRRIFAADAVPAPDIDGNRICTDCKHRPAL---NPP 57 P T E R + +CV C R+FA P D+ +C DC+ AL PP Sbjct: 48 PSTTEQRRRAALLAAKPRCVDICERLFAPRVEPPRDL----LCRDCRTARALAEQPP 98</p>		Score (bits)	E Value	Schiebs_21, function unknown, 62	32	0.74	EpicDab_22, function unknown, 106	30	2.2	Mbo4_11, minor capsid protein, 156	29	4.8	Donovan_48, function unknown, 130	29	4.8
	Score (bits)	E Value														
Schiebs_21, function unknown, 62	32	0.74														
EpicDab_22, function unknown, 106	30	2.2														
Mbo4_11, minor capsid protein, 156	29	4.8														
Donovan_48, function unknown, 130	29	4.8														
<p>Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>																

DECISION:

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-valeur < 10^{-4} et une couverture acceptable ?</p>	<p>Listez le meilleur hit Blastp pour chaque source :</p> <p>*** PhagesDB :</p> <p>Query= (61 letters)</p> <p>Distribution of 4 Blast Hits on the Query Sequence</p>  <p>Sequences producing significant alignments:</p> <table border="1"><thead><tr><th></th><th>Score (bits)</th><th>E Value</th></tr></thead><tbody><tr><td>Schiebs_21, function unknown, 62</td><td>32</td><td>0.74</td></tr><tr><td>EpicDab_22, function unknown, 106</td><td>30</td><td>2.2</td></tr><tr><td>Mbo4_11, minor capsid protein, 156</td><td>29</td><td>4.8</td></tr><tr><td>Donovan_48, function unknown, 130</td><td>29</td><td>4.8</td></tr></tbody></table> <p>>Schiebs_21, function unknown, 62 Length = 62</p> <p>Score = 31.6 bits (70), Expect = 0.74 Identities = 20/47 (42%), Positives = 24/47 (51%), Gaps = 6/47 (12%)</p> <p>Query: 12 PSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCK-HRPALNPP 57 P + H +C CRR FAA VP DG IC DC+ + A PP</p> <p>Sbjct: 11 PDRASAH--RCECCRKFAAGHVPT---DGLLICRDCRAEQEAAAPP 52</p> <p>>EpicDab_22, function unknown, 106 Length = 106</p> <p>Score = 30.0 bits (66), Expect = 2.2 Identities = 19/55 (34%), Positives = 25/55 (45%), Gaps = 6/55 (10%)</p> <p>Query: 5 PHTNEFRPSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCKHRPAL--NPP 57 P T E R + +CV C R+FA P D+ +C DC+ AL PP</p> <p>Sbjct: 48 PSTTEQRRALLAAKPRCVDLCERLFAPRVEPPRDL----LCRDCRTARALAQEPP 98</p> <p>*** nr :</p> <p>Query Length 61</p> <p>Other reports Distance tree of results Multiple alignment MSA viewer ?</p> <p>Compare these results against the new Clustered nr database ? BLAST</p> <p>Descriptions Graphic Summary Alignments Taxonomy</p> <p>hover to see the title click to show alignments <input checked="" type="checkbox"/> Show Conserved Domains Alignment Scores <input type="checkbox"/> < 40 <input type="checkbox"/> 40 - 50 <input type="checkbox"/> 50 - 80</p> <p>2 sequences selected ?</p> <p>No putative conserved domains have been detected</p> <p>Distribution of the top 2 Blast Hits on 2 subject sequences</p> 		Score (bits)	E Value	Schiebs_21, function unknown, 62	32	0.74	EpicDab_22, function unknown, 106	30	2.2	Mbo4_11, minor capsid protein, 156	29	4.8	Donovan_48, function unknown, 130	29	4.8
	Score (bits)	E Value														
Schiebs_21, function unknown, 62	32	0.74														
EpicDab_22, function unknown, 106	30	2.2														
Mbo4_11, minor capsid protein, 156	29	4.8														
Donovan_48, function unknown, 130	29	4.8														

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

hypothetical protein VH15_05895 [Corynebacterium ulcerans]

Sequence ID: [KKO87260.1](#) Length: 96 Number of Matches: 1

Range 1: 40 to 87 [GenPept](#) [Graphics](#)

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

Score	Expect	Method	Identities	Positives	Gaps
62.8 bits(151)	1e-10	Compositional matrix adjust.	27/49(55%)	35/49(71%)	1/49(2%)

Query	1	MPIDPHTNEFRPSQRDPHGRKCVKRRIFAADAVPAPDIDGNRICTDCK	49
		MP D + E+RP +RDP GR+C C R+FAA VP P +DG R+CTDC+	
Sbjct	40	MPWD-ESGEYRPYRRDPEGRRCNHCRNVFAAGCVPVPYLDGTRLCTDCR	87

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

hypothetical protein VN94_09850 [Corynebacterium diphtheriae]

Sequence ID: [KKA80634.1](#) Length: 64 Number of Matches: 1

Range 1: 5 to 49 [GenPept](#) [Graphics](#)

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

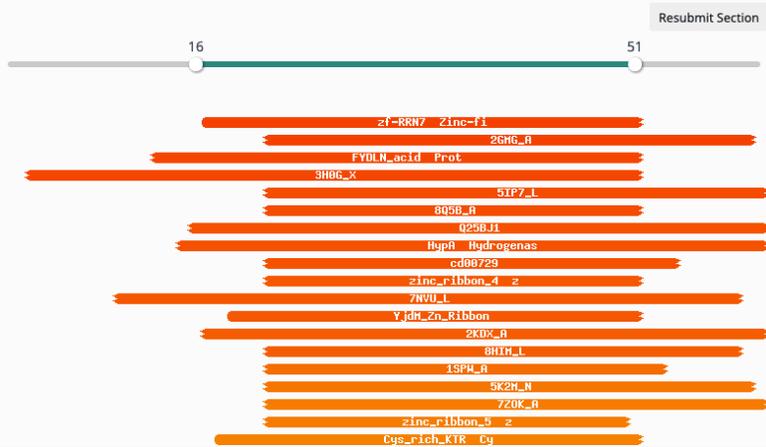
Score	Expect	Method	Identities	Positives	Gaps
45.8 bits(107)	2e-04	Compositional matrix adjust.	21/45(47%)	27/45(60%)	0/45(0%)

Query	7	TNEFRPSQRDPHGRKCVKRRIFAADAVPAPDIDGNRICTDCKHR	51
		EFRP +R+P G +C C FAA VP D +G RIC+ C+ R	
Sbjct	5	NGEFRPRRRNPDGIRCDDCNLTFAAGGVPPKDRNGERICSSCRQR	49

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): 3.76964

Visualization

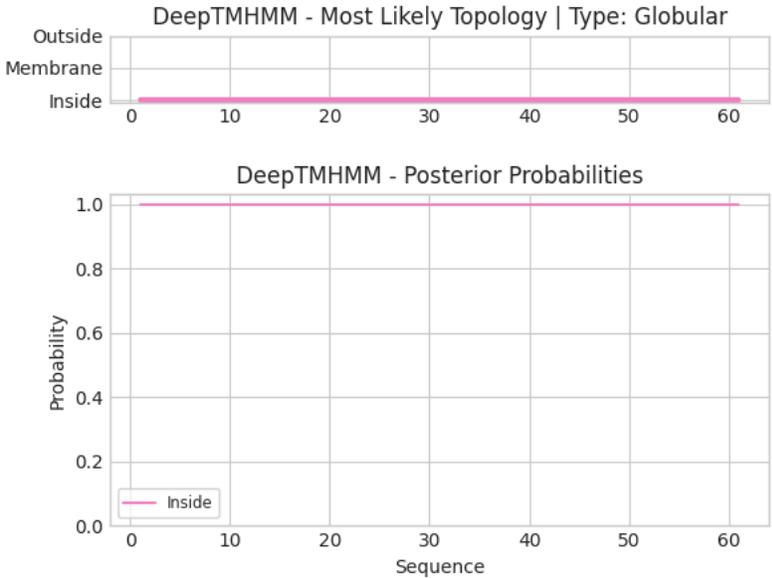


Vis	Hits	Aln	Select All	Forward	Forward Query A3M	Model using selection	Download HHR	Color Seqs	Wrap Seqs
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length	
<input type="checkbox"/>	1	PF11781.13 ; zf-RRN7 ; Zinc-finger of RNA-polymerase I-specific TFIIIB, Rrn7	94.14	0.1	25.32	1.7	30	32	
<input type="checkbox"/>	2	2GMG_A hypothetical protein PF0610; winged-helix like protein with metal binding site, Structural Genomics, PSI, Protein Struct	94.13	0.13	32.97	2.5	34	105	
<input type="checkbox"/>	3	PF09538.15 ; FYDLN_acid ; Protein of unknown function (FYDLN_acid)	93.97	0.1	34.43	1.9	34	107	
<input type="checkbox"/>	4	3H0G_X DNA-directed RNA polymerases I, II, and III subunit RPABC4; Transcription, Multi-protein complex, Polymerase, DNA-bindin	93.87	0.16	30.45	2.4	44	63	
<input type="checkbox"/>	5	SIP7_L DNA-directed RNA polymerases I, II, and III subunit RPABC4; transcription; 3.52A (Saccharomyces cerevisiae)	93.59	0.23	26.14	2.5	35	46	
<input type="checkbox"/>	6	8Q5B_A Small CPxCG-related zinc finger protein; zinc-finger, METAL BINDING PROTEIN; NMR (Haloferax volcanii DS2)	93.16	0.11	31.92	1	30	56	
<input type="checkbox"/>	7	Q25BJ1 Y04_HIS11 Uncharacterized protein ORF4 O5=His1 virus (isolate Australia/Victoria) OX=654912 GN=ORF4 PE=4 SV=1	92.74	0.3	24.68	2.1	47	48	
<input type="checkbox"/>	8	PF01155.24 ; HypA ; Hydrogenase/urease nickel incorporation, metallochaperone, hypA	92.69	0.24	30.47	2	42	113	
<input type="checkbox"/>	9	cd00729 rubredoxin_SM; Rubredoxin, Small Modular nonheme iron binding domain containing a [Fe(SCys)4] center, present in rubrery	92.66	0.31	22.03	1.9	27	34	
<input type="checkbox"/>	10	PF13717.11 ; zinc_ribbon_4 ; zinc-ribbon domain	92.53	0.22	22.43	1.3	31	37	
<input type="checkbox"/>	11	7NVU_L RNA polymerase II subunit K; Initiation, TRANSCRIPTION; HET: ZN; 2.5A (Sus scrofa)	92.37	0.42	27.15	2.6	45	58	
<input type="checkbox"/>	12	PF08274.17 ; Yjdm_Zn_Ribbon ; PhnA Zinc-Ribbon	92.33	0.22	23.8	1.2	28	30	

Seems to have at least 4 cys that could bind Zinc

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".

<p>Est-ce que ce gène code pour une protéine transmembranaire (TM) ?</p>	<h3>DeepTMHMM - Predictions</h3> <p>Predicted topologies can be downloaded in .gff3 format and .3line format</p>  <p>You can download the probabilities used to generate this plot here</p> <h4>Predicted Topologies</h4> <pre>>Unnamed GLOB</pre>
<p>Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?</p>	<p>hypothetical</p>
<p>DECISION:</p>	<p>NKF</p>

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranOPS
Gène #	24
Coordonnées du Stop	14146
Direction (For/Rev)	For

Gap/chevauchement avec un autre gène	Gap 135
Coordonnées du Start retenu	13877
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)?	NO
Y-a-t-il des éléments supportant un potentiel codant ?	<p>Red frame below</p> <p>GenMark.hmm prediction yanoPS complete sequence, 14870 bp including 14-bp 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 8/9</p> <p>The figure displays six tracks of genomic data for the yanoPS complete sequence (14870 bp). The top three tracks are labeled 'Direct Sequence' and the bottom three are 'Complementary Sequence'. The x-axis represents 'Nucleotide Position' from 12800 to 14800. A GeneMark.hmm prediction track at the top shows a black bar indicating a predicted coding region. A red box highlights a specific region in the Direct Sequence track around nucleotide 14000, which is also reflected in the Complementary Sequence tracks. A green box highlights another region around nucleotide 13600.</p>

<p>Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?</p>	<p>Non</p> <hr/> <p>Query Length 89</p> <hr/> <p>Other reports ?</p> <hr/> <p> No significant similarity found. For reasons why, click here</p>
<p>Est-ce que le candidat est en contradiction avec les principes d'annotation ?</p>	<p>We are filling a gap (546 bp)</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 ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : NA Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : NA
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 ?	
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 ?	
DECISION:	

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 : *** PhagesDB : no hits found

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

*** nr :

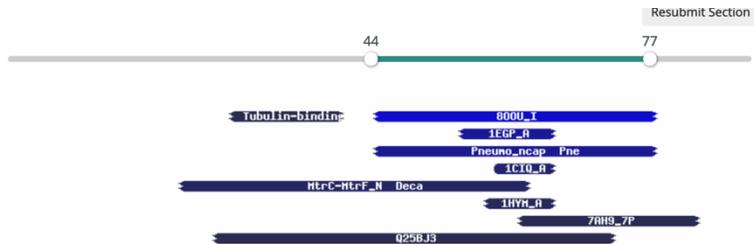
Query Length 89

Other reports ?



No significant similarity found. For reasons why, [click here](#)

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 ?



Hitlist

Show 25 Entries

Search:

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/> 1	800U_I	Nucleoprotein; D10-symmetry, N-RNA, nucleoprotein, RSV, VIRAL PROTEIN; 2.9A (Trichoplusia ni)	36.17	68	29.25	2.1	34	391
<input type="checkbox"/> 2	1EGP_A	EGLIN-C; PROTEINASE INHIBITOR; 2.0A (Hirudo medicinalis) SCOP: d.40.1.1	31.74	50	19.31	0.4	12	45
<input type="checkbox"/> 3	PF03246.18	; Pneumo_ncap ; Pneumovirus nucleocapsid protein	30.92	71	29.17	1.4	34	392
<input type="checkbox"/> 4	1CIQ_A	CHYMOTRYPSIN INHIBITOR 2; CLEAVED INHIBITOR, SERINE PROTEASE INHIBITOR; 2.2A (Hordeum vulgare) SCOP: d.40.1.1	27	52	18.53	0	8	40
<input type="checkbox"/> 5	PF22111.1	; MtrC-MtrF_N ; Decaheme cytochrome c	24.66	450	18.28	4	42	128

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".

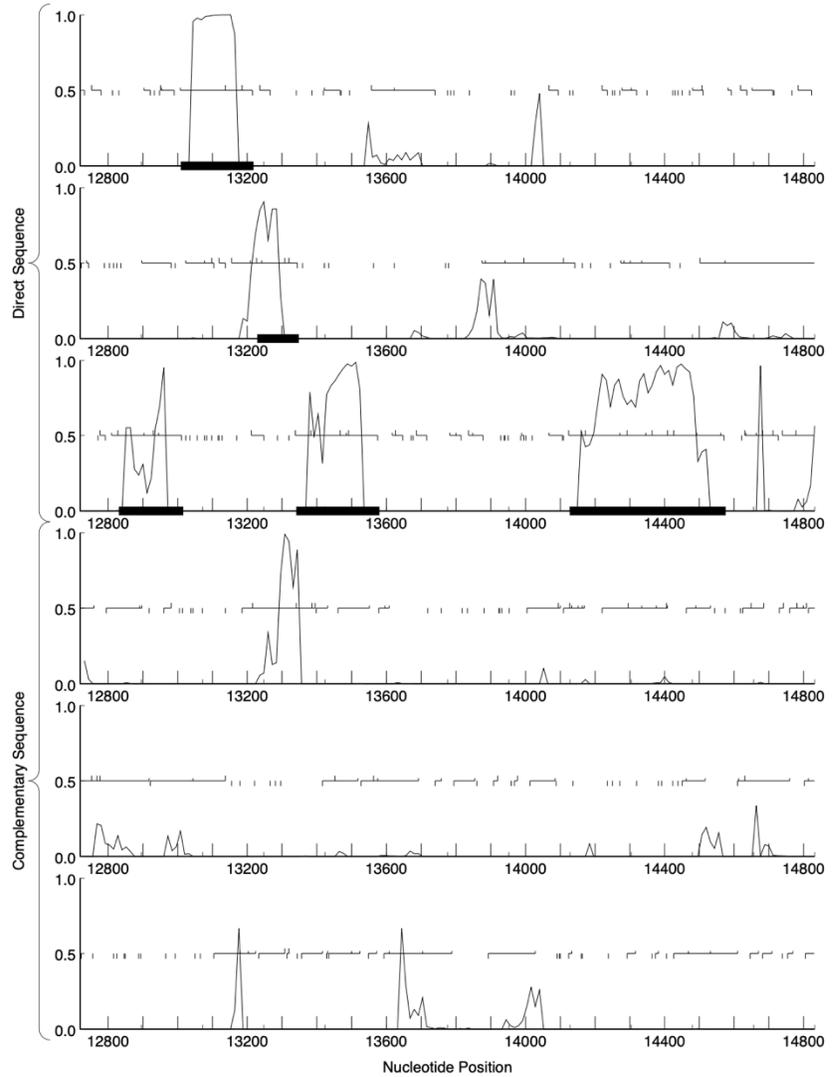
Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	25
Coordonnées du Stop	14576
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 546
Coordonnées du Start retenu	14 127
Fonction prédite	WhiB family transcription factor

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 ?	

GeneMark hmmprediction
 yanoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 8/9



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

Color Key for Alignment Scores

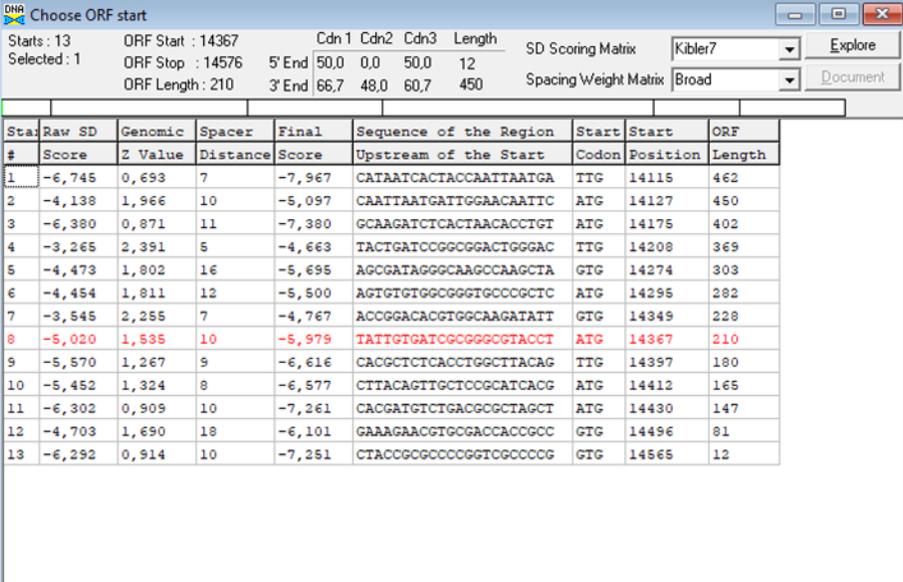
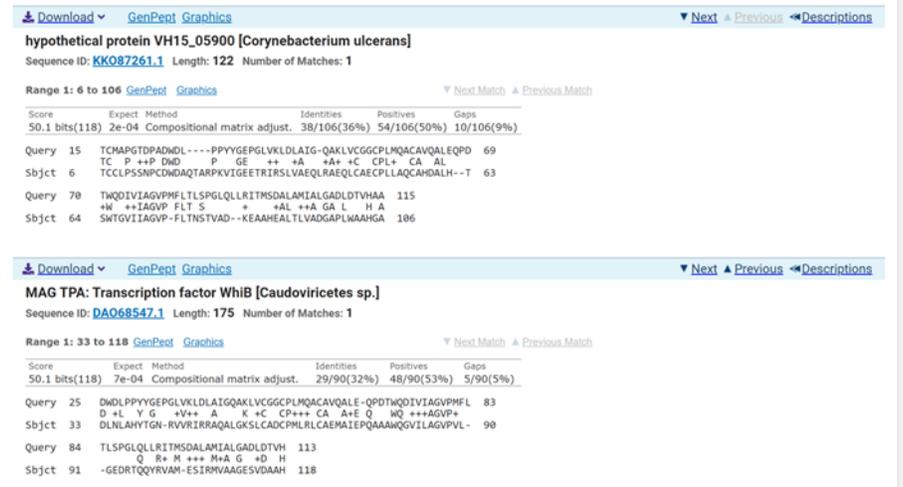


Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?

Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_23, function unknown, 149	308	3e-84
Trax_62, WhiB family transcription factor, 251	36	0.039
Neville_63, WhiB family transcription factor, 251	35	0.051
Rimmer_91, WhiB family transcription factor, 86	34	0.15
Petra64142_Draft_98, function unknown, 86	34	0.15
Pat3_90, function unknown, 86	34	0.15
mul6_54, function unknown, 171	34	0.15
Gator_91, WhiB family transcription factor, 86	34	0.15
DoctorDiddles_92, WhiB family transcription factor, 86	34	0.15
Daikon_Draft_94, function unknown, 86	34	0.15
Cookies_90, WhiB family transcription factor, 86	34	0.15

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

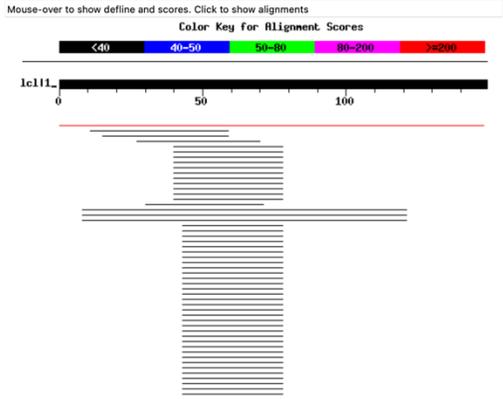
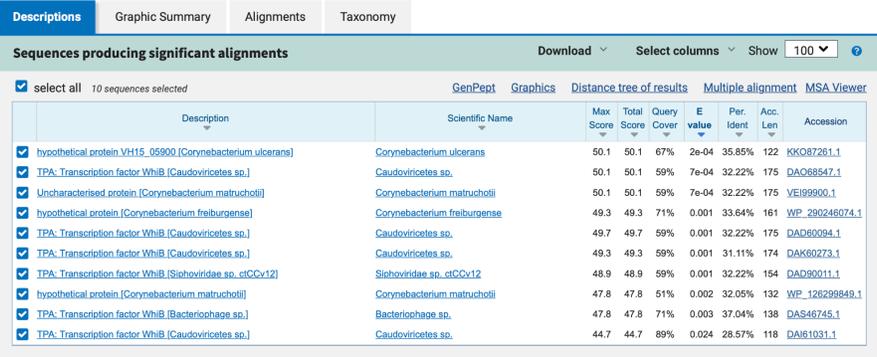
Décision #2 : Quel est le start du gène ?

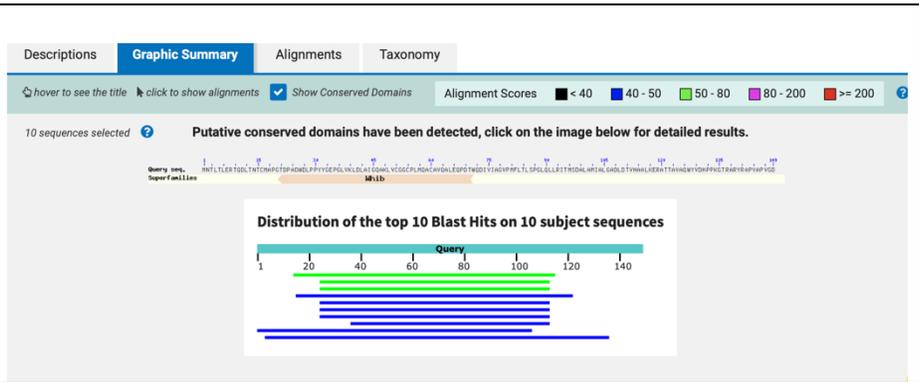
Collection des éléments de réponse	Rationnelle																																																																																																																																							
<p>Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p>Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 14 127 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 14 127</p>																																																																																																																																							
<p>Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	 <table border="1"> <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>-6,745</td><td>0,693</td><td>7</td><td>-7,967</td><td>CATAATCACTACCAATTAATGA</td><td>TTG</td><td>14115</td><td>462</td></tr> <tr><td>2</td><td>-4,138</td><td>1,966</td><td>10</td><td>-5,097</td><td>CAATTAATGATTGGACAATTC</td><td>ATG</td><td>14127</td><td>450</td></tr> <tr><td>3</td><td>-6,380</td><td>0,871</td><td>11</td><td>-7,380</td><td>GCAAGATCTCACTAACACCTGT</td><td>ATG</td><td>14175</td><td>402</td></tr> <tr><td>4</td><td>-3,265</td><td>2,391</td><td>5</td><td>-4,663</td><td>TACTGATCCGGGGACTGGGAC</td><td>TTG</td><td>14208</td><td>369</td></tr> <tr><td>5</td><td>-4,473</td><td>1,802</td><td>16</td><td>-5,695</td><td>AGCGATAGGGCAAGCCAAAGCTA</td><td>GTG</td><td>14274</td><td>303</td></tr> <tr><td>6</td><td>-4,454</td><td>1,811</td><td>12</td><td>-5,500</td><td>AGTGTGTGGCGGTGCCCGCTC</td><td>ATG</td><td>14295</td><td>282</td></tr> <tr><td>7</td><td>-3,545</td><td>2,255</td><td>7</td><td>-4,767</td><td>ACCGGACACGTGGCAAGATATT</td><td>ATG</td><td>14349</td><td>228</td></tr> <tr><td>8</td><td>-5,020</td><td>1,535</td><td>10</td><td>-5,979</td><td>TATTGTGATCGGGCGTACCT</td><td>ATG</td><td>14367</td><td>210</td></tr> <tr><td>9</td><td>-5,570</td><td>1,267</td><td>9</td><td>-6,616</td><td>CACGCTCTCACCTGGCTTACAG</td><td>TTG</td><td>14397</td><td>180</td></tr> <tr><td>10</td><td>-5,452</td><td>1,324</td><td>8</td><td>-6,577</td><td>CTTACAGTGTCTCCGCATCACG</td><td>ATG</td><td>14412</td><td>165</td></tr> <tr><td>11</td><td>-6,302</td><td>0,909</td><td>10</td><td>-7,261</td><td>CACGATGTCTGACGGCTAGCT</td><td>ATG</td><td>14430</td><td>147</td></tr> <tr><td>12</td><td>-4,703</td><td>1,690</td><td>18</td><td>-6,101</td><td>GAANGAAGCTGGACCCCGCC</td><td>GTG</td><td>14496</td><td>81</td></tr> <tr><td>13</td><td>-6,292</td><td>0,914</td><td>10</td><td>-7,251</td><td>CTACCGCGCCCGGTGCCCGCC</td><td>GTG</td><td>14565</td><td>12</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	-6,745	0,693	7	-7,967	CATAATCACTACCAATTAATGA	TTG	14115	462	2	-4,138	1,966	10	-5,097	CAATTAATGATTGGACAATTC	ATG	14127	450	3	-6,380	0,871	11	-7,380	GCAAGATCTCACTAACACCTGT	ATG	14175	402	4	-3,265	2,391	5	-4,663	TACTGATCCGGGGACTGGGAC	TTG	14208	369	5	-4,473	1,802	16	-5,695	AGCGATAGGGCAAGCCAAAGCTA	GTG	14274	303	6	-4,454	1,811	12	-5,500	AGTGTGTGGCGGTGCCCGCTC	ATG	14295	282	7	-3,545	2,255	7	-4,767	ACCGGACACGTGGCAAGATATT	ATG	14349	228	8	-5,020	1,535	10	-5,979	TATTGTGATCGGGCGTACCT	ATG	14367	210	9	-5,570	1,267	9	-6,616	CACGCTCTCACCTGGCTTACAG	TTG	14397	180	10	-5,452	1,324	8	-6,577	CTTACAGTGTCTCCGCATCACG	ATG	14412	165	11	-6,302	0,909	10	-7,261	CACGATGTCTGACGGCTAGCT	ATG	14430	147	12	-4,703	1,690	18	-6,101	GAANGAAGCTGGACCCCGCC	GTG	14496	81	13	-6,292	0,914	10	-7,251	CTACCGCGCCCGGTGCCCGCC	GTG	14565	12
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	-6,745	0,693	7	-7,967	CATAATCACTACCAATTAATGA	TTG	14115	462																																																																																																																																
2	-4,138	1,966	10	-5,097	CAATTAATGATTGGACAATTC	ATG	14127	450																																																																																																																																
3	-6,380	0,871	11	-7,380	GCAAGATCTCACTAACACCTGT	ATG	14175	402																																																																																																																																
4	-3,265	2,391	5	-4,663	TACTGATCCGGGGACTGGGAC	TTG	14208	369																																																																																																																																
5	-4,473	1,802	16	-5,695	AGCGATAGGGCAAGCCAAAGCTA	GTG	14274	303																																																																																																																																
6	-4,454	1,811	12	-5,500	AGTGTGTGGCGGTGCCCGCTC	ATG	14295	282																																																																																																																																
7	-3,545	2,255	7	-4,767	ACCGGACACGTGGCAAGATATT	ATG	14349	228																																																																																																																																
8	-5,020	1,535	10	-5,979	TATTGTGATCGGGCGTACCT	ATG	14367	210																																																																																																																																
9	-5,570	1,267	9	-6,616	CACGCTCTCACCTGGCTTACAG	TTG	14397	180																																																																																																																																
10	-5,452	1,324	8	-6,577	CTTACAGTGTCTCCGCATCACG	ATG	14412	165																																																																																																																																
11	-6,302	0,909	10	-7,261	CACGATGTCTGACGGCTAGCT	ATG	14430	147																																																																																																																																
12	-4,703	1,690	18	-6,101	GAANGAAGCTGGACCCCGCC	GTG	14496	81																																																																																																																																
13	-6,292	0,914	10	-7,251	CTACCGCGCCCGGTGCCCGCC	GTG	14565	12																																																																																																																																
<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>	<p>ORF la plus longue ☐ 462pb TTg 14115, gap 534pb ORF de 450 ☐ ATG 14127 ☐ 546pb</p>																																																																																																																																							
<p>Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	 <p>hypothetical protein VH15_05900 [Corynebacterium ulcerans] Sequence ID: KK087261.1 Length: 122 Number of Matches: 1</p> <p>Range 1: 6 to 106 GenPeot Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>50.1 bits(118)</td> <td>2e-04</td> <td>Compositional matrix adjust.</td> <td>38/106(36%)</td> <td>54/106(50%)</td> <td>10/106(9%)</td> </tr> </tbody> </table> <p>Query 15 TCMAGTDPADNDL...PPYGEPLVKLDLAI...GAKLVCGGCLMCAVQALEQPD 69 TC P ++P DMD P GE ++ SA +A+ +C CPL+ CA AL Sbjct 6 TCCLPSSNCDNDQAQARPKVIGETIRISLVAEQLRAEQLCAECPLLAQCAHDALH--T 63</p> <p>Query 70 TWQDIVIAGVPMFLTSPGLQLRITMSDALAMIALGADLDTVHA 115 +W ++IAGVP FLT S + +AL +A GA L H A Sbjct 64 SMTGVIIAGVP-FLTNSTVAD--KEAAHEALTLDVAGAPLMAAHA 106</p> <hr/> <p>MAG TPA: Transcription factor WhiB [Caudoviricetes sp.] Sequence ID: DA068547.1 Length: 175 Number of Matches: 1</p> <p>Range 1: 33 to 118 GenPeot Graphics Next Match Previous Match</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>50.1 bits(118)</td> <td>7e-04</td> <td>Compositional matrix adjust.</td> <td>29/90(32%)</td> <td>48/90(53%)</td> <td>5/90(5%)</td> </tr> </tbody> </table> <p>Query 25 DNDLPPYGEPLVKLDLAI...GAKLVCGGCLMCAVQALE-QPDTWQDIVIAGVPMFL 83 D +L Y G +H+ A + +C CP+++ CA +R+ Q HQ +++AGVP+ Sbjct 33 DLNLAHYTGN-RVVRIRRAQALGKSLCADCPLRLCAEMAIEPQAAANQGVILAGVPVL- 90</p> <p>Query 84 TSPGLQLRITMSDALAMIALGADLDTVH 113 Q R+ M+++ M+A G +D H Sbjct 91 -GEDRTQYRVAM-ESIRNVAAGESVDAAH 118</p>	Score	Expect	Method	Identities	Positives	Gaps	50.1 bits(118)	2e-04	Compositional matrix adjust.	38/106(36%)	54/106(50%)	10/106(9%)	Score	Expect	Method	Identities	Positives	Gaps	50.1 bits(118)	7e-04	Compositional matrix adjust.	29/90(32%)	48/90(53%)	5/90(5%)																																																																																																															
Score	Expect	Method	Identities	Positives	Gaps																																																																																																																																			
50.1 bits(118)	2e-04	Compositional matrix adjust.	38/106(36%)	54/106(50%)	10/106(9%)																																																																																																																																			
Score	Expect	Method	Identities	Positives	Gaps																																																																																																																																			
50.1 bits(118)	7e-04	Compositional matrix adjust.	29/90(32%)	48/90(53%)	5/90(5%)																																																																																																																																			

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

DECISION: ATG 14127, bon score, orf de grande taille

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>Sequences producing significant alignments:</p> <table border="1"> <thead> <tr> <th></th> <th>Score</th> <th>E Value</th> </tr> </thead> <tbody> <tr> <td>CyranoPS_Draft_23, function unknown, 149</td> <td>308</td> <td>3e-84</td> </tr> <tr> <td>Trax_62, WhiB family transcription factor, 251</td> <td>36</td> <td>0.039</td> </tr> <tr> <td>Neville_63, WhiB family transcription factor, 251</td> <td>35</td> <td>0.051</td> </tr> <tr> <td>Rimmer_91, WhiB family transcription factor, 86</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>Petra64142_Draft_98, function unknown, 86</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>Pat3_90, function unknown, 86</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>mul6_54, function unknown, 171</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>Gator_91, WhiB family transcription factor, 86</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>DoctorDiddles_92, WhiB family transcription factor, 86</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>Daikon_Draft_94, function unknown, 86</td> <td>34</td> <td>0.15</td> </tr> <tr> <td>Cookies_90, WhiB family transcription factor, 86</td> <td>34</td> <td>0.15</td> </tr> </tbody> </table> <p>*** nr :</p> 		Score	E Value	CyranoPS_Draft_23, function unknown, 149	308	3e-84	Trax_62, WhiB family transcription factor, 251	36	0.039	Neville_63, WhiB family transcription factor, 251	35	0.051	Rimmer_91, WhiB family transcription factor, 86	34	0.15	Petra64142_Draft_98, function unknown, 86	34	0.15	Pat3_90, function unknown, 86	34	0.15	mul6_54, function unknown, 171	34	0.15	Gator_91, WhiB family transcription factor, 86	34	0.15	DoctorDiddles_92, WhiB family transcription factor, 86	34	0.15	Daikon_Draft_94, function unknown, 86	34	0.15	Cookies_90, WhiB family transcription factor, 86	34	0.15
	Score	E Value																																			
CyranoPS_Draft_23, function unknown, 149	308	3e-84																																			
Trax_62, WhiB family transcription factor, 251	36	0.039																																			
Neville_63, WhiB family transcription factor, 251	35	0.051																																			
Rimmer_91, WhiB family transcription factor, 86	34	0.15																																			
Petra64142_Draft_98, function unknown, 86	34	0.15																																			
Pat3_90, function unknown, 86	34	0.15																																			
mul6_54, function unknown, 171	34	0.15																																			
Gator_91, WhiB family transcription factor, 86	34	0.15																																			
DoctorDiddles_92, WhiB family transcription factor, 86	34	0.15																																			
Daikon_Draft_94, function unknown, 86	34	0.15																																			
Cookies_90, WhiB family transcription factor, 86	34	0.15																																			



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 ?

**PFAM : WhiB ; Transcription factor WhiB
 PDB : Transcriptional regulator WhiB1; nitric oxide, sigmaA, iron-sulfur, tuberculosis, Wbl protein, SIGNALING PROTEIN; HET: S
 Uniprot : WHIB_BPMT4 Probable transcriptional regulator
 WhiBTM4 OS=Mycobacterium phage TM4 OX=88870 GN=whiBTM4
 PE=1 SV=1**



<input type="checkbox"/>	1	5OAY_A	Transcriptional regulator WhiB1; nitric oxide, sigmaA, iron-sulfur, tuberculosis, Wbl protein, SIGNALING PROTEIN; HET: S	99.73	4.5e-17	112.66	8.5	75	94
<input type="checkbox"/>	2	6ONU_A	Transcriptional regulator WhiB1; Iron-sulfur cluster, transcription regulation, redox-sensing, TRANSCRIPTION; HET: MSE,	99.65	1e-15	101.7	7.2	69	76
<input type="checkbox"/>	3	7KUG_C	Probable transcriptional regulator WhiB7; Redox-sensitive, Iron-sulfur cluster, transcription, activator; HET: SF4, MSE;	99.65	1.5e-15	101.76	7.6	67	79
<input type="checkbox"/>	4	Q9ZK29	WHIB_BPMT4 Probable transcriptional regulator WhiBTM4 OS=Mycobacterium phage TM4 OX=88870 GN=whiBTM4 PE=1 SV=1	99.65	1.1e-15	101.65	6.7	76	76
<input type="checkbox"/>	5	7KUF_A	Probable transcriptional regulator WhiB7; Redox-sensitive, Iron-sulfur cluster, transcription, activator; 2.6A (Mycobact	99.64	9.5e-16	105.57	5.9	81	92
<input type="checkbox"/>	6	8CWT_E	Redox- and pH-responsive transcriptional regulator WhiB3; redox sensor, transcriptional factor, TRANSCRIPTION; HET: TRS,	99.64	1.5e-15	104.14	6.7	76	90
<input type="checkbox"/>	7	8CYF_A	Redox- and pH-responsive transcriptional regulator WhiB3; redox sensor, transcriptional factor, protein-DNA complex, TRA	99.63	1.8e-15	106.32	6.8	77	102
<input type="checkbox"/>	8	8DY7_H	Transcriptional regulator WhiB; RNA polymerase, Transcription factor, Iron cluster, TRANSCRIPTION-TRANSFERASE-DNA comple	99.62	2.6e-15	102.62	6.2	75	87
<input type="checkbox"/>	9	7F7N_A	Transcriptional regulator WhiB4; apo-WhiB4, helix-turn-helix domain, intrinsically	99.57	8.7e-15	106.93	5.8	88	126

<input type="checkbox"/>	10	PF02467.20	; Whib ; Transcription factor WhiB	99.56	2.2e-14	92.75	6.4	64	65
<input type="checkbox"/>	11	8D5V_A	Probable transcriptional regulator WhiB6; redox sensor, transcriptional factor, protein-DNA complex, TRANSCRIPTION, TRAN	99.55	5.7e-14	101.08	8.1	86	116
<input type="checkbox"/>	12	PF08097.15	; Toxin_26 ; Conotoxin T-superfamily	55.61	7.1	17.02	-0.4	7	11
<input type="checkbox"/>	13	PF20056.3	; DUF6455 ; Family of unknown function (DUF6455)	54.95	45	23.17	3.2	23	85
<input type="checkbox"/>	14	PF10576.13	; EndIII_4Fe-25 ; Iron-sulfur binding domain of endonuclease III	53.7	7.6	19.28	-0.5	10	17
<input type="checkbox"/>	15	Q70LD4	Y223_AVF1Y Uncharacterized protein ORF223 OS=Acidianus filamentous virus 1 (isolate United States/Yellowstone) OX=654909	27.05	78	24.01	0.9	13	223
<input type="checkbox"/>	16	cd09637	Cas4_I-A_I-B_I-C_I-D_II-B; CRISPR/Cas system-associated protein Cas4, CRISPR (Clustered Regularly Interspaced Short Pal	26.79	42	23.77	-0.5	10	170
<input type="checkbox"/>	17	O64262	VG69_BPMD2 Gene 69 protein OS=Mycobacterium phage D29 OX=28369 GN=69 PE=4 SV=1	26.08	70	24.84	0.5	10	269
<input type="checkbox"/>	18	PF01930.21	; Cas_Cas4 ; Domain of unknown function DUF83	23.56	89	22.35	0.7	10	162
<input type="checkbox"/>	19	Q8QL42	Y207_SIRV1 Uncharacterized protein 207 OS=Sulfolobus islandicus rod-shaped virus 1 OX=157898 GN=207 PE=4 SV=1	23.52	83	23.6	0.5	10	207
<input type="checkbox"/>	20	7AE1_P	DNA-directed RNA polymerase III subunit RPC6; HUMAN, TRANSCRIPTION, SHORT RNAs; 2.8A (Homo sapiens)	22.39	66	28	-0.2	10	316

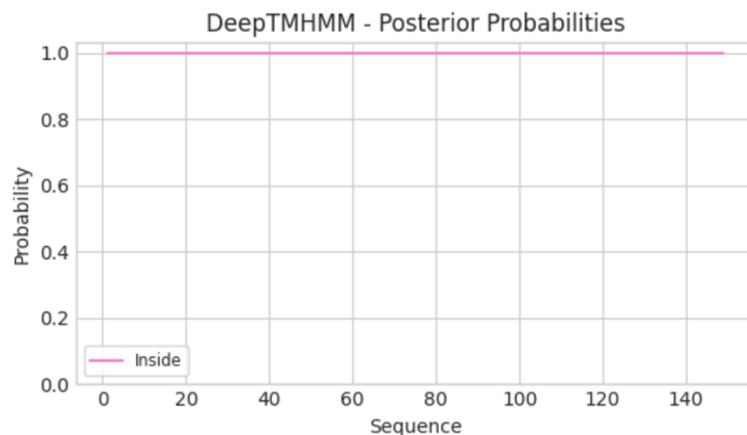
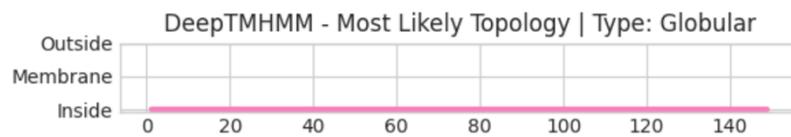
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

"Aucune synténie observée".

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

DeepTMHMM - Predictions

Predicted topologies can be downloaded in [.gff3 format](#) and [.3line format](#)



Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	Oui
DECISION:	WhiB family transcription factor

Student Gene Annotation Worksheet

Basic Phage Information	
Nom du Phage	CyranoPS
Gène #	26
Coordonnées du Stop	14851
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	
Coordonnées du Start retenu	14576
Fonction prédite	HNH endonuclease

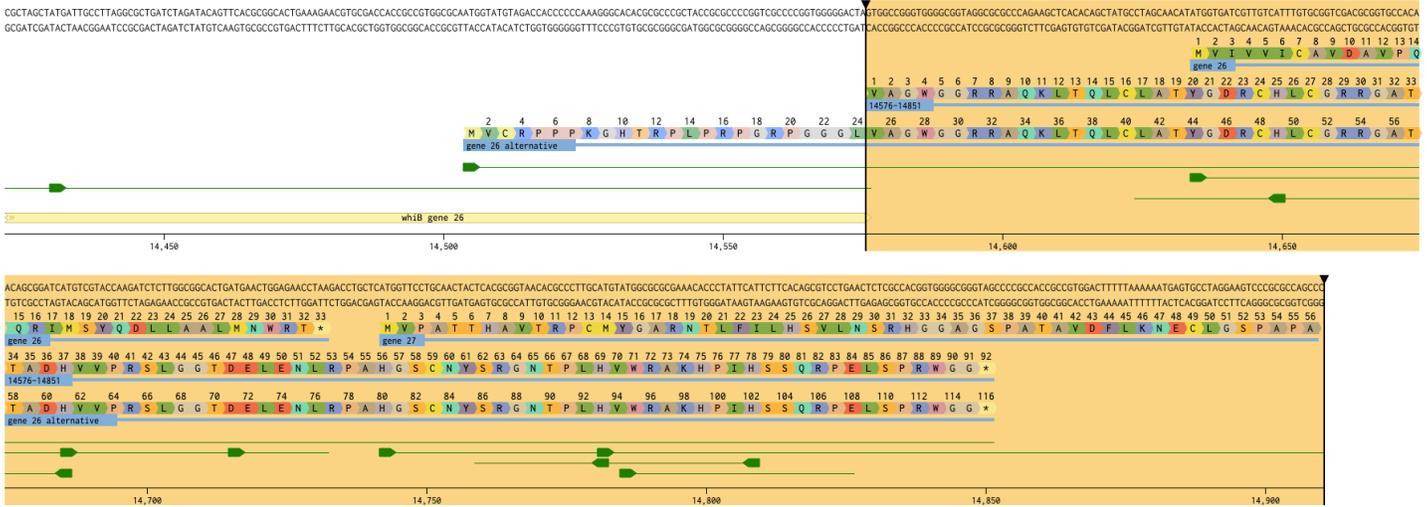
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	NO
Y-a-t-il des éléments supportant un potentiel codant ?	<p>Red frame below</p> <p>GeneMark.hmm prediction CyranoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGGCCAGCCC), Order 2, Window 96, Step 12, 8/9</p> <p>The figure displays a GeneMark.hmm prediction plot for the CyranoPS complete sequence (14910 bp). The plot shows the Direct Sequence (Y-axis, 0.0 to 1.0) versus the sequence position (X-axis, 12800 to 14800). A red box highlights a peak in the Direct Sequence plot at approximately 14576 bp, indicating a potential coding region. The plot also shows the predicted gene structure (black bars) and the predicted protein sequence (lines) for the region.</p>

--	--

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	non
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	This candidate gene was revealed when we analyzed the terminal large gap in the annotation
DECISION:	?

Question : after whiB, there are 335 bp until the end. Is there a gene in this region with low coding potential? BlastX was performed against Phagedb (region used as query is highlighted in orange)

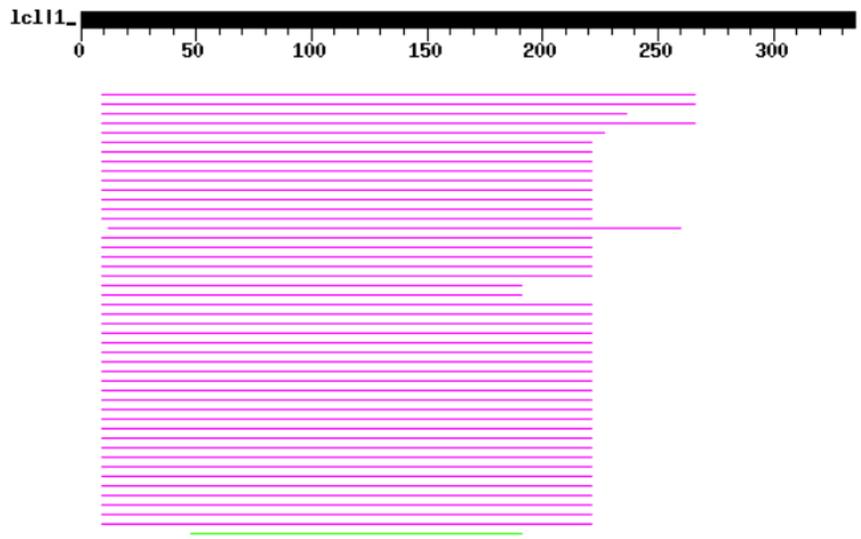
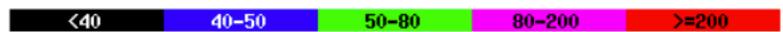


Query=
(335 letters)

Distribution of 100 Blast Hits on the Query Sequence

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

Color Key for Alignment Scores



>Schiebs_25, HNH endonuclease, 101
Length = 101

Score = 111 bits (277), Expect = 8e-25
Identities = 52/86 (60%), Positives = 58/86 (67%)
Frame = +1

Query: 10 WGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 189
WGR AQ LT L L TYGD CHLCGR GATTADH+VPRS GG D L NLRPAHG CN +R
Sbjct: 4 WGGRAAQALTALVLITYGDVCHLCGRPGATTADHLVPRSRRGGDSLGNLRPAHGPCNSAR 63

Query: 190 GNTPLHVWRRAKHPHSSQRPELSPRW 267
+ L W +HP+ + P S W
Sbjct: 64 QDMTLAEWFRRHPLTAPDTPPPSRDW 89

>RRH1_20, HNH endonuclease, 91
Length = 91

Score = 110 bits (274), Expect = 2e-24
Identities = 53/86 (61%), Positives = 62/86 (72%)
Frame = +1

Query: 10 WGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 189
W GR AQ+LT L LATYGD CHLCGR GATTADHVVPR GG + ++NLRPAH SCN R
Sbjct: 8 WSGRVAQRLTALTLATYGDVCHLCGRAGATTADHVVPRVQGGDNSIDNLRPAHLSCNSRR 67

Query: 190 GNTPLHVWRRAKHPHSSQRPELSPRW 267
G+ PL +R ++PI + P S RW
Sbjct: 68 GDMPLDQYRTRYPIVTGAPP--SRRW 91

Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
<p align="center">Quel start Glimmer et GeneMark suggèrent-ils ?</p>	<p align="center">Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : NA Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : NA</p>
<p align="center">Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?</p>	
<p align="center">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>	
<p align="center">Est-ce que le start est conservé chez les homologues voir Starterator ?</p>	
<p align="center">Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?</p>	
<p align="center">DECISION:</p>	

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

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

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

```
>Schiebs_25, HNH endonuclease, 101
Length = 101

Score = 112 bits (281), Expect = 2e-25
Identities = 53/89 (59%), Positives = 60/89 (67%)

Query: 1  MAGWGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCN 60
      M+ WGGR AQ LT L L TYGD CHLCGR GATTADH+VPRS GG D L NLRPAHG CN
Sbjct: 1  MSRWGGRAAQALTALVLITYGDVCHLCGRPGATTADHLVPRSRGGDDSLNLRPAHGPCN 60

Query: 61 YSRGNTPLHVWRAKHPHSSQRPELSPRW 89
      +R + L W +HP+ + P S W
Sbjct: 61 SARQDMTLAEWFRRHPLTAPDTPPPSRDW 89

>RRH1_20, HNH endonuclease, 91
Length = 91

Score = 110 bits (274), Expect = 2e-24
Identities = 53/86 (61%), Positives = 62/86 (72%), Gaps = 2/86 (2%)

Query: 4  WGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 63
      W GR AQ+LT L LATYGD CHLCGR GATTADHVVPR GG + ++NLRPAH SCN R
Sbjct: 8  WSGRVAQRLTALTALTYGDVCHLCGRAGATTADHVVPRVQGGDNSIDNLRPAHLSNCSRR 67

Query: 64 GNTPLHVWRAKHPHSSQRPELSPRW 89
      G+ PL +R ++PI + P S RW
Sbjct: 68 GDMPLDQYRTRYPIVTGAPP--SRRW 91

>GMA5_27, function unknown, 92
Length = 92

Score = 107 bits (268), Expect = 8e-24
Identities = 50/76 (65%), Positives = 57/76 (75%)

Query: 4  WGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 63
      WGG+ A+ LT L LATYG CHLCGR GA+TADHV+PRS GG D L+NLRPAH SCN SR
Sbjct: 7  WGGQVAKLTSLTALTYGRVCHLCGRPGASTADHVIPRSRGGRDALNLRPAHLSNCSQR 66

Query: 64 GNTPLHVWRAKHPHIS 79
      G+ L W A HP+ S
Sbjct: 67 GDMTLDEWFASHPLPS 82
```

*** nr :

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

HNH endonuclease [Corynebacterium striatum]
 Sequence ID: [HCTS225606.1](#) Length: 107 Number of Matches: 1

Range 1: 1 to 89 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
133 bits(335)	5e-38	Compositional matrix adjust.	62/89(70%)	72/89(80%)	0/89(0%)

```
Query 1  MAGWGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCN 60
      M+ WGGR+AQ LT+L LATYG RCHLCGR GATTADHVVPRSLGGTDEL NLRPAH SCN
Sbjct 1  MSRWGGRKAQGLTRLTLATYGTCHLCGRDATTADHVVPRSLGGTDELNLRPAHSSCN 60

Query 61 YSRGNTPLHVWRAKHPHSSQRPELSPRW 89
      Y R + PLH+WR ++P+++ SPRW
Sbjct 61 YKRQDMPLHIWRERYPMNAQAIQSRSPRW 89
```

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

MULTISPECIES: HNH endonuclease signature motif containing protein [unclassified Corynebacterium]
 Sequence ID: [WP_083314907.1](#) Length: 107 Number of Matches: 1

Range 1: 1 to 89 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
132 bits(333)	1e-37	Compositional matrix adjust.	61/89(69%)	72/89(80%)	0/89(0%)

```
Query 1  MAGWGGRRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCN 60
      M+ WGGR+AQ LT+L LATYG RC+LCGR GATTADHVVPRSLGGTDEL NLRPAH SCN
Sbjct 1  MSRWGGRKAQGLTRLTLATYGTTCYLCGRDATTADHVVPRSLGGTDELNLRPAHSSCN 60

Query 61 YSRGNTPLHVWRAKHPHSSQRPELSPRW 89
      Y R + PLH+WR ++P+++ SPRW
Sbjct 61 YKRQDMPLHIWRERYPMNAQAIQSRSPRW 89
```

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: 75
Query MSA diversity (Neff): 12.7961

Visualization



Resubmit Section

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
<input type="checkbox"/>	7ENH_A	CRISPR-associated endonuclease Cas9; Inhibitor, Complex, VIRAL PROTEIN; HET: NI; 2.097A (Staphylococcus aureus)	97.8	0.0003	35.41	4.9	64	166
<input type="checkbox"/>	5H0M_A	HNH endonuclease; Thermophilic bacteriophage, HNH Endonuclease, DNA nicking, HYDROLASE; 1.52A (Geobacillus virus E2)	97.61	0.00012	34.74	1.9	61	130
<input type="checkbox"/>	8D2K_A	CRISPR-associated endonuclease, Csn1 family; Cas9, AcCas9, Crispr, RNA BINDING PROTEIN, RNA BINDING PROTEIN-DNA-RNA comp	97.56	0.00085	43.37	5.3	64	1138
<input type="checkbox"/>	4H9D_C	HNH endonuclease; Structural Genomics, PSI-Biology, Protein Structure Initiative, Northeast Structural Genomics Consorti	97.54	0.0032	29.26	6.1	88	112
<input type="checkbox"/>	PF09665.15	; RE_Alw26IDE ; Type II restriction endonuclease (RE_Alw26IDE)	97.52	0.001	39.62	4.9	79	511
<input type="checkbox"/>	cd09643	Csn1; CRISPR/Cas system-associated protein Cas9. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and	97.5	0.0022	40.14	6.4	64	1049
<input type="checkbox"/>	8CTL_D	IscB; CRISPR, IscB, HEARO RNA, omega RNA, RNA BINDING PROTEIN-RNA-DNA complex; 3.1A (synthetic construct)	97.38	0.00087	39.69	3.6	54	496
<input type="checkbox"/>	4OGE_A	HNH endonuclease domain protein; CRISPR-Cas, Cas9, HNH, RuvC, RNA-guided DNA endonuclease, cytoplasmic, Hydrolase; HET:	97.35	0.0012	42.64	4.1	60	1101
<input type="checkbox"/>	P13340	END7_BPT4 Recombination endonuclease VII O5=Enterobacteria phage T4	97.3	0.0019	32.67	3.8	61	157

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 ?

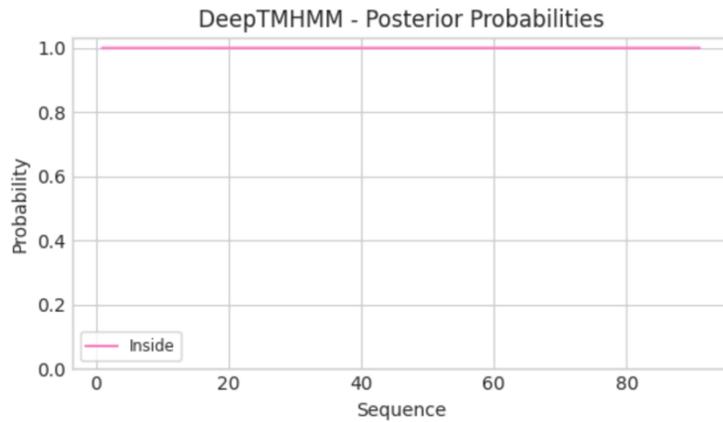
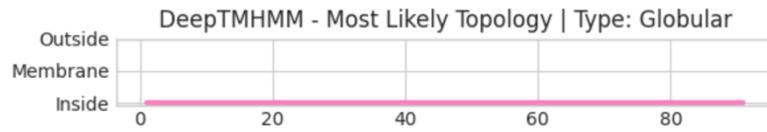
"Aucune synténie observée".

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

No

DeepTMHMM - Predictions

Predicted topologies can be downloaded in [.gff3 format](#) and [.3line format](#)



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

oui

DECISION:

HNH endonuclease (although coding potential low with GeneMark)