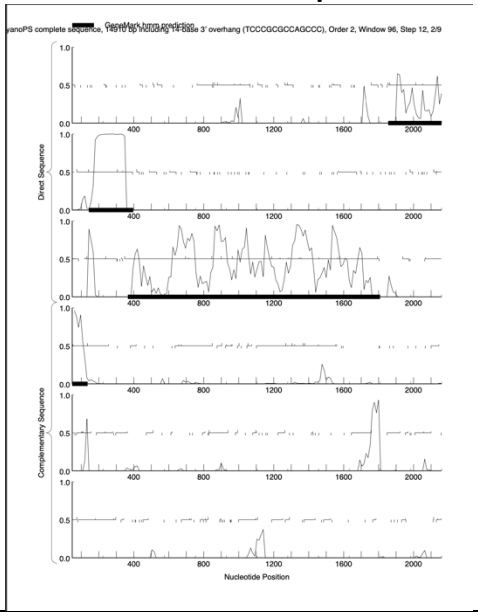
Student Gene Annotation Worksheet – **Phage CyranoPS** – University Paris-Saclay

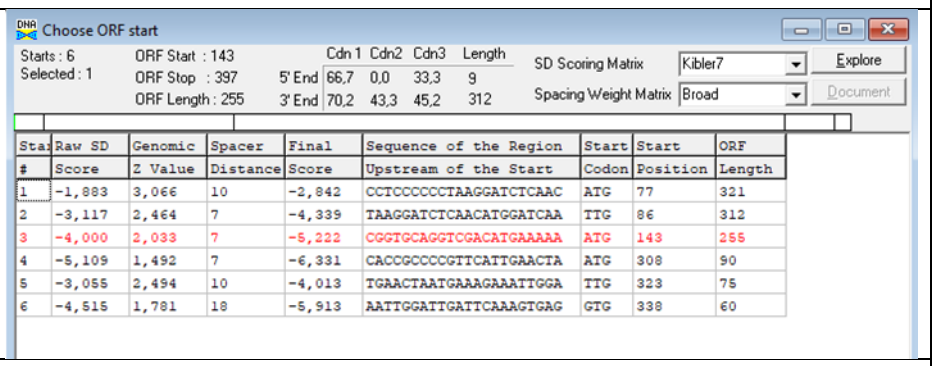
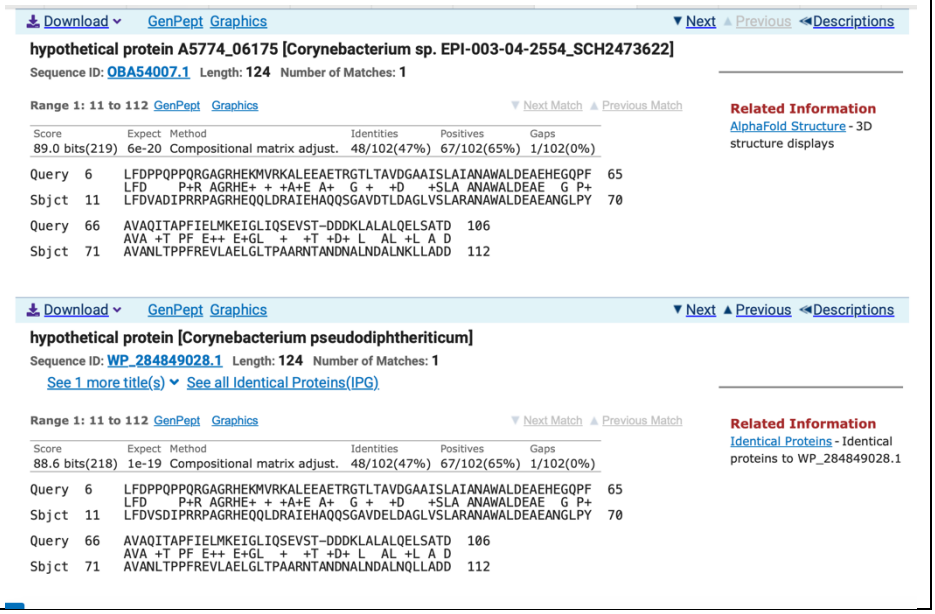
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**1**
Coordonnées du Stop	**397**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Ove 60**
Coordonnées du Start retenu	**77**
Fonction prédite	**terminase, small subunit**

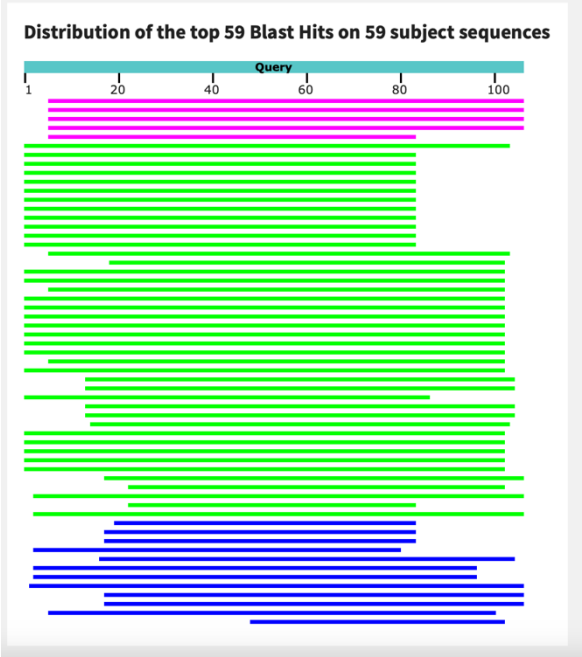
Décision #1 : Est-ce un gène ?

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 ?	**GeneMarkS montre un potentiel codant.** 

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**NR:** select all 59 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 A5774_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 — OBA54007.1 hypothetical protein [Corynebacterium pseudodiphtheriticum] — Corynebacterium pseudodiphtheriticum — 88.6 — 88.6 — 95% — 1e-19 — 47.06% — 124 — WP_284849028.1 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 hypothetical protein [Corynebacterium propinquum] — Corynebacterium propinquum — 84.7 — 84.7 — 95% — 3e-18 — 46.08% — 125 — WP_284594237.1 hypothetical protein [Corynebacterium pseudodiphtheriticum] — Corynebacterium pseudodiphtheriticum — 82.4 — 82.4 — 73% — 3e-17 — 51.28% — 124 — WP_284587129.1 hypothetical protein [Corynebacterium freiburgense] — Corynebacterium freiburgense — 75.9 — 75.9 — 97% — 8e-15 — 43.27% — 113 — WP_051255733.1 hypothetical protein [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 75.1 — 75.1 — 78% — 1e-14 — 44.58% — 106 — WP_041627971.1 hypothetical protein [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 75.1 — 75.1 — 78% — 1e-14 — 44.58% — 106 — WP_003852540.1 hypothetical protein [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 74.7 — 74.7 — 78% — 2e-14 — 44.58% — 114 — WP_241879851.1 hypothetical protein [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 74.7 — 74.7 — 78% — 2e-14 — 44.58% — 113 — WP_241880226.1 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) Distribution of 21 Blast Hits on the Query Sequence Mouse-over to show defline and scores. Click to show alignments Color Key for Alignment Scores <40 — 40-50 — 50-80 — 80-200 — >=200 lcl|1_ 0 50 100 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 McGonagall_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 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 ?

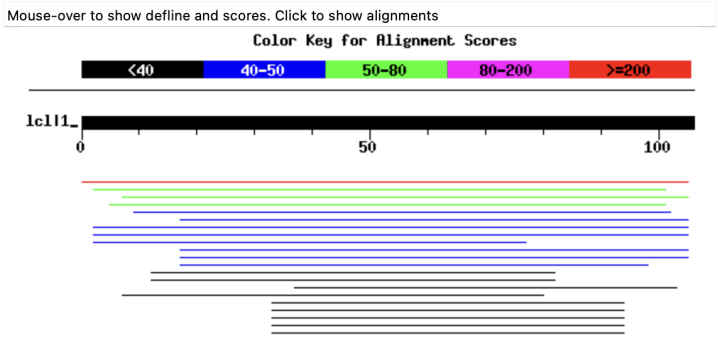
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 77** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 143**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 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 CCTCCCCCCTAAGGATCTCAAC ATG 77 321 2 -3,117 2,464 7 -4,339 TAAGGATCTCAACATGGATCAA TTG 86 312 3 -4,000 2,033 7 -5,222 CGGTGCAGGTCGACATGAAAAA ATG 143 255 4 -5,109 1,492 7 -6,331 CACCGCCCCGTTCATTGAACTA ATG 308 90 5 -3,055 2,494 10 -4,013 TGAACTAATGAAAGAAATTGGA TTG 323 75 6 -4,515 1,781 18 -5,913 AATTGGATTGATTCAAAGTGAG GTG 338 60
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	**ATG 77 = ORF la plus longue 321pb** **ATG 143 = ORF de 255pb** **TTG 86 = ORF de 312 pb** **-> Favoriser celui de glimmer à 77 : ORF le plus long**
Est-ce que le start est conservé chez les homologues voir Starterator ?	**Pas d'information sur Starterator** Starterator | **Orpham, no data**
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	**Nr :** 



Phage DB :

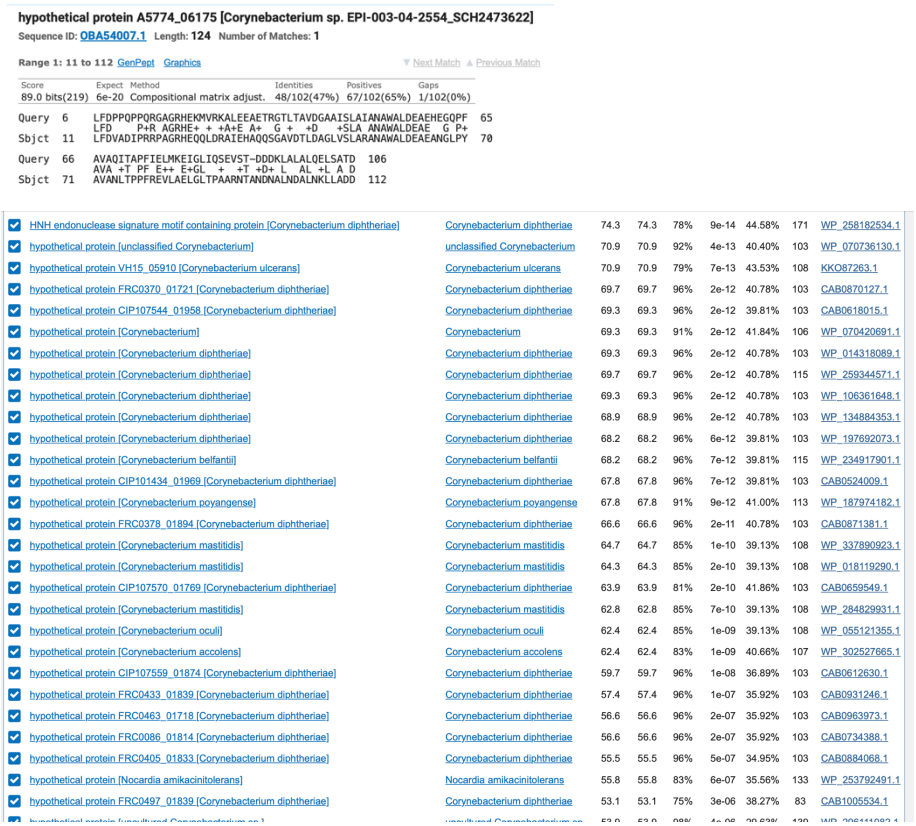
```
>Schiebs_1, terminase small subunit, 135
          Length = 135

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

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

Query: 68  AQITAPFIELMKEIGLIQSEVSTD-DDKLALALQELSATD 106
           A++  P +  ++E  L      T  DD +   L +L++ D
Sbjct: 77  AKLIEPMVNALREARLTPDARQTSVDDSIKELLGDLASAD 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   QLRLFDPPQP---PQRGAGRHEKMVRKALEEAETRGTLTAVDGAAISLAIANAWALDEAE 59
           Q  +FD  +P   P R  GRH +   +AL  A+    +T VD A  ++  A+AWALD  E
Sbjct: 13  QAAMFDAGEPVATPPRH-GRHSEATERALAAAKAADLITDVDEALAAVVRASAWALDRFE 71
```

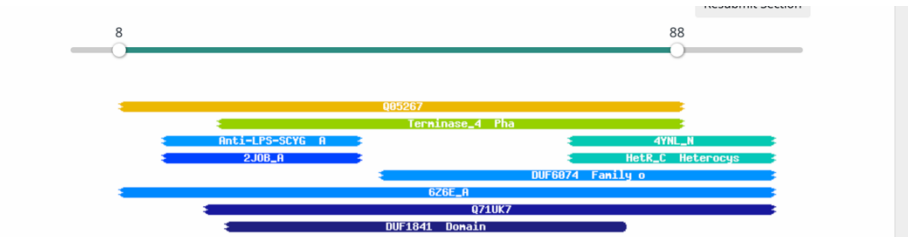
Distribution of 21 Blast Hits on the Query Sequence

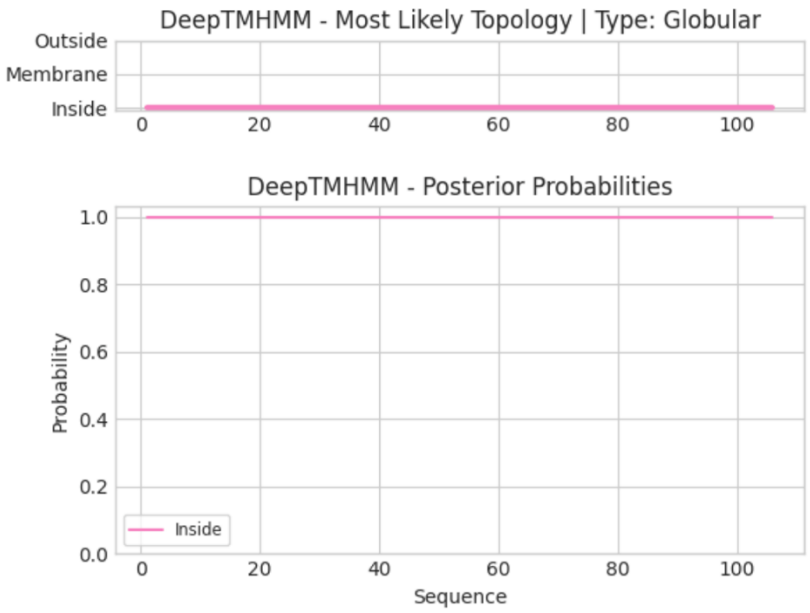


DECISION:	**ATG 77 : meilleur score et ORF la plus longue .**

Décision #3 : Quelle est la function 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 :** **(q6: s11) :** >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 DPPQPPQRGAGRHEKMVRKALEEAETRGTLTAVDGAAISLAIANAWALDEAEHEGQPFAV 67 DP PP GRH V +AL A G + VD A +++ A AW+LD E + QP+ Sbjct: 20 DPATPPP---GRHSAAVERALGAARDAGLVDDVDEALLTIVTAGAWSLDTFERQNQPYGP 76 Query: 68 AQITAPFIELMKEIGLIQSEVSTD-DDKLALALQELSATD 106 A++ P + ++E L T DD + L +L++ D Sbjct: 77 AKLIEPMVNALREARLTPDARQTSVDDSIKELLGDLASAD 116 ***** nr :** 
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors	**"Pas de correspondance pertinente", les probabilités sont toutes inferieures a 90%.**

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** 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 2 PF05119.16 ; Terminase_4 ; Phage terminase, small subunit 78.91 32 23.14 8.4 67 108 3 4YNL_N Heterocyst differentiation control protein; Heterocyst differentiation, Transcription factor, TRANSCRIPTION; 2.1A {Nosto 59.47 2.9 29.58 -2 30 90 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 gene suivant code pour une « terminase large subunit »**
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**NON**

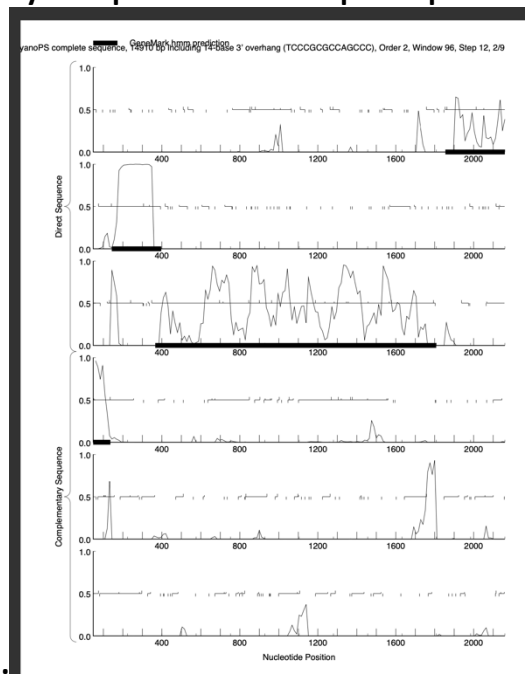
	DeepTMHMM - Predictions Predicted topologies can be downloaded in .gff3 format and .3line format 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**OUI**
DECISION:	**terminase, small subunit**

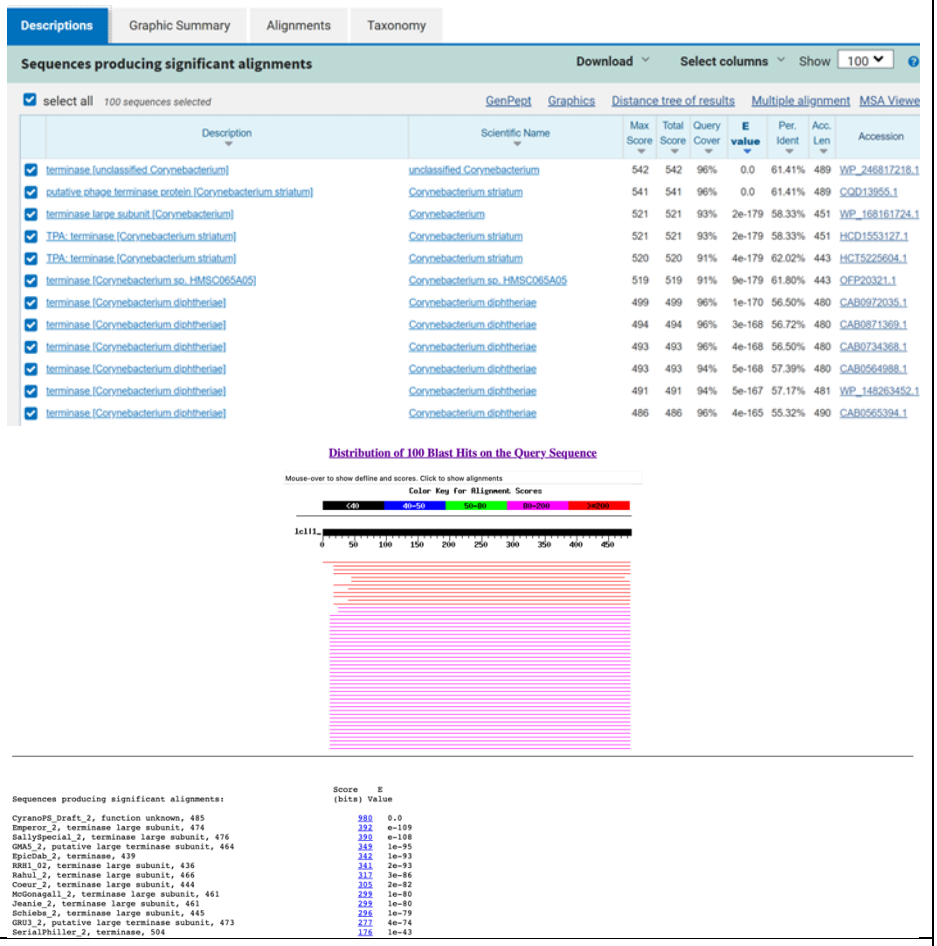
Student Gene Annotation Worksheet

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 ?	**Il y a un potentiel codant prédit par la carte de Genemark** . 

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 Viewe 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 Sequences producing significant alignments: Score (bits) E Value
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**

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
terminase [unclassified Corynebacterium]	unclassified Corynebacterium	542	542	96%	0.0	61.41%	489	WP_246817218.1
putative phage terminase protein [Corynebacterium striatum]	Corynebacterium striatum	541	541	96%	0.0	61.41%	489	CQD13955.1
terminase large subunit [Corynebacterium]	Corynebacterium	521	521	93%	2e-179	58.33%	451	WP_168161724.1
TPA: terminase [Corynebacterium striatum]	Corynebacterium striatum	521	521	93%	2e-179	58.33%	451	HCD1553127.1
TPA: terminase [Corynebacterium striatum]	Corynebacterium striatum	520	520	91%	4e-179	62.02%	443	HCT5225604.1
terminase [Corynebacterium sp. HMSC065A05]	Corynebacterium sp. HMSC065A05	519	519	91%	9e-179	61.80%	443	OFP20321.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	499	499	96%	1e-170	56.50%	480	CAB0972035.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	494	494	96%	3e-168	56.72%	480	CAB0871369.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	493	493	96%	4e-168	56.50%	480	CAB0734368.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	493	493	94%	5e-168	57.39%	480	CAB0564988.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	491	491	94%	5e-167	57.17%	481	WP_148263452.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	486	486	96%	4e-165	55.32%	490	CAB0565394.1

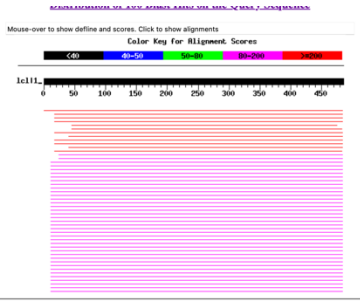
Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_2, function unknown, 485	980	0.0
Emperor_2, terminase large subunit, 474	392	e-109
SallySpecial_2, terminase large subunit, 476	390	e-108
GMA5_2, putative large terminase subunit, 464	349	1e-95
EpicDab_2, terminase, 439	342	1e-93
RRH1_02, terminase large subunit, 436	341	2e-93
Rahul_2, terminase large subunit, 466	317	3e-86
Coeur_2, terminase large subunit, 444	305	2e-82
McGonagall_2, terminase large subunit, 461	299	1e-80
Jeanie_2, terminase large subunit, 461	299	1e-80
Schiebs_2, terminase large subunit, 445	296	1e-79
GRU3_2, putative large terminase subunit, 473	277	4e-74
SerialPhiller_2, terminase, 504	176	1e-43

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) : 351** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 366**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	Choose ORF start Starts : 21 Selected : 1 — ORF Start : 966 — ORF Stop : 1808 — ORF Length : 843 — Cdn 1 Cdn2 Cdn3 Length — 5' End 0,0 66,7 66,7 9 — 3' End 61,7 49,1 50,9 1449 — SD Scoring Matrix: Kibler7 — Spacing Weight Matrix: Broad — Explore — Document Sta# | Raw SD Score | Genomic Z Value | Spacer Distance | Final Score | Sequence of the Region Upstream of the Start | Start Codon | Start Position | ORF Length 1 | -3,933 | 2,066 | 14 | -5,030 | CAAAGTGAGGTGAGCACAGACG | ATG | 351 | 1458 2 | -4,568 | 1,756 | 11 | -5,568 | GTGAGCACAGACGATGACAAGC | TTG | 360 | 1449 3 | -4,568 | 1,756 | 17 | -5,869 | ACAGACGATGACAAGCTTGCAC | 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 | ATTACATTCTGAACAGTCGGAC | ATG | 849 | 960 7 | -5,278 | 1,409 | 16 | -6,500 | TGAAGCGCAAGCTAGCGCACTC | ATG | 903 | 906 8 | -4,265 | 1,904 | 17 | -5,566 | GCAAACGATTATCACTAGCACA | ATG | 966 | 843 9 | -5,500 | 1,301 | 13 | -6,571 | ACATCACCCGGCATATGGCCTA | TTG | 1119 | 690 10 | -5,419 | 1,341 | 15 | -6,574 | AACCGCTGGTTTTGCCCGCGCT | ATG | 1191 | 618 11 | -7,444 | 0,352 | 15 | -8,599 | TGATTCCCTAATCCCGCTGCAG | GTG | 1239 | 570 12 | -5,358 | 1,370 | 6 | -6,659 | TGTCATTCCCCCTGATGTATTT | GTG | 1287 | 522 13 | -4,877 | 1,605 | 12 | -5,923 | GGCAATCGCTGCAGCTGCTTGG | GTG | 1350 | 459 14 | -3,684 | 2,187 | 7 | -4,906 | TGCAGCTGCTTGGGTGGACGGC | GTG | 1359 | 450 15 | -5,077 | 1,507 | 15 | -6,232 | ACCCGCAGGCCAAACCACCCTA | GTG | 1461 | 348 16 | -4,224 | 1,923 | 15 | -5,379 | CGCAGAAGCAACCCGCGCGAAC | TTG | 1518 | 291 17 | -5,720 | 1,193 | 8 | -6,845 | TCTCGCCACCGCCACGGTTGAA | GTG | 1569 | 240 18 | -3,444 | 2,304 | 8 | -4,568 | CACGGTTGAAGTGTGGGACAGA | ATG | 1581 | 228 19 | -5,094 | 1,499 | 12 | -6,139 | GCTTGAAGCTGCAACCCTTGCT | TTG | 1755 | 54 20 | -6,896 | 0,620 | 7 | -8,118 | TGCAACCCTTGCTTTGCATGGC | TTG | 1764 | 45 21 | -4,023 | 2,022 | 7 | -5,245 | CCCCGTACCCGCACAGAAACCA | ATG | 1797 | 12
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	**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 ; cehavauchement 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 ?	**PhageDB.** >Emperor_2, terminase large subunit, 474 Length = 474 Score = 392 bits (1008), Expect = e-109 Identities = 219/475 (46%), Positives = 298/475 (62%), Gaps = 14/475 (2%) Query: 18 PRFITEW-PDAPSVGSKVAKVQRSLGSEPLPWQVLAAHAVGARNPDGTPRFPFVLVSVPR 76 PR++T PD S G +A+ LG +PWQ AA +G + +G +P V+++V R Sbjct: 4 PRYLTPRNPDRRSFGPNIARTSALLGRPFMPWQHAAAGVIGEVDAEGRFAYPLVIMTVQR 63

	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) Range 1: 1 to 450 GenPept Graphics ▼ Next Match ▲ Previous Matc Score Expect Method Identities Positives Gaps 521 bits(1343) 2e-179 Compositional matrix adjust. 266/456(58%) 337/456(73%) 7/456(1%) Query 30 VGSKVAKVQRSLGSEPLPWQVLAAHAVGARNPDGTPRFPFVLVSVPRQAGKTRASWSWLY 89 +GSK+ +V + LG P+PWQ LAAH +GAR P+G P++PF++++VPRQ+GKT + + Sbjct 1 MGSKIEQVSKYLGRTPMPWQRLAAHIIGARLPNGRPKWPFIVITVPRQSGKTTICSAVQF 60
DECISION:	**351 : meilleur ORF, score bon, ATG**

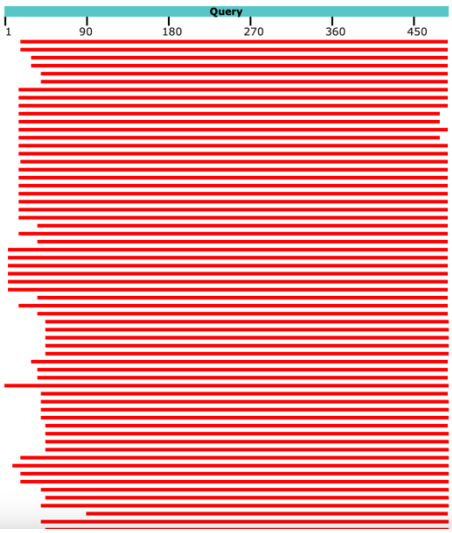
Décision #3 : Quelle est la function de la protéine putative ?

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 :**  Score E Sequences producing significant alignments: (bits) Value CyranoPS_Draft_2, function unknown, 485 980 0.0 Emperor_2, terminase large subunit, 474 392 e-109 SallySpecial_2, terminase large subunit, 476 390 e-108 GMA5_2, putative large terminase subunit, 464 349 1e-95 EpicDab_2, terminase, 439 342 1e-93 RRH1_02, terminase large subunit, 436 341 2e-93 Rahul_2, terminase large subunit, 466 317 3e-86 Coeur_2, terminase large subunit, 444 305 2e-82 McGonagall_2, terminase large subunit, 461 299 1e-80 ***** nr :** **(q#: s#) : [alignment] ; e-value :**

terminase [Corynebacterium diphtheriae]

No putative conserved domains have been detected

Distribution of the top 100 Blast Hits on 100 subject sequences



putative phage terminase protein [Corynebacterium striatum]	Corynebacterium striatum	541	541	96%	0.0	61.41%	489	CQD13955.1
terminase large subunit [Corynebacterium]	Corynebacterium	521	521	93%	2e-179	58.33%	451	WP_168161724.1
TPA: terminase [Corynebacterium striatum]	Corynebacterium striatum	521	521	93%	2e-179	58.33%	451	HCD1553127.1
TPA: terminase [Corynebacterium striatum]	Corynebacterium striatum	520	520	91%	4e-179	62.02%	443	HCT5225604.1
terminase [Corynebacterium sp. HMSC065A05]	Corynebacterium sp. HMSC065A05	519	519	91%	9e-179	61.80%	443	OFP20321.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	499	499	96%	1e-170	56.50%	480	CAB0972035.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	494	494	96%	3e-168	56.72%	480	CAB0871369.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	493	493	96%	4e-168	56.50%	480	CAB0734368.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	493	493	94%	5e-168	57.39%	480	CAB0564988.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	491	491	94%	5e-167	57.17%	481	WP_148263452.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	486	486	96%	4e-165	55.32%	490	CAB0565394.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	483	483	94%	4e-164	58.04%	481	CAB0659529.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	483	483	96%	5e-164	56.93%	480	CAB0919103.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	482	482	96%	1e-163	56.50%	480	CAB0733954.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	481	481	96%	2e-163	56.84%	480	CAB0519177.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	481	481	96%	2e-163	56.72%	480	WP_241008567.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	481	481	96%	3e-163	56.72%	480	CAB0913794.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	480	480	96%	5e-163	56.93%	480	WP_342351212.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	480	480	96%	8e-163	56.29%	480	CAB0519844.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	480	480	96%	9e-163	56.29%	480	CAB1005517.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	479	479	96%	1e-162	56.72%	481	CAB0884063.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	478	478	96%	5e-162	56.72%	481	CAB0612654.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	474	474	92%	5e-161	56.70%	445	RLP10252.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	474	474	96%	2e-160	56.08%	480	CAB0567705.1
hypothetical protein CDBH8_1793 [Corynebacterium diphtheriae BH8]	Corynebacterium diphtheriae BH8	470	470	92%	2e-159	56.25%	445	AEX49311.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	468	468	98%	6e-158	54.98%	481	WP_258174496.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	468	468	98%	8e-158	54.98%	485	PSA85850.1
terminase [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	466	466	98%	2e-157	54.77%	481	WP_226813674.1

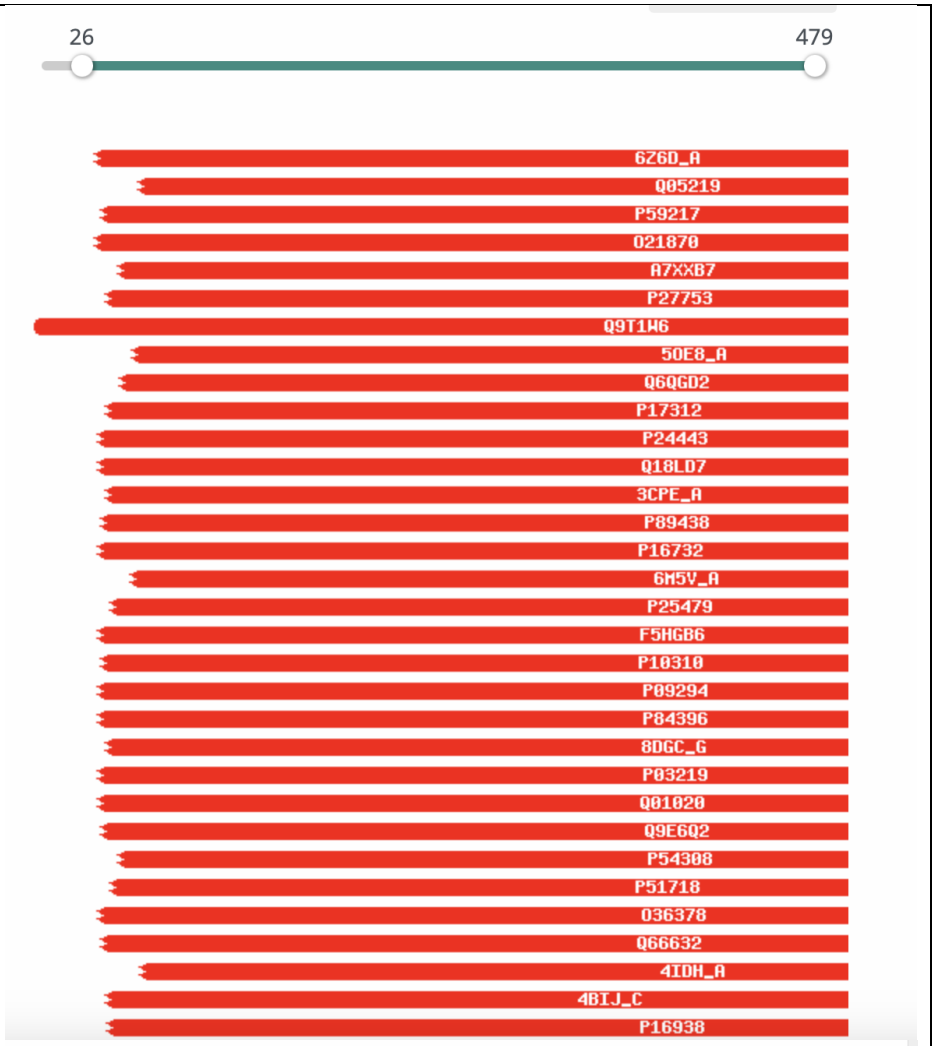
MULTISPECIES: terminase [unclassified Corynebacterium]

Sequence ID: WP_246817218.1 **Length:** 489 **Number of Matches:** 1

Range 1: 26 to 488 GenPept Graphics ▼ Next Match ▲ Previous Match

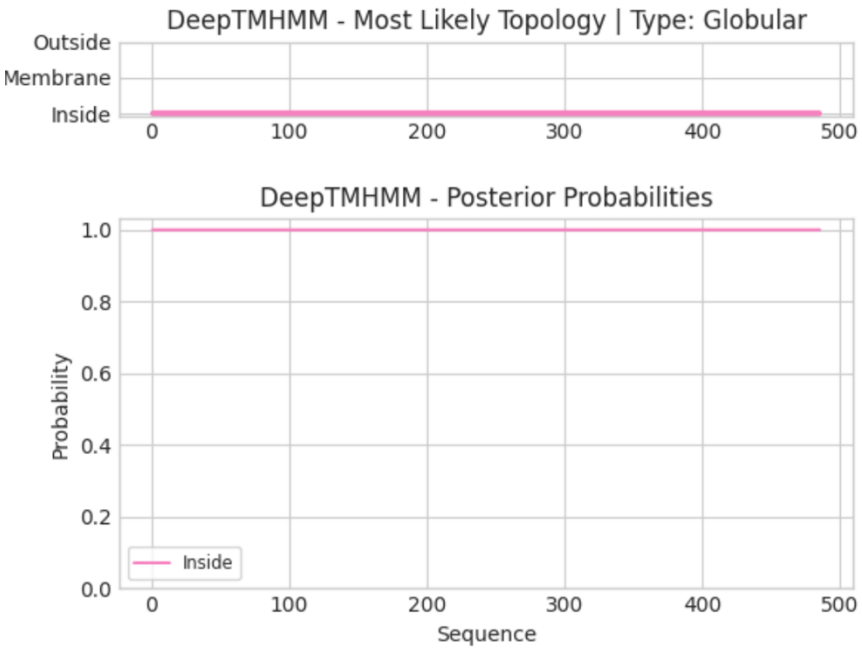
Score	Expect	Method	Identities	Positives	Gaps
542 bits(1396)	0.0	Compositional matrix adjust.	288/469(61%)	344/469(73%)	8/469(1%)

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 >= 90% et une couverture acceptable ?



Vis Hits Aln | Select All | Forward | Forward Query A3M | Model using selection | Download HHR | Color Seqs | Wrap Seqs

	Hits	Aln						
☐	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
☐	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
☐	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
☐	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
☐	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
☐	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
☐	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
☐	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
☐	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
☐	10	P17312	TERL_BPT4 Terminase, large subunit OS=Enterobacteria phage T4 OX=10665	100	3.3e-33	278.83	41.8	406	610

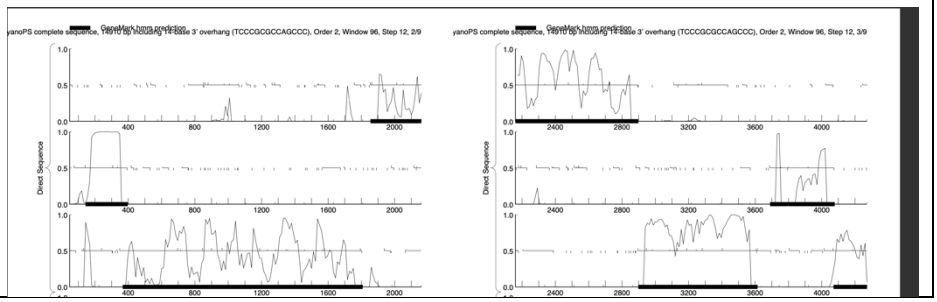
	PDB : Terminase large subunit; genome packaging, bacteriophage, ATPase, nuclease, VIRAL PROTEIN; HET: BR; 2.2A {Enterobacteria phage HK97} **UNIPROT : Q05219 VG13_BPML5 Gene 13 protein OS=Mycobacterium phage L5 OX=31757 GN=13 PE=3 SV=1** **Pfam : TerL_nuclease ; Terminase large subunit, endonuclease domain** **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.**
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 gène avant code une terminase small subunit. Et le gene apres code une « portal protein »**
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**OUI**
DECISION:	**Terminase, large subunit**

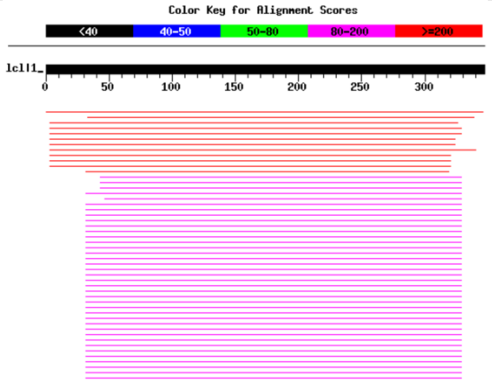
Student Gene Annotation Worksheet

Basic Phage Information

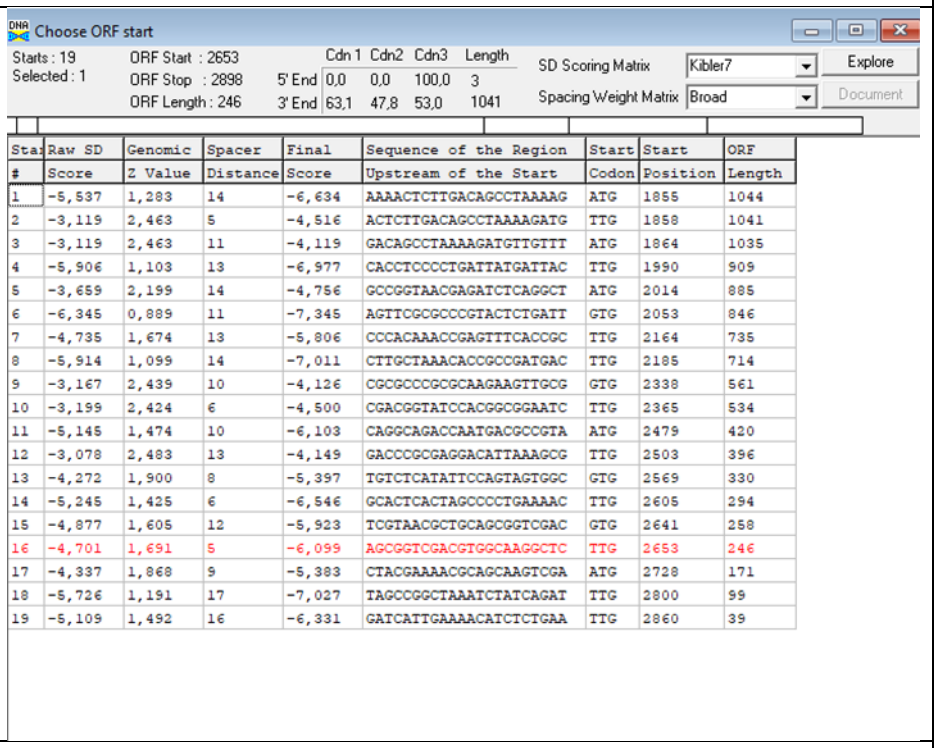
Nom du Phage	CyranoPS
Gène #	3
Coordonnées du Stop	2898
Direction (For/Rev)	For
Gap/chevauchement avec un autre gène	Gap 55
Coordonnées du Start retenu	1864
Fonction prédite	portal protein

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 ?	GeneMarkS montre un potentiel codant : 

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 MSLLDRIRKLGQLPQLLATGTAPLSVWAGD-HLEQVTSPDYDYLPVTRSQAMAIPAVVRA 62 M L+DR+R+L LPQ A S WA HL PD + + R AM IPA++RA Sbjct 1 MGLIDRLRRLTSLPQQFAAAQYG-SPWANSAHLAIADFPDLEDSAINRDYAMQIPALLRA 59 Query 63 RTLIVTTVARCPIVSDTDEITPGWLHGIAPLGGHPQTEFHRLLNTADDLLFHGSAAWAIE 122 R L+VTT+AR P+ + + P W+ G P G PQTEF+R+L TADDL F+G +AWAIE Sbjct 60 RNLVVTTIARVPLTA-VEGTAPAWIAGTVPSTGKPQTEFYRMLWTADDLFFYGRSAWAIE 118 Query 123 RDELGNVTAAVHIPRSLWERTADGTIIVDGVPAPAQEVAVFDGIHGGILTHGARALRDAS 182 RD G T A+HIP +LW +A+G +IVDG A EV F GIH GIL HGA ALRDA Sbjct 119 RDAQGVPTVAIHIPLALWNYSAEGDVIVDGEVADPAEVCTFSGIHAGILEHGADALRDAH 178 Query 183 NILSAAARVADTPAALIELRQTNDAVMTREDIKALIQGYVDARRGKNGGVSYSSSGVEVH 242 + A+ARVAD PAAL ELRQTNDA++ ++I ALI YV AR+G+ GVSYSS GVEV Sbjct 179 RLNRASARVADAPAALTELRQTNDAILNDDEIDALIARYVQARQGRRHGVSYSSPGVEVV 238 Query 243 EHSLAPENLLIEGRNAAAVDVARLLGVPAPFIDATVGGTSLSYENAASRMTELITFGVAP 302 EHSLAPENLLI GRNAA VD+ARL +PAPFIDATVGGTSLSYEN+ASRMTELI FG++P Sbjct 239 EHSLAPENLLIAGRNAATVDIARLANLPAPFIDATVGGTSLSYENSASRMTELIAFGLSP 298 Query 303 LLAAISSRLNLSDLTPVG-TVSFDTAQIIENISELLPTNNTTN 344 LAAI++RLNL+D+ P G + FDT +I+++I PT T N Sbjct 299 TLAAITARLNLADMNPAGAKIRFDTERILDDI----PTAATPN 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
Quel start Glimmer et GeneMark suggèrent-ils ?	**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**
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 ?	**. 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**
Est-ce que le start est conservé chez les homologues voir Starterator ?	**–**

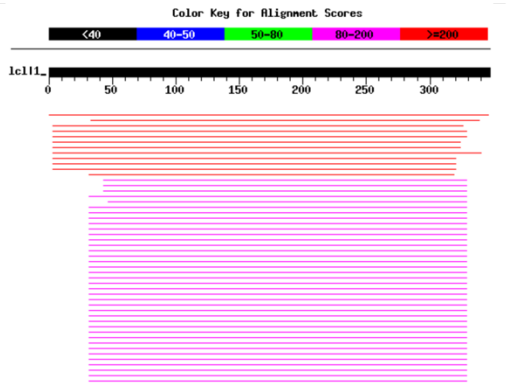
Choose ORF start

Starts : 19 Selected : 1 | ORF Start : 2653 | ORF Stop : 2898 | ORF Length : 246 | Cdn 1 Cdn2 Cdn3 Length | 5' End 0,0 0,0 100,0 3 | 3' End 63,1 47,8 53,0 1041 | SD Scoring Matrix: Kibler7 | Spacing Weight Matrix: Broad | Explore | Document

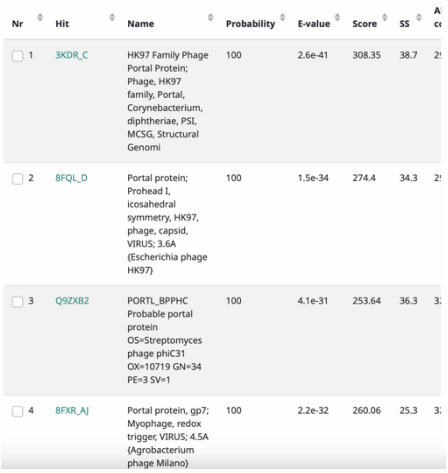
Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-5,537	1,283	14	-6,634	AAAACTCTTGACAGCCTAAAAG	ATG	1855	1044
2	-3,119	2,463	5	-4,516	ACTCTTGACAGCCTAAAAGATG	TTG	1858	1041
3	-3,119	2,463	11	-4,119	GACAGCCTAAAAGATGTTGTTT	ATG	1864	1035
4	-5,906	1,103	13	-6,977	CACCTCCCCTGATTATGATTAC	TTG	1990	909
5	-3,659	2,199	14	-4,756	GCCGGTAACGAGATCTCAGGCT	ATG	2014	885
6	-6,345	0,889	11	-7,345	AGTTCGCGCCCGTACTCTGATT	GTG	2053	846
7	-4,735	1,674	13	-5,806	CCCACAAACCGAGTTTCACCGC	TTG	2164	735
8	-5,914	1,099	14	-7,011	CTTGCTAAACACCGCCGATGAC	TTG	2185	714
9	-3,167	2,439	10	-4,126	CGCGCCCGCGCAAGAAGTTGCG	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	GACCCGCGAGGACATTAAAGCG	TTG	2503	396
13	-4,272	1,900	8	-5,397	TGTCTCATATTCCAGTAGTGGC	GTG	2569	330
14	-5,245	1,425	6	-6,546	GCACTCACTAGCCCCTGAAAAC	TTG	2605	294
15	-4,877	1,605	12	-5,923	TCGTAACGCTGCAGCGGTCGAC	GTG	2641	258
16	-4,701	1,691	5	-6,099	AGCGGTCGACGTGGCAAGGCTC	TTG	2653	246
17	-4,337	1,868	9	-5,383	CTACGAAAACGCAGCAAGTCGA	ATG	2728	171
18	-5,726	1,191	17	-7,027	TAGCCGGCTAAATCTATCAGAT	TTG	2800	99
19	-5,109	1,492	16	-6,331	GATCATTGAAAACATCTCTGAA	TTG	2860	39

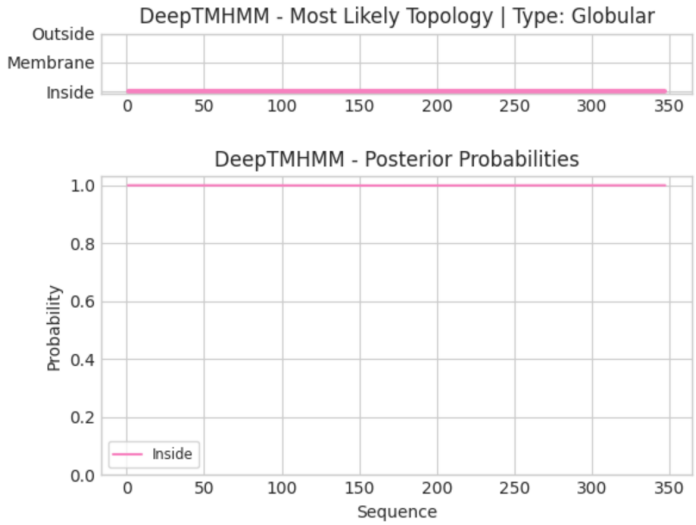
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	**START 1855 pas conservé :** **phage portal protein [Corynebacterium striatum]** Sequence ID: HCD1553128.1 Length: 368 Number of Matches: 1 See 4 more title(s) See all Identical Proteins(IPG) Range 1: 1 to 337 GenPept Graphics ▼ 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-HLEQVTSPDYDYLPVTRSQAMAIPAVVRA 62 M L++R+R+L LPQ LA S WA HL PD + + R AM IPA++RA Sbjct 1 MGLINRLRRLTSLPQQLAAAQYG-SPWANSAHLAIADFPDLEDSAIHRDYAMQIPALLRA 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 HLEQVTSPDYDYL----PVTRSQAMAIPAVVRARTLIVTTVARCPIVS---DTDEITPGW 86 HLE + D L PVTR++AMA+PA+ RAR L+ T+A P+ + + +++ W Sbjct: 26 HLEPIVWADVAGLATHQPVTRAEAMAVPALARARNLLAGTIAGIPLRAVRGEGEDVPLSW 85 Query: 87 LH---GIAPLGGHPQTEFHRLLNTADDLLFHGSAAWAIERDELGNVTAAVHIPRSLWERT 143 L GIA + F R+L+T DDLLF+G + W + RD GNV A H+P S W Sbjct: 86 LDRSDGIA-------SPFQRMLHTTDDLLFYGISLWGLRRDFDGNVVRAAHVPMSRWTTD 138 Query: 144 ADGTIIVDGVPAPAQEVAVFDGIHGGILTHGARALRDASNILSAAARVADTPAALIELRQ 203 +G I++DG P ++EV V GIH GILT+G +R A N++ A + + P A IEL Q Sbjct: 139 TEGYILIDGEPVSSREVCVIPGIHEGILTYGRTTIRHAGNLIRHADKAVENPNAYIELHQ 198 Query: 204 TNDAVMTREDIKALIQGYVDARRGKNGGVSYSSSGVEVHEHSLAPENLLIEGRNAAAVDV 263 +DA MT DI+ L+ +V ARRG NGGV++++ G+ V EH E LL EGRNAAA+D+ Sbjct: 199 ESDAPMTEPDIEKLVNRWVKARRGDNGGVAFTNKGIRVIEHGSPKELLLTEGRNAAAIDI 258 Query: 264 ARLLGVPAPFIDATVGGTSLSYENAASRMTELITFGVAPLLAAISSRLNLSDLTPVGT-V 322 ARL G+PA +DA GTS++Y+N+ +RM EL+TFG+AP++ AI +RL D+ GT V Sbjct: 259 ARLTGLPASLVDAAQSGTSVTYQNSQARMAELVTFGLAPIMGAIVARLGQDDIVARGTRV 318 Query: 323 SFDTAQIIENISELLPTN 340 FD +EN LPT+ Sbjct: 319 EFD----LENALAALPTS 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-GDHLEQVTSPDYDYLPVTRSQAMAIPAVVRA 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 I 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-HLEQVTSPDYDYLPVTRSQAMAIPAVVRA 59 M L++R+R+L LPQ LA S WA HL PD + + R AM IPA++RA Sbjct 1 MGLINRLRRLTSLPQQLAAAQYG-SPWANSAHLAIADFPDLEDSAIHRDYAMQIPALLRA 59
DECISION:	**1864 : ATG ; bon score ; le plus conservé ; ORf de grande taille. .**

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

Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	**Listez le meilleur hit Blastp pour chaque source :** ***** PhagesDB :** **(q#: s#) : [alignment] ; e-value :**  Sequences producing significant alignments: Score (bits) E Value 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 ***** nr :** **(q#: s#) : [alignment] ; e-value :**

$^{-4}$

Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	  Nr Hit Name Probability E-value Score SS 1 3KDR_C HK97 Family Phage Portal Protein; Phage, HK97 family, Portal, Corynebacterium, diphtheriae, PSI, MCSG, Structural Genomi 100 2.6e-41 308.35 38.7 2 8FQL_D Portal protein; Prohead I, icosahedral symmetry, HK97, phage, capsid, VIRUS; 3.6A {Escherichia phage HK97} 100 1.5e-34 274.4 34.3 3 Q9ZXB2 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 4 8FXR_AJ Portal protein, gp7; Myophage, redox trigger, VIRUS; 4.5A {Agrobacterium phage Milano} 100 2.2e-32 260.06 25.3 **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**

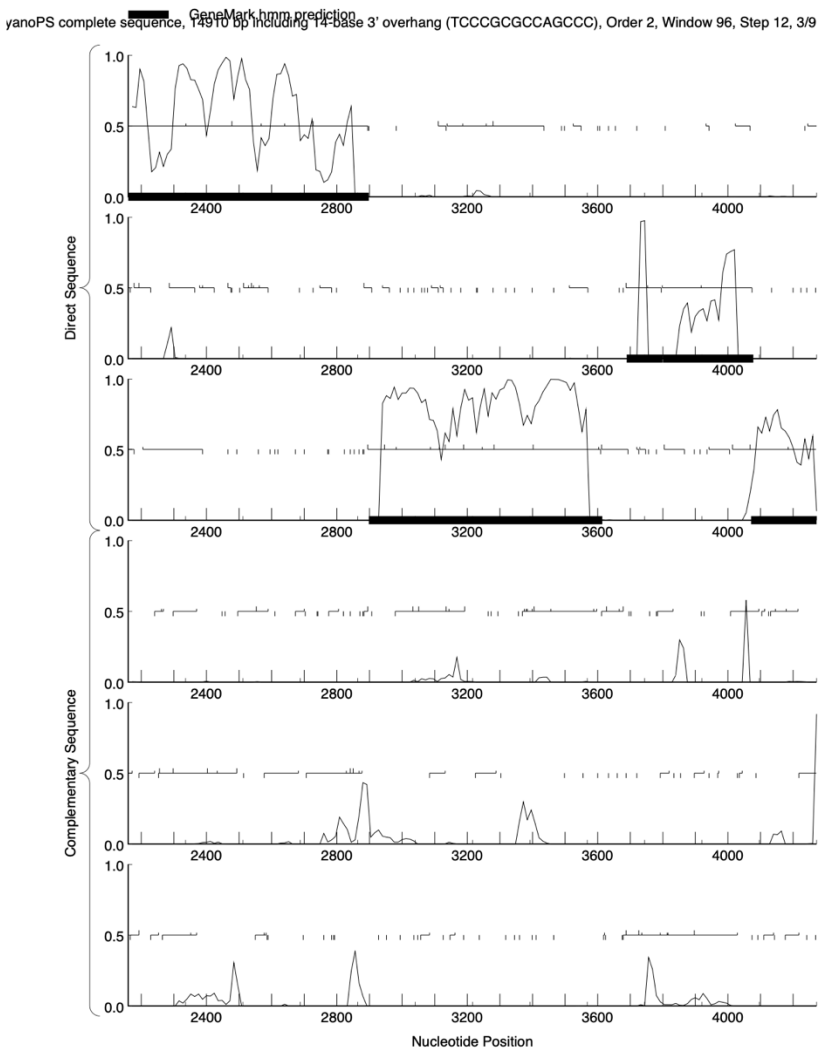
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**oui**
DECISION:	**portal protein**

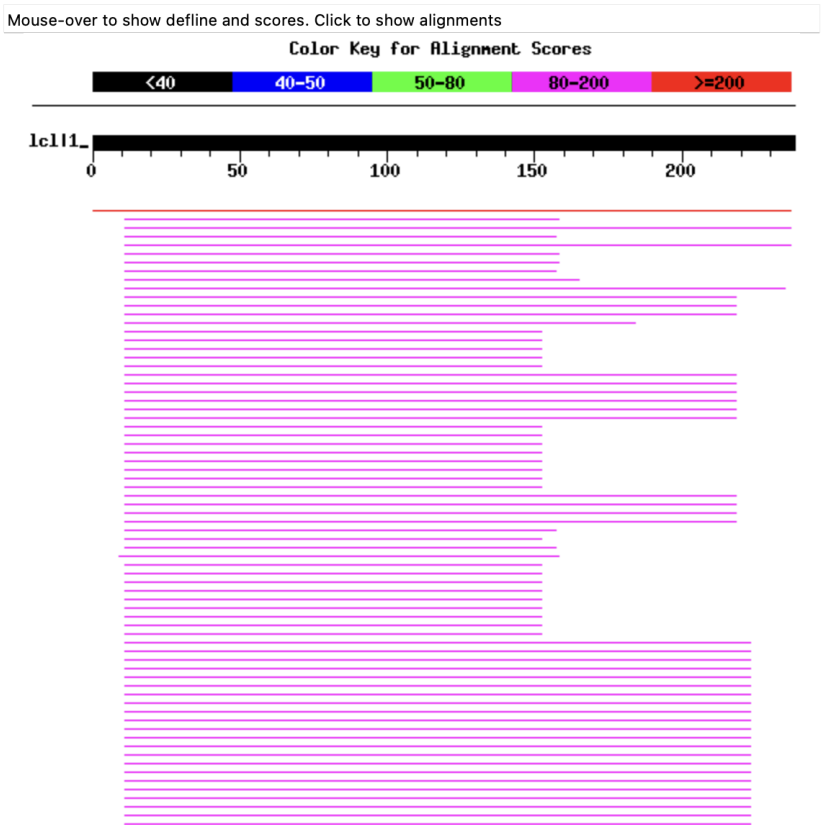
Student Gene Annotation Worksheet

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**

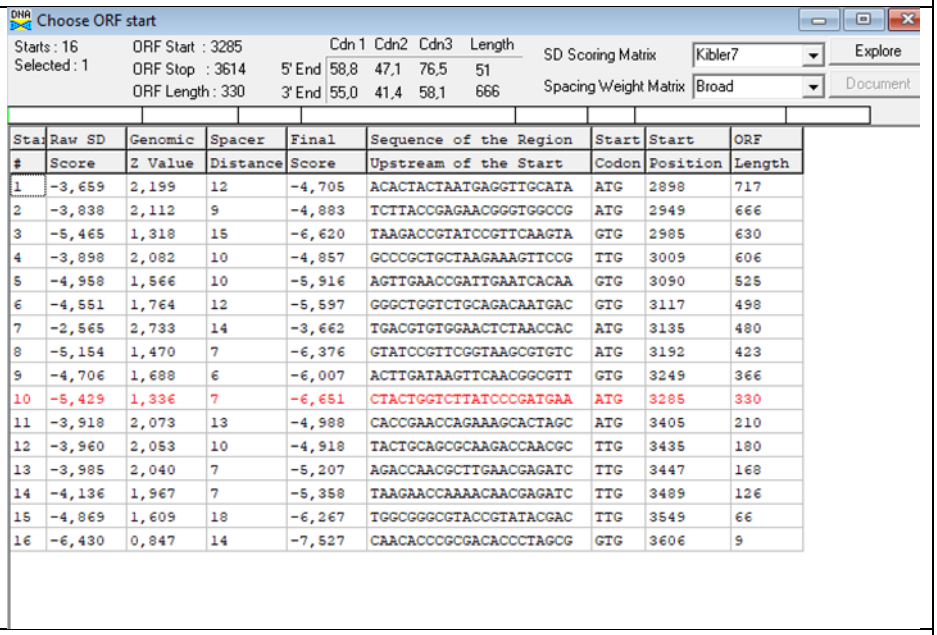
Décision #1 : Est-ce un gène ?

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**

Y-a-t-il des éléments supportant un potentiel codant ?	**Oui il y a un potentiel codant avec GeneMarkS** 

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**Distribution of 100 Blast Hits on the Query Sequence**  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 RYKVQPGVLTENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEPIE 62 RY+ Q GV +ENGWPM ++ VS VVPAAK+VPL AG+VATILNAW+I +NREVEPI Sbjct 2 RYQAQIGVPSENGWPMCSSASCVSEMVVPAAKRVPLRAGDVATILNAWLILYNREVEPIT 61 Query 63 SQVWGWSADNDVWNSNHMSGTAIDIGAPKYPFGKRVMPAATKAKVRALLDKFNGVVYWGA 122 SQVWGWSADNDVWNSNHMSGTA+DIGAPKYP+G+R MP ATKAKVRALL KF GV++WGA Sbjct 62 SQVWGWSADNDVWNSNHMSGTAVDIGAPKYPWGQRTMPPATKAKVRALLAKFEGVIFWGA 121 Query 123 YWSYPDEMHFQIGLPPSNAKVAELANKLNGGYLNIYGTDTTEPESTSMSLTAAQDQRLNE 182 W YPDEMH+QIGLPPS+ +V A +LN GYL Y DT P+ SM+ D L + Sbjct 122 DWDYPDEMHYQIGLPPSDPRVHAFAERLNNGYLGAY-ADTPAPKGNSMN---DDDLYLRD 177 Query 183 ILENTKTIKNQNNEILKQLGQPGGWPQGGGRTVYDLISAVAEIEGVPNTRDTLA 236 + LG GWPQ GGRTV D ++A+ +P D A Sbjct 178 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 MTRYKVQPGVLTENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEP 60 MTRY+ Q GV +ENGWPM ++ VS VVPAA+ VPL AG+VATILNAW+I +NR VEP Sbjct 1 MTRYQAQIGVPSENGWPMCSSAACVSETVVPAARAVPLRAGDVATILNAWLILYNRLVEP 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 ?	
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 :**

Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-3,659	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	TGACGTGTGGAACTCTAACCAC	ATG	3135	480
8	-5,154	1,470	7	-6,376	GTATCCGTTCGGTAAGCGTGTC	ATG	3192	423
9	-4,706	1,688	6	-6,007	ACTTGATAAGTTCAACGGCGTT	GTG	3249	366
10	-5,429	1,336	7	-6,651	CTACTGGTCTTATCCCGATGAA	ATG	3285	330
11	-3,918	2,073	13	-4,988	CACCGAACCAGAAAGCACTAGC	ATG	3405	210
12	-3,960	2,053	10	-4,918	TACTGCAGCGCAAGACCAACGC	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	CAACACCCGCGACACCCTAGCG	GTG	3606	9

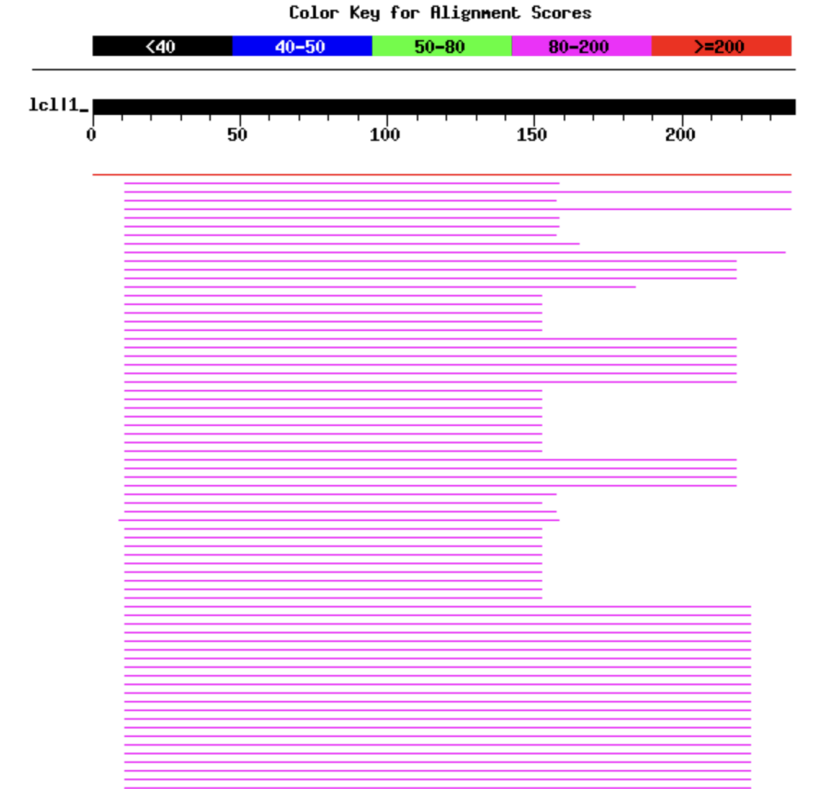
	MAG TPA: Chitinase A [Siphoviridae sp. ctCCv12] Sequence ID: DAD89994.1 Length: 230 Number of Matches: 1 Range 1: 1 to 227 GenPept Graphics ▼ Next Match ▲ Previous M Score Expect Method Identities Positives Gaps 259 bits(662) 2e-83 Compositional matrix adjust. 134/236(57%) 160/236(67%) 9/236(3%) Query 1 MTRYKVQPGVLTENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEP 60 MTRY+ Q GV +ENGWPM ++ VS VVPAA+ VPL AG+VATILNAW+I +NR VEP Sbjct 1 MTRYQAQIGVPSENGWPMCSSAACVSETVVPAARAVPLRAGDVATILNAWLILYNRLVEP 60 Query 61 IESQVWGWSADNDVWNSNHMSGTAIDIGAPKYPFGKRVMPAATKAKVRALLDKFNGVVYW 120 I SQVWGWSADNDVWNSNHMSGTA+DIGAPKYP+G+R MP ATKAKVR LL KF GVV+W Sbjct 61 ITSQVWGWSADNDVWNSNHMSGTAVDIGAPKYPWGQRTMPPATKAKVRTLLAKFEGVVFW 120 Query 121 GAYWSYPDEMHFQIGLPPSNAKVAELANKLNGGYLNIYGTDTTEPESTSMSLTAAQDQRL 180 GA W YPDEMH+QIGLPPS+ +V A +LN GYL Y D +A + + Sbjct 121 GADWDYPDEMHYQIGLPPSDPRVHAFAERLNNGYLGAYPADDAP--------SAPKGNPM 172 Query 181 NEILENTKTIKNQNNEILKQLGQPGGWPQGGGRTVYDLISAVAEIEGVPNTRDTLA 236 N+ + +K Q LG GWPQ GGRTV D ++A+ +P D A Sbjct 173 NDDDLYLRDLKAQMTGA-PGLGAYPGWPQLGGRTVVDALAAIGAKLEIPGFADLKA 227 >RRH1_05, putative lysin, 244 Length = 244 Score = 184 bits (467), Expect = 2e-46 Identities = 85/148 (57%), Positives = 108/148 (72%) Query: 12 TENGWPMVNADKTVSVQVVPAAKKVPLLAGNVATILNAWIIAFNREVEPIESQVWGWSAD 71 +ENGW M N D+TV+V+VVP + P+ G ATILNAW+I ++R VEPI SQVWGWS D Sbjct: 12 SENGWRMCNHDETVTVKVVPGSVGAPIRRGAPATILNAWLIEYDRHVEPITSQVWGWSRD 71 Query: 72 NDVWNSNHMSGTAIDIGAPKYPFGKRVMPAATKAKVRALLDKFNGVVYWGAYWSYPDEMH 131 NDV NSNH++GTA+DI APKYP+G+RVMP +VRA L++F G V+WGA WS DEMH Sbjct: 72 NDVANSNHLAGTAVDINAPKYPWGRRVMPRERILRVRAALNRFEGSVFWGADWSRADEMH 131 Query: 132 FQIGLPPSNAKVAELANKLNGGYLNIYG 159 +Q+ + +V A KL GG+L IYG Sbjct: 132 YQLAWREGDPRVDAFARKLEGGHLGIYG 159 >Rahul_5, lysin A, L-Ala-D-Glu peptidase domain, 260 Length = 260 Score = 178 bits (452), Expect = 9e-45 Identities = 104/245 (42%), Positives = 134/245 (54%), Gaps = 18/245 (7%) Query: 12 TENGWPMVNADKTVSV---QVVPAAKKVPLLAGNVATILNAWIIAFNREVEPIESQVWGW 68 +ENGW MVN D V +P P+ AG +TIL AW+ ++R VEP+ S VWGW Sbjct: 12 SENGWRMVNRDGCVVAGPSMGLPYTNTAPIRAGVASTILVAWLAWYHRNVEPLASPVWGW 71
DECISION:	**2898, conserve chez certains homologues, bon score ATG**

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

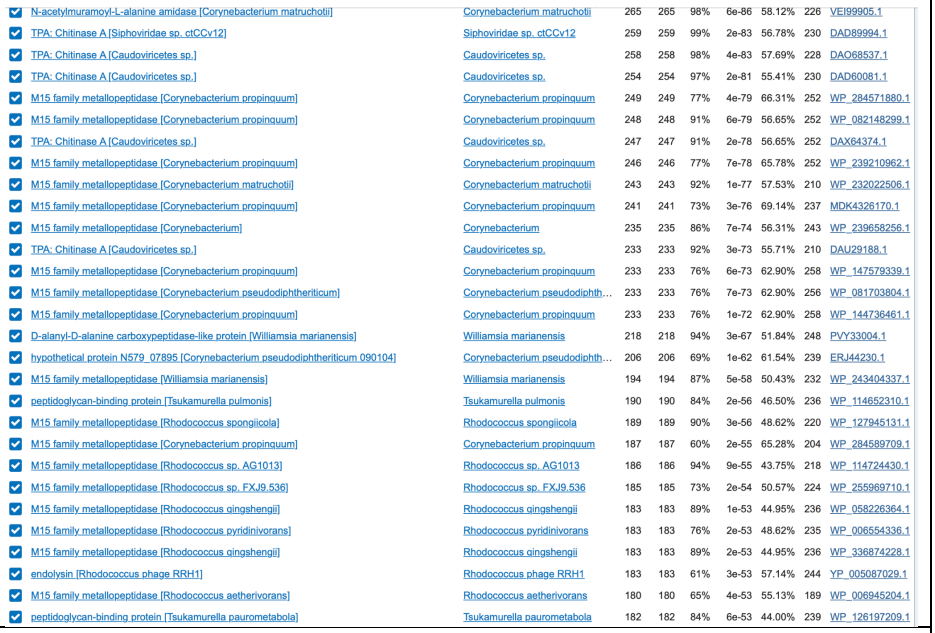
Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	**Listez le meilleur hit Blastp pour chaque source :** ***** PhagesDB :** **(q#: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

Distribution of 100 Blast Hits on the Query Sequence

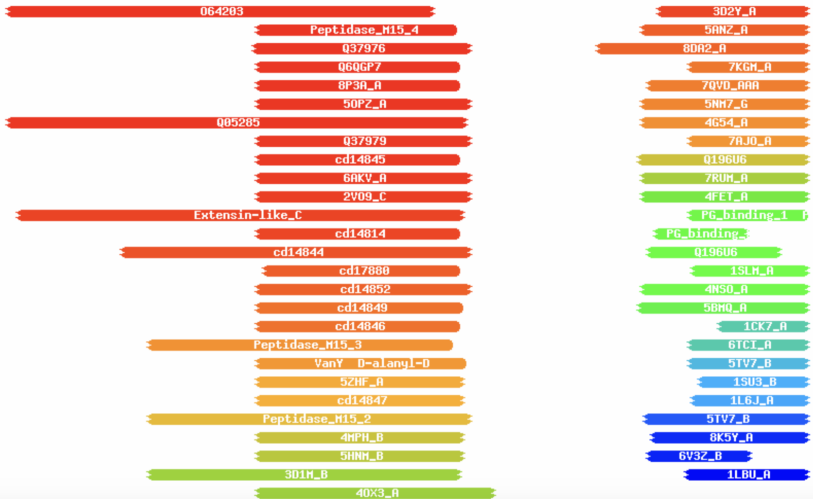
Mouse-over to show defline and scores. Click to show alignments

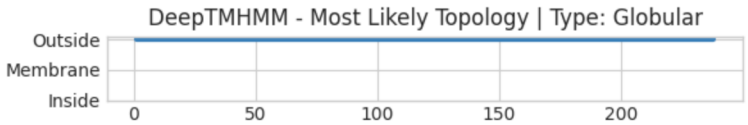
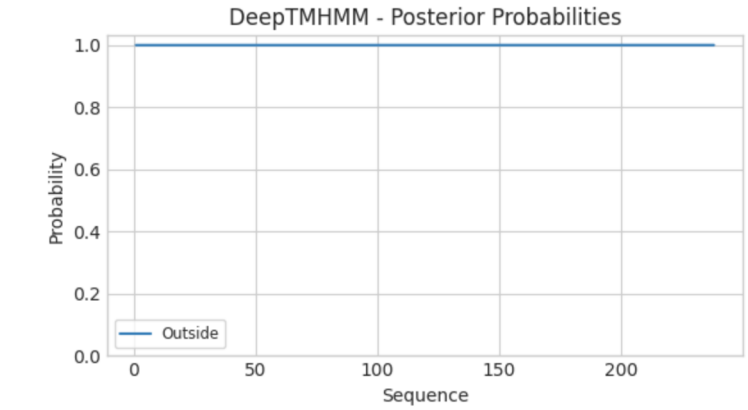


Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_4, function unknown, 238	488	e-138
RRH1_05, putative lysin, 244	184	2e-46
Rahul_5, lysin A, L-Ala-D-Glu peptidase domain, 260	178	9e-45
Dardanus_25, lysin A, 342	177	3e-44
Coeur_5, lysin A, L-Ala-D-Glu peptidase domain, 260	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

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

N-acetylmuramoyl-L-alanine amidase [Corynebacterium matruchotii]	Corynebacterium matruchotii	265	265	98%	6e-86	58.12%	226	VEI99905.1
TPA: Chitinase A [Siphoviridae sp. ctCCv12]	Siphoviridae sp. ctCCv12	259	259	99%	2e-83	56.78%	230	DAD89994.1
TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	258	258	98%	4e-83	57.69%	228	DAO68537.1
TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	254	254	97%	2e-81	55.41%	230	DAD60081.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	249	249	77%	4e-79	66.31%	252	WP_284571880.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	248	248	91%	6e-79	56.65%	252	WP_082148299.1
TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	247	247	91%	2e-78	56.65%	252	DAX64374.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	246	246	77%	7e-78	65.78%	252	WP_239210962.1
M15 family metallopeptidase [Corynebacterium matruchotii]	Corynebacterium matruchotii	243	243	92%	1e-77	57.53%	210	WP_232022506.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	241	241	73%	3e-76	69.14%	237	MDK4326170.1
M15 family metallopeptidase [Corynebacterium]	Corynebacterium	235	235	86%	7e-74	56.31%	243	WP_239658256.1
TPA: Chitinase A [Caudoviricetes sp.]	Caudoviricetes sp.	233	233	92%	3e-73	55.71%	210	DAU29188.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	233	233	76%	6e-73	62.90%	258	WP_147579339.1
M15 family metallopeptidase [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphth…	233	233	76%	7e-73	62.90%	256	WP_081703804.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	233	233	76%	1e-72	62.90%	258	WP_144736461.1
D-alanyl-D-alanine carboxypeptidase-like protein [Williamsia marianensis]	Williamsia marianensis	218	218	94%	3e-67	51.84%	248	PVY33004.1
hypothetical protein N579_07895 [Corynebacterium pseudodiphtheriticum 090104]	Corynebacterium pseudodiphth…	206	206	69%	1e-62	61.54%	239	ERJ44230.1
M15 family metallopeptidase [Williamsia marianensis]	Williamsia marianensis	194	194	87%	5e-58	50.43%	232	WP_243404337.1
peptidoglycan-binding protein [Tsukamurella pulmonis]	Tsukamurella pulmonis	190	190	84%	2e-56	46.50%	236	WP_114652310.1
M15 family metallopeptidase [Rhodococcus spongiicola]	Rhodococcus spongiicola	189	189	90%	3e-56	48.62%	220	WP_127945131.1
M15 family metallopeptidase [Corynebacterium propinquum]	Corynebacterium propinquum	187	187	60%	2e-55	65.28%	204	WP_284589709.1
M15 family metallopeptidase [Rhodococcus sp. AG1013]	Rhodococcus sp. AG1013	186	186	94%	9e-55	43.75%	218	WP_114724430.1
M15 family metallopeptidase [Rhodococcus sp. FXJ9.536]	Rhodococcus sp. FXJ9.536	185	185	73%	2e-54	50.57%	224	WP_255969710.1
M15 family metallopeptidase [Rhodococcus qingshengii]	Rhodococcus qingshengii	183	183	89%	1e-53	44.95%	236	WP_058226364.1
M15 family metallopeptidase [Rhodococcus pyridinivorans]	Rhodococcus pyridinivorans	183	183	76%	2e-53	48.62%	235	WP_006554336.1
M15 family metallopeptidase [Rhodococcus qingshengii]	Rhodococcus qingshengii	183	183	89%	2e-53	44.95%	236	WP_336874228.1
endolysin [Rhodococcus phage RRH1]	Rhodococcus phage RRH1	183	183	61%	3e-53	57.14%	244	YP_005087029.1
M15 family metallopeptidase [Rhodococcus aetherivorans]	Rhodococcus aetherivorans	180	180	65%	4e-53	55.13%	189	WP_006945204.1
peptidoglycan-binding protein [Tsukamurella paurometabola]	Tsukamurella paurometabola	182	182	84%	6e-53	44.00%	239	WP_126197209.1

	Vis Hits Aln Select All Forward Forward Query A3M Model using selection Download HHR Color Seqs Wrap Seqs 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 2 PF13539.10 ; Peptidase_M15_4 ; D-alanyl-D-alanine carboxypeptidase 99.12 3.2e-10 80.22 5.8 58 81 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 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 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 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 7 Q05285 ENLYS_BPML5 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 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 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 1 127 
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**

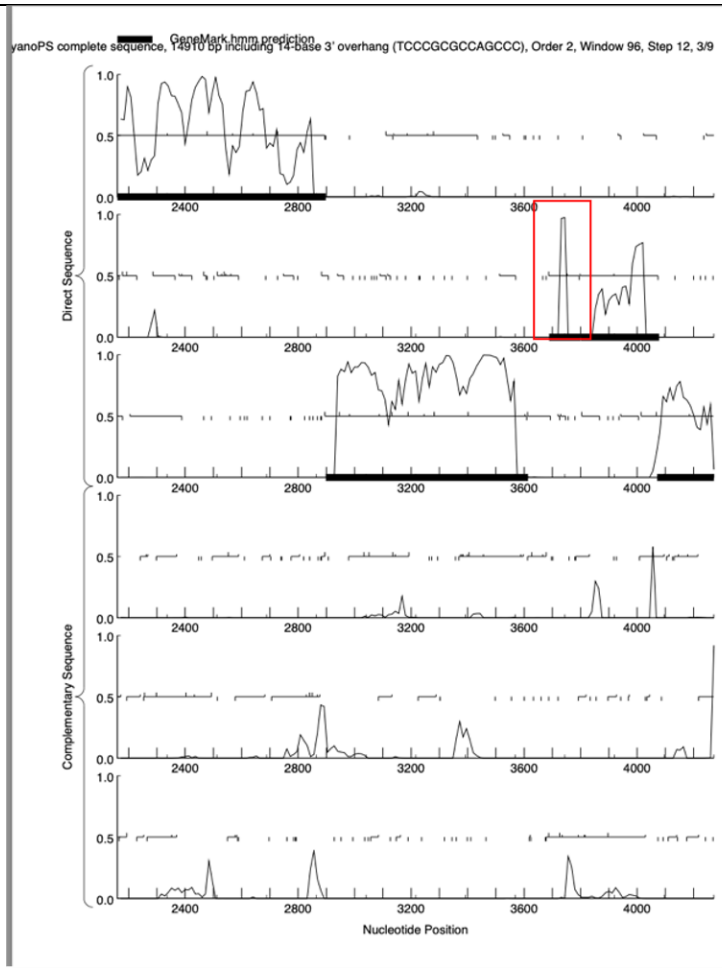
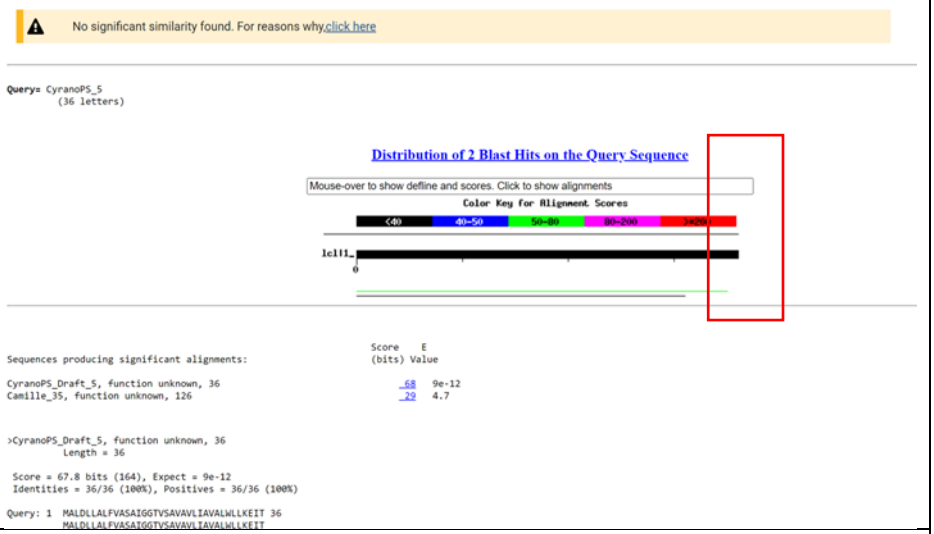
	  # Phobius prediction --- **Prediction of UNNAMED** ID UNNAMED FT TOPO_DOM 1 238 NON CYTOPLASMIC. //
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**OUI**
DECISION:	**Endolysine**

Student Gene Annotation Worksheet

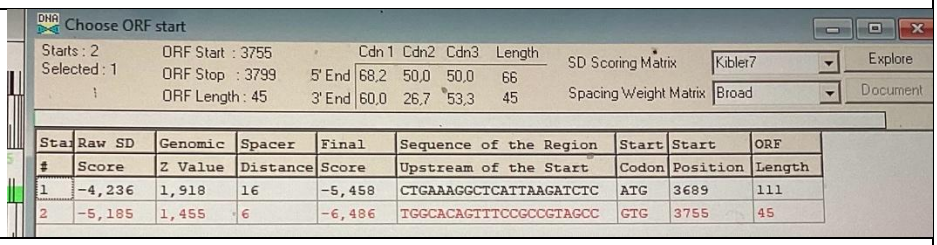
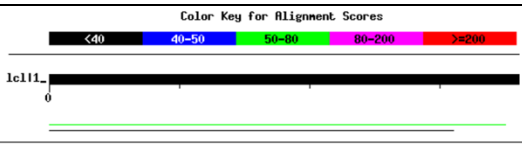
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**5**
Coordonnées du Stop	**3799**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Gap de 74**
Coordonnées du Start retenu	**3689**
Fonction prédite	**NKF**

Décision #1 : Est-ce un gène ?

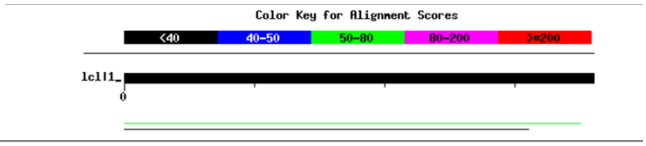
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	***Original GeneMark call @bp 3689***
Y-a-t-il des éléments supportant un potentiel codant ?	**GeneMarkS : potentiel codant détecté**

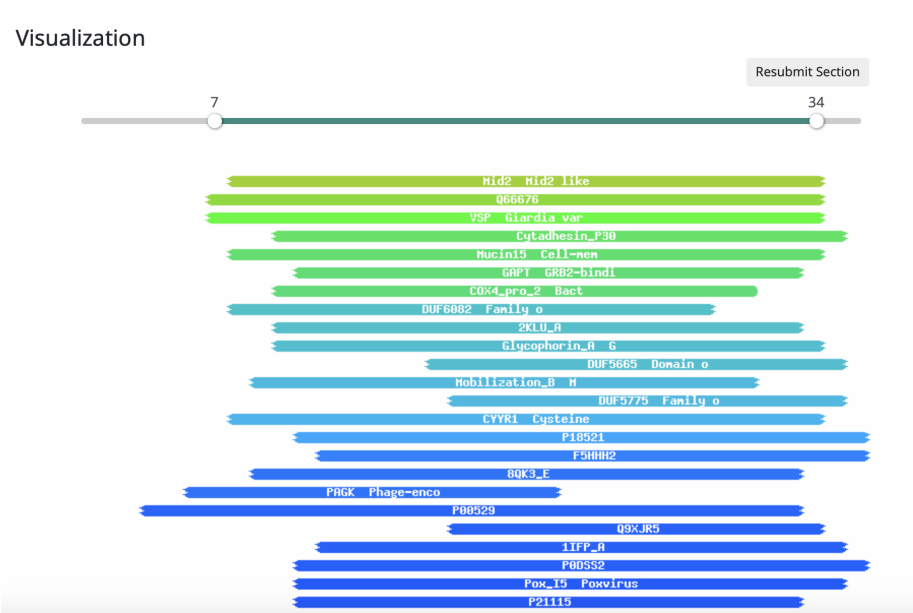
	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**NON** 
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**Conflit au niveau du potentiel codant chez GeneMark et Glimmer**
DECISION:	**si on retire ce gene le gap entre les ORF 5 et ORF 7 sera énorme, alors on le garde**

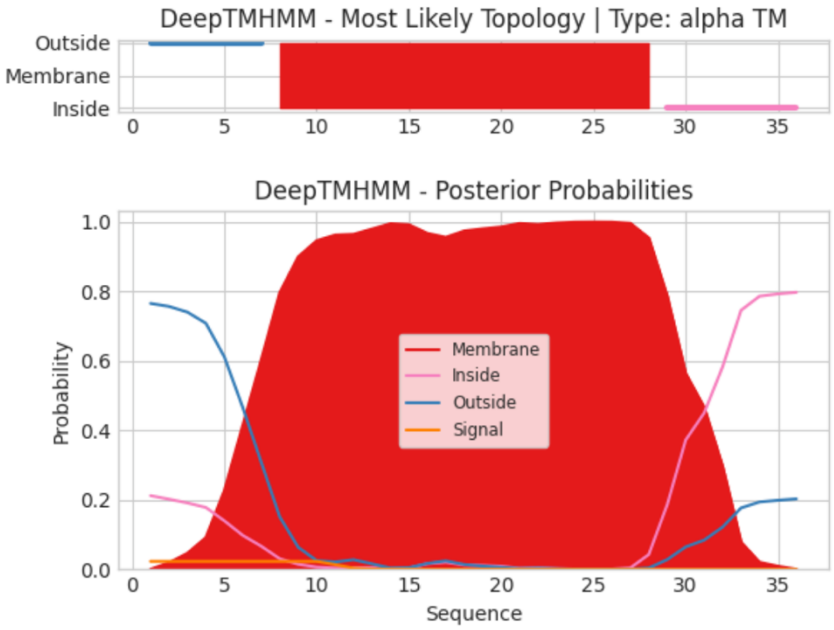
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :**
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 ?	**ORF : ATG3689 la plus longue = 111pb meilleur score , gap de 74pb**
Est-ce que le start est conservé chez les homologues voir Starterator ?	**NA**
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	 Score E Sequences producing significant alignments: (bits) Value CyranoPS_Draft_5, function unknown, 36 68 9e-12 Camille_35, function unknown, 126 29 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 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT Sbjct: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36 >Camille_35, function unknown, 126 Length = 126 Score = 28.9 bits (63), Expect = 4.7 Identities = 15/32 (46%), Positives = 23/32 (71%) Query: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLL 32 +AL L+A ++ SA G VAV++AVALW++ Sbjct: 29 VALILIAGYLESAPLGAALWVAVILAVALWIV 60
DECISION:	**Indiquez le start retenu et justifiez brièvement.**

Décision #3 : Quelle est la function 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 :**  Sequences producing significant alignments: Score (bits) E Value CyranoPS_Draft_5, function unknown, 36 68 9e-12 Camille_35, function unknown, 126 29 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 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT Sbjct: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLLKEIT 36 >Camille_35, function unknown, 126 Length = 126 Score = 28.9 bits (63), Expect = 4.7 Identities = 15/32 (46%), Positives = 23/32 (71%) Query: 1 MALDLLALFVASAIGGTVSAVAVLIAVALWLL 32 +AL L+A ++ SA G VAV++AVALW++ Sbjct: 29 VALILIAGYLESAPLGAALWVAVILAVALWIV 60 ***** nr :** **RIEN** ⚠ No significant similarity found.
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 ?	**"Pas de correspondance pertinente"**

	 **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**

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**OUI** **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 ?	**non**
DECISION:	**Membrane protein**

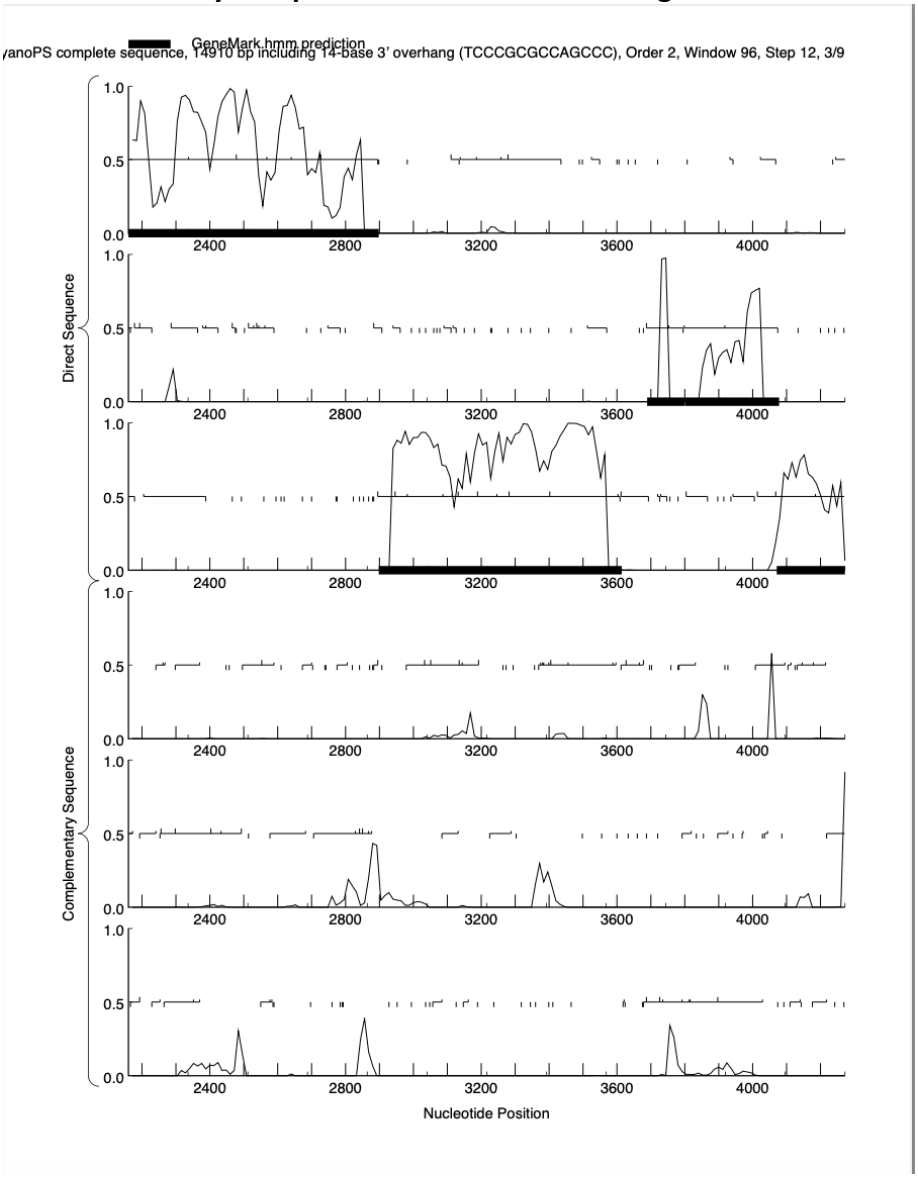
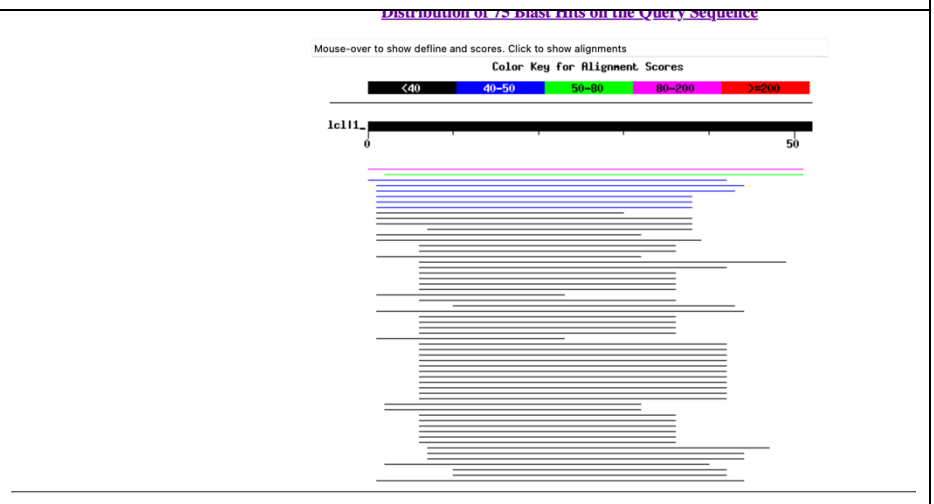
Student Gene Annotation Worksheet

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

Y-a-t-il des éléments supportant un potentiel codant ?	**GeneMarkS : il ya du potentiel codant tout le long** 
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 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

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 ?	Starts : 2 Selected : 1 ORF Start : 3920 ORF Stop : 4078 ORF Length : 159 Cdn 1 Cdn2 Cdn3 Length 5' End 65,0 47,5 52,5 120 3' End 60,4 45,3 45,3 159 SD Scoring Matrix Kibler7 Explore Spacing Weight Matrix Broad Document 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 GCTACTAAAGGAAATCACGTAA GTG 3800 279 2 -6,367 0,878 13 -7,437 CATAGCCCGCGAGTTCGGCTTA 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**

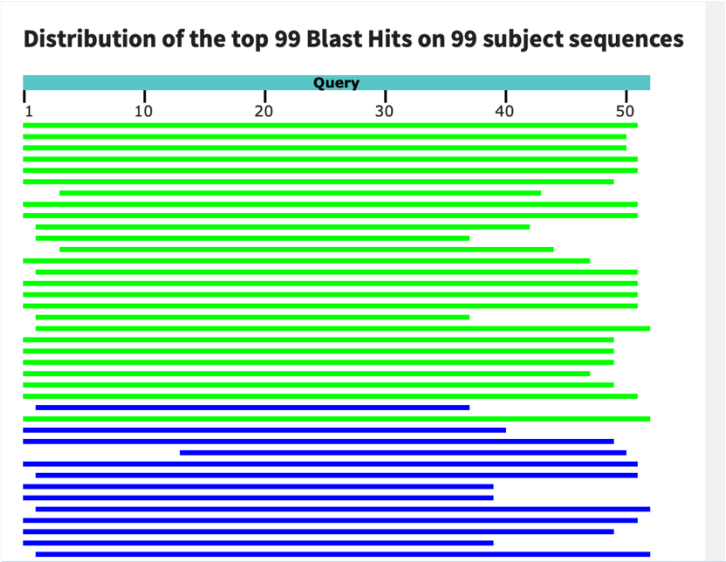
Décision #3 : Quelle est la function de la protéine putative ?

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 :**

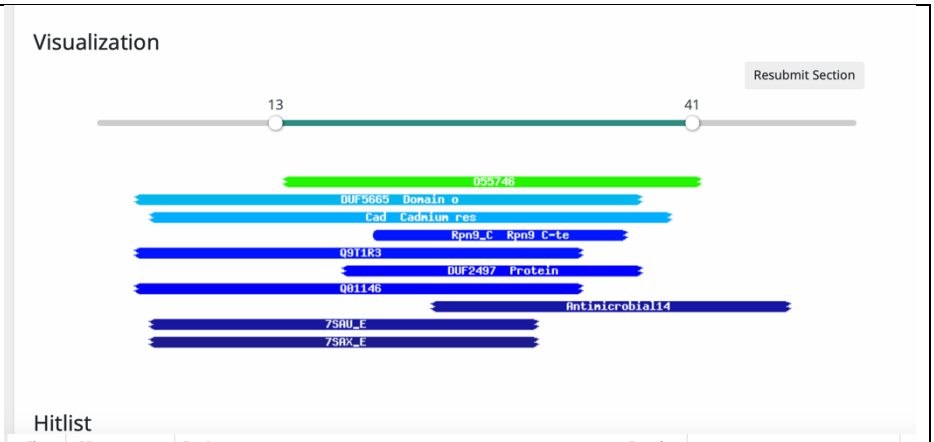
une e-value < 10^{-4} et une couverture acceptable ?

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

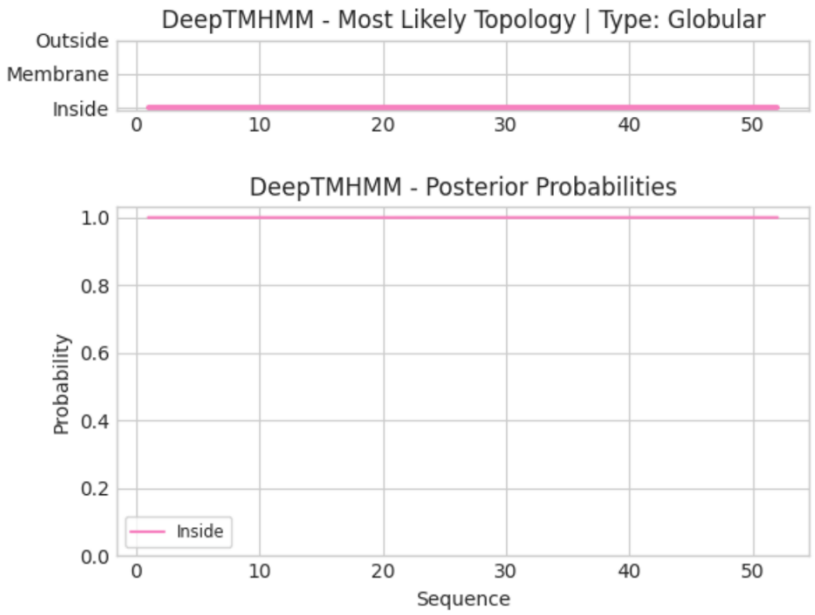
***** nr : il n' y a que des hypothetical protein meme plus bas que sur la capture d'ecran**



Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
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
hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	58.2	58.2	96%	4e-09	52.00%	90	WP_095275390.1
TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	58.2	58.2	96%	6e-09	54.00%	93	DAI73899.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	57.8	57.8	98%	8e-09	54.72%	91	WP_284587136.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	57.8	57.8	98%	8e-09	54.72%	91	WP_284849032.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	54.3	54.3	94%	2e-07	46.94%	91	WP_150114276.1
hypothetical protein [Propionivibrio sp.]	Propionivibrio sp.	53.9	53.9	76%	2e-07	57.50%	96	MBK7565154.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	53.9	53.9	98%	3e-07	49.02%	102	HDV6349140.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	53.9	53.9	98%	3e-07	49.02%	102	HCG2976162.1
hypothetical protein U2A4042570135 [Corynebacterium striatum]	Corynebacterium striatum	52.4	52.4	78%	9e-07	60.98%	97	CQD13966.1
hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	52.4	52.4	69%	1e-06	66.67%	97	WP_083314906.1
TPA: hypothetical protein [Caudoviricetes sp.]	Caudoviricetes sp.	52.0	52.0	78%	1e-06	58.54%	100	DAX19516.1
hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	51.6	51.6	90%	2e-06	53.06%	89	WP_035111211.1
hypothetical protein [Nocardioides massiliensis]	Nocardioides massiliensis	52.4	52.4	96%	2e-06	48.00%	132	WP_220138296.1
hypothetical protein [Tomitella biformata]	Tomitella biformata	51.6	51.6	98%	2e-06	42.31%	94	WP_024794453.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	50.4	50.4	98%	3e-06	41.18%	68	HAT1137148.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	50.1	50.1	98%	4e-06	41.18%	68	HAT1153063.1
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 crible HHPred avec une proba >= 90% et une couverture acceptable ?	 Hitlist Show 25 Entries Search: **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 ?	

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
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
2	PF18910.4	; DUF5665 ; Domain of unknown function (DUF5665)	55.86	76	18.44	3.6	35	58
3	PF03596.17	; Cad ; Cadmium resistance transporter	54.7	70	22.09	3.9	36	191
4	PF18261.5	; Rpn9_C ; Rpn9 C-terminal helix	42.08	36	16.23	0.8	18	33
5	Q9T1R3	VP35_BPAPS Putative DNA transfer protein p35 OS=Acyrthosiphon pisum secondary endosymbiont phage 1 OX=67571 GN=35 PE=3 S	41.79	140	24.08	4.3	31	625

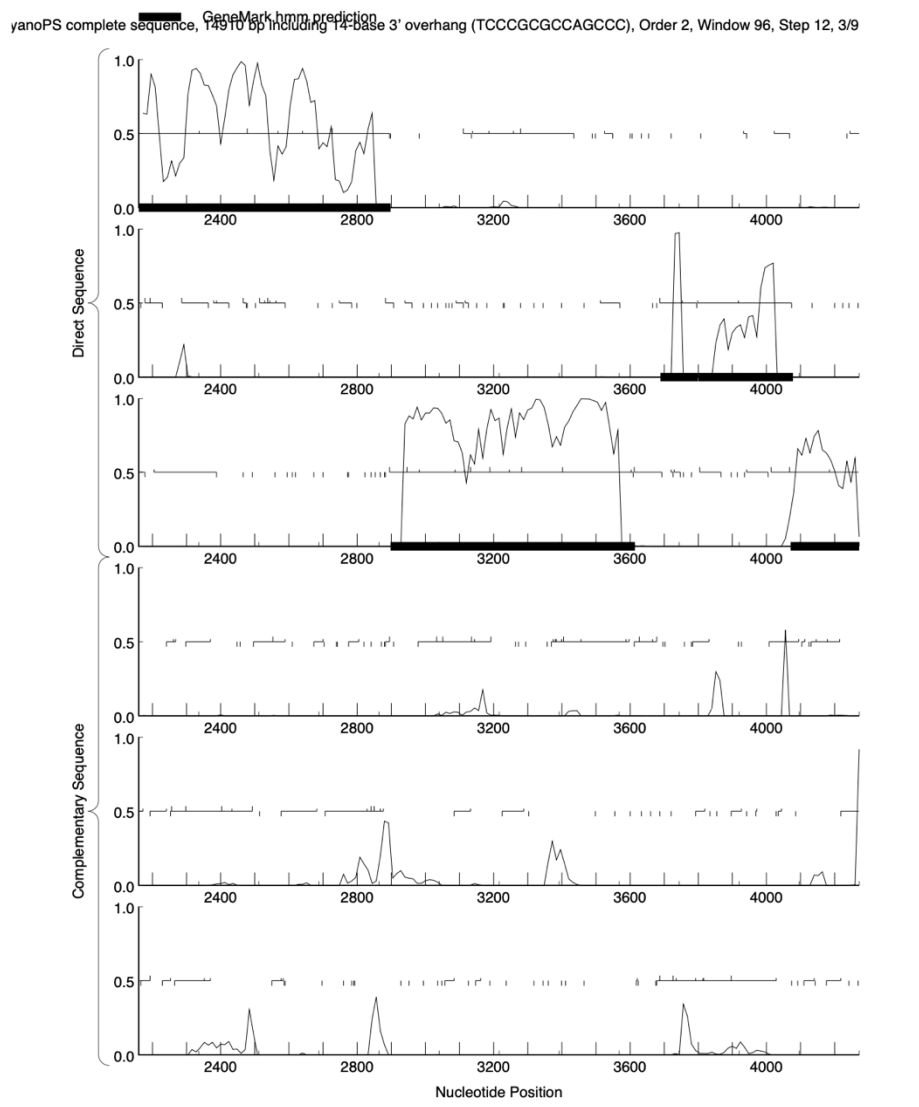
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Non**
DECISION:	**NKF**

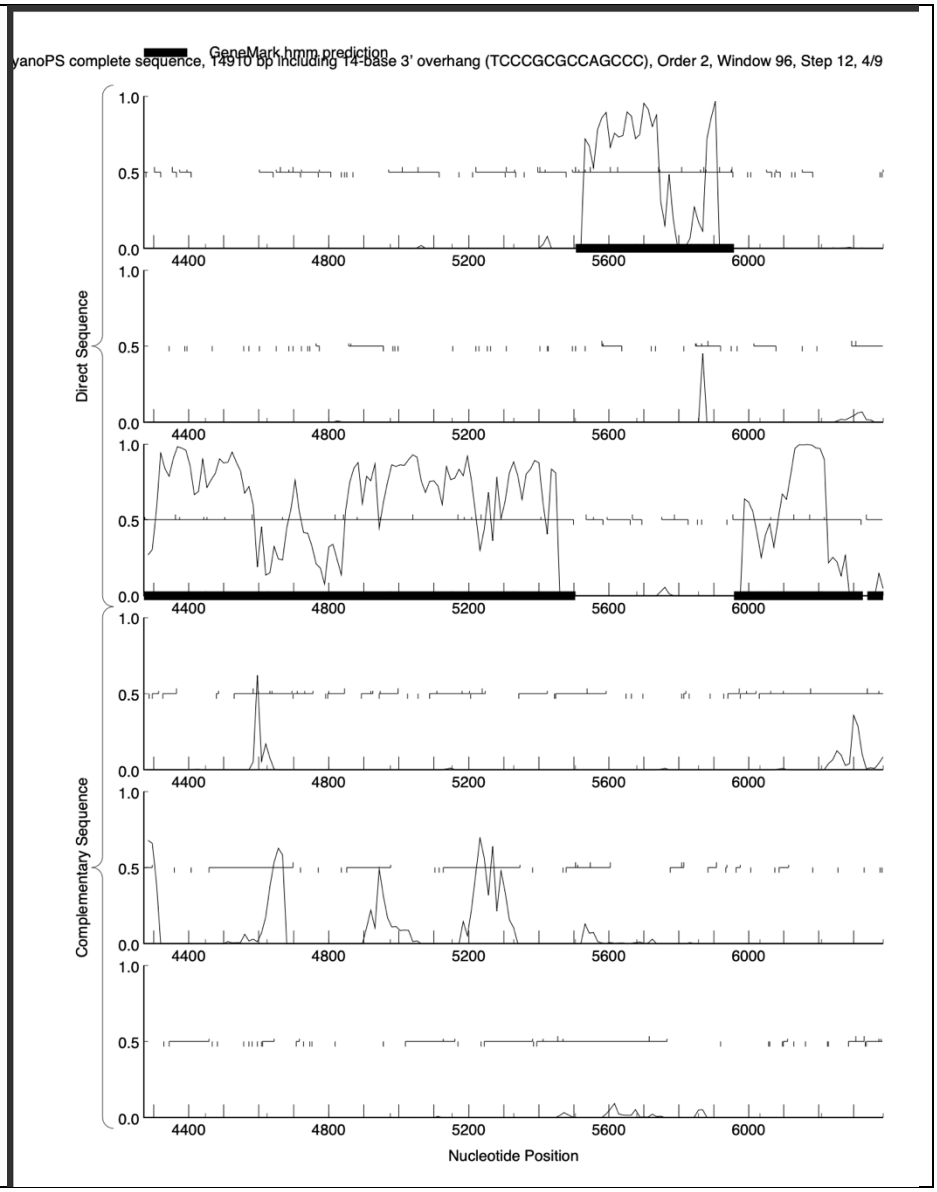
Student Gene Annotation Worksheet

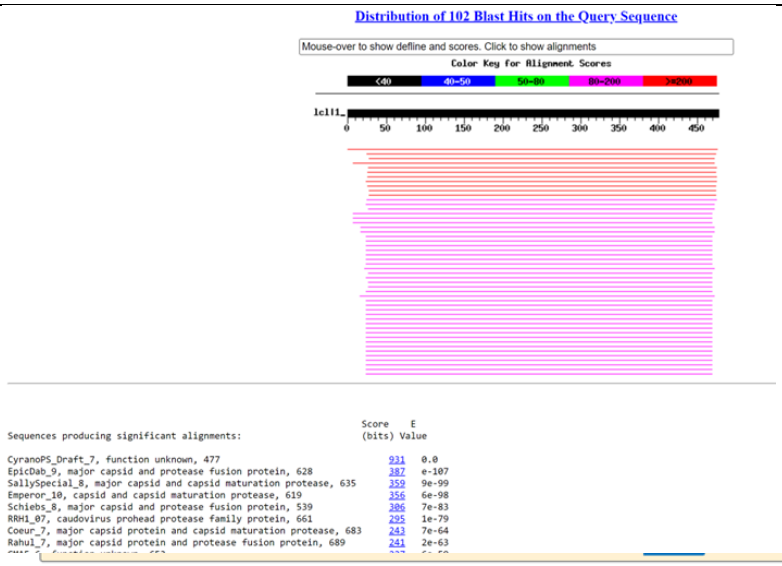
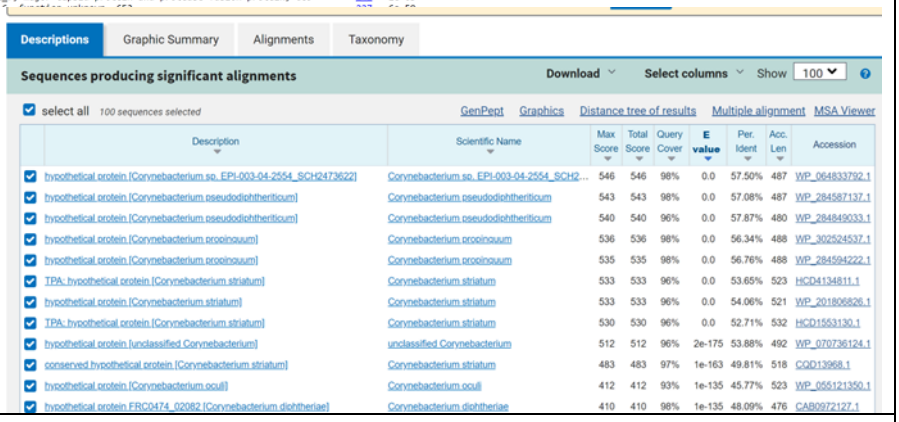
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**7**
Coordonnées du Stop	**5504**
Direction (For/Rev)	**for**
Gap/chevauchement avec un autre gène	**Ove 8**
Coordonnées du Start retenu	**4071**
Fonction prédite	**major capsid and protease fusion protein**

Décision #1 : Est-ce un gène ?

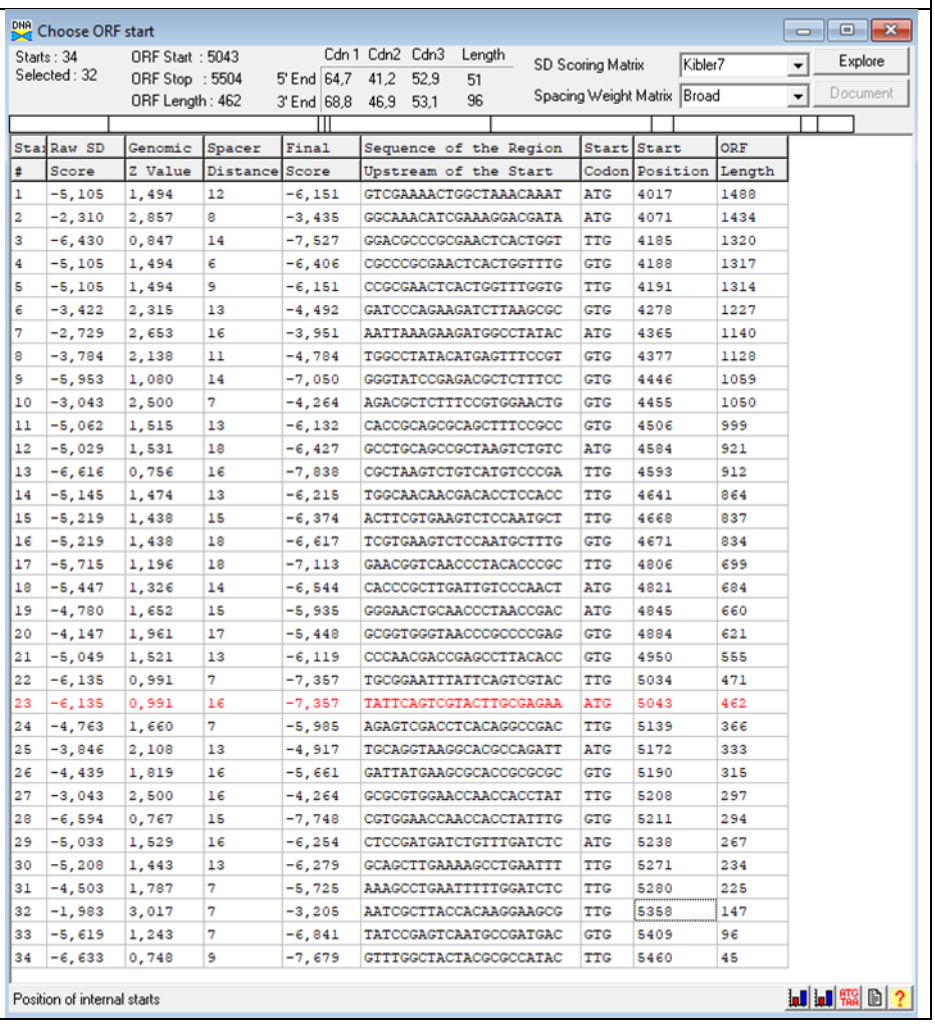
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
Y-a-t-il des éléments supportant un potentiel codant ?	**GeneMarkS :**


GeneMark.hmm prediction
yanoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 3/9
Direct Sequence
Complementary Sequence
1.0
0.5
0.0
2400
2800
3200
3600
4000
Nucleotide Position


GeneMark.hmm prediction
yanoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 4/9
Direct Sequence
Complementary Sequence
1.0
0.5
0.0
4400
4800
5200
5600
6000
Nucleotide Position

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 Sequences producing significant alignments: Score (bits) E Value CyranoPS_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, 635 359 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, 661 295 1e-79 Coeur_7, major capsid protein and capsid maturation protease, 683 243 7e-64 Rahul_7, major capsid protein and protease fusion protein, 689 241 2e-63 
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) : 4071** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 4071**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	

Choose ORF start

Starts : 34 Selected : 32 | ORF Start : 5043 | ORF Stop : 5504 | ORF Length : 462

	Cdn 1	Cdn2	Cdn3	Length
5' End	64,7	41,2	52,9	51
3' End	68,8	46,9	53,1	96

SD Scoring Matrix: Kibler7 — Explore

Spacing Weight Matrix: Broad — Document

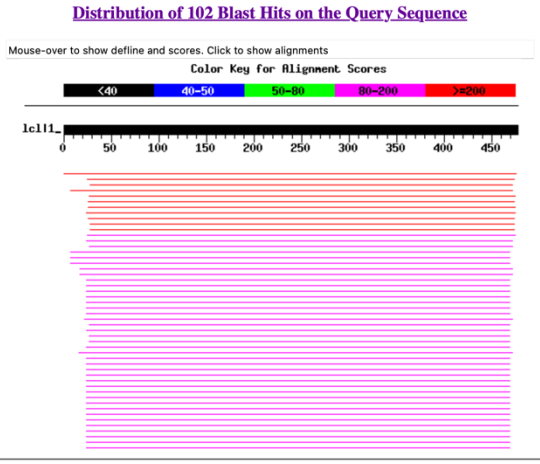
Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-5,105	1,494	12	-6,151	GTCGAAAACTGGCTAAACAAAT	ATG	4017	1488
2	-2,310	2,857	8	-3,435	GGCAAACATCGAAAGGACGATA	ATG	4071	1434
3	-6,430	0,847	14	-7,527	GGACGCCCGCGAACTCACTGGT	TTG	4185	1320
4	-5,105	1,494	6	-6,406	CGCCCGCGAACTCACTGGTTTG	GTG	4188	1317
5	-5,105	1,494	9	-6,151	CCGCGAACTCACTGGTTTGGTG	TTG	4191	1314
6	-3,422	2,315	13	-4,492	GATCCCAGAAGATCTTAAGCGC	GTG	4278	1227
7	-2,729	2,653	16	-3,951	AATTAAAGAAGATGGCCTATAC	ATG	4365	1140
8	-3,784	2,138	11	-4,784	TGGCCTATACATGAGTTTCCGT	GTG	4377	1128
9	-5,953	1,080	14	-7,050	GGGTATCCGAGACGCTCTTTCC	GTG	4446	1059
10	-3,043	2,500	7	-4,264	AGACGCTCTTTCCGTGGAACTG	GTG	4455	1050
11	-5,062	1,515	13	-6,132	CACCGCAGCGCAGCTTTCCGCC	GTG	4506	999
12	-5,029	1,531	18	-6,427	GCCTGCAGCCGCTAAGTCTGTC	ATG	4584	921
13	-6,616	0,756	16	-7,838	CGCTAAGTCTGTCATGTCCCGA	TTG	4593	912
14	-5,145	1,474	13	-6,215	TGGCAACAACGACACCTCCACC	TTG	4641	864
15	-5,219	1,438	15	-6,374	ACTTCGTGAAGTCTCCAATGCT	TTG	4668	837
16	-5,219	1,438	18	-6,617	TCGTGAAGTCTCCAATGCTTTG	GTG	4671	834
17	-5,715	1,196	18	-7,113	GAACGGTCAACCCTACACCCGC	TTG	4806	699
18	-5,447	1,326	14	-6,544	CACCCGCTTGATTGTCCCAACT	ATG	4821	684
19	-4,780	1,652	15	-5,935	GGGAACTGCAACCCTAACCGAC	ATG	4845	660
20	-4,147	1,961	17	-5,448	GCGGTGGGTAACCCGCCCCGAG	GTG	4884	621
21	-5,049	1,521	13	-6,119	CCCAACGACCGAGCCTTACACC	GTG	4950	555
22	-6,135	0,991	7	-7,357	TGCGGAATTTATTCAGTCGTAC	TTG	5034	471
23	-6,135	0,991	16	-7,357	TATTCAGTCGTACTTGCGAGAA	ATG	5043	462
24	-4,763	1,660	7	-5,985	AGAGTCGACCTCACAGGCCGAC	TTG	5139	366
25	-3,846	2,108	13	-4,917	TGCAGGTAAGGCACGCCAGATT	ATG	5172	333
26	-4,439	1,819	16	-5,661	GATTATGAAGCGCACCGCGCGC	GTG	5190	315
27	-3,043	2,500	16	-4,264	GCGCGTGGAACCAACCACCTAT	TTG	5208	297
28	-6,594	0,767	15	-7,748	CGTGGAACCAACCACCTATTTG	GTG	5211	294
29	-5,033	1,529	16	-6,254	CTCCGATGATCTGTTTGATCTC	ATG	5238	267
30	-5,208	1,443	13	-6,279	GCAGCTTGAAAAGCCTGAATTT	TTG	5271	234
31	-4,503	1,787	7	-5,725	AAAGCCTGAATTTTTGGATCTC	TTG	5280	225
32	-1,983	3,017	7	-3,205	AATCGCTTACCACAAGGAAGCG	TTG	5358	147
33	-5,619	1,243	7	-6,841	TATCCGAGTCAATGCCGATGAC	GTG	5409	96
34	-6,633	0,748	9	-7,679	GTTTGGCTACTACGCGCCATAC	TTG	5460	45

Position of internal starts

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 SILTASQSVTATRYIQPETLTAS---------ADEDARELTGLVLPFGKPGRTSRGKLSV 74 +++TAS ++I P LTAS +E+AREL+GLVLPF KPG+T+RG SV Sbjct 3 NLITASAHNQPCKFILPADLTASQATTDDDTTTEENARELSGLVLPFNKPGKTNRGVFSV 62 Query 75 TASALQIPEDLKRVKLYRDHSDVGGTPVGYATAAEIKEDGLYMSFRVGATPDGDAALVDV 134 A+++PEDL RVKLYRDHS+ GTPVG ATAAEIK+DG++MSF VGAT DGDAAL DV Sbjct 63 NKDAIELPEDLGRVKLYRDHSNDNGTPVGRATAAEIKDDGIHMSFHVGATADGDAALADV 122 Query 135 SEGIRDALSVELVSPQISGSQITAAQLSAVAIVAVPAYEDARVQSTPKPAAAKSVMSRLV 194 +EGIRDALSVEL P+I+ +TAA+L+AVA+VA+PAY+DARV T +++ S + Sbjct 123 TEGIRDALSVELDDPEITDGIVTAARLTAVALVALPAYDDARV--TAAAGKPRNLASTVT 180 Query 195 TRDVITAGNNDTSTLTLREVSNALVARLQGDSTHEDLTAALGEITVSGNPKVTSPTWLGE 254 T I A ++ T+ L + +N +V + + H D+TAALGEIT + NP V + WLGE Sbjct 181 TH--INASKHEEKTMNLTKAANTIVTGMTS-ADHSDITAALGEITNANNPAVDNVQWLGE 237 Query 255 LWNGQPYTRLIVPTMGTATLTDMTLSGWRWVTRPEVDDYAGNLAEIPTNTPTTEPYTVEA 314 LW+G + R I+PT+ + LT + L GWRW T P VDDYAGNL EIPTN+P TE VEA Sbjct 238 LWSGARFQRQIIPTLSSQPLTGIKLRGWRWKTLPTVDDYAGNLQEIPTNSPATEAIEVEA 297 Query 315 KRLAGGHKLDRKYIDFPNAEFIQSYLREMARSYAEKTDARAATYIVGESTTVLESTSQAD 374 KR+AGGHKLDRKYIDFP+A FIQSYL EM SYA K+D +AA ++V E+ E T Q + Sbjct 298 KRIAGGHKLDRKYIDFPDAGFIQSYLAEMTNSYAMKSDEKAADFVVAEAKKKKEKTKQPN 357
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 ?	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 NLITASAHNQPCKFILPADLTASQATTDDDTTTEENARELSGLVLPFNKPGKTNRGVFSV 62 Query 57 TASALQIPEDLKRVKLYRDHSDVGGTPVGYATAAEIKEDGLYMSFRVGATPDGDAALVDV 116 A+++PEDL RVKLYRDHS+ GTPVG ATAAEIK+DG++MSF VGAT DGDAAL DV Sbjct 63 NKDAIELPEDLGRVKLYRDHSNDNGTPVGRATAAEIKDDGIHMSFHVGATADGDAALADV 122 Query 117 SEGIRDALSVELVSPQISGSQITAAQLSAVAIVAVPAYEDARVQSTPKPAAAKSVMSRLV 176 +EGIRDALSVEL P+I+ +TAA+L+AVA+VA+PAY+DARV T +++ S + Sbjct 123 TEGIRDALSVELDDPEITDGIVTAARLTAVALVALPAYDDARV--TAAAGKPRNLASTVT 180 Query 177 TRDVITAGNNDTSTLTLREVSNALVARLQGDSTHEDLTAALGEITVSGNPKVTSPTWLGE 236 T I A ++ T+ L + +N +V + + H D+TAALGEIT + NP V + WLGE Sbjct 181 TH--INASKHEEKTMNLTKAANTIVTGMTS-ADHSDITAALGEITNANNPAVDNVQWLGE 237 Query 237 LWNGQPYTRLIVPTMGTATLTDMTLSGWRWVTRPEVDDYAGNLAEIPTNTPTTEPYTVEA 296 LW+G + R I+PT+ + LT + L GWRW T P VDDYAGNL EIPTN+P TE VEA Sbjct 238 LWSGARFQRQIIPTLSSQPLTGIKLRGWRWKTLPTVDDYAGNLQEIPTNSPATEAIEVEA 297 Query 297 KRLAGGHKLDRKYIDFPNAEFIQSYLREMARSYAEKTDARAATYIVGESTTVLESTSQAD 356 KR+AGGHKLDRKYIDFP+A FIQSYL EM SYA K+D +AA ++V E+ E T Q + Sbjct 298 KRIAGGHKLDRKYIDFPDAGFIQSYLAEMTNSYAMKSDEKAADFVVAEAKKKKEKTKQPN 357 Query 357 LLRAAGKARQIMKRTARVEPTTYLVNSDDLFDLMDITQLEKPEFLDLLGVDPAAFVDDPA 416 LL AA KARQI+K+ R+EPT +LVN DDLF LMDIT ++ PE+L LLGV+P F+ D Sbjct 358 LLHAAAKARQIIKKQVRMEPTAFLVNPDDLFKLMDITTMDNPEYLKLLGVEPEKFIADDK 417 Query 417 VPAGRVIAYHKEALKFGELPGSPIRVNADDVARGGTDNALFGYYAPYLEDSRGLVAVPFG 476 VPA VIAYHK+A+ FGELPGSPIRVNA DVARGG+D+ALFGY+A ++ D+RG+V V F Sbjct 418 VPANSVIAYHKDAITFGELPGSPIRVNAVDVARGGSDHALFGYWAAFVSDARGIVQVEFN 477
DECISION:	**4071**

Décision #3 : Quelle est la function 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 :**  Sequences producing significant alignments: Score (bits) E Value CyranoPS_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, 635 359 9e-99 Emperor_10, capsid and capsid maturation protease, 619 356 6e-98 ***** nr :** **(q#: s#) : [alignment] ; e-value :**

Ig-like domain-containing protein [Rhodococcus sp. IEGM 1318]	Rhodococcus sp. IEGM 1318	331	331	92%	4e-103	40.87%	591	WP_317825412.1
hypothetical protein [Gordonia westfalica]	Gordonia westfalica	331	331	93%	8e-103	43.63%	619	WP_074850014.1
hypothetical protein [Corynebacterium matruchotii]	Corynebacterium matruchotii	323	323	95%	1e-101	40.13%	480	WP_126299854.1
TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	321	321	95%	1e-100	39.70%	480	DAK55032.1
TPA: prohead serine protease [Siphoviridae sp. ctCCv12]	Siphoviridae sp. ctCCv12	319	319	95%	6e-100	39.70%	482	DAD90001.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	318	318	93%	4e-99	38.56%	498	HAT6525545.1
hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	313	313	94%	2e-97	39.75%	485	WP_290246076.1
TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	311	311	98%	6e-97	39.58%	481	DAU29195.1
TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	310	310	95%	3e-96	39.44%	482	DAD60100.1
major head protein [Gordonia phage SallySpecial]	Gordonia phage SallySpecial	314	314	93%	3e-96	43.80%	635	YP_010674629.1
hypothetical protein [Corynebacterium amycolatum]	Corynebacterium amycolatum	301	301	60%	2e-95	51.74%	303	MCQ9175172.1
hypothetical protein [Corynebacterium amycolatum]	Corynebacterium amycolatum	303	303	64%	3e-95	49.68%	360	MCQ9177229.1
hypothetical protein [Nocardia salmonicida]	Nocardia salmonicida	298	298	93%	4e-91	38.01%	524	WP_328659909.1
hypothetical protein VH15_05935 [Corynebacterium ulcerans]	Corynebacterium ulcerans	290	290	79%	1e-89	42.15%	389	KKO87266.1
TPA: prohead serine protease [Caudoviricetes sp.]	Caudoviricetes sp.	289	289	93%	5e-88	37.17%	512	DAI73898.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	285	285	94%	9e-87	34.76%	493	WP_201816512.1
TPA: prohead protease [Caudoviricetes sp.]	Caudoviricetes sp.	279	409	84%	1e-83	49.10%	544	DAL07352.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	276	399	84%	2e-82	48.75%	559	WP_284900640.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	273	395	84%	3e-81	48.03%	547	WP_284895768.1
hypothetical protein [Corynebacterium auriscanis]	Corynebacterium auriscanis	271	271	97%	7e-81	35.14%	510	WP_282939128.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	269	326	71%	2e-80	46.95%	481	WP_302526986.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	267	392	84%	2e-79	49.46%	536	HAT1137147.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	267	392	84%	3e-79	49.46%	536	HAT6642577.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	265	390	84%	2e-78	49.10%	536	HAT1171110.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	261	261	93%	5e-77	34.58%	504	WP_284790713.1
hypothetical protein FRC0370_01633 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	259	259	93%	9e-77	36.36%	493	CAB0869439.1
hypothetical protein [Corynebacterium rouxii]	Corynebacterium rouxii	259	259	93%	1e-76	35.54%	493	WP_155873811.1
hypothetical protein [Corynebacterium ulcerans]	Corynebacterium ulcerans	259	259	93%	1e-76	35.23%	487	WP_013912151.1

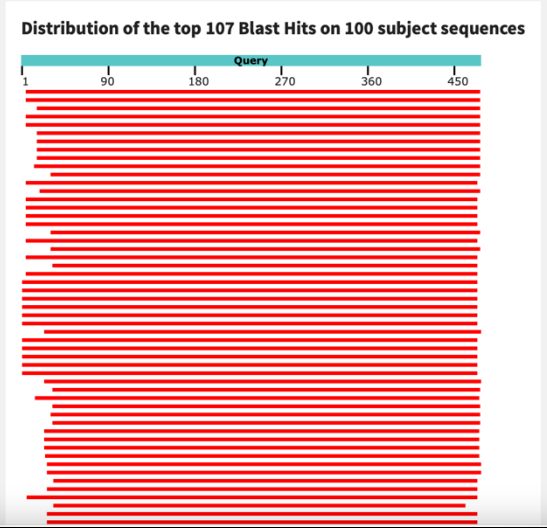
Descriptions | Graphic Summary | Alignments | Taxonomy

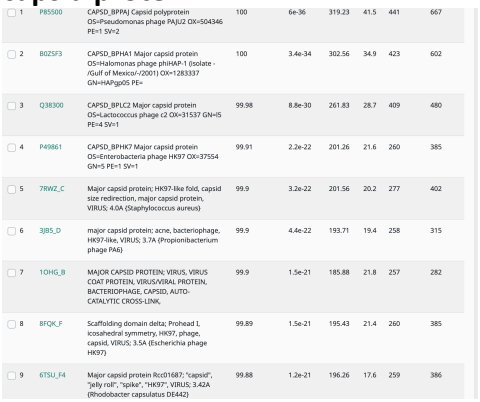
Sequences producing significant alignments Download Select columns Show 100

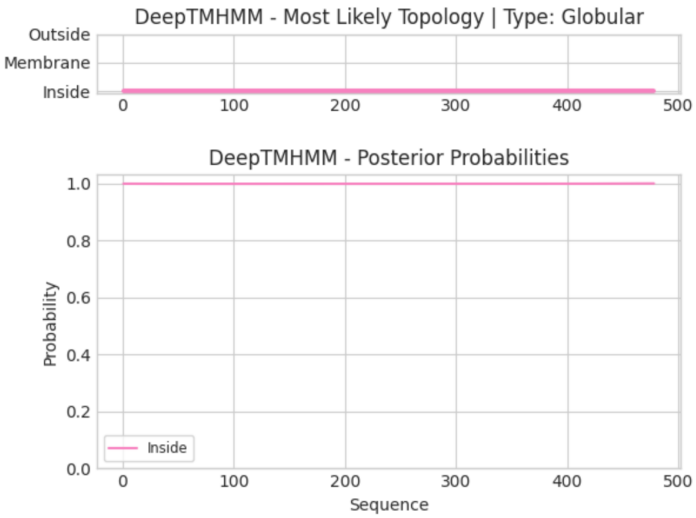
select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewe

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
hypothetical protein [Corynebacterium sp. EPI-003-04-2554_SCH2473622]	Corynebacterium sp. EPI-003-04-255...	546	546	98%	0.0	57.50%	487	WP_064833792.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	543	543	98%	0.0	57.08%	487	WP_284587137.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	540	540	96%	0.0	57.87%	480	WP_284849033.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	536	536	98%	0.0	56.34%	488	WP_302524537.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	535	535	98%	0.0	56.76%	488	WP_284594222.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	533	533	96%	0.0	53.65%	523	HCD4134811.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	533	533	96%	0.0	54.06%	521	WP_201806826.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	530	530	96%	0.0	52.71%	532	HCD1553130.1
hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	512	512	96%	2e-175	53.88%	492	WP_070736124.1
conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	483	483	97%	1e-163	49.81%	518	CQD13968.1
hypothetical protein [Corynebacterium oculi]	Corynebacterium oculi	412	412	93%	1e-135	45.77%	523	WP_055121350.1
hypothetical protein FRC0474_02082 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	410	410	98%	1e-135	48.09%	476	CAB0972127.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	411	411	95%	2e-135	45.97%	524	WP_337890918.1
hypothetical protein [Corynebacterium belfantii]	Corynebacterium belfantii	409	409	98%	4e-135	47.88%	476	WP_197692068.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	409	409	98%	5e-135	48.09%	476	WP_182001105.1

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



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 ?	Visualization **Pfam : L_lactis_ph-MCP ; Lactococcus lactis bacteriophage major capsid protein** 
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	**gene precedent est de fonction inconnue** **gene suivant est une head to tail adaptator**

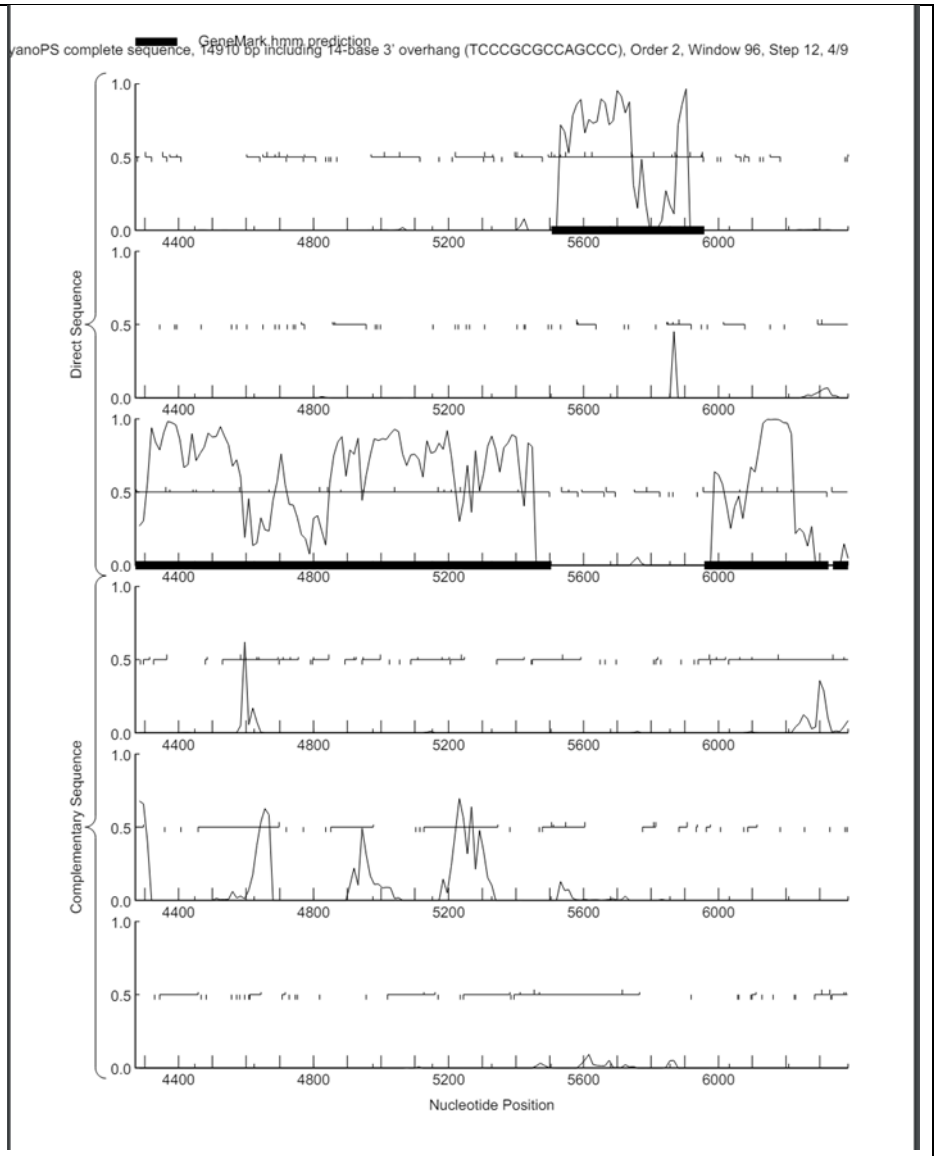
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format  You can download the probabilities used to generate this plot here
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**oui**
DECISION:	**major capsid and protease fusion protein**

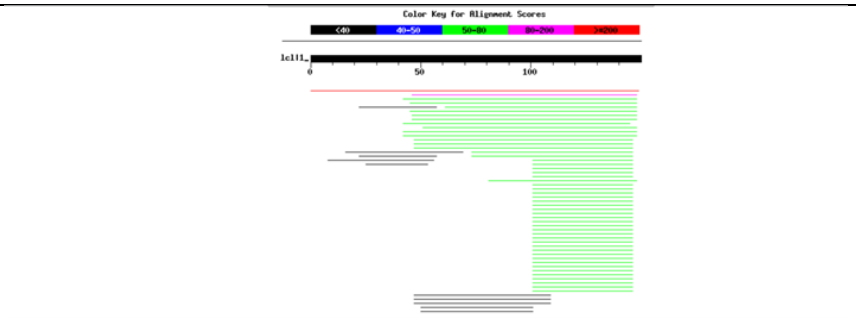
Student Gene Annotation Worksheet

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**

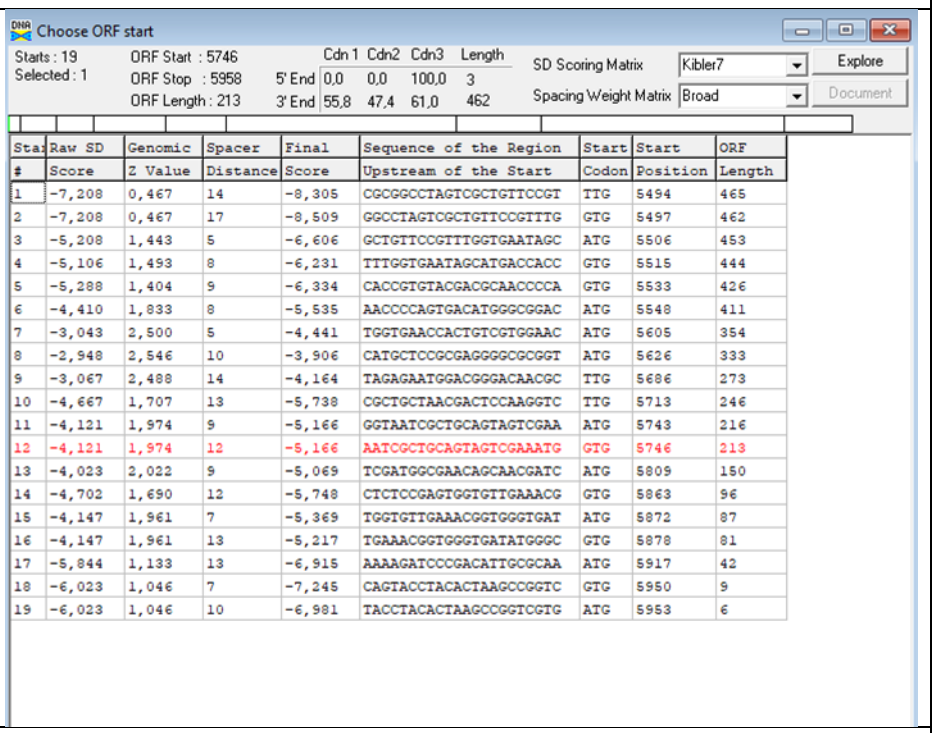
Décision #1 : Est-ce un gène ?

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: 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 Rahul_8, head-to-tail adaptor, 114 70 1e-12 GRU3_7, function unknown, 113 70 2e-12 GMA5_7, function unknown, 113 67 1e-11 RRH1_08, function unknown, 118 64 1e-10 Adala_6, head-to-tail adaptor, 114 62 7e-10 SerialPhiller_6, head-to-tail adaptor, 111 57 2e-08 McGonagall_8, function unknown, 107 57 2e-08 **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 diphtheriae] Corynebacterium diphtheriae 110 110 67% 6e-28 53.47% 103 WP_003852555.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 108 108 68% 3e-27 53.40% 100 WP_134884367.1 hypothetical protein [Corynebacterium belfantii] Corynebacterium belfantii 107 107 68% 6e-27 52.43% 100 WP_197692067.1 hypothetical protein FRC0497_01833 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 107 107 68% 6e-27 52.43% 100 CAB1005428.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 107 107 68% 1e-26 52.43% 100 WP_014318095.1 hypothetical protein CIP107527_01916 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 107 107 68% 1e-26 52.43% 100 CAB0565497.1 hypothetical protein CIP107524_01827 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 106 106 68% 2e-26 52.43% 100 CAB0565169.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 106 106 68% 2e-26 52.43% 100 WP_088298561.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 105 105 68% 5e-26 51.46% 100 WP_106361643.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 105 105 68% 5e-26 51.46% 100 WP_182001104.1 hypothetical protein FRC0474_02083 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 105 105 68% 5e-26 51.46% 100 CAB0972152.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 105 105 68% 6e-26 50.49% 100 WP_205917741.1 hypothetical protein FRC0370_01714 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 104 104 68% 1e-25 51.46% 100 CAB0870082.1 hypothetical protein [Corynebacterium] Corynebacterium 100 100 68% 5e-24 47.57% 102 WP_064833790.1 hypothetical protein [Corynebacterium pseudodiphtheriticum] Corynebacterium pseudodiphtheriticum 100 100 68% 8e-24 47.57% 102 WP_284587138.1 hypothetical protein [Corynebacterium glaucum] Corynebacterium glaucum 100 100 53% 8e-24 60.00% 103 WP_301925782.1 hypothetical protein [Corynebacterium propinquum] Corynebacterium propinquum 99.8 99.8 58% 1e-23 56.82% 103 WP_239210970.1 hypothetical protein [Corynebacterium hadale] Corynebacterium hadale 99.4 99.4 68% 1e-23 53.85% 103 WP_095275388.1 hypothetical protein [Corynebacterium propinquum] Corynebacterium propinquum 99.0 99.0 58% 2e-23 55.68% 103 WP_284571883.1 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 ?

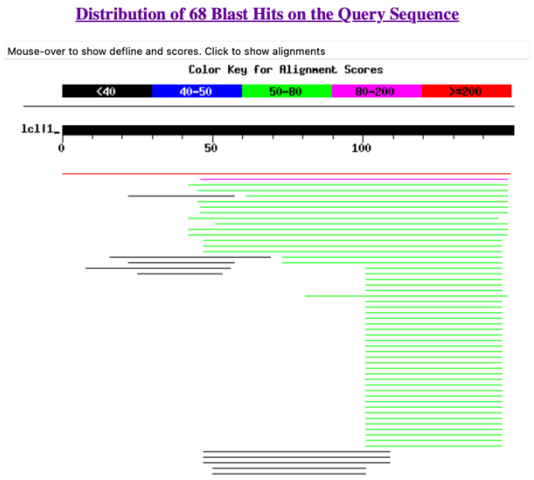
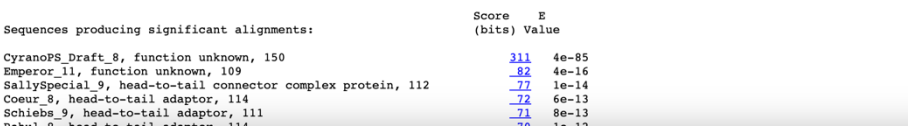
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 5506** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 5506**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
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**

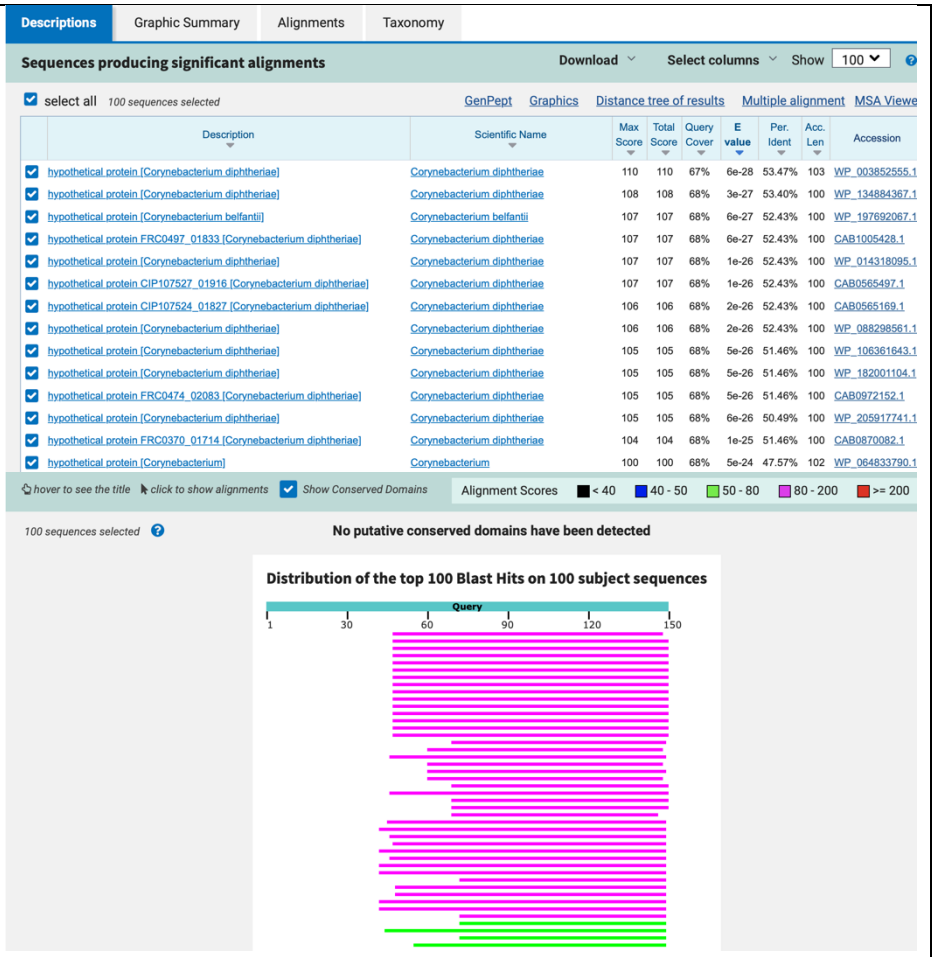
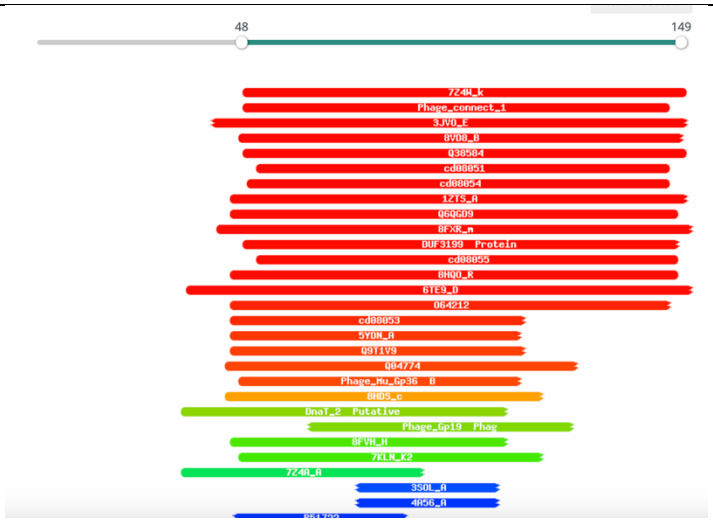
Choose ORF start

Starts : 19 Selected : 1 | ORF Start : 5746 | ORF Stop : 5958 | ORF Length : 213 | Cdn 1 Cdn2 Cdn3 Length | 5' End 0,0 0,0 100,0 3 | 3' End 55,8 47,4 61,0 462 | SD Scoring Matrix: Kibler7 | Spacing Weight Matrix: Broad | Explore | Document

Sta #	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	-7,208	0,467	14	-8,305	CGCGGCCTAGTCGCTGTTCCGT	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	AACCCCAGTGACATGGGCGGAC	ATG	5548	411
7	-3,043	2,500	5	-4,441	TGGTGAACCACTGTCGTGGAAC	ATG	5605	354
8	-2,948	2,546	10	-3,906	CATGCTCCGCGAGGGGCGCGGT	ATG	5626	333
9	-3,067	2,488	14	-4,164	TAGAGAATGGACGGGACAACGC	TTG	5686	273
10	-4,667	1,707	13	-5,738	CGCTGCTAACGACTCCAAGGTC	TTG	5713	246
11	-4,121	1,974	9	-5,166	GGTAATCGCTGCAGTAGTCGAA	ATG	5743	216
12	-4,121	1,974	12	-5,166	AATCGCTGCAGTAGTCGAAATG	GTG	5746	213
13	-4,023	2,022	9	-5,069	TCGATGGCGAACAGCAACGATC	ATG	5809	150
14	-4,702	1,690	12	-5,748	CTCTCCGAGTGGTGTTGAAACG	GTG	5863	96
15	-4,147	1,961	7	-5,369	TGGTGTTGAAACGGTGGGTGAT	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	TACCTACACTAAGCCGGTCGTG	ATG	5953	6

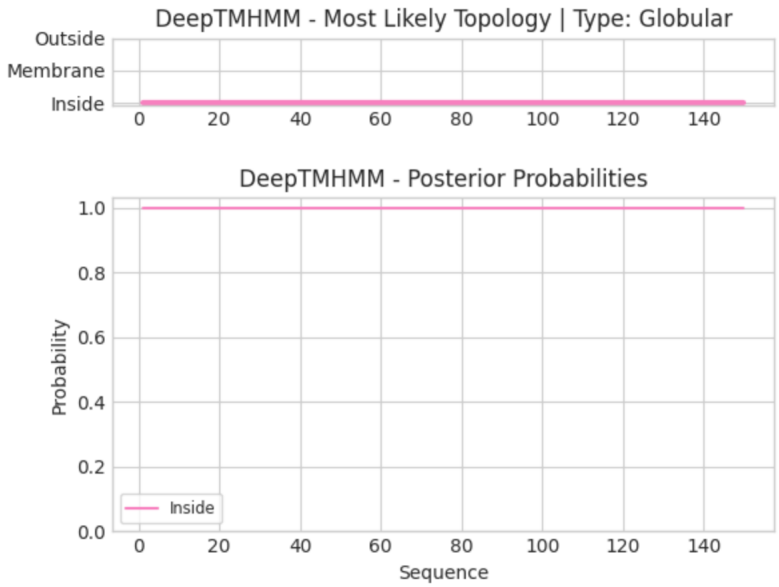
Décision #3 : Quelle est la function de la protéine putative ?

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

	
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_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_BPSPP Head completion protein gp15 OS=Bacillus phage SPP1 OX=10724 GN=15 PE=1 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 gene avant est une major capsid**

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	P85500	CAPSD_BPPAJ Capsid polyprotein OS=Pseudomonas phage PAJU2 OX=504346 PE=1 SV=2	100	6e-36	319.23	41.5	441	667
2	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
3	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
4	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
5	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
6	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
7	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
8	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

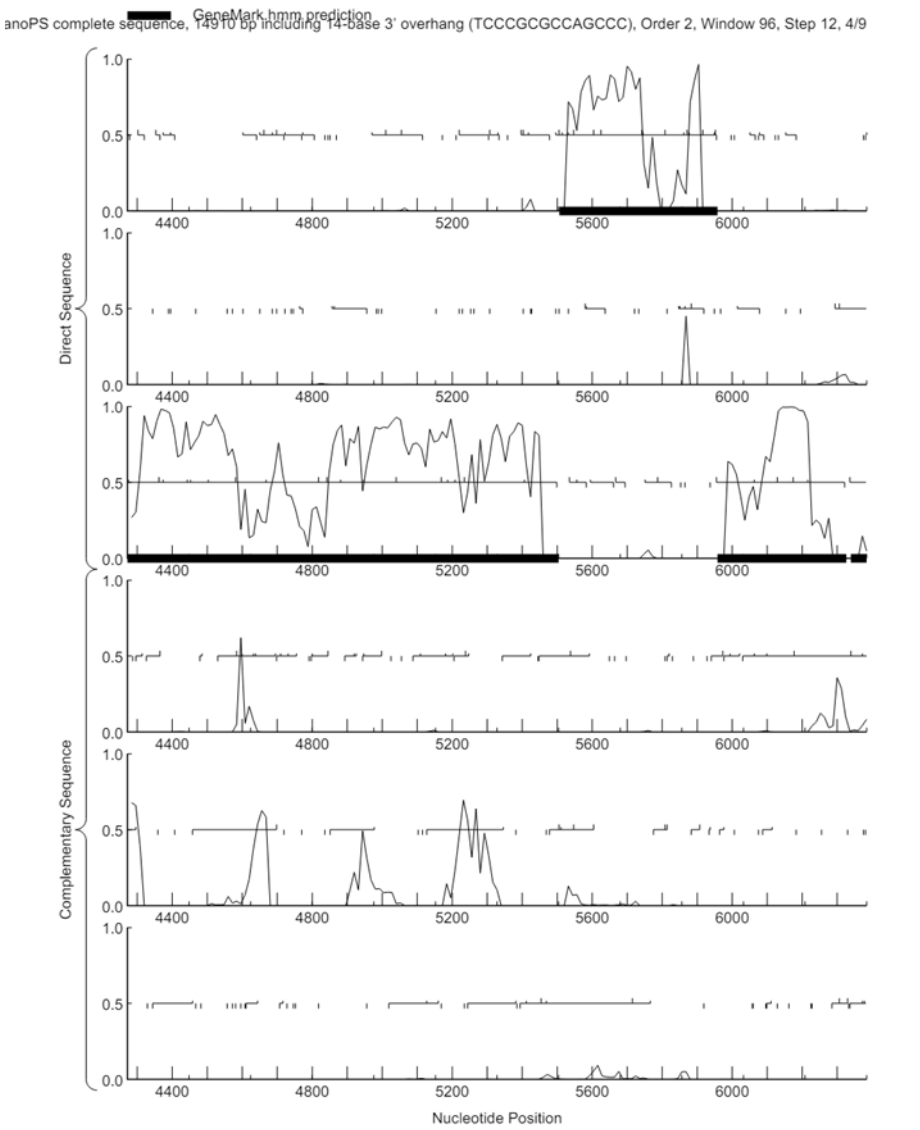
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	 **Si la réponse est OUI, supportez votre propos par les prédictions d'au moins deux programmes de prédiction de segments TM (voir TMHMM ou Phoebius par ex).**
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**oui**
DECISION:	**head-to-tail adaptor**

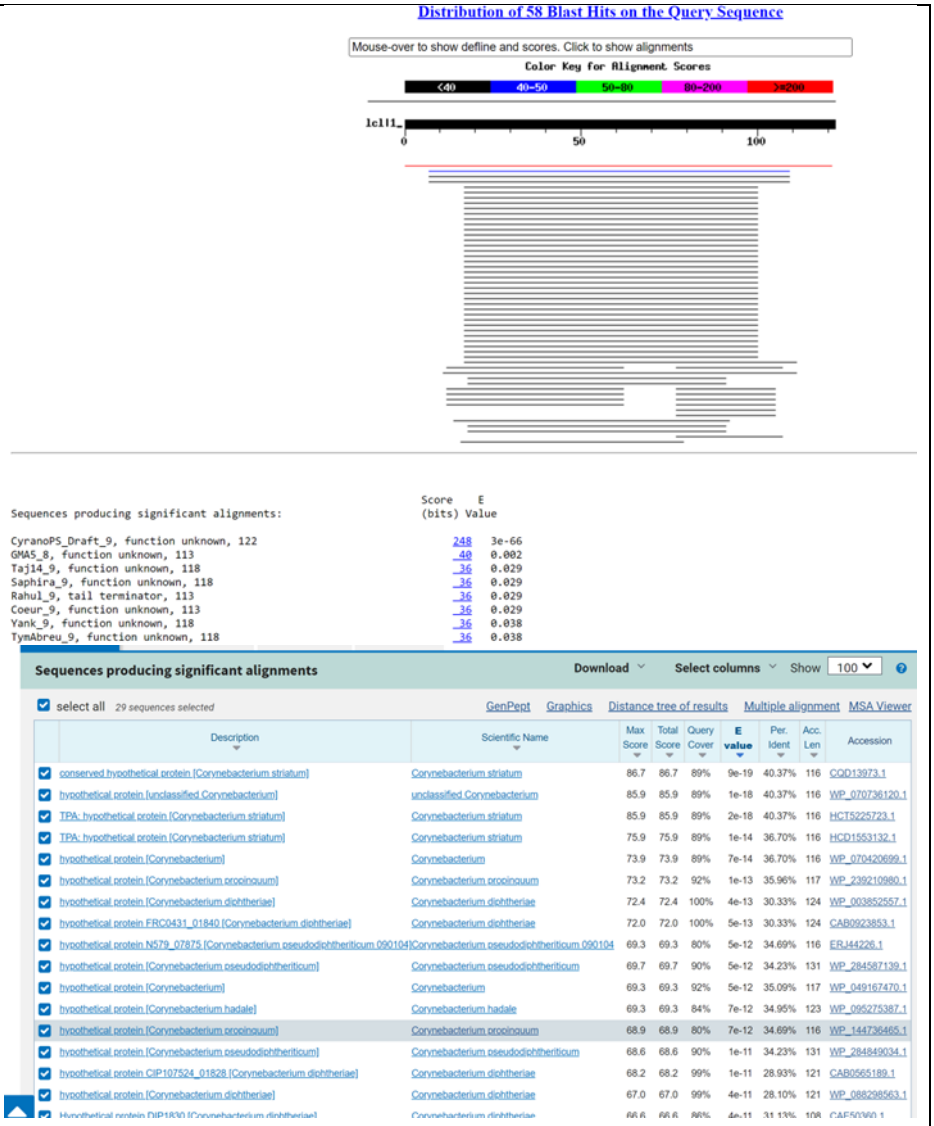
Student Gene Annotation Worksheet

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**

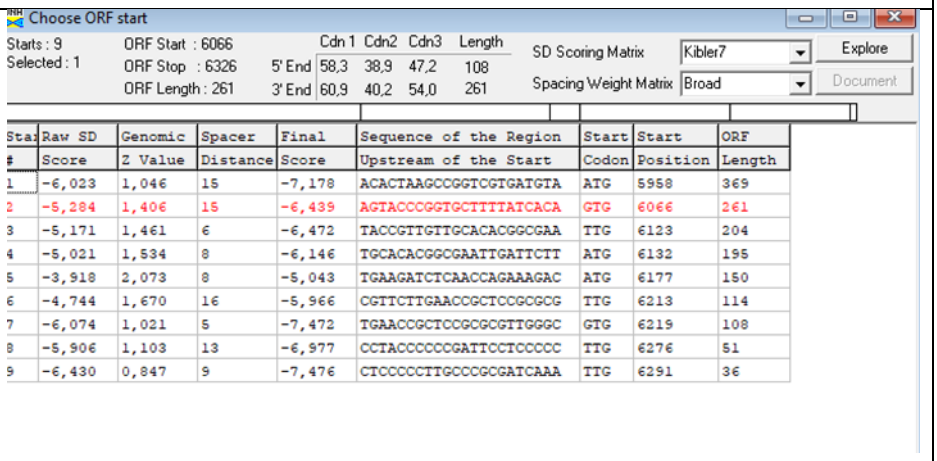
Coordonnées du Start retenu	**5958**
Fonction prédite	**Tail terminator**

Décision #1 : Est-ce un gène ?

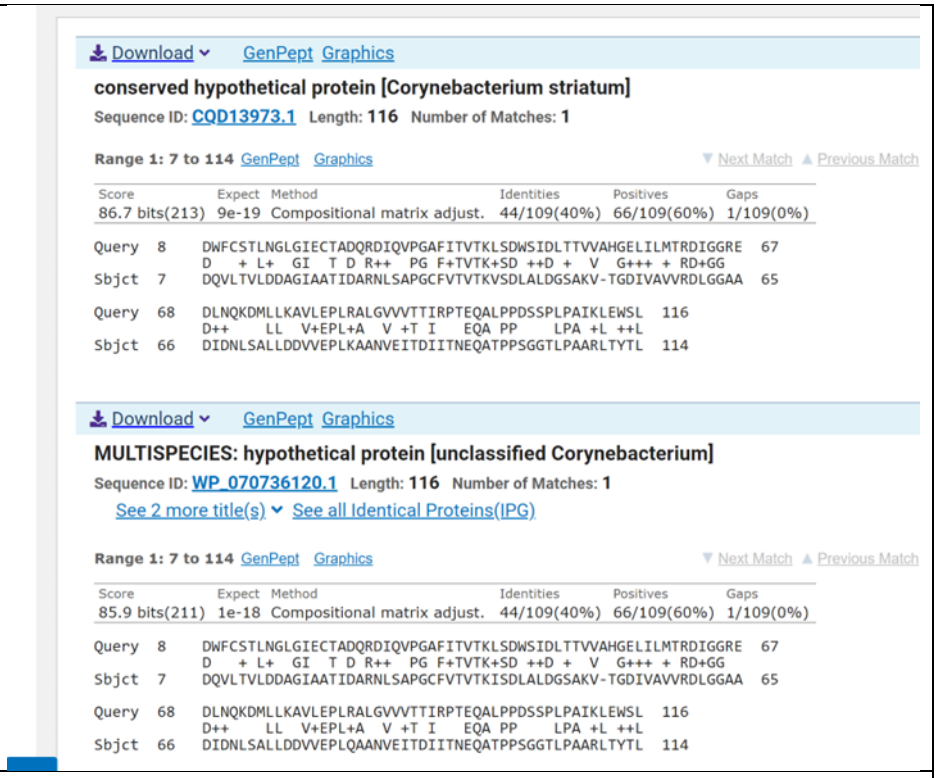
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
Y-a-t-il des éléments supportant un potentiel codant ?	

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 Sequences producing significant alignments: Score (bits) E Value CyranoPS_Draft_9, function unknown, 122 248 3e-66 GMA5_8, function unknown, 113 40 0.002 Taj14_9, function unknown, 118 36 0.029 Saphira_9, function unknown, 118 36 0.029 Rahul_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 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 conserved hypothetical protein [Corynebacterium striatum] — Corynebacterium striatum — 86.7 — 86.7 — 89% — 9e-19 — 40.37% — 116 — CQD13973.1 hypothetical protein [unclassified Corynebacterium] — unclassified Corynebacterium — 85.9 — 85.9 — 89% — 1e-18 — 40.37% — 116 — WP_070736120.1 TPA: hypothetical protein [Corynebacterium striatum] — Corynebacterium striatum — 85.9 — 85.9 — 89% — 2e-18 — 40.37% — 116 — HCT5225723.1 TPA: hypothetical protein [Corynebacterium striatum] — Corynebacterium striatum — 75.9 — 75.9 — 89% — 1e-14 — 36.70% — 116 — HCD1553132.1 hypothetical protein [Corynebacterium] — Corynebacterium — 73.9 — 73.9 — 89% — 7e-14 — 36.70% — 116 — WP_070420699.1 hypothetical protein [Corynebacterium propinquum] — Corynebacterium propinquum — 73.2 — 73.2 — 92% — 1e-13 — 35.96% — 117 — WP_239210980.1 hypothetical protein [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 72.4 — 72.4 — 100% — 4e-13 — 30.33% — 124 — WP_003852557.1 hypothetical protein FRC0431_01840 [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 72.0 — 72.0 — 100% — 5e-13 — 30.33% — 124 — CAB0923853.1 hypothetical protein N579_07875 [Corynebacterium pseudodiphtheriticum 090104] — Corynebacterium pseudodiphtheriticum 090104 — 69.3 — 69.3 — 80% — 5e-12 — 34.69% — 116 — ERJ44226.1 hypothetical protein [Corynebacterium pseudodiphtheriticum] — Corynebacterium pseudodiphtheriticum — 69.7 — 69.7 — 90% — 5e-12 — 34.23% — 131 — WP_284587139.1 hypothetical protein [Corynebacterium] — Corynebacterium — 69.3 — 69.3 — 92% — 5e-12 — 35.09% — 117 — WP_049167470.1 hypothetical protein [Corynebacterium hadale] — Corynebacterium hadale — 69.3 — 69.3 — 84% — 7e-12 — 34.95% — 123 — WP_095275387.1 hypothetical protein [Corynebacterium propinquum] — Corynebacterium propinquum — 68.9 — 68.9 — 80% — 7e-12 — 34.69% — 116 — WP_144736465.1 hypothetical protein [Corynebacterium pseudodiphtheriticum] — Corynebacterium pseudodiphtheriticum — 68.6 — 68.6 — 90% — 1e-11 — 34.23% — 131 — WP_284849034.1 hypothetical protein CIP107524_01828 [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 68.2 — 68.2 — 99% — 1e-11 — 28.93% — 121 — CAB0565189.1 hypothetical protein [Corynebacterium diphtheriae] — Corynebacterium diphtheriae — 67.0 — 67.0 — 99% — 4e-11 — 28.10% — 121 — WP_088298563.1
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) : 5958** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 5958**
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 ?	***ORF la plus longue*** ***ORF la plus longue avec un ATG*** ***Score de -7.170***
Est-ce que le start est conservé chez les homologues voir Starterator ?	–

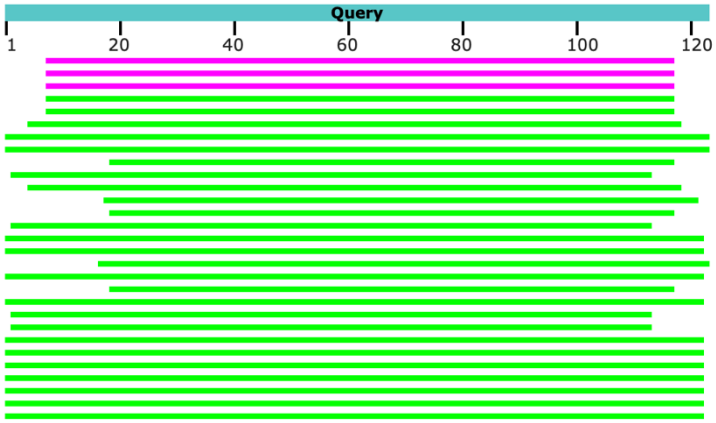
Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-6,023	1,046	15	-7,178	ACACTAAGCCGGTCGTGATGTA	ATG	5958	369
2	-5,284	1,406	15	-6,439	AGTACCCGGTGCTTTTATCACA	GTG	6066	261
3	-5,171	1,461	6	-6,472	TACCGTTGTTGCACACGGCGAA	TTG	6123	204
4	-5,021	1,534	8	-6,146	TGCACACGGCGAATTGATTCTT	ATG	6132	195
5	-3,918	2,073	8	-5,043	TGAAGATCTCAACCAGAAAGAC	ATG	6177	150
6	-4,744	1,670	16	-5,966	CGTTCTTGAACCGCTCCGCGCG	TTG	6213	114
7	-6,074	1,021	5	-7,472	TGAACCGCTCCGCGCGTTGGGC	GTG	6219	108
8	-5,906	1,103	13	-6,977	CCTACCCCCCGATTCCTCCCCC	TTG	6276	51
9	-6,430	0,847	9	-7,476	CTCCCCCTTGCCCGCGATCAAA	TTG	6291	36

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	
DECISION:	**5958**

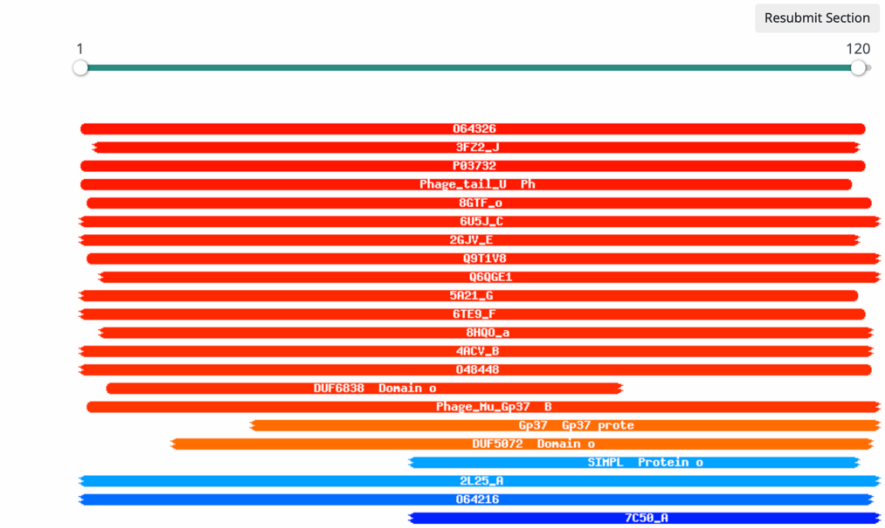
Décision #3 : Quelle est la function de la protéine putative ?

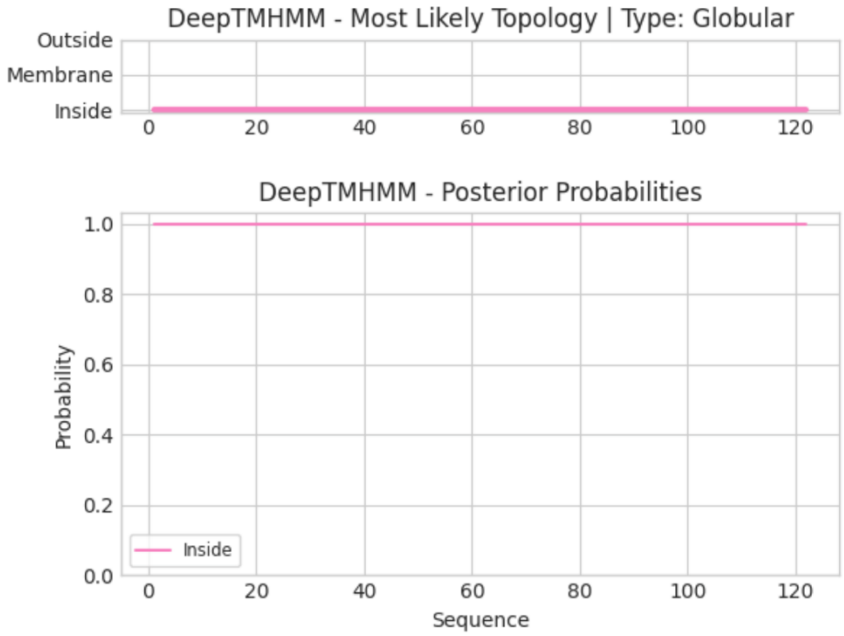
Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	**Listez le meilleur hit Blastp pour chaque source :** ***** PhagesDB :** **RIEN**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

Distribution of the top 29 Blast Hits on 29 subject sequences



conserved hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	86.7	86.7	89%	9e-19	40.37%	116	CQD13973.1
hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	85.9	85.9	89%	1e-18	40.37%	116	WP_070736120.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	85.9	85.9	89%	2e-18	40.37%	116	HCT5225723.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	75.9	75.9	89%	1e-14	36.70%	116	HCD1553132.1
hypothetical protein [Corynebacterium]	Corynebacterium	73.9	73.9	89%	7e-14	36.70%	116	WP_070420699.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	73.2	73.2	92%	1e-13	35.96%	117	WP_239210980.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	72.4	72.4	100%	4e-13	30.33%	124	WP_003852557.1
hypothetical protein FRC0431_01840 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	72.0	72.0	100%	5e-13	30.33%	124	CAB0923853.1
hypothetical protein N579_07875 [Corynebacterium pseudodiphtheriticum 090104]	Corynebacterium pseudodiphtheriticum 090104	69.3	69.3	80%	5e-12	34.69%	116	ERJ44226.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	69.7	69.7	90%	5e-12	34.23%	131	WP_284587139.1
hypothetical protein [Corynebacterium]	Corynebacterium	69.3	69.3	92%	5e-12	35.09%	117	WP_049167470.1
hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	69.3	69.3	84%	7e-12	34.95%	123	WP_095275387.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	68.9	68.9	80%	7e-12	34.69%	116	WP_144736465.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	68.6	68.6	90%	1e-11	34.23%	131	WP_284849034.1
hypothetical protein CIP107524_01828 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	68.2	68.2	99%	1e-11	28.93%	121	CAB0565189.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	67.0	67.0	99%	4e-11	28.10%	121	WP_088298563.1
Hypothetical protein DIP1830 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	66.6	66.6	86%	4e-11	31.13%	108	CAE50360.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	67.0	67.0	99%	5e-11	28.10%	121	WP_342351207.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	66.2	66.2	80%	7e-11	33.67%	116	WP_284596343.1
hypothetical protein [Corynebacterium belfantii]	Corynebacterium belfantii	66.2	66.2	99%	1e-10	28.10%	121	WP_197692066.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	66.2	66.2	90%	1e-10	33.33%	131	WP_284594218.1
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
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	65.1	65.1	99%	2e-10	27.27%	121	WP_106361642.1
hypothetical protein FRC0474_02084 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	64.7	64.7	99%	4e-10	27.27%	121	CAB0972182.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	64.3	64.3	99%	4e-10	26.45%	121	WP_014318096.1
hypothetical protein FRC0087_01817 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	64.3	64.3	99%	5e-10	26.45%	121	CAB0733833.1
hypothetical protein FRC0497_01832 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.9	63.9	99%	7e-10	26.45%	121	CAB1005424.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	63.5	63.5	99%	8e-10	26.45%	121	WP_205917719.1
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 >= 90% et une couverture acceptable ?	Visualization  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 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 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 4 PF06141.15 ; Phage_tail_U ; Phage minor tail protein U 97.71 0.0062 40.55 12 115 130 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 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 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 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 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 10 5A21_G TAIL-TO-HEAD JOINING PROTEIN GP17; VIRAL PROTEIN, VIRAL INFECTION, TAILED 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**

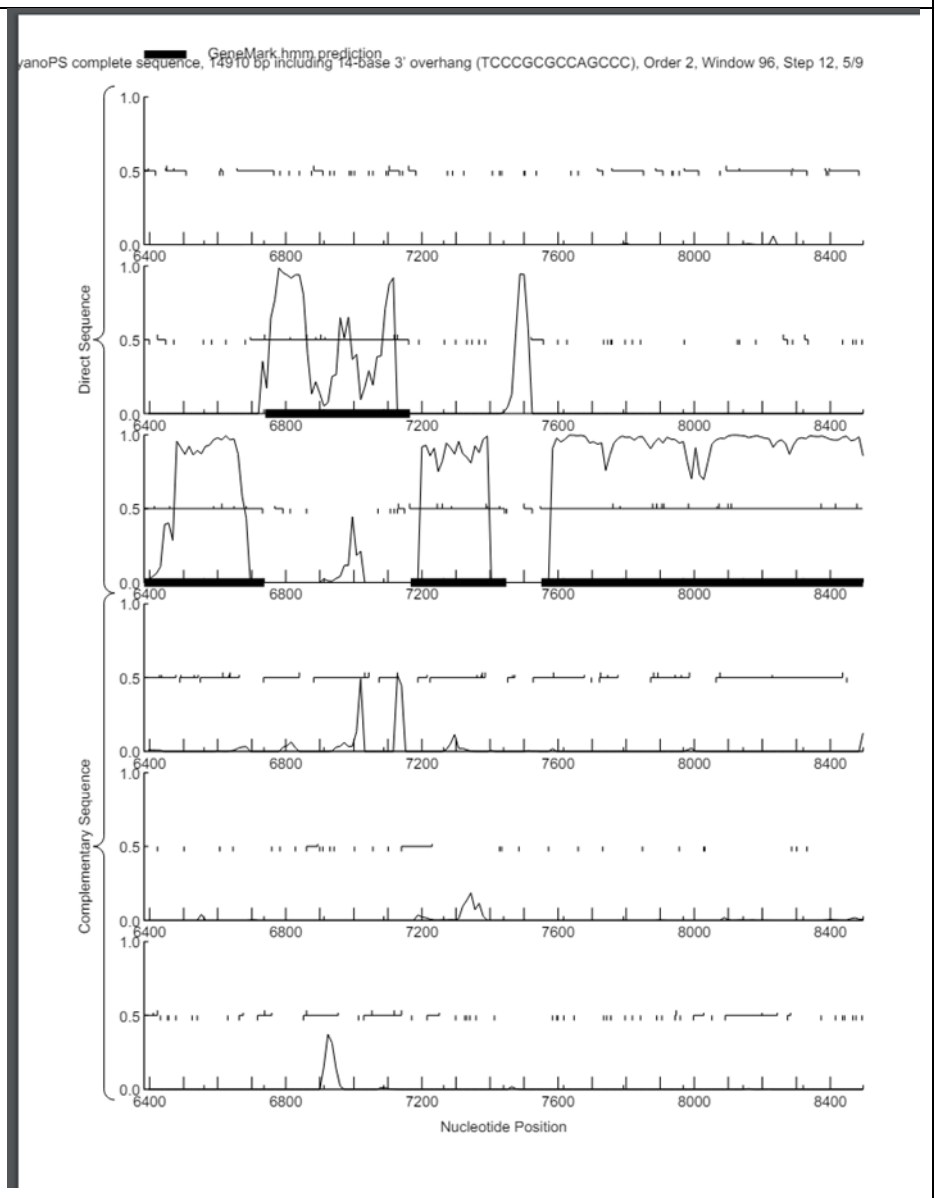
qui montre une forte conservation de l'ordre des gènes ?	
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**non** **DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Oui**
DECISION:	**Tail terminator**

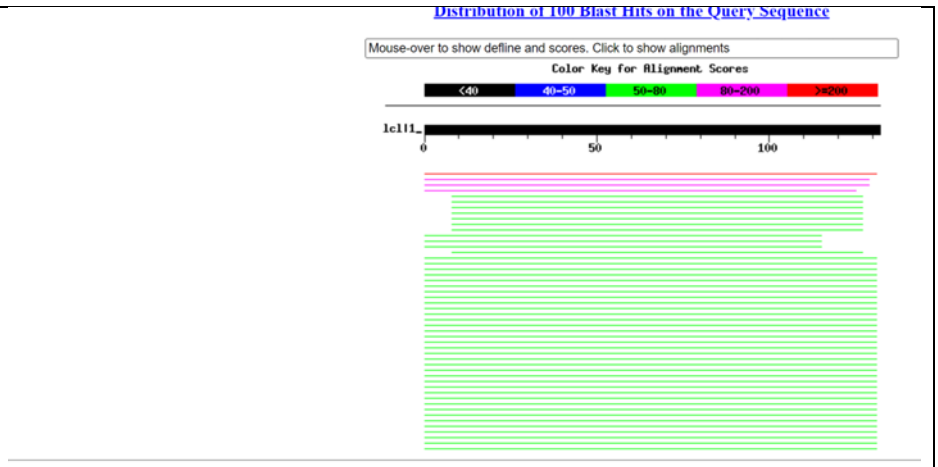
Student Gene Annotation Worksheet

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**

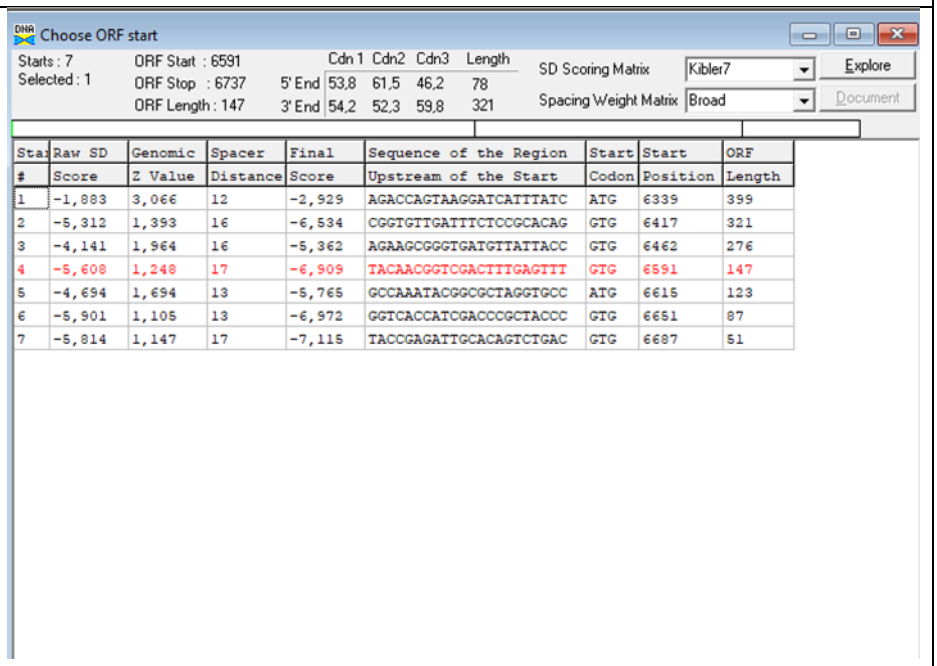
Coordonnées du Start retenu	**6339**
Fonction prédite	**Major Tail protein**

Décision #1 : Est-ce un gène ?

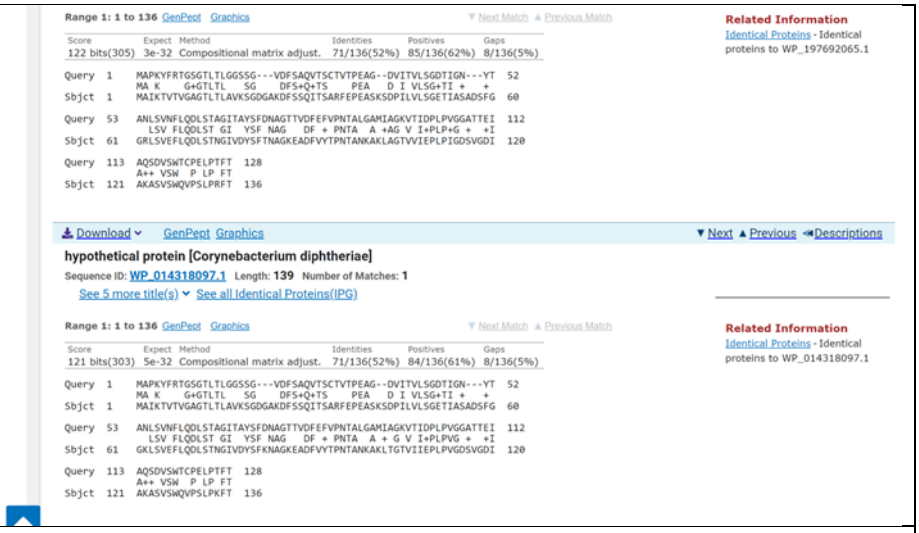
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
Y-a-t-il des éléments supportant un potentiel codant ?	

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 Sequences producing significant alignments: Score (bits) E 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 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_197692065.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 121 121 96% 5e-32 52.21% 139 WP_014318097.1 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 CAB0612796.1 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 HCD1553133.1 hypothetical protein [Corynebacterium sp. HMSC065A05] Corynebacterium sp. HMSC065A05 114 114 93% 2e-29 52.34% 136 WP_070776423.1 putative tail protein [Corynebacterium striatum] Corynebacterium striatum 114 114 93% 3e-29 52.34% 135 CQD13976.1 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 45.26% 137 WP_049167469.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 108 108 96% 3e-27 46.72% 140 WP_003852559.1 hypothetical protein [Corynebacterium propinquum] Corynebacterium propinquum 108 108 100% 7e-27 44.53% 137 WP_239210982.1 hypothetical protein [Corynebacterium] Corynebacterium 106 106 100% 4e-26 43.80% 137 WP_021353241.1 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**

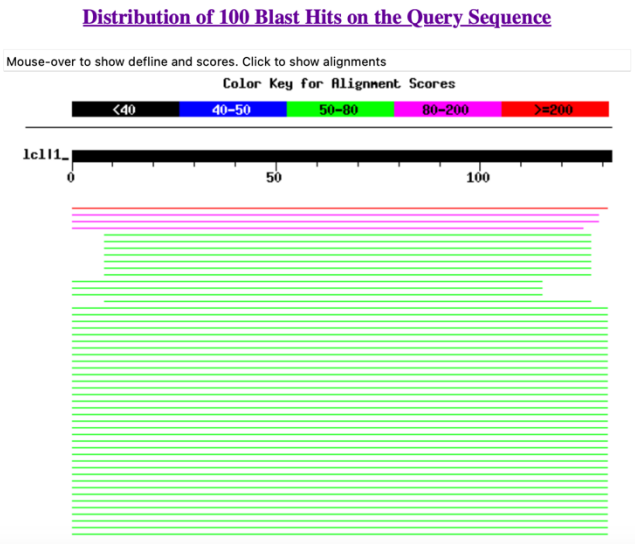
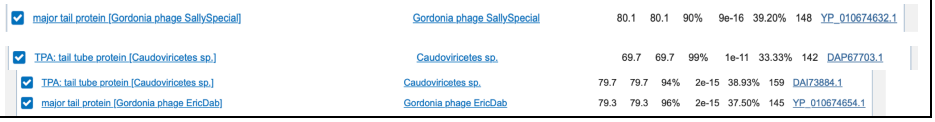
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**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**
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 ?	*ORF le plus long* *Score de -2* *ATG*
Est-ce que le start est conservé chez les homologues voir Starterator ?	–

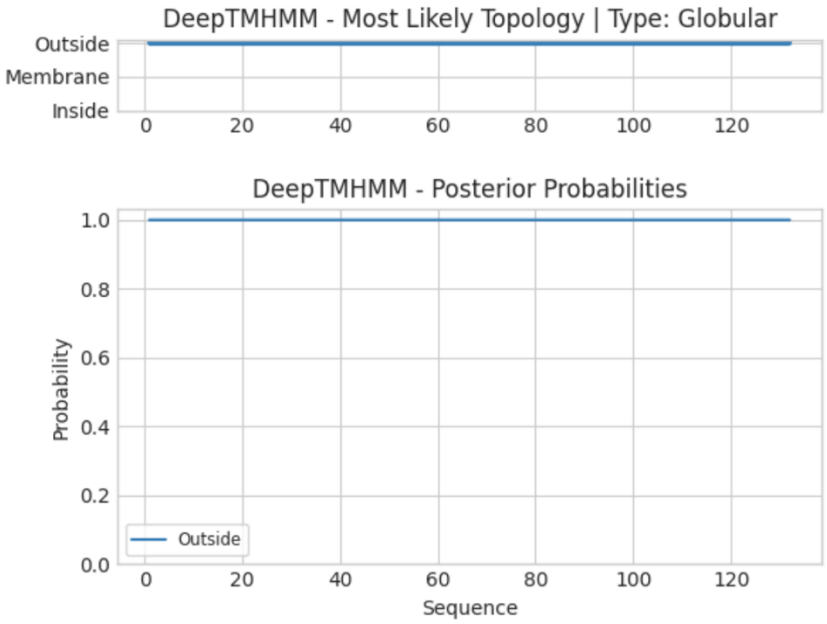
Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-1,883	3,066	12	-2,929	AGACCAGTAAGGATCATTTATC	ATG	6339	399
2	-5,312	1,393	16	-6,534	CGGTGTTGATTTCTCCGCACAG	GTG	6417	321
3	-4,141	1,964	16	-5,362	AGAAGCGGGTGATGTTATTACC	GTG	6462	276
4	-5,608	1,248	17	-6,909	TACAACGGTCGACTTTGAGTTT	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

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	
DECISION:	**6339**

Décision #3 : Quelle est la function 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 :**  ***** nr :** **(q#: s#) : [alignment] ; e-value :** 

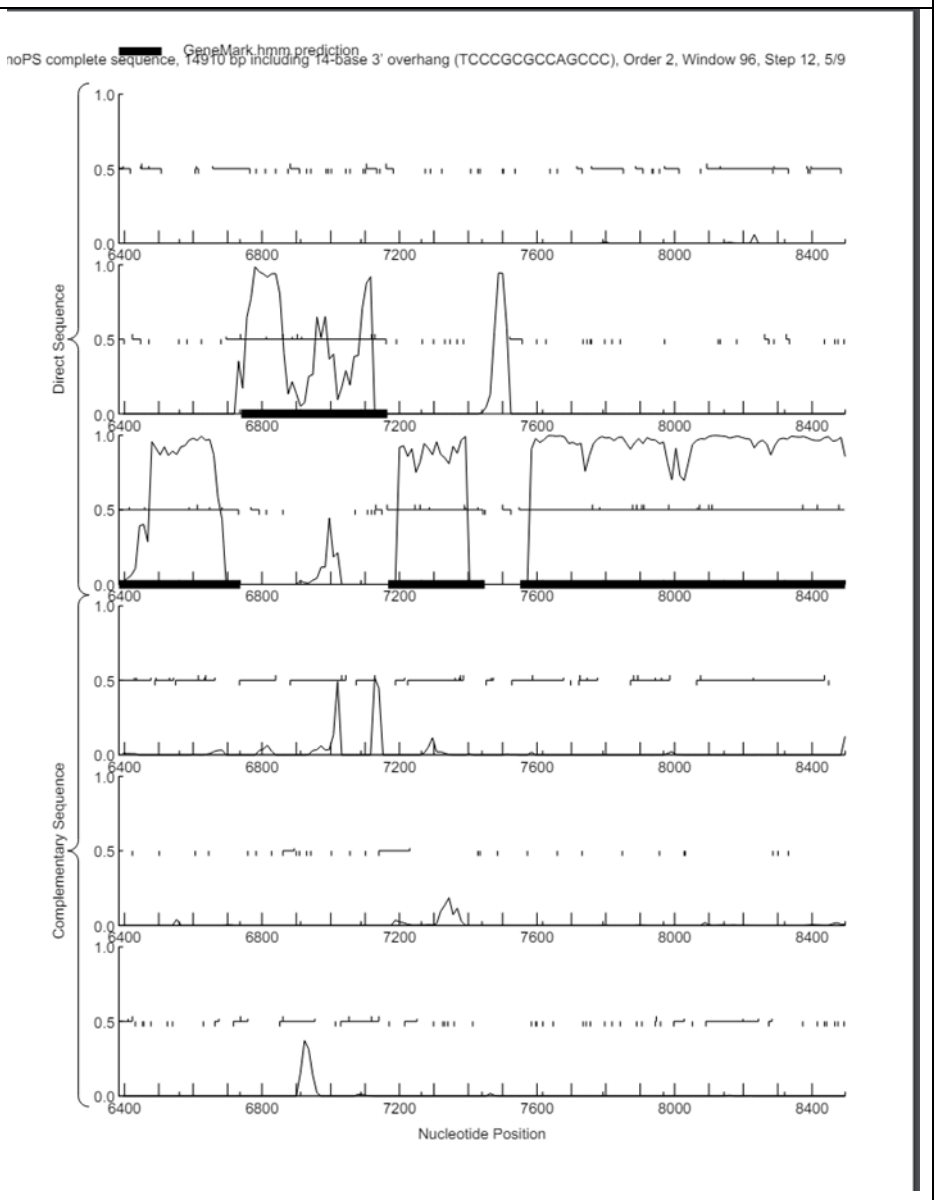
	Sequences producing significant alignments Download Select columns Show 100 select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession hypothetical protein [Corynebacterium belfantii] | Corynebacterium belfantii | 122 | 122 | 96% | 3e-32 | 52.21% | 139 | WP_197692065.1 hypothetical protein [Corynebacterium diphtheriae] | Corynebacterium diphtheriae | 121 | 121 | 96% | 5e-32 | 52.21% | 139 | WP_014318097.1 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 | CAB0612796.1 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 | HCD1553133.1 hypothetical protein [Corynebacterium sp. HMSC065A05] | Corynebacterium sp. HMSC065A05 | 114 | 114 | 93% | 2e-29 | 52.34% | 136 | WP_070776423.1 putative tail protein [Corynebacterium striatum] | Corynebacterium striatum | 114 | 114 | 93% | 3e-29 | 52.34% | 135 | CQD13976.1 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
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 ?	Visualization Resubmit Section 6 132 **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, Hub 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**
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**

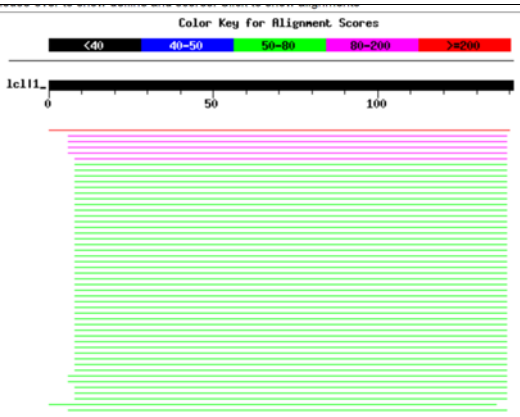
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Non**
DECISION:	**Major Tail protein**

Student Gene Annotation Worksheet

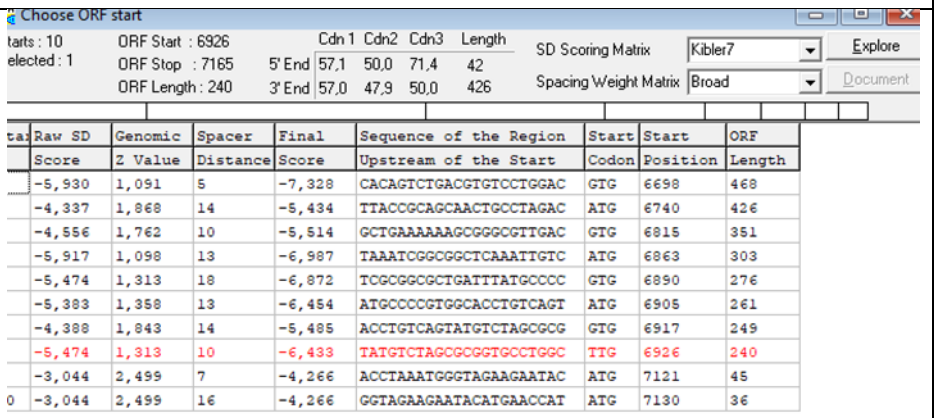
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**11**
Coordonnées du Stop	**7165**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Gap 2**
Coordonnées du Start retenu	**6740**
Fonction prédite	**NKF**

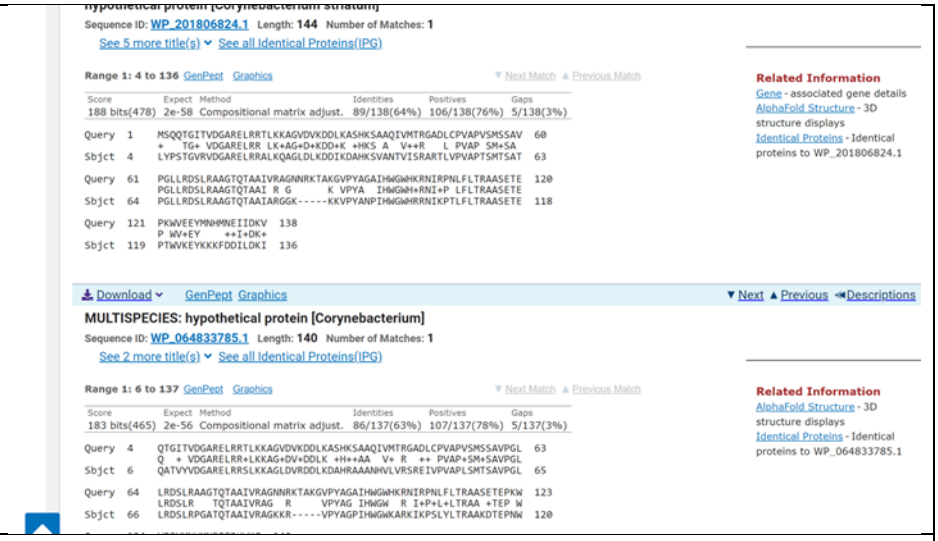
Décision #1 : Est-ce un gène ?

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 ?	

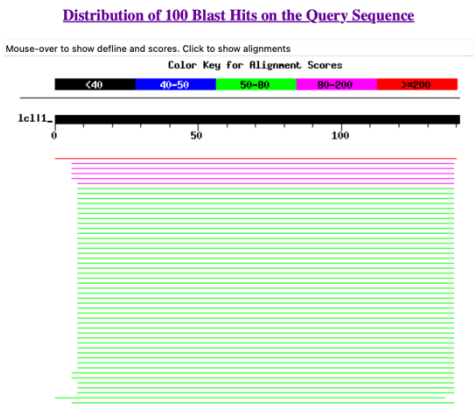
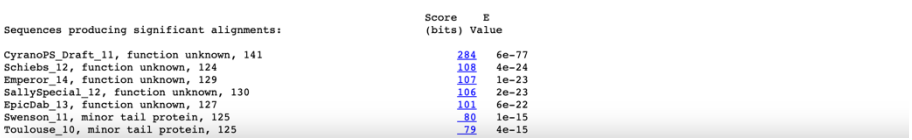
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 Sequences producing significant alignments: Score (bits) E Value CyranoPS_Draft_11, function unknown, 141 284 6e-77 Schiebs_12, function unknown, 124 108 4e-24 Emperor_14, function unknown, 129 107 1e-23 SallySpecial_12, function unknown, 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 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 hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 188 188 97% 2e-58 64.49% 144 WP_201806824.1 hypothetical protein [Corynebacterium] Corynebacterium 183 183 97% 2e-56 62.77% 140 WP_064833785.1 hypothetical protein [Corynebacterium pseudodiphtheriticum] Corynebacterium pseudodiphth... 183 183 97% 2e-56 62.77% 140 WP_284849035.1 hypothetical protein [Corynebacterium propinquum] Corynebacterium propinquum 182 182 97% 8e-56 62.77% 140 WP_284594216.1 hypothetical protein [Corynebacterium hadale] Corynebacterium hadale 180 180 95% 3e-55 64.18% 142 WP_095275385.1 hypothetical protein [Corynebacterium sp. HMSC06C06] Corynebacterium sp. HMSC06C06 180 180 87% 5e-55 69.92% 142 WP_246817189.1 putative phage protein [Corynebacterium striatum] Corynebacterium striatum 177 177 95% 9e-54 60.74% 147 CQD13979.1 hypothetical protein [unclassified Corynebacterium] unclassified Corynebacterium 175 175 95% 4e-53 60.74% 147 WP_246817216.1 TPA: hypothetical protein [Corynebacterium striatum] Corynebacterium striatum 172 172 92% 4e-52 61.07% 138 HCT5225725.1 hypothetical protein HMPREF2998_07920 [Corynebacterium sp. HMSC065A05] Corynebacterium sp. HMSC065... 171 171 92% 1e-51 61.07% 138 OFP20320.1 hypothetical protein [Corynebacterium diphtheriae] Corynebacterium diphtheriae 162 162 99% 5e-48 57.75% 137 WP_010935362.1 hypothetical protein CIP107541_01712 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 162 162 99% 7e-48 57.75% 152 CAB0610097.1 hypothetical protein FRC0431_01842 [Corynebacterium diphtheriae] Corynebacterium diphtheriae 160 160 99% 3e-47 57.04% 152 CAB0923859.1 hypothetical protein [Corynebacterium propinquum] Corynebacterium propinquum 159 159 96% 6e-47 53.68% 144 WP_239210984.1 hypothetical protein [Corynebacterium] Corynebacterium 159 159 94% 1e-46 54.89% 144 WP_065420615.1 hypothetical protein W5M_08714 [Corynebacterium diphtheriae bv. intermedius str. NCTC 5011] Corynebacterium diphtheriae 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**

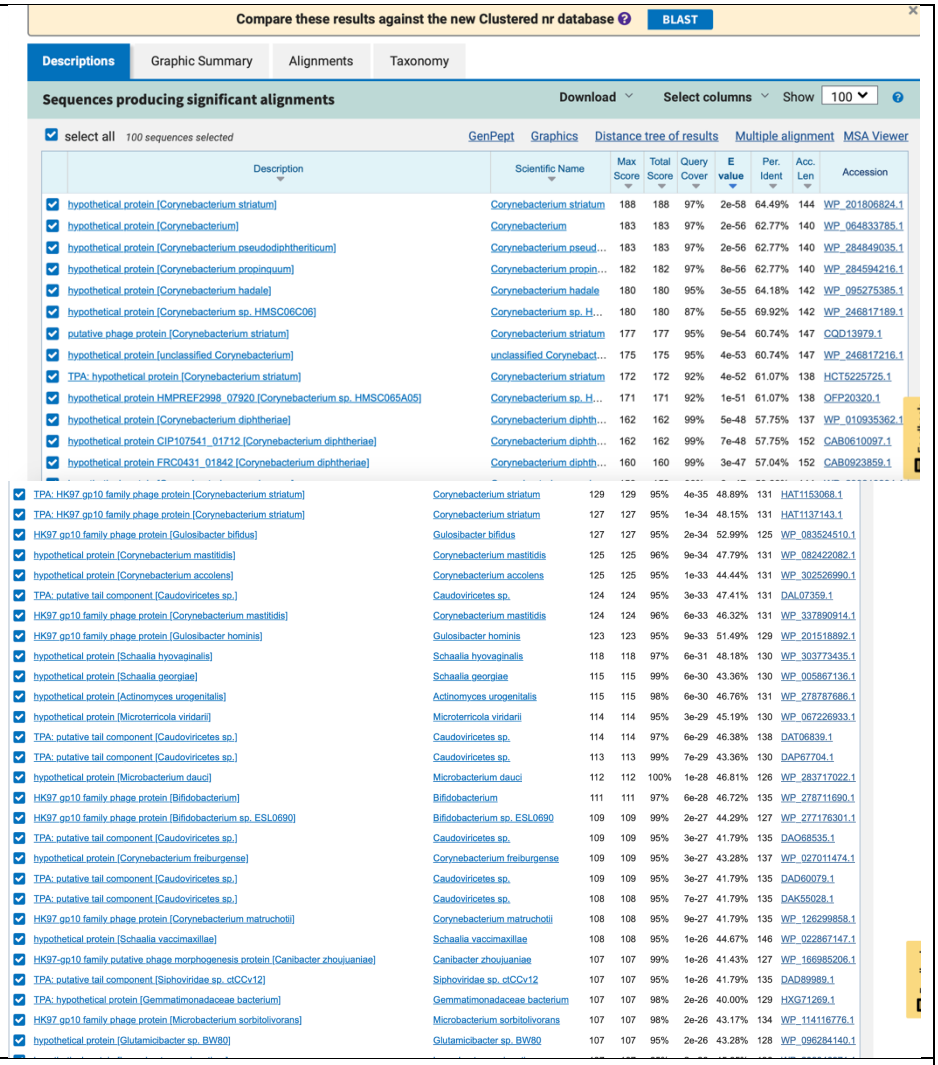
Décision #2 : Quel est le start du gène ?

Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**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**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 Raw SD Score / Genomic Z Value / Spacer Distance / Final Score / Sequence of the Region Upstream of the Start / Start Codon / Start Position / ORF Length -5,930 1,091 5 -7,328 CACAGTCTGACGTGTCCTGGAC 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 TAAATCGGCGGCTCAAATTGTC ATG 6863 303 -5,474 1,313 18 -6,872 TCGCGGCGCTGATTTATGCCCC GTG 6890 276 -5,383 1,358 13 -6,454 ATGCCCCGTGGCACCTGTCAGT ATG 6905 261 -4,388 1,843 14 -5,485 ACCTGTCAGTATGTCTAGCGCG GTG 6917 249 -5,474 1,313 10 -6,433 TATGTCTAGCGCGGTGCCTGGC TTG 6926 240 -3,044 2,499 7 -4,266 ACCTAAATGGGTAGAAGAATAC ATG 7121 45 -3,044 2,499 16 -4,266 GGTAGAAGAATACATGAACCAT ATG 7130 36
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	***Plus long ORF avec un ATG***
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:	**6740**

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

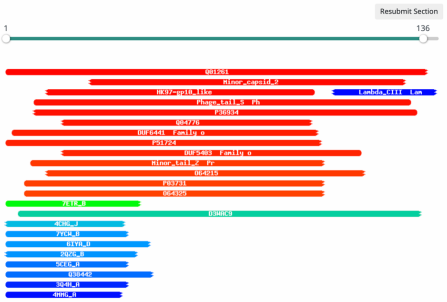
Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	**Listez le meilleur hit Blastp pour chaque source :** ***** PhagesDB :**   *****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



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

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
2	PF11114.12	; Minor_capsid_2 ; Minor capsid protein	99.42	2.8e-12	74.8	7.1	83	105
3	PF04883.16	; HK97-gp10_like ; Bacteriophage HK97-gp10, putative tail-component	99.13	5.5e-10	60.68	5.7	72	74
4	PF05069.17	; Phage_tail_S ; Phage virion morphogenesis family	98.64	1.4e-7	58.5	4.7	113	145
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
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
7	PF20039.3	; DUF6441 ; Family of unknown function (DUF6441)	97.48	0.0015	43.84	6.3	87	215
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
9	PF17395.6	; DUF5403 ; Family of unknown function (DUF5403)	97.02	0.0026	36.93	3.4	80	93
10	PF06763.15	; Minor_tail_Z ; Prophage minor tail protein Z (GPZ)	95.33	0.44	31.31	7.2	85	182
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

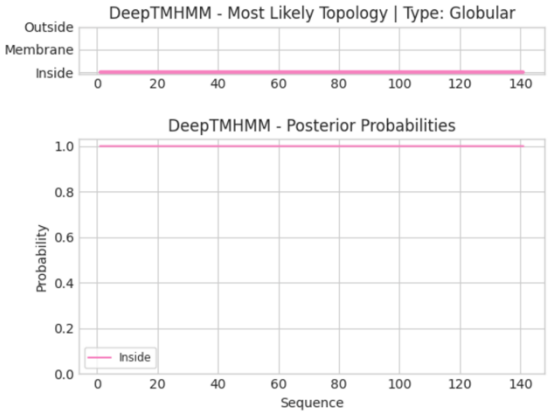
Visualization



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 synthénie

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



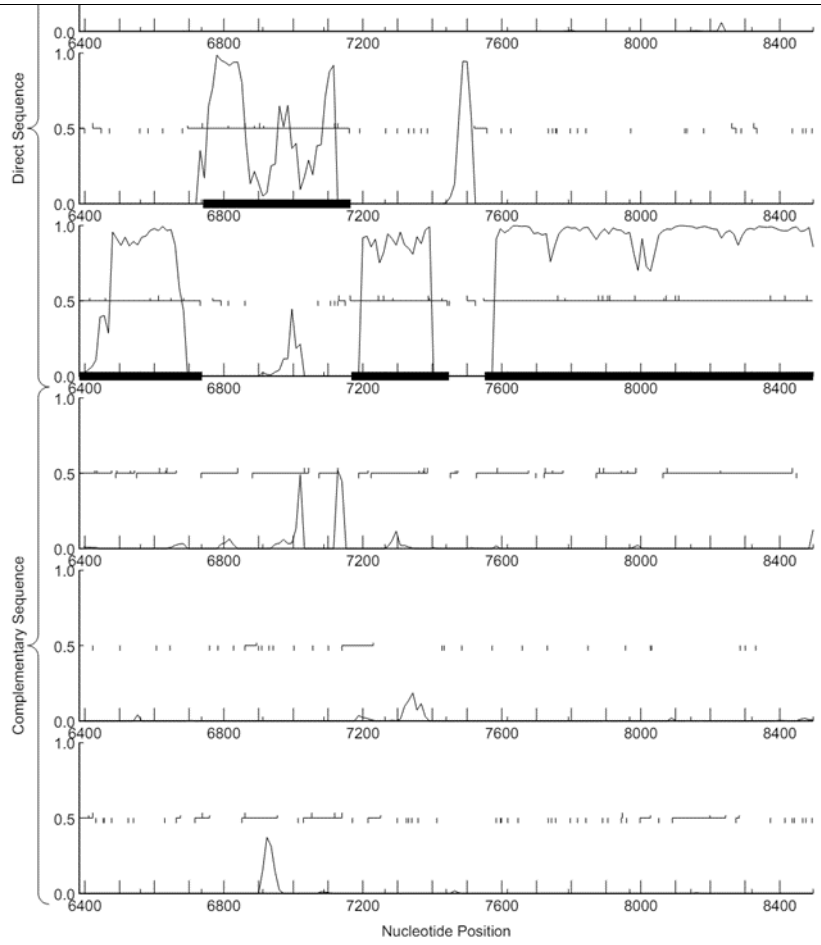
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Non**
DECISION:	**NKF**

Student Gene Annotation Worksheet

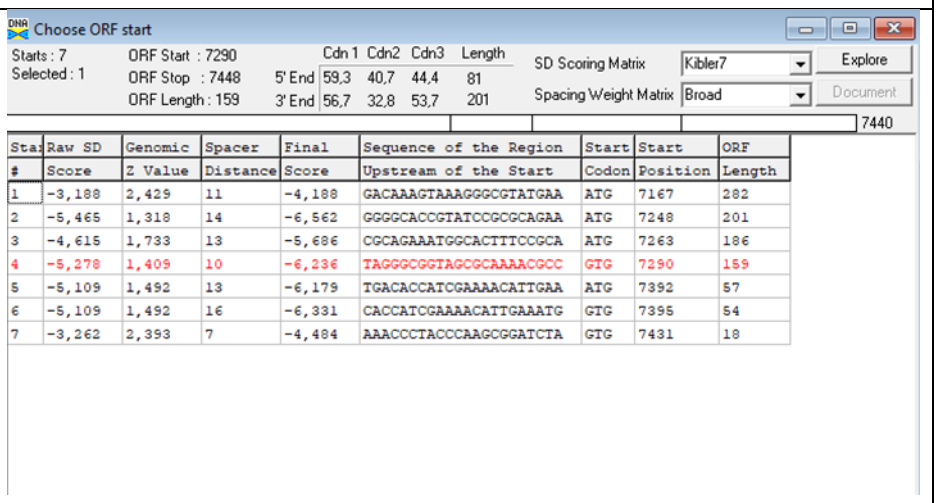
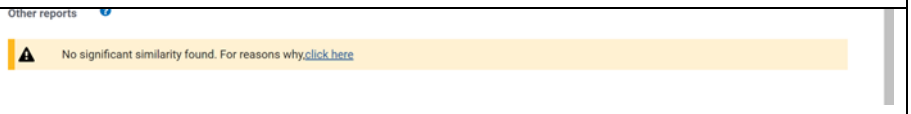
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**12**
Coordonnées du Stop	**7448**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Gap 1**
Coordonnées du Start retenu	**7167**
Fonction prédite	**NKF**

Décision #1 : Est-ce un gène ?

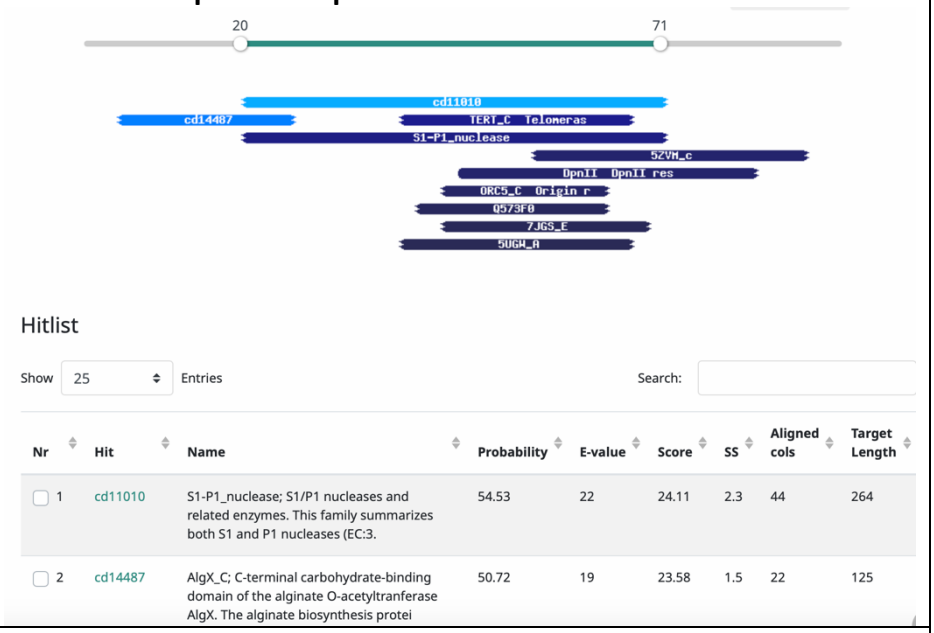
Collection des éléments de réponse	**Rationnelle**
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**

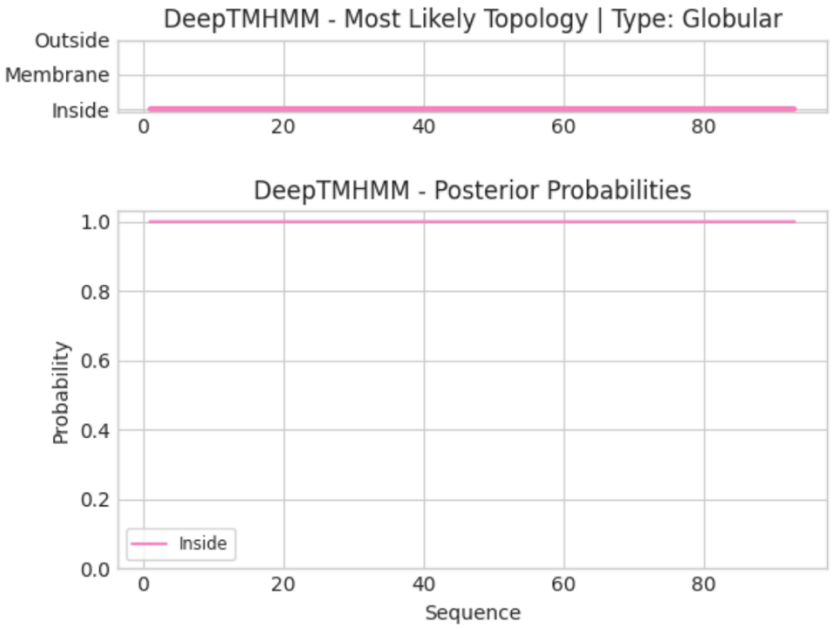
Y-a-t-il des éléments supportant un potentiel codant ?	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	Other reports No significant similarity found. For reasons why,click here Query= CyranoPS_12 (93 letters) Distribution of 3 Blast Hits on the Query Sequence Mouse-over to show defline and scores. Click to show alignments Color Key for Alignment Scores <40 40-50 50-80 80-200 >=200 lcl|1_ 0 50 Sequences producing significant alignments: Score (bits) E Value CyranoPS_Draft_12, function unknown, 93 186 2e-47 KeAlii_2, terminase large subunit, 565 28 7.9 DrSierra_2, terminase large subunit, 568 28 7.9 >CyranoPS_Draft_12, function unknown, 93 Length = 93 Score = 186 bits (472), Expect = 2e-47 Identities = 93/93 (100%), Positives = 93/93 (100%) Query: 1 MAIQEVTYLDGKKETATISQGHRIRAEMALSAMGRAVAQNAVTWLNVAAFFAVVNPKKFN 60 MAIQEVTYLDGKKETATISQGHRIRAEMALSAMGRAVAQNAVTWLNVAAFFAVVNPKKFN Sbjct: 1 MAIQEVTYLDGKKETATISQGHRIRAEMALSAMGRAVAQNAVTWLNVAAFFAVVNPKKFN 60 Query: 61 FDEFLTWSDTIENIEMVEEQVNPTQADLVLDPQ 93 FDEFLTWSDTIENIEMVEEQVNPTQADLVLDPQ Sbjct: 61 FDEFLTWSDTIENIEMVEEQVNPTQADLVLDPQ 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 ?	 Sta# / Raw SD Score / Genomic Z Value / Spacer Distance / Final Score / Sequence of the Region Upstream of the Start / Start Codon / Start Position / ORF Length 1 -3,188 2,429 11 -4,188 GACAAAGTAAAGGGCGTATGAA ATG 7167 282 2 -5,465 1,318 14 -6,562 GGGGCACCGTATCCGCGCAGAA ATG 7248 201 3 -4,615 1,733 13 -5,686 CGCAGAAATGGCACTTTCCGCA ATG 7263 186 4 -5,278 1,409 10 -6,236 TAGGGCGGTAGCGCAAAACGCC GTG 7290 159 5 -5,109 1,492 13 -6,179 TGACACCATCGAAAACATTGAA ATG 7392 57 6 -5,109 1,492 16 -6,331 CACCATCGAAAACATTGAAATG GTG 7395 54 7 -3,262 2,393 7 -4,484 AAACCCTACCCAAGCGGATCTA 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 ?	***Meilleur score,*** ***ORF le plus long*** ***ORF le plus long avec un ATG***
Est-ce que le start est conservé chez les homologues voir Starterator ?	–
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	
DECISION:	**7167**

Décision #3 : Quelle est la function 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 :** **RIEN** ***** nr :** **(q#: s#) : [alignment] ; e-value :** **Note : Vous avez peut-être déjà trouvé ces informations à partir de la décision d'annotation #2. Il suffit de fournir une seule correspondance de chaque base de données.**
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	**"Pas de correspondance pertinente"**  Hitlist Show 25 Entries Search: 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-acetyltranferase AlgX. The alginate biosynthesis protei 50.72 19 23.58 1.5 22 125
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 inconu mais le gene suivant est une tape measure protein**

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 ?	**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.**
DECISION:	**NKF**

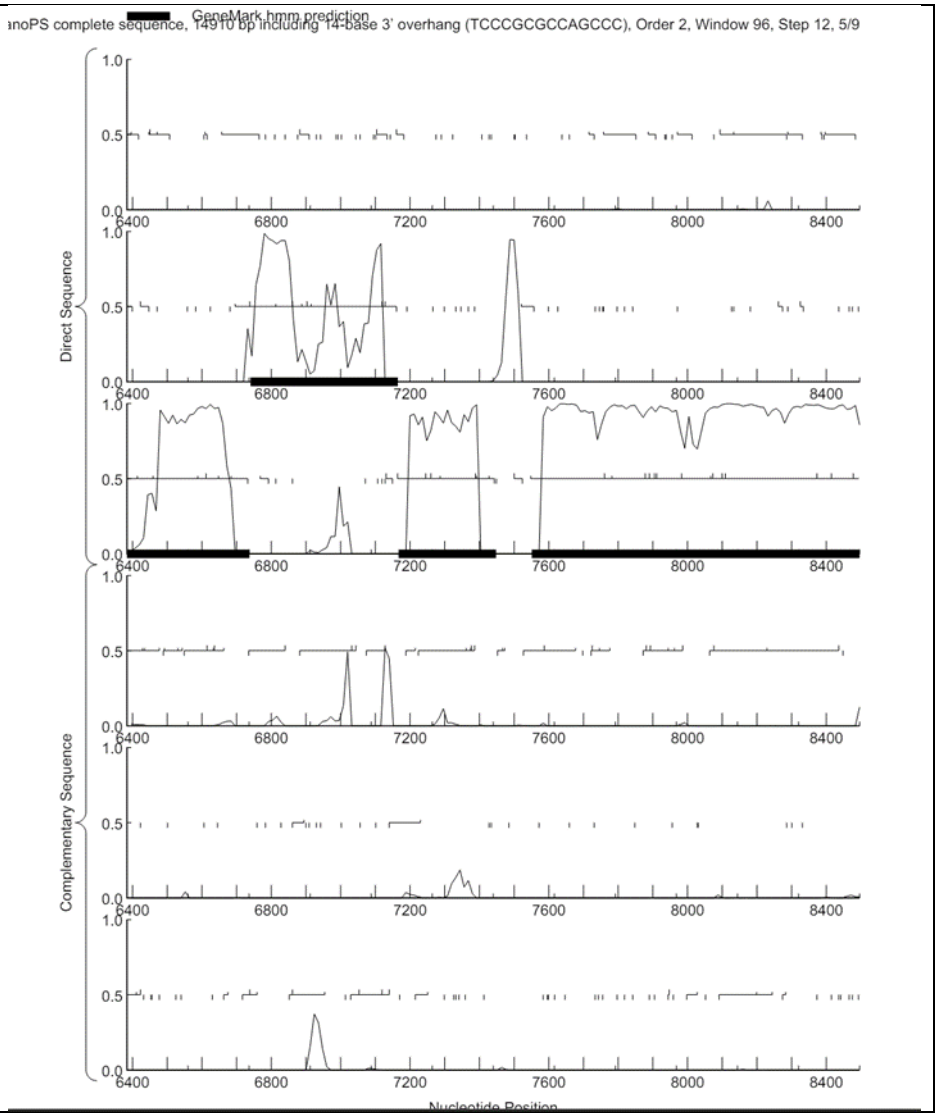
Student Gene Annotation Worksheet

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**

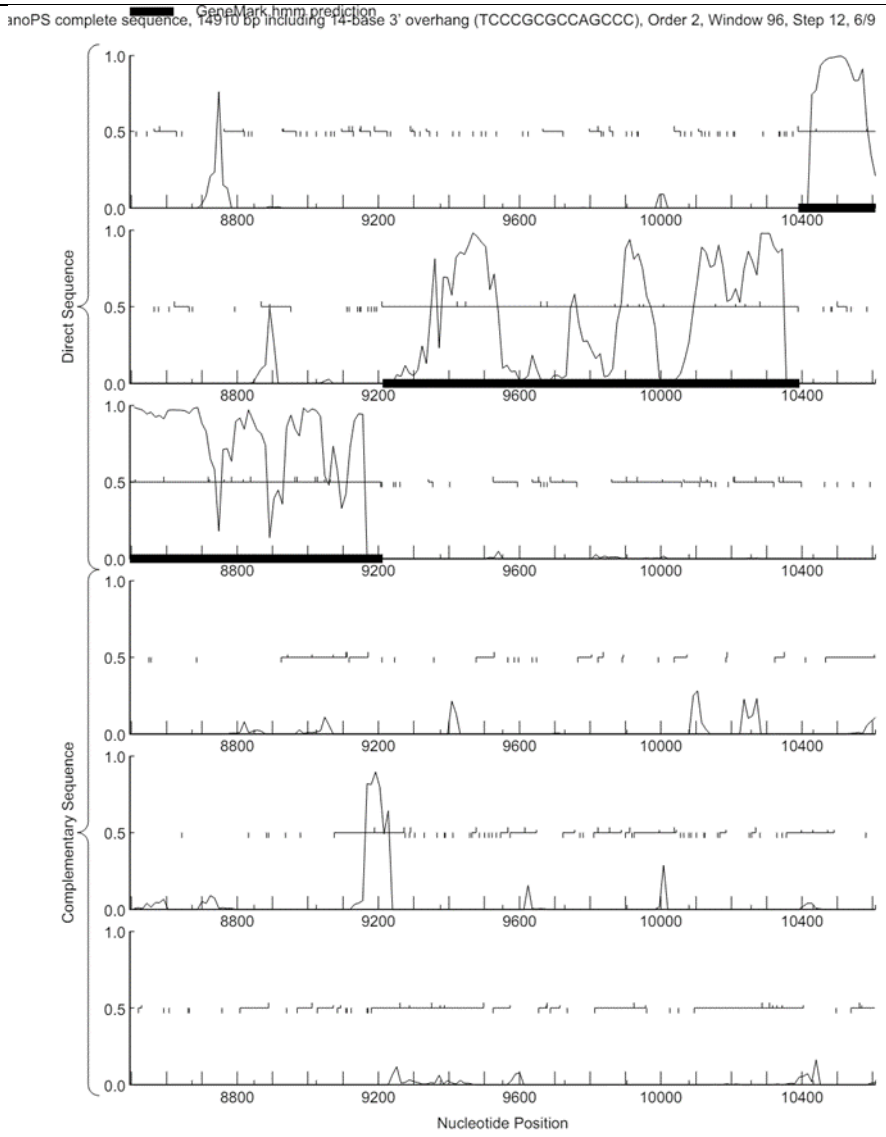
Décision #1 : Est-ce un gène ?

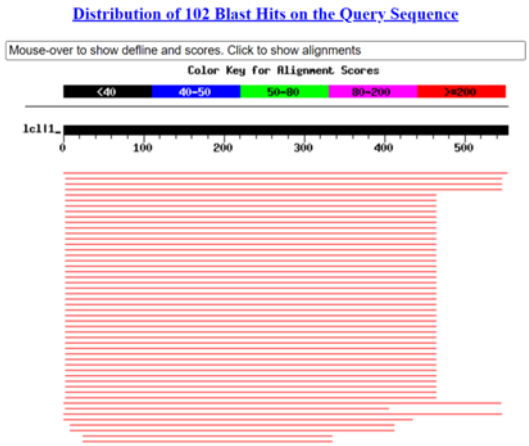
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**

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



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





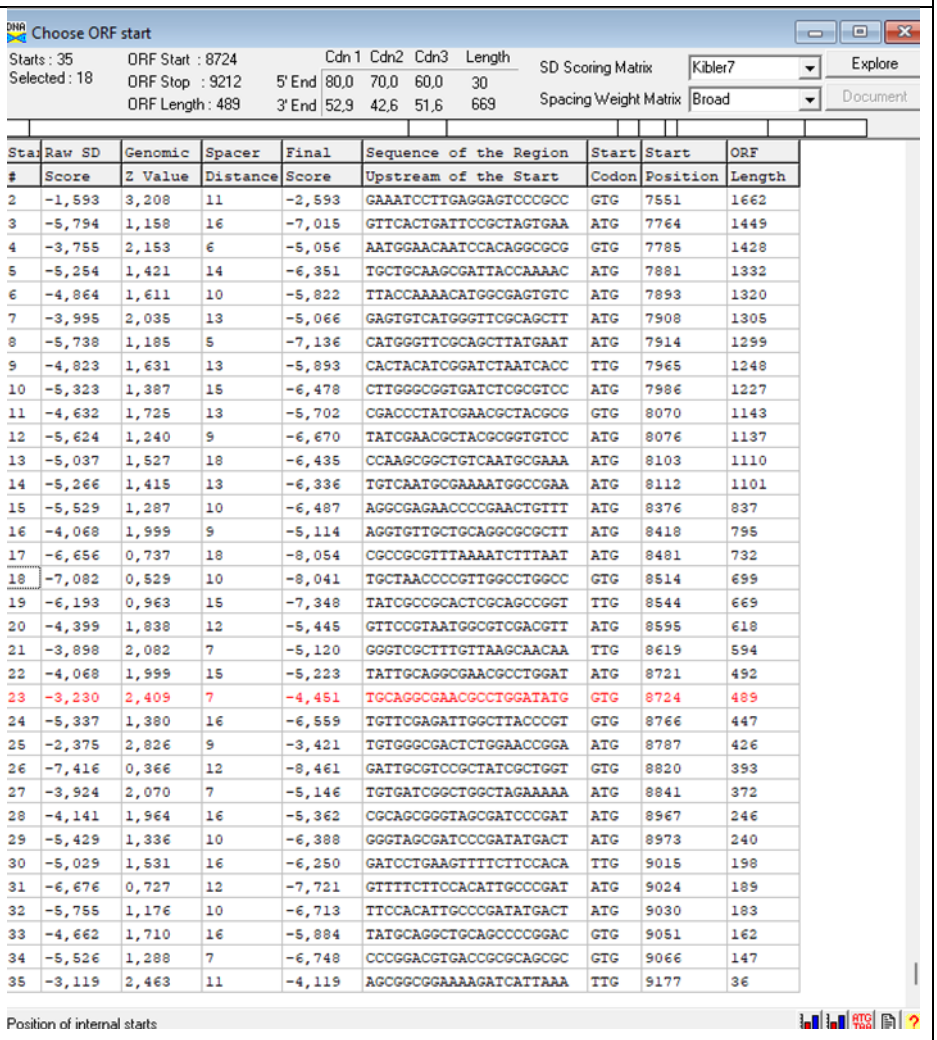
Sequences producing significant alignments:

	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

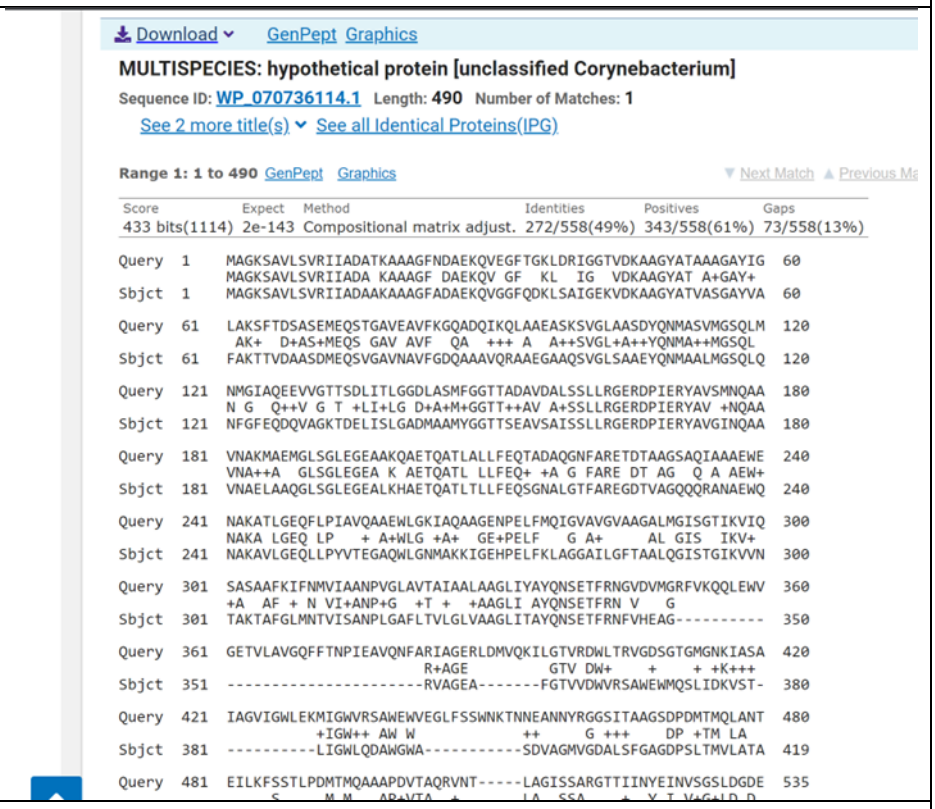
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**NON**
DECISION:	**OUI**

hypothetical protein [Corynebacterium]	Corynebacterium	444	444	100%	1e-147	49.10%	492	WP_070420702.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	436	436	100%	3e-144	48.92%	490	HCT5225727.1
hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	433	433	100%	2e-143	48.75%	490	WP_070736114.1
putative phage protein [Corynebacterium striatum]	Corynebacterium striatum	433	433	100%	3e-143	48.75%	490	CQD13985.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	432	432	100%	7e-143	48.74%	492	HCD1553136.1
hypothetical protein CIP107524_01832 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	391	391	80%	2e-126	53.63%	523	CAB0565259.1
hypothetical protein FRC0370_01709 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	389	389	96%	1e-125	47.54%	523	CAB0870047.1
hypothetical protein FRC0378_01882 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	388	388	96%	3e-125	48.09%	523	CAB0871289.1
hypothetical protein FRC0474_02088 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	388	388	80%	4e-125	53.19%	523	CAB0972304.1
hypothetical protein D9R17_02160 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	387	387	80%	5e-125	53.19%	503	RLP10245.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	387	387	80%	6e-125	53.19%	523	WP_182001102.1
hypothetical protein FRC0430_01824 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	387	387	80%	6e-125	53.19%	523	CAB0913975.1
hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	386	386	100%	7e-125	46.32%	488	WP_095275383.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	386	386	96%	2e-124	47.72%	523	WP_088298569.1
hypothetical protein [Corynebacterium glaucum]	Corynebacterium glaucum	385	385	100%	2e-124	42.58%	489	WP_301925793.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	385	385	80%	5e-124	52.75%	523	WP_205917723.1
hypothetical protein CIP107521_01967 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	384	384	96%	1e-123	47.54%	523	CAB0567951.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	385	385	99%	6e-123	39.35%	625	HAT6525551.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	382	382	98%	2e-122	43.49%	554	WP_284594213.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	380	380	98%	1e-121	44.50%	551	WP_284596340.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	378	378	80%	2e-121	53.41%	523	WP_342351204.1
hypothetical protein [Corynebacterium]	Corynebacterium	379	379	99%	3e-121	43.09%	551	WP_064833146.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	379	379	98%	3e-121	43.67%	554	WP_284849037.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	379	379	98%	4e-121	43.65%	554	WP_302524538.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	377	377	80%	5e-121	53.41%	523	WP_106361639.1
hypothetical protein [Winkia]	Winkia	284	284	67%	9e-86	44.25%	462	WP_004806609.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	289	358	100%	1e-85	46.40%	627	WP_201816501.1
hypothetical protein [Propionimicrobium lymphophilum]	Propionimicrobium lymphophilum	283	283	79%	5e-84	40.81%	556	WP_285112190.1
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
hypothetical protein GCM10023233_27360 [Brevibacterium otitidis]	Brevibacterium otitidis	280	280	79%	5e-83	39.87%	539	BFF07767.1
hypothetical protein [Schaalia georgiae]	Schaalia georgiae	277	277	68%	2e-82	47.45%	495	WP_005867126.1
hypothetical protein [Corynebacterium poyangense]	Corynebacterium poyangense	278	278	64%	2e-82	43.77%	529	WP_187974193.1
phage tail tape measure protein [Corynebacterium oculi]	Corynebacterium oculi	278	338	55%	7e-81	63.11%	680	WP_150114274.1
hypothetical protein [uncultured Leifsonia sp.]	uncultured Leifsonia sp.	273	273	53%	2e-80	52.88%	525	WP_314148240.1
tape measure protein [Arthrobacter phage Swenson]	Arthrobacter phage Swenson	273	273	62%	4e-79	49.14%	665	ASR83993.1
hypothetical protein [Gordonia westfalica]	Gordonia westfalica	269	269	98%	5e-79	37.01%	527	WP_074850009.1
tail length tape measure protein [Gordonia phage EricDab]	Gordonia phage EricDab	266	266	72%	3e-78	41.63%	513	YP_010674657.1
hypothetical protein [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	269	269	67%	5e-78	49.62%	636	WP_064317077.1
hypothetical protein [Gulosibacter bifidus]	Gulosibacter bifidus	266	266	97%	1e-77	38.53%	536	WP_066057107.1
hypothetical protein [Microbacterium sp. A(2022)]	Microbacterium sp. A(2022)	263	263	57%	2e-77	49.54%	477	WP_311878682.1
tail length tape measure protein [Arthrobacter phage Decurro]	Arthrobacter phage Decurro	268	268	62%	3e-77	49.14%	665	YP_009191306.1
tape measure protein [Arthrobacter phage Jessica]	Arthrobacter phage Jessica	268	268	62%	3e-77	49.14%	665	ALF00769.1
tape measure protein [Arthrobacter phage Muttlie]	Arthrobacter phage Muttlie	268	268	62%	4e-77	49.14%	665	ALY09769.1
tape measure protein [Arthrobacter phage Toulouse]	Arthrobacter phage Toulouse	268	268	62%	4e-77	49.14%	665	ALY10652.1
tape measure protein [Arthrobacter phage Moloch]	Arthrobacter phage Moloch	268	268	62%	4e-77	49.14%	665	ALY09743.1
tape measure protein [Arthrobacter phage Elkhorn]	Arthrobacter phage Elkhorn	265	265	62%	3e-76	49.14%	665	ASR83520.1
tape measure protein [Arthrobacter phage KylieMac]	Arthrobacter phage KylieMac	265	265	62%	3e-76	49.14%	665	ASR83547.1
hypothetical protein [Brevibacterium sp. UMB1308A]	Brevibacterium sp. UMB1308A	261	261	73%	1e-75	41.49%	555	WP_284926673.1
hypothetical protein [Microbacterium sp. cx-55]	Microbacterium sp. cx-55	259	259	71%	2e-75	42.45%	524	WP_223721111.1
tape measure protein [Arthrobacter phage Saphira]	Arthrobacter phage Saphira	263	263	62%	3e-75	48.56%	665	QGH75234.1

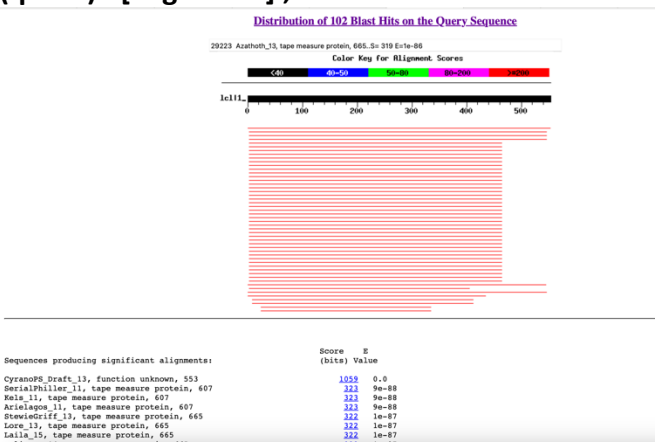
Décision #2 : Quel est le start du gène ?

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) :7551** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :7551**
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 ?	*Plus long orf mais pas avec un ATG* *Meilleur score*
Est-ce que le start est conservé chez les	–

Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
2	-1,593	3,208	11	-2,593	GAAATCCTTGAGGAGTCCCGCC	GTG	7551	1662
3	-5,794	1,158	16	-7,015	GTTCACTGATTCCGCTAGTGAA	ATG	7764	1449
4	-3,755	2,153	6	-5,056	AATGGAACAATCCACAGGCGCG	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	GAGTGTCATGGGTTCGCAGCTT	ATG	7908	1305
8	-5,738	1,185	5	-7,136	CATGGGTTCGCAGCTTATGAAT	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	CGACCCTATCGAACGCTACGCG	GTG	8070	1143
12	-5,624	1,240	9	-6,670	TATCGAACGCTACGCGGTGTCC	ATG	8076	1137
13	-5,037	1,527	18	-6,435	CCAAGCGGCTGTCAATGCGAAA	ATG	8103	1110
14	-5,266	1,415	13	-6,336	TGTCAATGCGAAAATGGCCGAA	ATG	8112	1101
15	-5,529	1,287	10	-6,487	AGGCGAGAACCCCGAACTGTTT	ATG	8376	837
16	-4,068	1,999	9	-5,114	AGGTGTTGCTGCAGGCGCGCTT	ATG	8418	795
17	-6,656	0,737	18	-8,054	CGCCGCGTTTAAAATCTTTAAT	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	GTTCCGTAATGGCGTCGACGTT	ATG	8595	618
21	-3,898	2,082	7	-5,120	GGGTCGCTTTGTTAAGCAACAA	TTG	8619	594
22	-4,068	1,999	15	-5,223	TATTGCAGGCGAACGCCTGGAT	ATG	8721	492
23	-3,230	2,409	7	-4,451	TGCAGGCGAACGCCTGGATATG	GTG	8724	489
24	-5,337	1,380	16	-6,559	TGTTCGAGATTGGCTTACCCGT	GTG	8766	447
25	-2,375	2,826	9	-3,421	TGTGGGCGACTCTGGAACCGGA	ATG	8787	426
26	-7,416	0,366	12	-8,461	GATTGCGTCCGCTATCGCTGGT	GTG	8820	393
27	-3,924	2,070	7	-5,146	TGTGATCGGCTGGCTAGAAAAA	ATG	8841	372
28	-4,141	1,964	16	-5,362	CGCAGCGGGTAGCGATCCCGAT	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	GTTTTCTTCCACATTGCCCGAT	ATG	9024	189
32	-5,755	1,176	10	-6,713	TTCCACATTGCCCGATATGACT	ATG	9030	183
33	-4,662	1,710	16	-5,884	TATGCAGGCTGCAGCCCCGGAC	GTG	9051	162
34	-5,526	1,288	7	-6,748	CCCGGACGTGACCGCGCAGCGC	GTG	9066	147
35	-3,119	2,463	11	-4,119	AGCGGCGGAAAAGATCATTAAA	TTG	9177	36

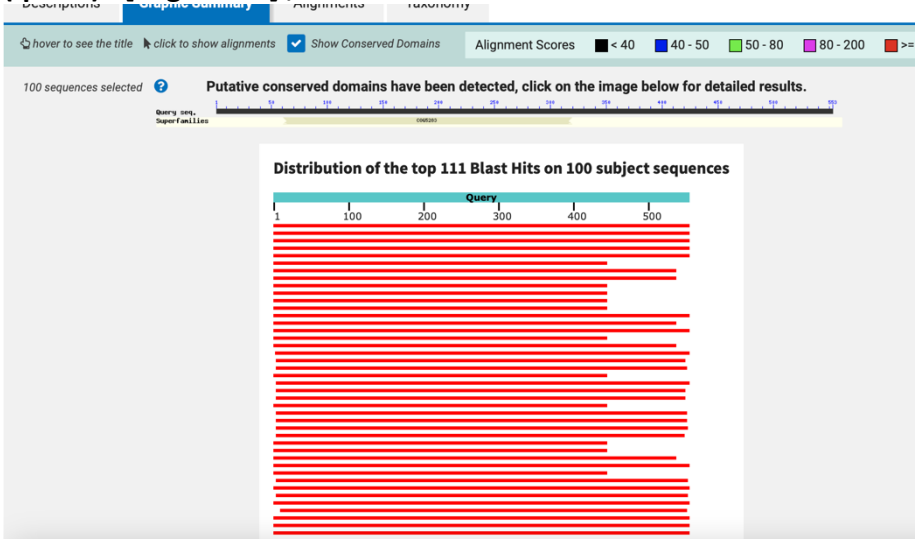
homologues voir Starterator ?	
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	
DECISION:	**7551**

Décision #3 : Quelle est la function 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 :** 

***** nr :**

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

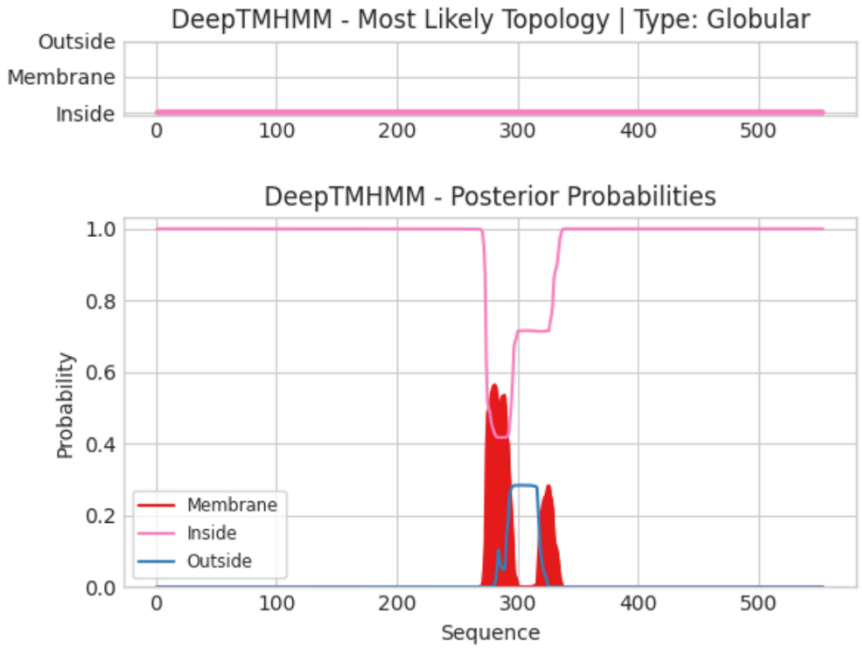


TPA: tail tape measure [Siphoviridae sp. ctCCv12]	Siphoviridae sp. ctCCv12	327	327	70%	2e-102	47.04%	466	DAD89990.1
hypothetical protein [Corynebacterium matruchotii]	Corynebacterium matruchotii	321	321	70%	7e-100	46.27%	466	WP_126299860.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	308	308	49%	2e-96	63.27%	333	WP_302526992.1
hypothetical protein [Schaalia hyovaginalis]	Schaalia hyovaginalis	311	311	64%	1e-95	50.95%	489	WP_320772602.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	308	308	54%	1e-95	53.38%	412	DAO68536.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	307	307	60%	3e-95	50.00%	414	DAK55025.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	307	307	60%	3e-95	50.00%	414	DAD60080.1
hypothetical protein [Schaalia hyovaginalis]	Schaalia hyovaginalis	308	308	64%	1e-94	50.41%	489	WP_303843687.1
hypothetical protein [Actinomyces polynesiensis]	Actinomyces polynesiensis	308	308	68%	2e-94	48.21%	491	WP_052450542.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	303	303	56%	9e-94	51.56%	412	DAU29185.1
hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	307	307	78%	5e-92	44.75%	675	WP_051255731.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	304	364	78%	8e-92	56.51%	587	HAT1331008.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	303	364	78%	1e-91	56.51%	597	HAT1419171.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	305	305	54%	1e-91	55.45%	628	WP_284900637.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	303	364	78%	1e-91	56.51%	598	HAT1137141.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	304	357	81%	2e-91	55.45%	628	WP_284895773.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	301	362	99%	1e-90	56.19%	616	HAT6642583.1
hypothetical protein Cocul_00066 [Corynebacterium oculi]	Corynebacterium oculi	303	363	65%	3e-90	60.13%	715	KQB84936.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	300	300	53%	3e-90	58.11%	598	HAT1153070.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	303	362	65%	3e-90	59.80%	713	WP_156801759.1
hypothetical protein [Mycobacterium sp.]	Mycobacterium sp.	295	295	65%	2e-89	48.83%	500	MDF2826866.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	301	360	65%	2e-89	59.80%	715	WP_337890912.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	298	298	55%	3e-89	56.07%	601	DAL07358.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	299	359	65%	1e-88	59.67%	715	WP_284829925.1
TPA: tail tape measure [Bacteriophage sp.]	Bacteriophage sp.	298	298	76%	3e-88	43.69%	730	DAS46749.1
hypothetical protein [Actinomyces urogenitalis]	Actinomyces urogenitalis	285	285	95%	8e-86	39.93%	491	WP_278787689.1
hypothetical protein [Winkia]	Winkia	284	284	67%	9e-86	44.25%	462	WP_004806609.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	289	358	100%	1e-85	46.40%	627	WP_201816501.1
hypothetical protein [Propionimicrobium lymphophilum]	Propionimicrobium lymphophilum	283	283	79%	5e-84	40.81%	556	WP_285112190.1
hypothetical protein [Actinomyces urogenitalis]	Actinomyces urogenitalis	285	285	95%	8e-86	39.93%	491	WP_278787689.1
hypothetical protein [Winkia]	Winkia	284	284	67%	9e-86	44.25%	462	WP_004806609.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	289	358	100%	1e-85	46.40%	627	WP_201816501.1
hypothetical protein [Propionimicrobium lymphophilum]	Propionimicrobium lymphophilum	283	283	79%	5e-84	40.81%	556	WP_285112190.1
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
hypothetical protein GCM10023233_27360 [Brevibacterium otitidis]	Brevibacterium otitidis	280	280	79%	5e-83	39.87%	539	BFF07767.1
hypothetical protein [Schaalia georgiae]	Schaalia georgiae	277	277	68%	2e-82	47.45%	495	WP_005867126.1
hypothetical protein [Corynebacterium poyangense]	Corynebacterium poyangense	278	278	64%	2e-82	43.77%	529	WP_187974193.1
phage tail tape measure protein [Corynebacterium oculi]	Corynebacterium oculi	278	338	55%	7e-81	63.11%	680	WP_150114274.1
hypothetical protein [uncultured Leifsonia sp.]	uncultured Leifsonia sp.	273	273	53%	2e-80	52.88%	525	WP_314148240.1
tape measure protein [Arthrobacter phage Swenson]	Arthrobacter phage Swenson	273	273	62%	4e-79	49.14%	665	ASR83993.1
hypothetical protein [Gordonia westfalica]	Gordonia westfalica	269	269	98%	5e-79	37.01%	527	WP_074850009.1
tail length tape measure protein [Gordonia phage EricDab]	Gordonia phage EricDab	266	266	72%	3e-78	41.63%	513	YP_010674657.1
hypothetical protein [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	269	269	67%	5e-78	49.62%	636	WP_064317077.1
hypothetical protein [Gulosibacter bifidus]	Gulosibacter bifidus	266	266	97%	1e-77	38.53%	536	WP_066057107.1
hypothetical protein [Microbacterium sp. A(2022)]	Microbacterium sp. A(2022)	263	263	57%	2e-77	49.54%	477	WP_311878682.1
tail length tape measure protein [Arthrobacter phage Decurro]	Arthrobacter phage Decurro	268	268	62%	3e-77	49.14%	665	YP_009191306.1
tape measure protein [Arthrobacter phage Jessica]	Arthrobacter phage Jessica	268	268	62%	3e-77	49.14%	665	ALF00769.1
tape measure protein [Arthrobacter phage Muttlie]	Arthrobacter phage Muttlie	268	268	62%	4e-77	49.14%	665	ALY09769.1
tape measure protein [Arthrobacter phage Toulouse]	Arthrobacter phage Toulouse	268	268	62%	4e-77	49.14%	665	ALY10652.1
tape measure protein [Arthrobacter phage Moloch]	Arthrobacter phage Moloch	268	268	62%	4e-77	49.14%	665	ALY09743.1
tape measure protein [Arthrobacter phage Elkhorn]	Arthrobacter phage Elkhorn	265	265	62%	3e-76	49.14%	665	ASR83520.1
tape measure protein [Arthrobacter phage KylieMac]	Arthrobacter phage KylieMac	265	265	62%	3e-76	49.14%	665	ASR83547.1
hypothetical protein [Brevibacterium sp. UMB1308A]	Brevibacterium sp. UMB1308A	261	261	73%	1e-75	41.49%	555	WP_284926673.1
hypothetical protein [Microbacterium sp. cx-55]	Microbacterium sp. cx-55	259	259	71%	2e-75	42.45%	524	WP_223721111.1
tape measure protein [Arthrobacter phage Saphira]	Arthrobacter phage Saphira	263	263	62%	3e-75	48.56%	665	QGH75234.1
tape measure protein [Arthrobacter phage Taj14]	Arthrobacter phage Taj14	262	262	62%	5e-75	48.56%	665	ASM62381.1

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

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	Q9ZXA5	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
2	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
3	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
4	Q6QIA5	TMP_BPBMU 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
5	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
6	Q24LI1	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
7	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

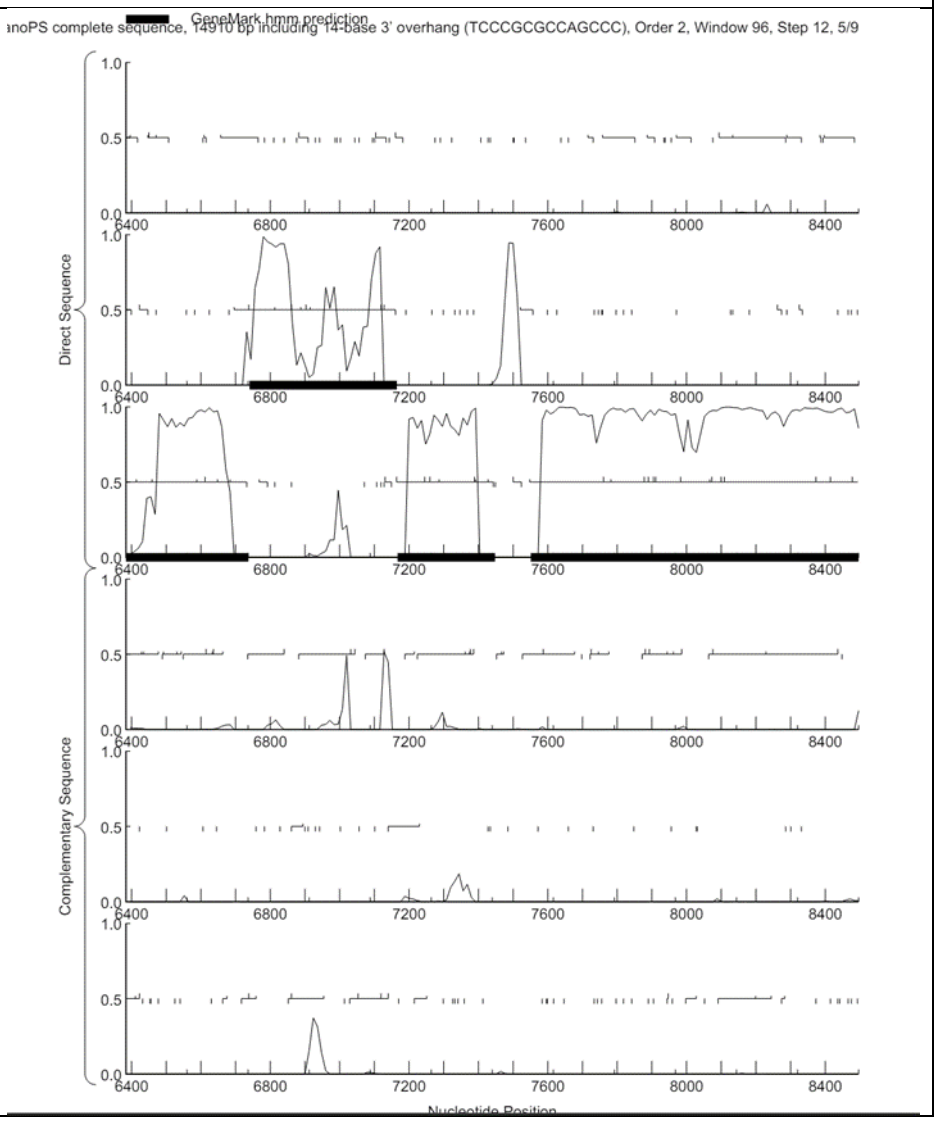
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**NON ?** 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Oui.**
DECISION:	**tape measure protein**

Student Gene Annotation Worksheet

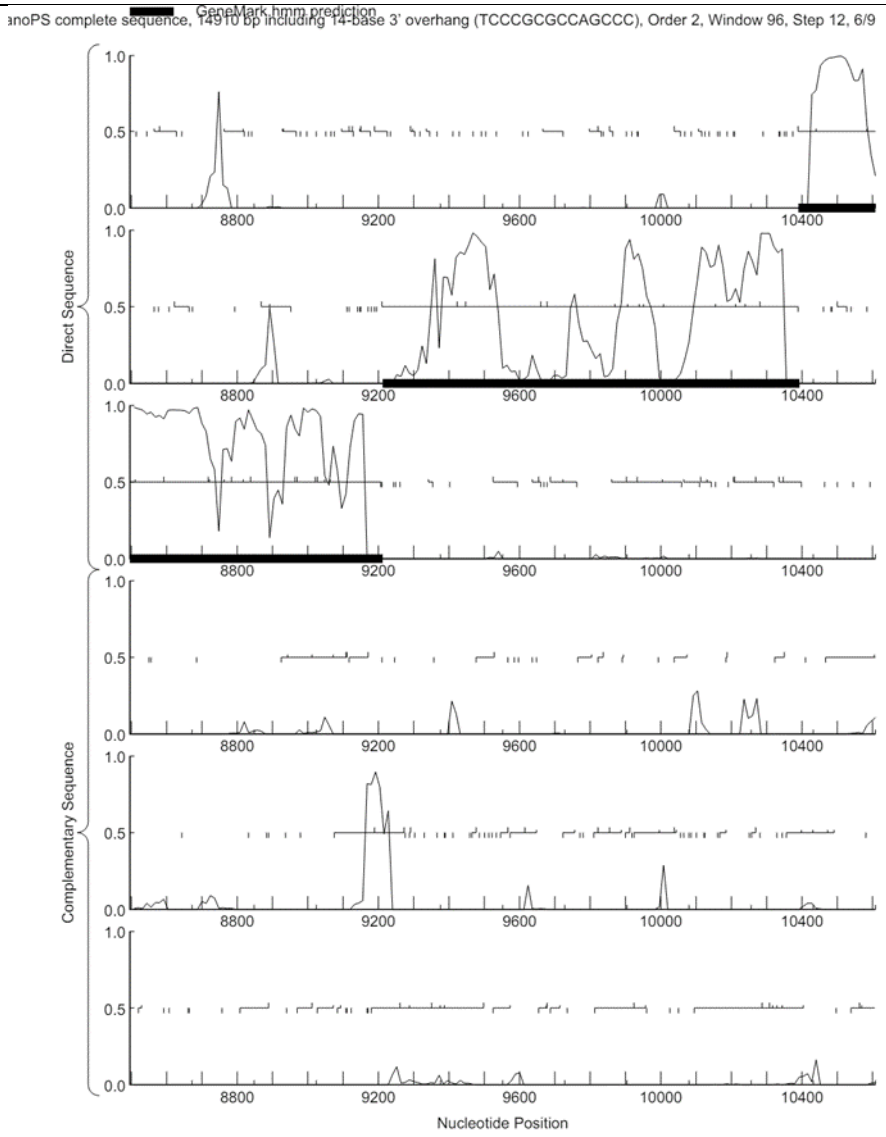
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**13**
Coordonnées du Stop	**9212**
Direction (For/Rev)	**for**
Gap/chevauchement avec un autre gène	**Gap 2**

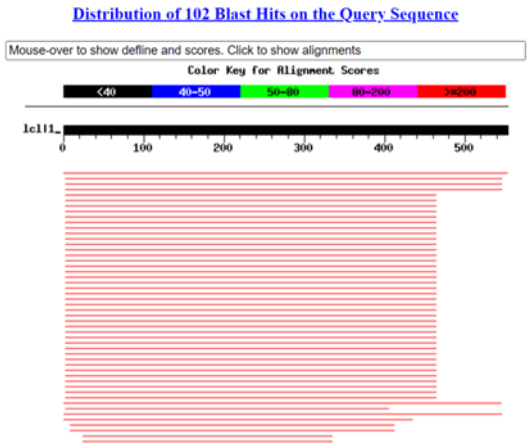
Coordonnées du Start retenu	**7551**
Fonction prédite	**tape measure protein**

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 ?	

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





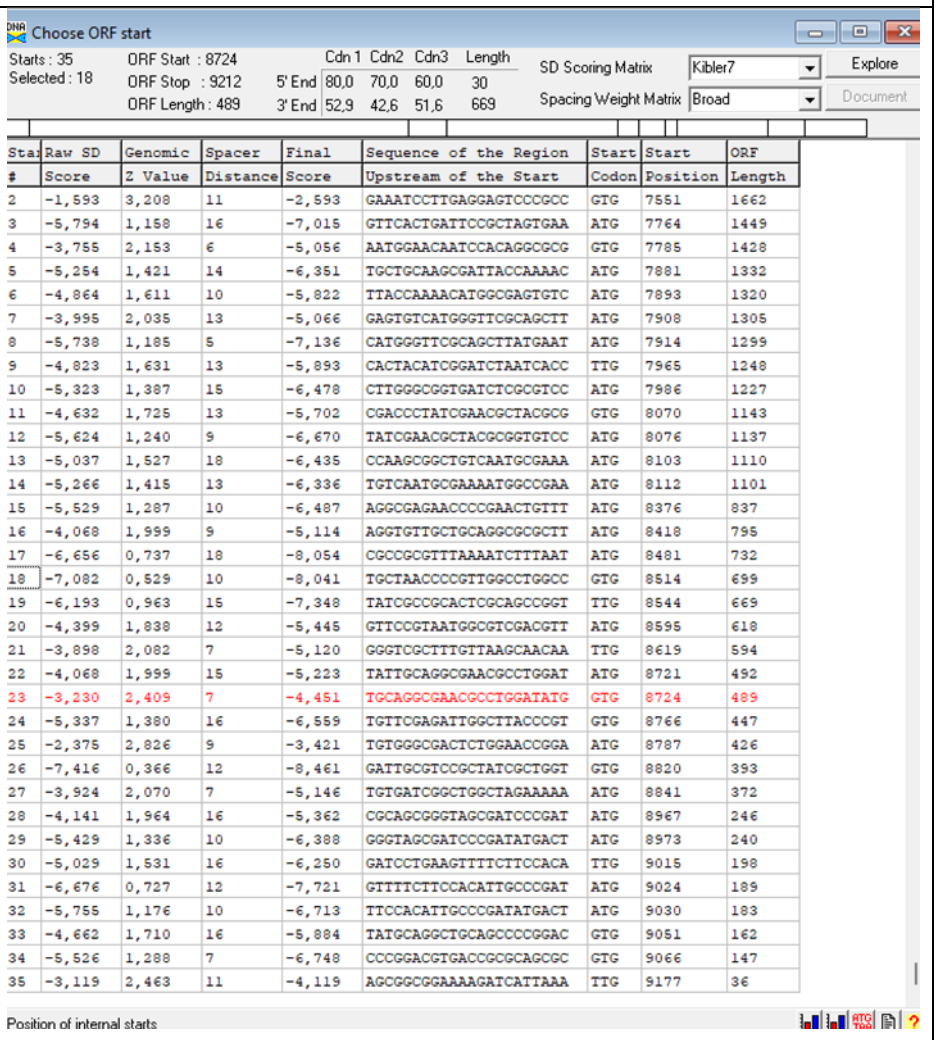
Sequences producing significant alignments:

	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

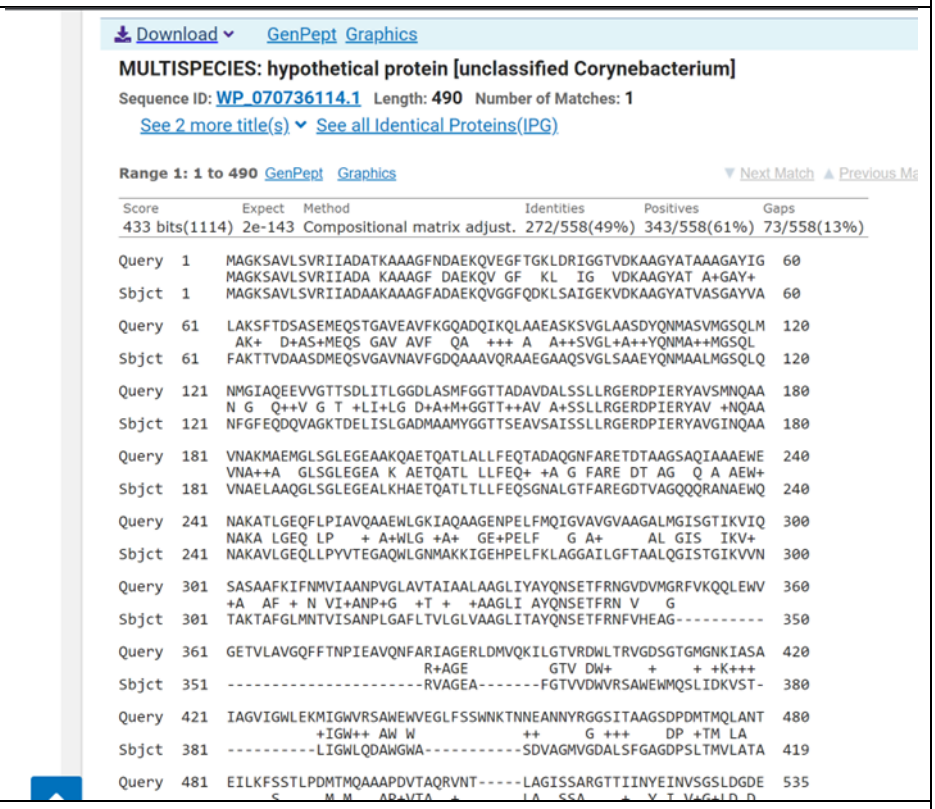
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**NON**
DECISION:	**OUI**

hypothetical protein [Corynebacterium]	Corynebacterium	444	444	100%	1e-147	49.10%	492	WP_070420702.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	436	436	100%	3e-144	48.92%	490	HCT5225727.1
hypothetical protein [unclassified Corynebacterium]	unclassified Corynebacterium	433	433	100%	2e-143	48.75%	490	WP_070736114.1
putative phage protein [Corynebacterium striatum]	Corynebacterium striatum	433	433	100%	3e-143	48.75%	490	CQD13985.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	432	432	100%	7e-143	48.74%	492	HCD1553136.1
hypothetical protein CIP107524_01832 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	391	391	80%	2e-126	53.63%	523	CAB0565259.1
hypothetical protein FRC0370_01709 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	389	389	96%	1e-125	47.54%	523	CAB0870047.1
hypothetical protein FRC0378_01882 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	388	388	96%	3e-125	48.09%	523	CAB0871289.1
hypothetical protein FRC0474_02088 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	388	388	80%	4e-125	53.19%	523	CAB0972304.1
hypothetical protein D9R17_02160 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	387	387	80%	5e-125	53.19%	503	RLP10245.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	387	387	80%	6e-125	53.19%	523	WP_182001102.1
hypothetical protein FRC0430_01824 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	387	387	80%	6e-125	53.19%	523	CAB0913975.1
hypothetical protein [Corynebacterium hadale]	Corynebacterium hadale	386	386	100%	7e-125	46.32%	488	WP_095275383.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	386	386	96%	2e-124	47.72%	523	WP_088298569.1
hypothetical protein [Corynebacterium glaucum]	Corynebacterium glaucum	385	385	100%	2e-124	42.58%	489	WP_301925793.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	385	385	80%	5e-124	52.75%	523	WP_205917723.1
hypothetical protein CIP107521_01967 [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	384	384	96%	1e-123	47.54%	523	CAB0567951.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	385	385	99%	6e-123	39.35%	625	HAT6525551.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	382	382	98%	2e-122	43.49%	554	WP_284594213.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	380	380	98%	1e-121	44.50%	551	WP_284596340.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	378	378	80%	2e-121	53.41%	523	WP_342351204.1
hypothetical protein [Corynebacterium]	Corynebacterium	379	379	99%	3e-121	43.09%	551	WP_064833146.1
hypothetical protein [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudodiphtheriticum	379	379	98%	3e-121	43.67%	554	WP_284849037.1
hypothetical protein [Corynebacterium propinquum]	Corynebacterium propinquum	379	379	98%	4e-121	43.65%	554	WP_302524538.1
hypothetical protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	377	377	80%	5e-121	53.41%	523	WP_106361639.1
hypothetical protein [Winkia]	Winkia	284	284	67%	9e-86	44.25%	462	WP_004806609.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	289	358	100%	1e-85	46.40%	627	WP_201816501.1
hypothetical protein [Propionimicrobium lymphophilum]	Propionimicrobium lymphophilum	283	283	79%	5e-84	40.81%	556	WP_285112190.1
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
hypothetical protein GCM10023233_27360 [Brevibacterium otitidis]	Brevibacterium otitidis	280	280	79%	5e-83	39.87%	539	BFF07767.1
hypothetical protein [Schaalia georgiae]	Schaalia georgiae	277	277	68%	2e-82	47.45%	495	WP_005867126.1
hypothetical protein [Corynebacterium poyangense]	Corynebacterium poyangense	278	278	64%	2e-82	43.77%	529	WP_187974193.1
phage tail tape measure protein [Corynebacterium oculi]	Corynebacterium oculi	278	338	55%	7e-81	63.11%	680	WP_150114274.1
hypothetical protein [uncultured Leifsonia sp.]	uncultured Leifsonia sp.	273	273	53%	2e-80	52.88%	525	WP_314148240.1
tape measure protein [Arthrobacter phage Swenson]	Arthrobacter phage Swenson	273	273	62%	4e-79	49.14%	665	ASR83993.1
hypothetical protein [Gordonia westfalica]	Gordonia westfalica	269	269	98%	5e-79	37.01%	527	WP_074850009.1
tail length tape measure protein [Gordonia phage EricDab]	Gordonia phage EricDab	266	266	72%	3e-78	41.63%	513	YP_010674657.1
hypothetical protein [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	269	269	67%	5e-78	49.62%	636	WP_064317077.1
hypothetical protein [Gulosibacter bifidus]	Gulosibacter bifidus	266	266	97%	1e-77	38.53%	536	WP_066057107.1
hypothetical protein [Microbacterium sp. A(2022)]	Microbacterium sp. A(2022)	263	263	57%	2e-77	49.54%	477	WP_311878682.1
tail length tape measure protein [Arthrobacter phage Decurro]	Arthrobacter phage Decurro	268	268	62%	3e-77	49.14%	665	YP_009191306.1
tape measure protein [Arthrobacter phage Jessica]	Arthrobacter phage Jessica	268	268	62%	3e-77	49.14%	665	ALF00769.1
tape measure protein [Arthrobacter phage Muttlie]	Arthrobacter phage Muttlie	268	268	62%	4e-77	49.14%	665	ALY09769.1
tape measure protein [Arthrobacter phage Toulouse]	Arthrobacter phage Toulouse	268	268	62%	4e-77	49.14%	665	ALY10652.1
tape measure protein [Arthrobacter phage Moloch]	Arthrobacter phage Moloch	268	268	62%	4e-77	49.14%	665	ALY09743.1
tape measure protein [Arthrobacter phage Elkhorn]	Arthrobacter phage Elkhorn	265	265	62%	3e-76	49.14%	665	ASR83520.1
tape measure protein [Arthrobacter phage KylieMac]	Arthrobacter phage KylieMac	265	265	62%	3e-76	49.14%	665	ASR83547.1
hypothetical protein [Brevibacterium sp. UMB1308A]	Brevibacterium sp. UMB1308A	261	261	73%	1e-75	41.49%	555	WP_284926673.1
hypothetical protein [Microbacterium sp. cx-55]	Microbacterium sp. cx-55	259	259	71%	2e-75	42.45%	524	WP_223721111.1
tape measure protein [Arthrobacter phage Saphira]	Arthrobacter phage Saphira	263	263	62%	3e-75	48.56%	665	QGH75234.1

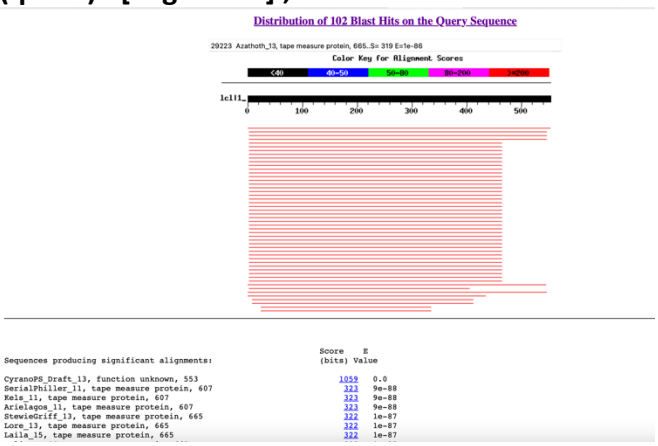
Décision #2 : Quel est le start du gène ?

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) :7551** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :7551**
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 ?	*Plus long orf mais pas avec un ATG* *Meilleur score*
Est-ce que le start est conservé chez les	–

Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
2	-1,593	3,208	11	-2,593	GAAATCCTTGAGGAGTCCCGCC	GTG	7551	1662
3	-5,794	1,158	16	-7,015	GTTCACTGATTCCGCTAGTGAA	ATG	7764	1449
4	-3,755	2,153	6	-5,056	AATGGAACAATCCACAGGCGCG	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	GAGTGTCATGGGTTCGCAGCTT	ATG	7908	1305
8	-5,738	1,185	5	-7,136	CATGGGTTCGCAGCTTATGAAT	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	CGACCCTATCGAACGCTACGCG	GTG	8070	1143
12	-5,624	1,240	9	-6,670	TATCGAACGCTACGCGGTGTCC	ATG	8076	1137
13	-5,037	1,527	18	-6,435	CCAAGCGGCTGTCAATGCGAAA	ATG	8103	1110
14	-5,266	1,415	13	-6,336	TGTCAATGCGAAAATGGCCGAA	ATG	8112	1101
15	-5,529	1,287	10	-6,487	AGGCGAGAACCCCGAACTGTTT	ATG	8376	837
16	-4,068	1,999	9	-5,114	AGGTGTTGCTGCAGGCGCGCTT	ATG	8418	795
17	-6,656	0,737	18	-8,054	CGCCGCGTTTAAAATCTTTAAT	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	GTTCCGTAATGGCGTCGACGTT	ATG	8595	618
21	-3,898	2,082	7	-5,120	GGGTCGCTTTGTTAAGCAACAA	TTG	8619	594
22	-4,068	1,999	15	-5,223	TATTGCAGGCGAACGCCTGGAT	ATG	8721	492
23	-3,230	2,409	7	-4,451	TGCAGGCGAACGCCTGGATATG	GTG	8724	489
24	-5,337	1,380	16	-6,559	TGTTCGAGATTGGCTTACCCGT	GTG	8766	447
25	-2,375	2,826	9	-3,421	TGTGGGCGACTCTGGAACCGGA	ATG	8787	426
26	-7,416	0,366	12	-8,461	GATTGCGTCCGCTATCGCTGGT	GTG	8820	393
27	-3,924	2,070	7	-5,146	TGTGATCGGCTGGCTAGAAAAA	ATG	8841	372
28	-4,141	1,964	16	-5,362	CGCAGCGGGTAGCGATCCCGAT	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	GTTTTCTTCCACATTGCCCGAT	ATG	9024	189
32	-5,755	1,176	10	-6,713	TTCCACATTGCCCGATATGACT	ATG	9030	183
33	-4,662	1,710	16	-5,884	TATGCAGGCTGCAGCCCCGGAC	GTG	9051	162
34	-5,526	1,288	7	-6,748	CCCGGACGTGACCGCGCAGCGC	GTG	9066	147
35	-3,119	2,463	11	-4,119	AGCGGCGGAAAAGATCATTAAA	TTG	9177	36

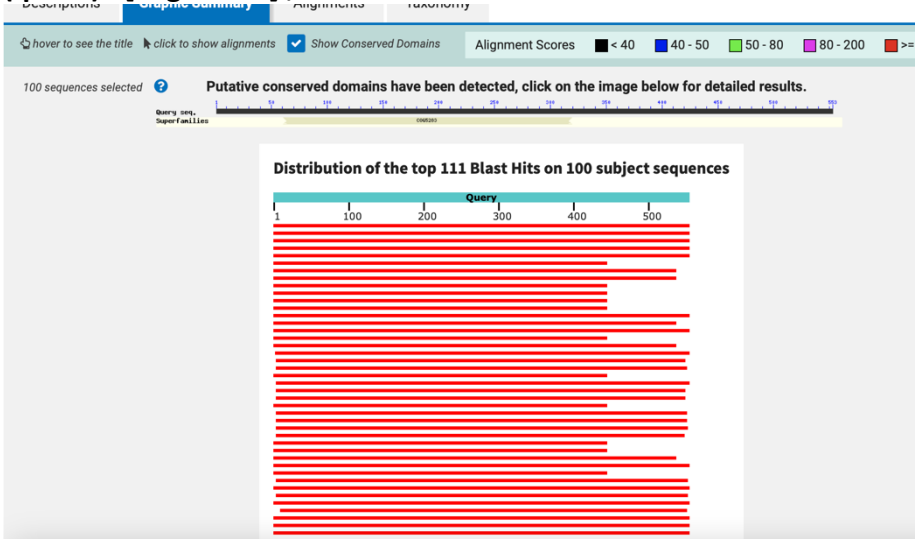
homologues voir Starterator ?	
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	
DECISION:	**7551**

Décision #3 : Quelle est la function 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 :** 

***** nr :**

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

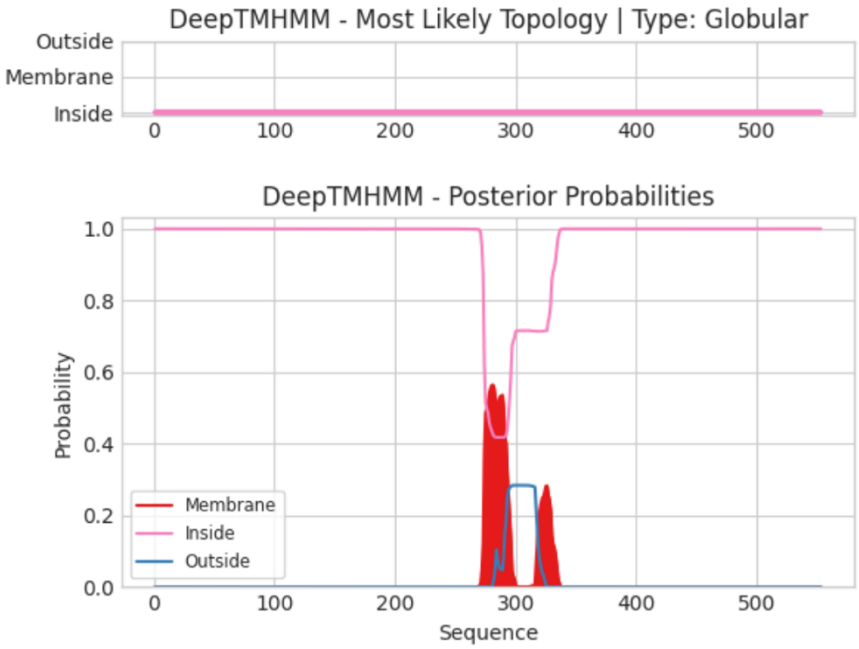


TPA: tail tape measure [Siphoviridae sp. ctCCv12]	Siphoviridae sp. ctCCv12	327	327	70%	2e-102	47.04%	466	DAD89990.1
hypothetical protein [Corynebacterium matruchotii]	Corynebacterium matruchotii	321	321	70%	7e-100	46.27%	466	WP_126299860.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	308	308	49%	2e-96	63.27%	333	WP_302526992.1
hypothetical protein [Schaalia hyovaginalis]	Schaalia hyovaginalis	311	311	64%	1e-95	50.95%	489	WP_320772602.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	308	308	54%	1e-95	53.38%	412	DAO68536.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	307	307	60%	3e-95	50.00%	414	DAK55025.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	307	307	60%	3e-95	50.00%	414	DAD60080.1
hypothetical protein [Schaalia hyovaginalis]	Schaalia hyovaginalis	308	308	64%	1e-94	50.41%	489	WP_303843687.1
hypothetical protein [Actinomyces polynesiensis]	Actinomyces polynesiensis	308	308	68%	2e-94	48.21%	491	WP_052450542.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	303	303	56%	9e-94	51.56%	412	DAU29185.1
hypothetical protein [Corynebacterium freiburgense]	Corynebacterium freiburgense	307	307	78%	5e-92	44.75%	675	WP_051255731.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	304	364	78%	8e-92	56.51%	587	HAT1331008.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	303	364	78%	1e-91	56.51%	597	HAT1419171.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	305	305	54%	1e-91	55.45%	628	WP_284900637.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	303	364	78%	1e-91	56.51%	598	HAT1137141.1
hypothetical protein [Corynebacterium accolens]	Corynebacterium accolens	304	357	81%	2e-91	55.45%	628	WP_284895773.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	301	362	99%	1e-90	56.19%	616	HAT6642583.1
hypothetical protein Cocul_00066 [Corynebacterium oculi]	Corynebacterium oculi	303	363	65%	3e-90	60.13%	715	KQB84936.1
TPA: hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	300	300	53%	3e-90	58.11%	598	HAT1153070.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	303	362	65%	3e-90	59.80%	713	WP_156801759.1
hypothetical protein [Mycobacterium sp.]	Mycobacterium sp.	295	295	65%	2e-89	48.83%	500	MDF2826866.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	301	360	65%	2e-89	59.80%	715	WP_337890912.1
TPA: tail tape measure [Caudoviricetes sp.]	Caudoviricetes sp.	298	298	55%	3e-89	56.07%	601	DAL07358.1
hypothetical protein [Corynebacterium mastitidis]	Corynebacterium mastitidis	299	359	65%	1e-88	59.67%	715	WP_284829925.1
TPA: tail tape measure [Bacteriophage sp.]	Bacteriophage sp.	298	298	76%	3e-88	43.69%	730	DAS46749.1
hypothetical protein [Actinomyces urogenitalis]	Actinomyces urogenitalis	285	285	95%	8e-86	39.93%	491	WP_278787689.1
hypothetical protein [Winkia]	Winkia	284	284	67%	9e-86	44.25%	462	WP_004806609.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	289	358	100%	1e-85	46.40%	627	WP_201816501.1
hypothetical protein [Propionimicrobium lymphophilum]	Propionimicrobium lymphophilum	283	283	79%	5e-84	40.81%	556	WP_285112190.1
hypothetical protein [Actinomyces urogenitalis]	Actinomyces urogenitalis	285	285	95%	8e-86	39.93%	491	WP_278787689.1
hypothetical protein [Winkia]	Winkia	284	284	67%	9e-86	44.25%	462	WP_004806609.1
hypothetical protein [Corynebacterium striatum]	Corynebacterium striatum	289	358	100%	1e-85	46.40%	627	WP_201816501.1
hypothetical protein [Propionimicrobium lymphophilum]	Propionimicrobium lymphophilum	283	283	79%	5e-84	40.81%	556	WP_285112190.1
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
hypothetical protein GCM10023233_27360 [Brevibacterium otitidis]	Brevibacterium otitidis	280	280	79%	5e-83	39.87%	539	BFF07767.1
hypothetical protein [Schaalia georgiae]	Schaalia georgiae	277	277	68%	2e-82	47.45%	495	WP_005867126.1
hypothetical protein [Corynebacterium poyangense]	Corynebacterium poyangense	278	278	64%	2e-82	43.77%	529	WP_187974193.1
phage tail tape measure protein [Corynebacterium oculi]	Corynebacterium oculi	278	338	55%	7e-81	63.11%	680	WP_150114274.1
hypothetical protein [uncultured Leifsonia sp.]	uncultured Leifsonia sp.	273	273	53%	2e-80	52.88%	525	WP_314148240.1
tape measure protein [Arthrobacter phage Swenson]	Arthrobacter phage Swenson	273	273	62%	4e-79	49.14%	665	ASR83993.1
hypothetical protein [Gordonia westfalica]	Gordonia westfalica	269	269	98%	5e-79	37.01%	527	WP_074850009.1
tail length tape measure protein [Gordonia phage EricDab]	Gordonia phage EricDab	266	266	72%	3e-78	41.63%	513	YP_010674657.1
hypothetical protein [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	269	269	67%	5e-78	49.62%	636	WP_064317077.1
hypothetical protein [Gulosibacter bifidus]	Gulosibacter bifidus	266	266	97%	1e-77	38.53%	536	WP_066057107.1
hypothetical protein [Microbacterium sp. A(2022)]	Microbacterium sp. A(2022)	263	263	57%	2e-77	49.54%	477	WP_311878682.1
tail length tape measure protein [Arthrobacter phage Decurro]	Arthrobacter phage Decurro	268	268	62%	3e-77	49.14%	665	YP_009191306.1
tape measure protein [Arthrobacter phage Jessica]	Arthrobacter phage Jessica	268	268	62%	3e-77	49.14%	665	ALF00769.1
tape measure protein [Arthrobacter phage Muttlie]	Arthrobacter phage Muttlie	268	268	62%	4e-77	49.14%	665	ALY09769.1
tape measure protein [Arthrobacter phage Toulouse]	Arthrobacter phage Toulouse	268	268	62%	4e-77	49.14%	665	ALY10652.1
tape measure protein [Arthrobacter phage Moloch]	Arthrobacter phage Moloch	268	268	62%	4e-77	49.14%	665	ALY09743.1
tape measure protein [Arthrobacter phage Elkhorn]	Arthrobacter phage Elkhorn	265	265	62%	3e-76	49.14%	665	ASR83520.1
tape measure protein [Arthrobacter phage KylieMac]	Arthrobacter phage KylieMac	265	265	62%	3e-76	49.14%	665	ASR83547.1
hypothetical protein [Brevibacterium sp. UMB1308A]	Brevibacterium sp. UMB1308A	261	261	73%	1e-75	41.49%	555	WP_284926673.1
hypothetical protein [Microbacterium sp. cx-55]	Microbacterium sp. cx-55	259	259	71%	2e-75	42.45%	524	WP_223721111.1
tape measure protein [Arthrobacter phage Saphira]	Arthrobacter phage Saphira	263	263	62%	3e-75	48.56%	665	QGH75234.1
tape measure protein [Arthrobacter phage Taj14]	Arthrobacter phage Taj14	262	262	62%	5e-75	48.56%	665	ASM62381.1

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

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	Q9ZXA5	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
2	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
3	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
4	Q6QIA5	TMP_BPBMU 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
5	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
6	Q24LI1	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
7	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

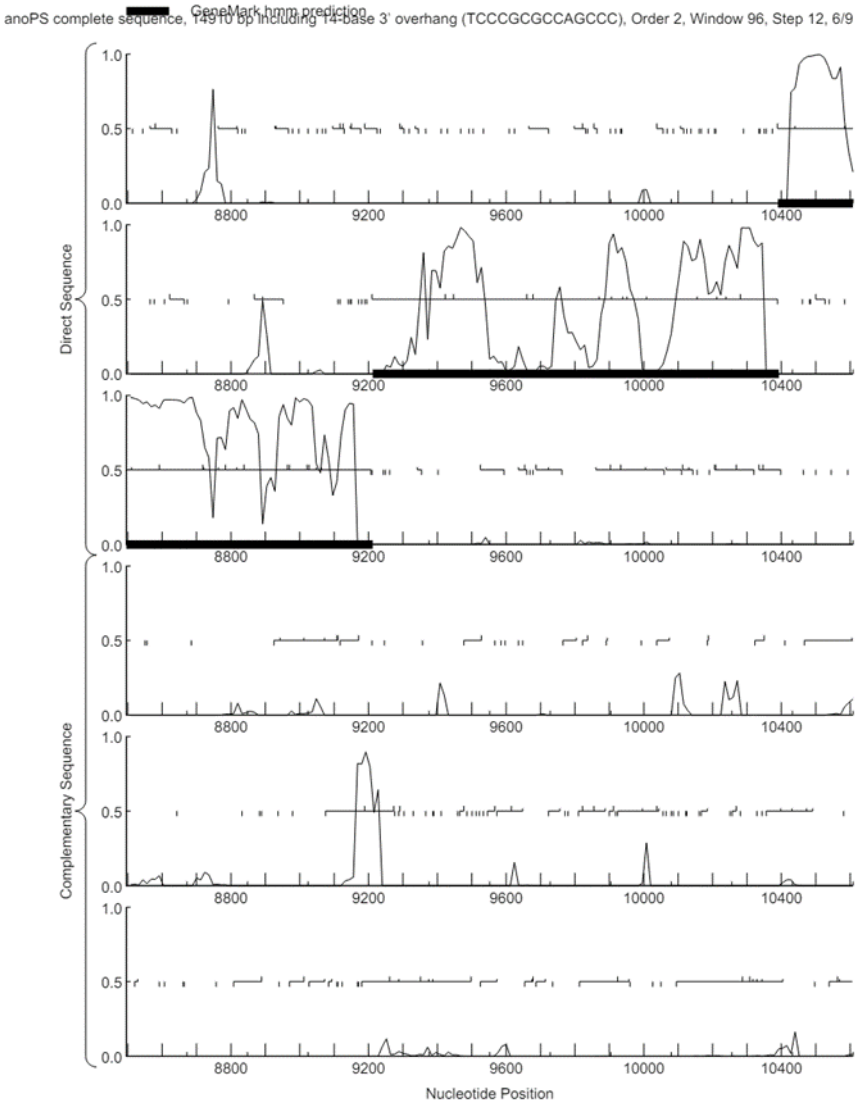
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**NON ?** 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Oui.**
DECISION:	**tape measure protein**

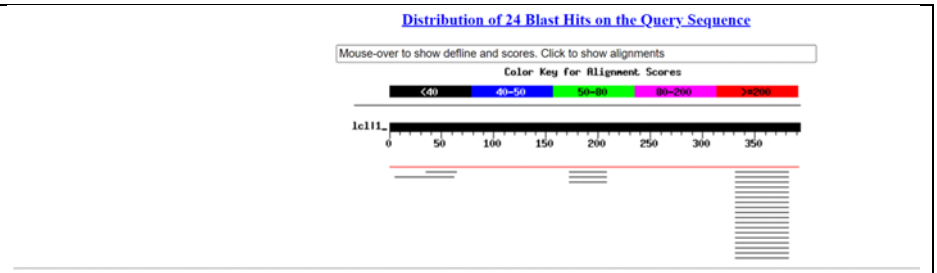
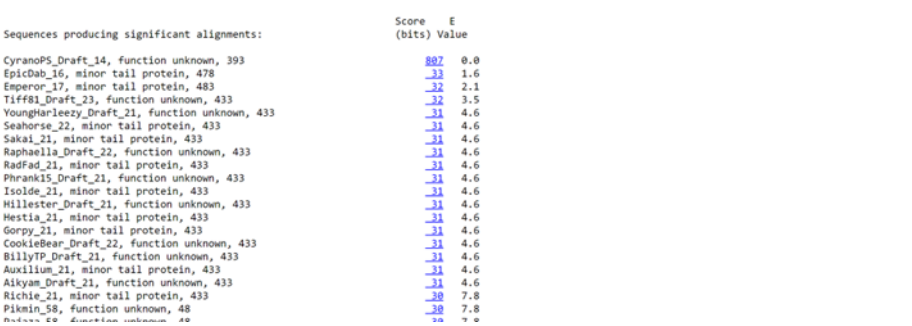
Student Gene Annotation Worksheet

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**

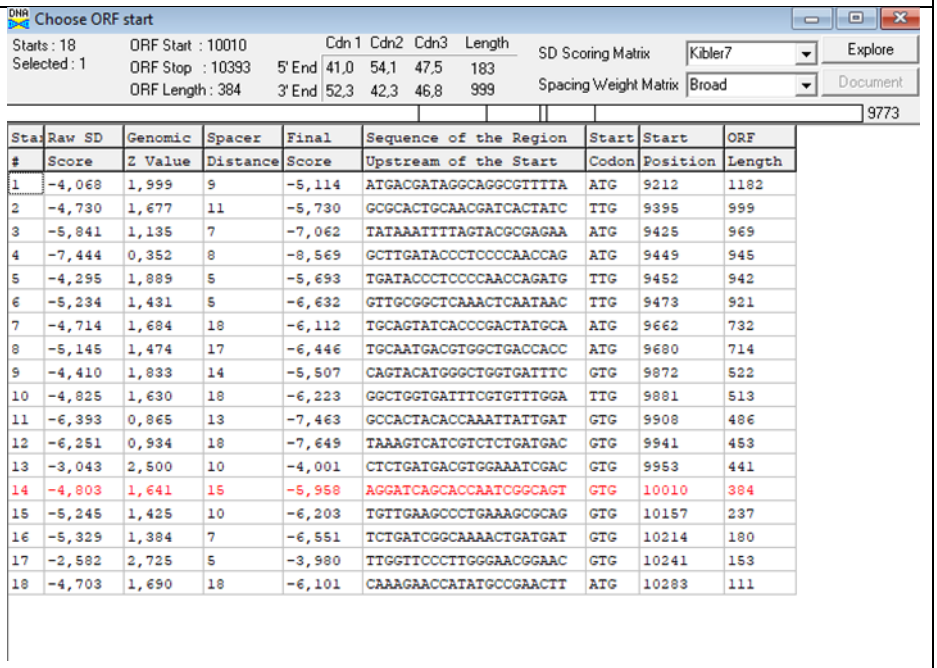
Fonction prédite	**Minor tail protein**

Décision #1 : Est-ce un gène ?

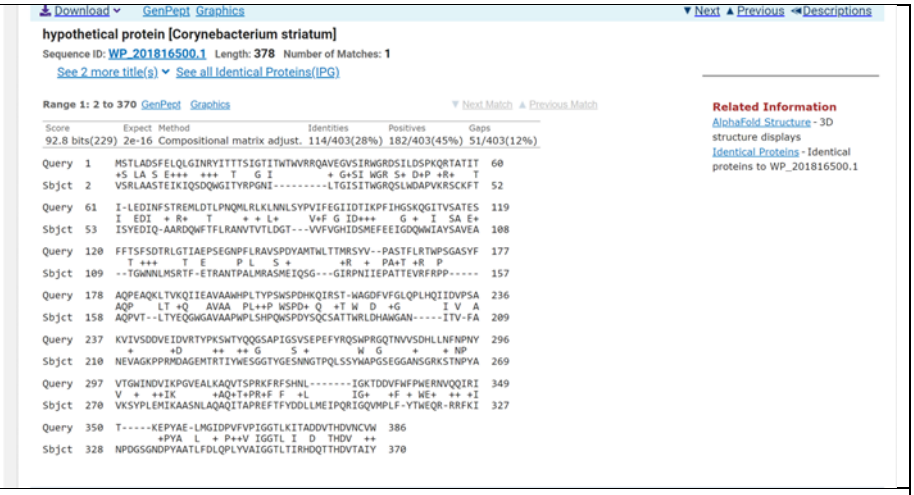
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
Y-a-t-il des éléments supportant un potentiel codant ?	

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	  TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 84.3 84.3 91% 2e-13 26.23% 408 DAS66265.1 TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 84.0 84.0 91% 2e-13 25.97% 408 DAS95148.1 TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 84.0 84.0 91% 2e-13 25.65% 408 DAW81670.1 TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 83.6 83.6 91% 3e-13 25.97% 409 DAK21463.1 TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 83.6 83.6 91% 4e-13 25.71% 408 DAJ35125.1 TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 83.6 83.6 91% 4e-13 25.72% 408 DAV44110.1 TPA: actin-like protein [Caudoviricetes sp.] Caudoviricetes sp. 83.6 83.6 91% 4e-13 25.71% 411 DAW69952.1 TPA: actin-like protein [Caudoviricetes sp.] Caudoviricetes sp. 83.2 83.2 91% 4e-13 25.65% 412 DAP14260.1 TPA: hypothetical protein [Caudoviricetes sp.] Caudoviricetes sp. 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 sp.] Caudoviricetes sp. 82.0 82.0 91% 1e-12 25.97% 378 DAQ14881.1 hypothetical protein [Actinomyces graevenitzii] Actinomyces graevenitzii 81.6 81.6 91% 1e-12 26.49% 378 MBF0930126.1
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**NON**
DECISION:	**OUI**

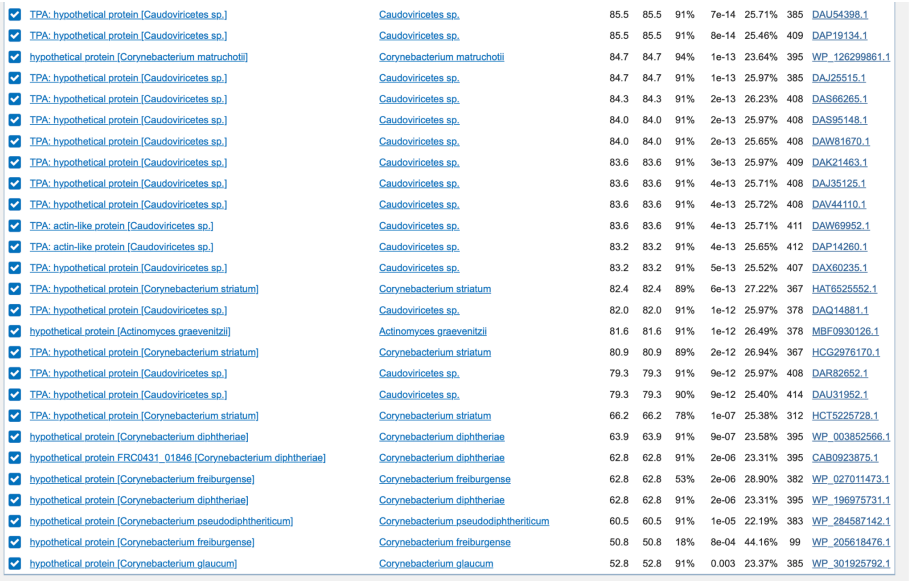
Décision #2 : Quel est le start du gène ?

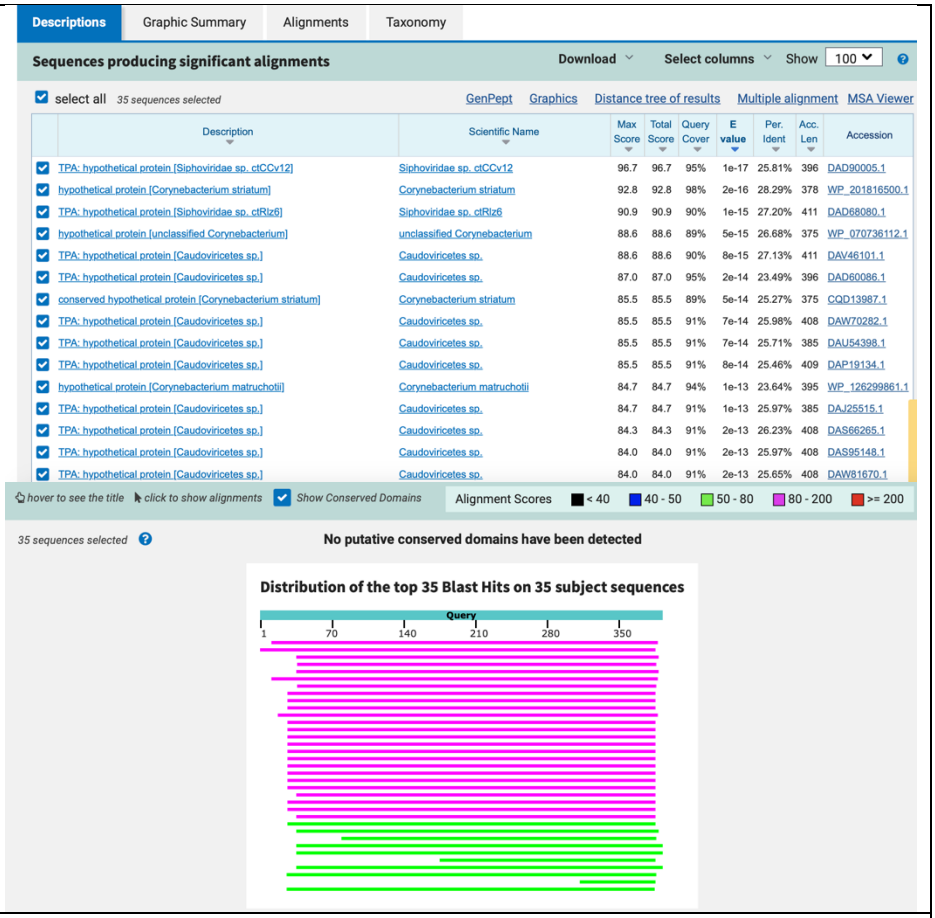
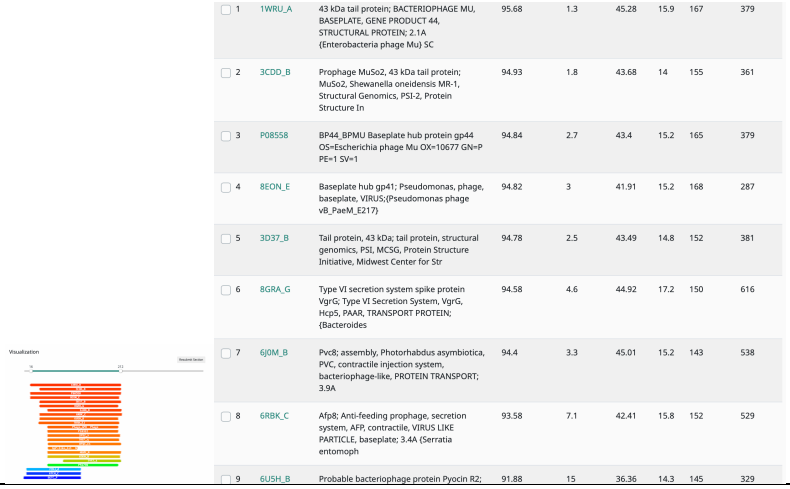
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :9212** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :9212**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	***ORF le plus long*** ***ATG*** ***Score : -5***
Est-ce que le start est conservé chez les homologues voir Starterator ?	–

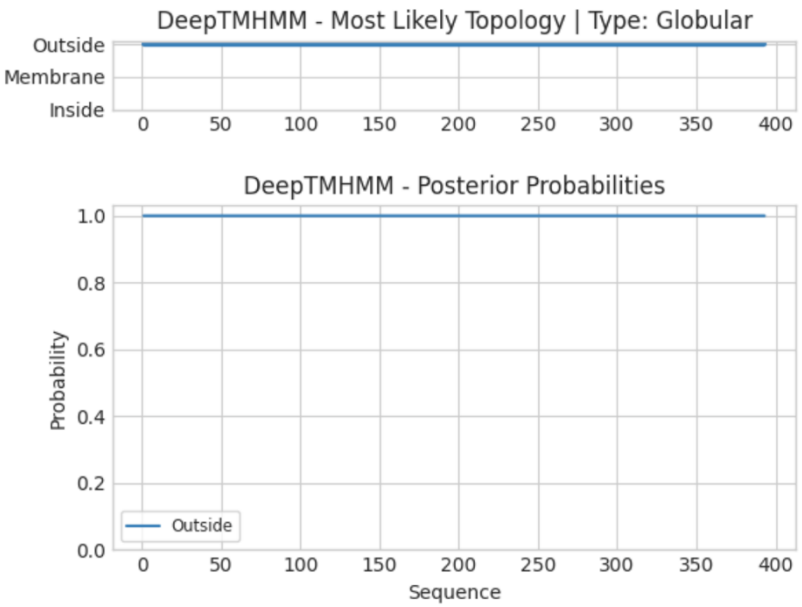
Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-4,068	1,999	9	-5,114	ATGACGATAGGCAGGCGTTTTA	ATG	9212	1182
2	-4,730	1,677	11	-5,730	GCGCACTGCAACGATCACTATC	TTG	9395	999
3	-5,841	1,135	7	-7,062	TATAAATTTTAGTACGCGAGAA	ATG	9425	969
4	-7,444	0,352	8	-8,569	GCTTGATACCCTCCCCAACCAG	ATG	9449	945
5	-4,295	1,889	5	-5,693	TGATACCCTCCCCAACCAGATG	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	CAGTACATGGGCTGGTGATTTC	GTG	9872	522
10	-4,825	1,630	18	-6,223	GGCTGGTGATTTCGTGTTTGGA	TTG	9881	513
11	-6,393	0,865	13	-7,463	GCCACTACACCAAATTATTGAT	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	AGGATCAGCACCAATCGGCAGT	GTG	10010	384
15	-5,245	1,425	10	-6,203	TGTTGAAGCCCTGAAAGCGCAG	GTG	10157	237
16	-5,329	1,384	7	-6,551	TCTGATCGGCAAAACTGATGAT	GTG	10214	180
17	-2,582	2,725	5	-3,980	TTGGTTCCCTTGGGAACGGAAC	GTG	10241	153
18	-4,703	1,690	18	-6,101	CAAAGAACCATATGCCGAACTT	ATG	10283	111

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	
DECISION:	**9212**

Décision #3 : Quelle est la function 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 :** **RIEN** ***** nr :** 

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

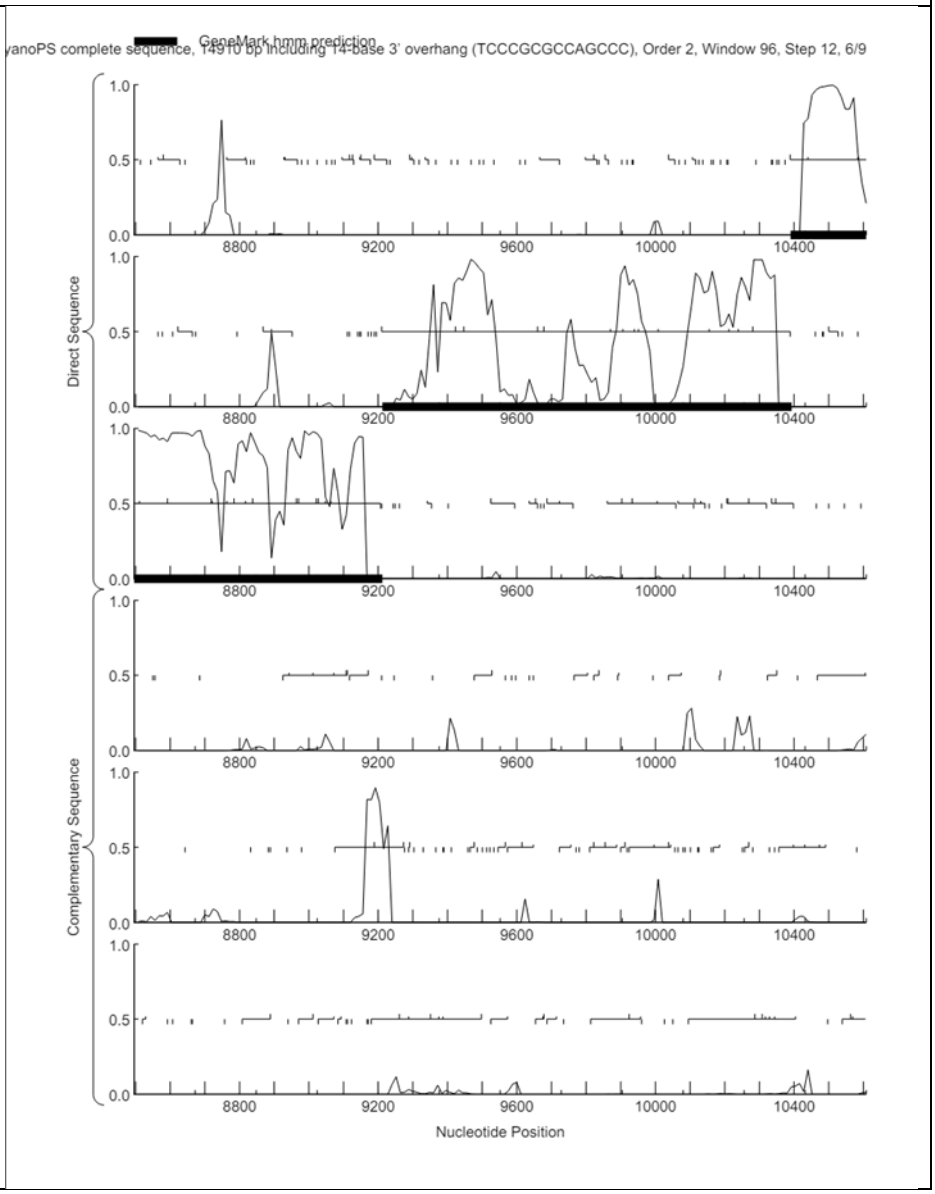
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**non**
DECISION:	**Minor tail protein**

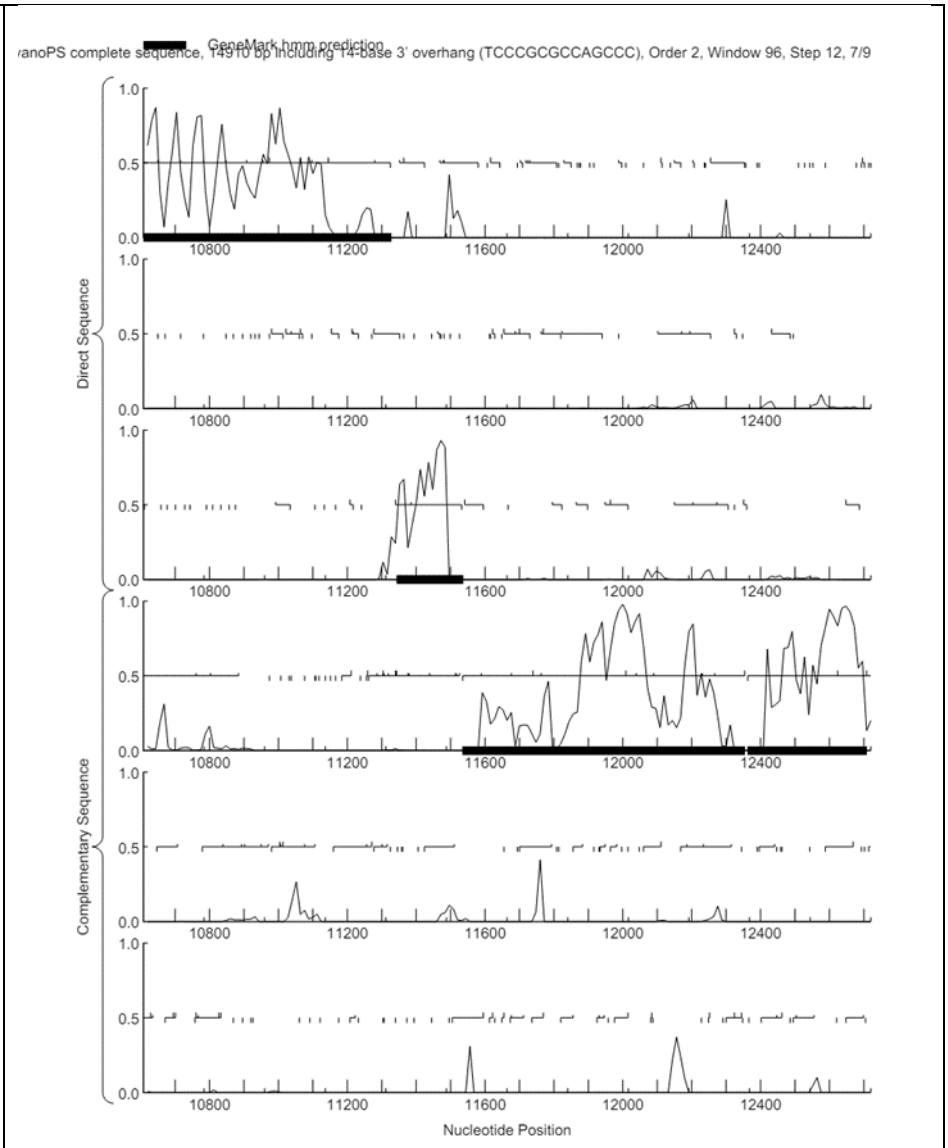
Student Gene Annotation Worksheet

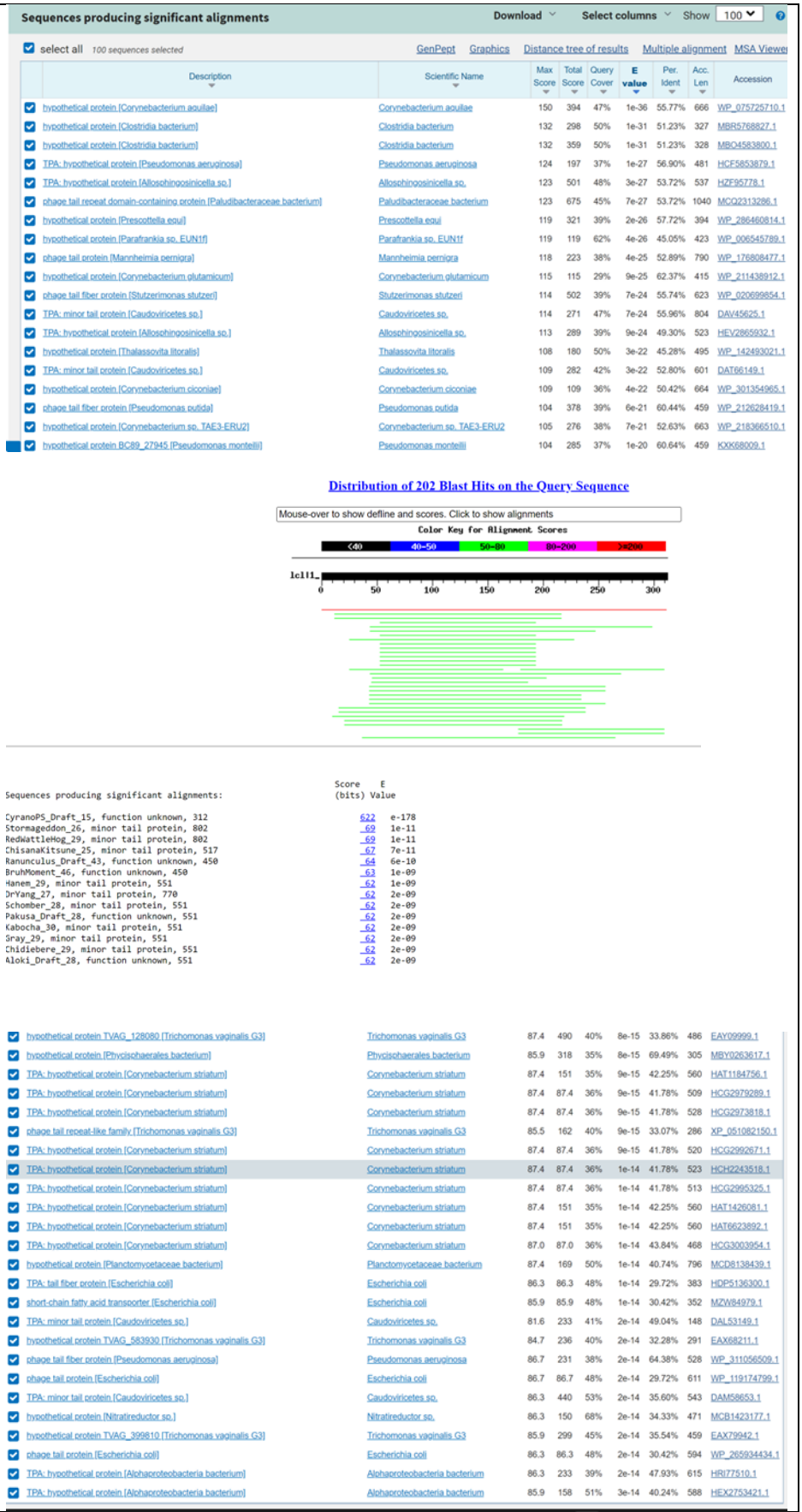
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**15**
Coordonnées du Stop	**11328**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Ove 4**

Coordonnées du Start retenu	**10390**
Fonction prédite	**Minor tail protein**

Décision #1 : Est-ce un gène ?

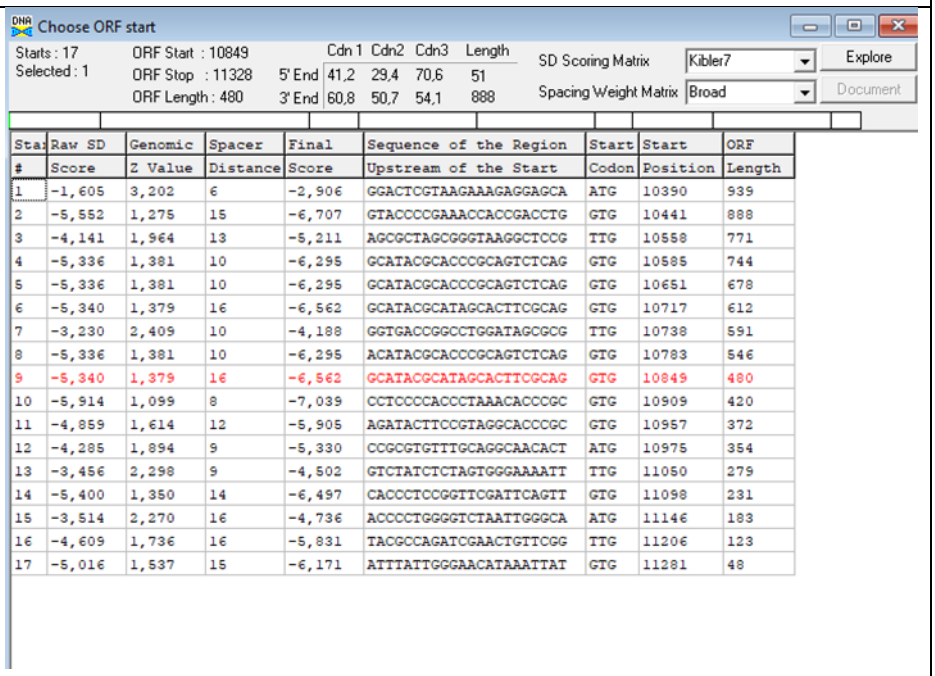
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
Y-a-t-il des éléments supportant un potentiel codant ?	


GeneMark hmm prediction
anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 7/9
Direct Sequence
Complementary Sequence
1.0
0.5
0.0
10800
11200
11600
12000
12400
Nucleotide Position

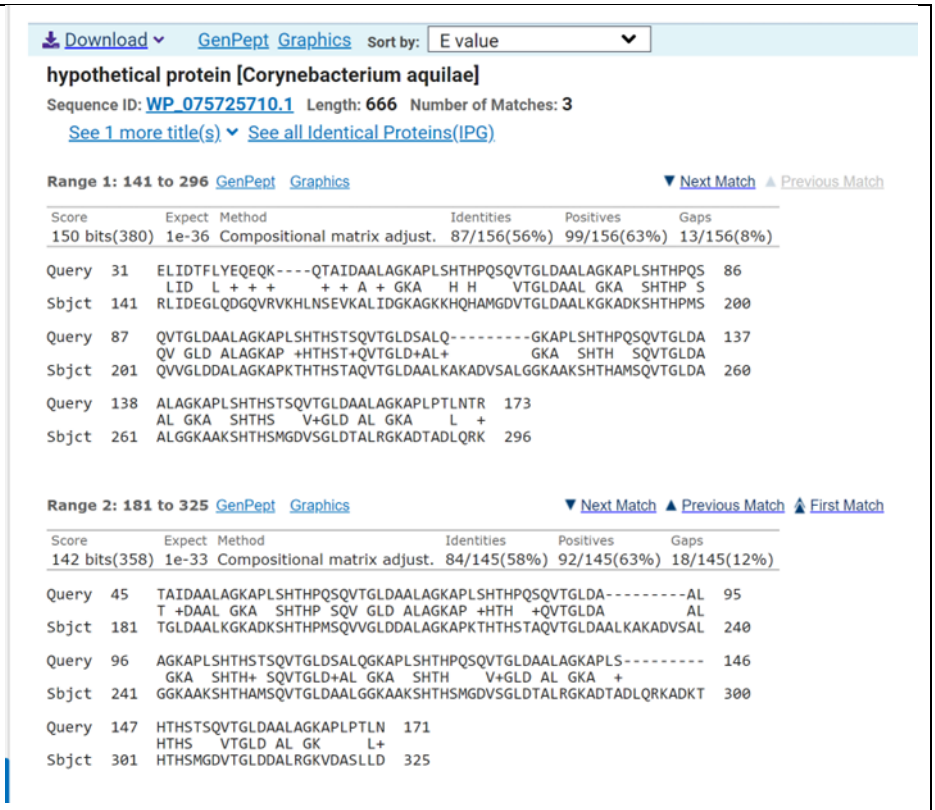
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**NON**

DECISION:	**OUI**

Décision #2 : Quel est le start du gène ?

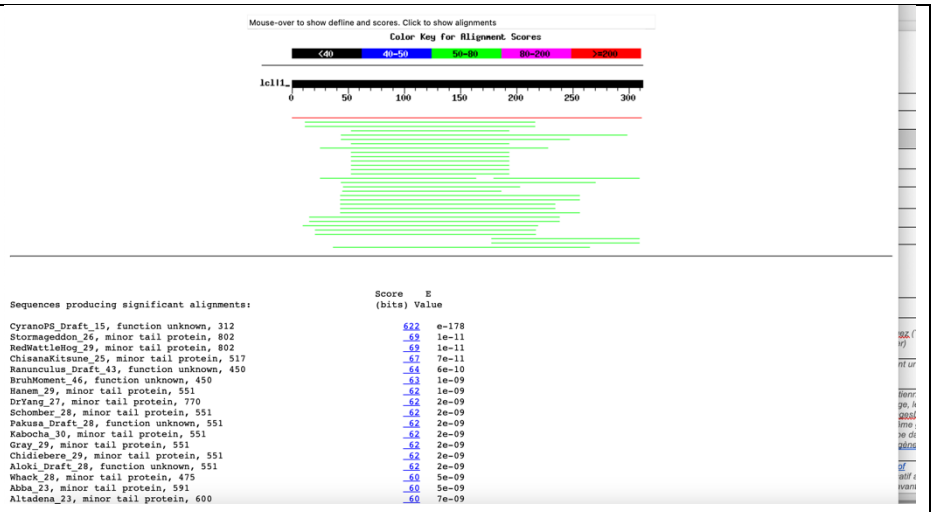
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) :10390** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :10390**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	***ORF le plus long (avec un ATG)*** ***Score de -2***
Est-ce que le start est conservé chez les homologues voir Starterator ?	–

Sta #	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	-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	GCATACGCACCCGCAGTCTCAG	GTG	10585	744
5	-5,336	1,381	10	-6,295	GCATACGCACCCGCAGTCTCAG	GTG	10651	678
6	-5,340	1,379	16	-6,562	GCATACGCATAGCACTTCGCAG	GTG	10717	612
7	-3,230	2,409	10	-4,188	GGTGACCGGCCTGGATAGCGCG	TTG	10738	591
8	-5,336	1,381	10	-6,295	ACATACGCACCCGCAGTCTCAG	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	AGATACTTCCGTAGGCACCCGC	GTG	10957	372
12	-4,285	1,894	9	-5,330	CCGCGTGTTTGCAGGCAACACT	ATG	10975	354
13	-3,456	2,298	9	-4,502	GTCTATCTCTAGTGGGAAAATT	TTG	11050	279
14	-5,400	1,350	14	-6,497	CACCCTCCGGTTCGATTCAGTT	GTG	11098	231
15	-3,514	2,270	16	-4,736	ACCCCTGGGGTCTAATTGGGCA	ATG	11146	183
16	-4,609	1,736	16	-5,831	TACGCCAGATCGAACTGTTCGG	TTG	11206	123
17	-5,016	1,537	15	-6,171	ATTTATTGGGAACATAAATTAT	GTG	11281	48

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	 Download GenPept Graphics Sort by: E value **hypothetical protein [Corynebacterium aquilae]** Sequence ID: WP_075725710.1 Length: 666 Number of Matches: 3 See 1 more title(s) See all Identical Proteins(IPG) **Range 1: 141 to 296** GenPept Graphics ▼ Next Match ▲ Previous Match Score Expect Method Identities Positives Gaps 150 bits(380) 1e-36 Compositional matrix adjust. 87/156(56%) 99/156(63%) 13/156(8%) Query 31 ELIDTFLYEQEQK----QTAIDAALAGKAPLSHTHPQSQVTGLDAALAGKAPLSHTHPQS 86 LID L + + + + + A + GKA H H VTGLDAAL GKA SHTHP S Sbjct 141 RLIDEGLQDGQVRVKHLNSEVKALIDGKAGKKHQHAMGDVTGLDAALKGKADKSHTHPMS 200 Query 87 QVTGLDAALAGKAPLSHTHSTSQVTGLDSALQ---------GKAPLSHTHPQSQVTGLDA 137 QV GLD ALAGKAP +HTHST+QVTGLD+AL+ GKA SHTH SQVTGLDA Sbjct 201 QVVGLDDALAGKAPKTHTHSTAQVTGLDAALKAKADVSALGGKAAKSHTHAMSQVTGLDA 260 Query 138 ALAGKAPLSHTHSTSQVTGLDAALAGKAPLPTLNTR 173 AL GKA SHTHS V+GLD AL GKA L + Sbjct 261 ALGGKAAKSHTHSMGDVSGLDTALRGKADTADLQRK 296 **Range 2: 181 to 325** GenPept Graphics ▼ Next Match ▲ Previous Match First Match Score Expect Method Identities Positives Gaps 142 bits(358) 1e-33 Compositional matrix adjust. 84/145(58%) 92/145(63%) 18/145(12%) Query 45 TAIDAALAGKAPLSHTHPQSQVTGLDAALAGKAPLSHTHPQSQVTGLDA---------AL 95 T +DAAL GKA SHTHP SQV GLD ALAGKAP +HTH +QVTGLDA AL Sbjct 181 TGLDAALKGKADKSHTHPMSQVVGLDDALAGKAPKTHTHSTAQVTGLDAALKAKADVSAL 240 Query 96 AGKAPLSHTHSTSQVTGLDSALQGKAPLSHTHPQSQVTGLDAALAGKAPLS--------- 146 GKA SHTH+ SQVTGLD+AL GKA SHTH V+GLD AL GKA + Sbjct 241 GGKAAKSHTHAMSQVTGLDAALGGKAAKSHTHSMGDVSGLDTALRGKADTADLQRKADKT 300 Query 147 HTHSTSQVTGLDAALAGKAPLPTLN 171 HTHS VTGLD AL GK L+ Sbjct 301 HTHSMGDVTGLDDALRGKVDASLLD 325 **Remarque : si vous proposez plusieurs starts, indiquez pourquoi et fournissez l'information pour chaque start.**
DECISION:	**10390**

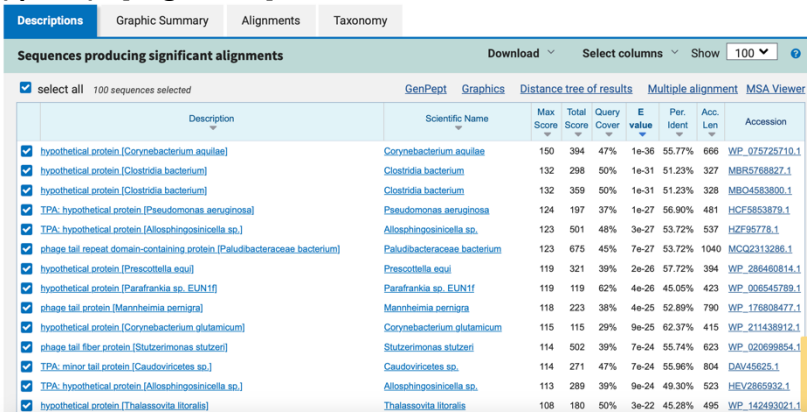
Décision #3 : Quelle est la function de la protéine putative ?

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 :**



***** nr :**

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

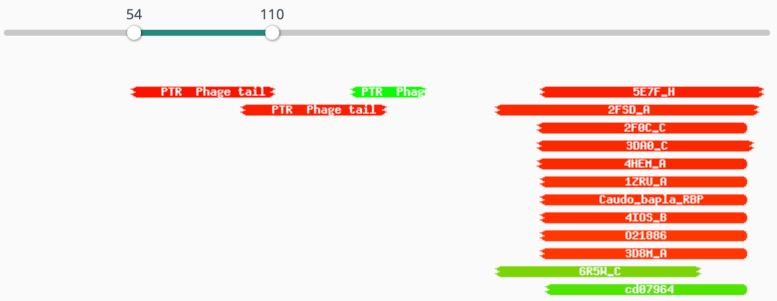


Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	**Pfam : PTR ; Phage tail repeat like** **Cd proba de 75 % : RBP-H; Head domain of virus receptor-binding 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**

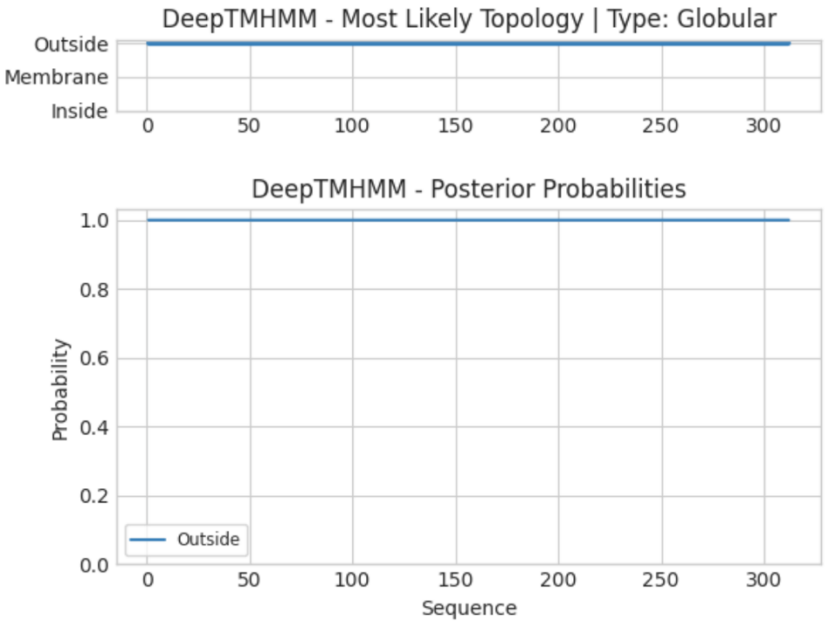
1	PF12789.11	; PTR ; Phage tail repeat like	97.95	0.000013	55.49	2.3	57	60
2	PF12789.11	; PTR ; Phage tail repeat like	97.6	0.00013	50.43	2.9	58	60
3	5E7F_H	Major structural protein 1; bacteriophages, Lactococcus lactis, Siphoviridae, receptor binding protein, single-chain nan	97.38	0.0034	54.48	9.3	84	174
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
5	2F0C_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
6	3DA0_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
7	4HEM_A	BPP; Alpha-beta, Phage Receptor Binding Protein, Llama vHH, viral protein, CELL ADHESION-IMMUNE SYSTEM complex; 1.65A {L	96.2	0.1	44.52	9.1	80	163
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
9	PF08931.14	; Caudo_bapla_RBP ; Lactophage receptor-binding protein C-terminal head domain	95.73	0.26	37.96	8.8	79	101
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

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 ?	**"Aucune synténie observée".**

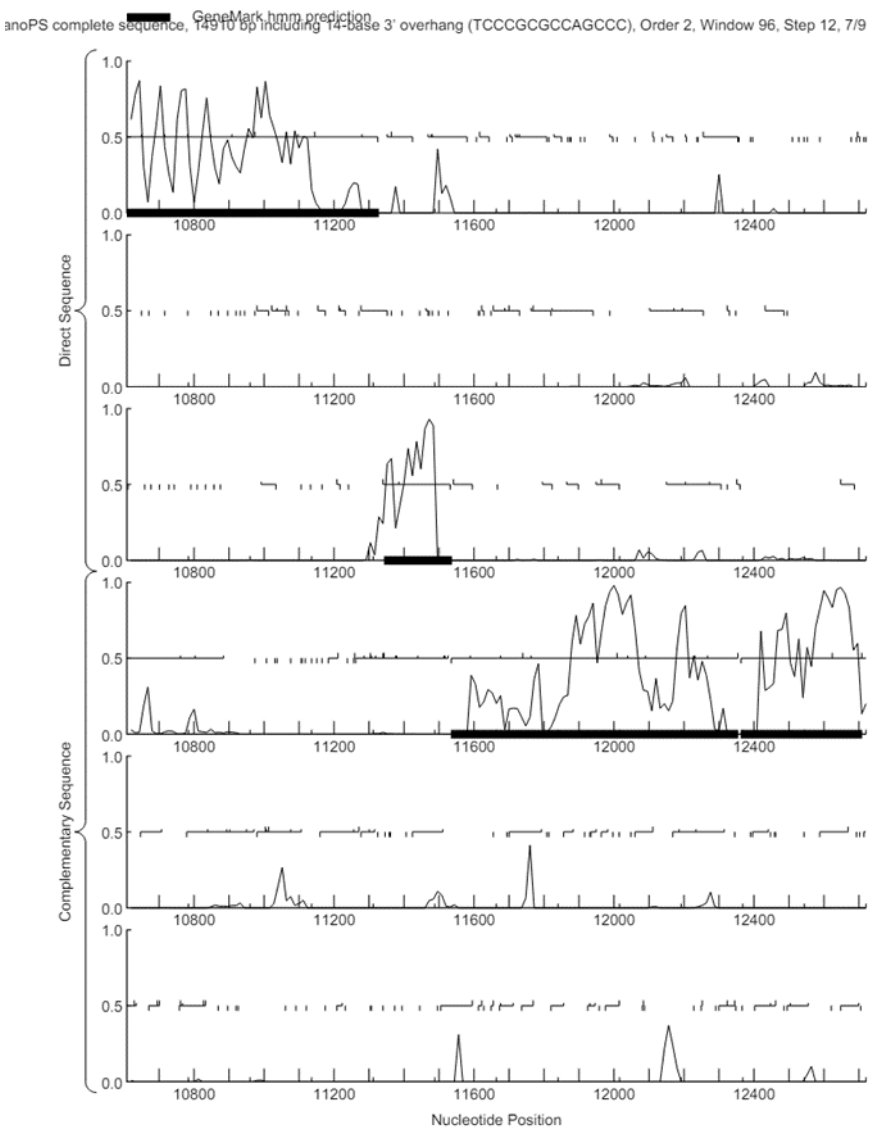
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**oui**
DECISION:	**Minor tail protein**

Student Gene Annotation Worksheet

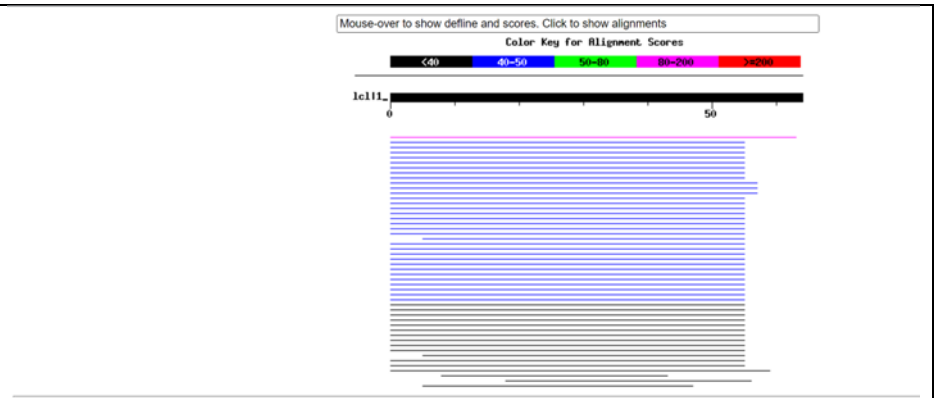
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**16**
Coordonnées du Stop	**11537**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Ove 16**
Coordonnées du Start retenu	**11343**

Fonction prédite	NKF

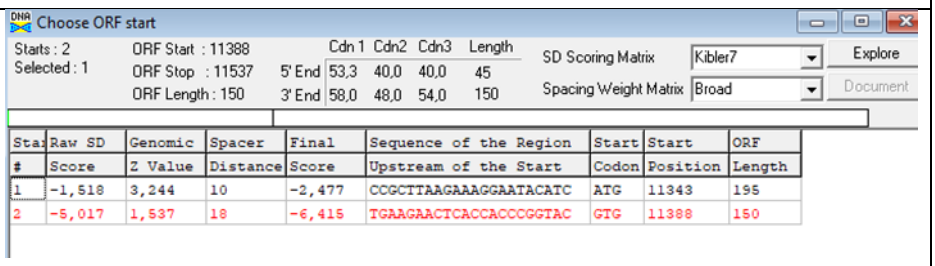
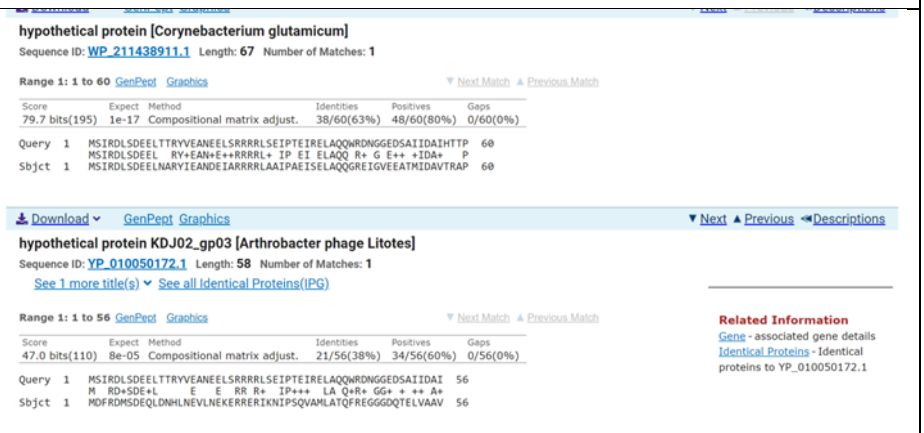
Décision #1 : Est-ce un gène ?

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 ?	

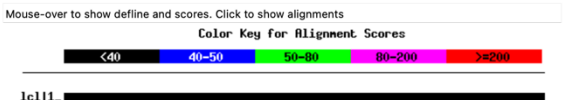
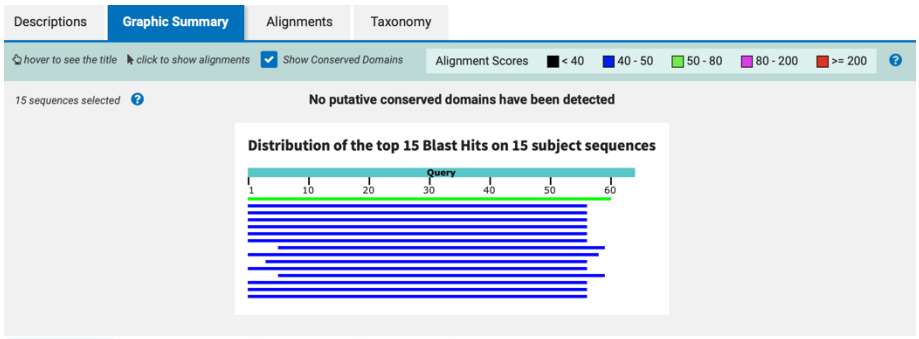
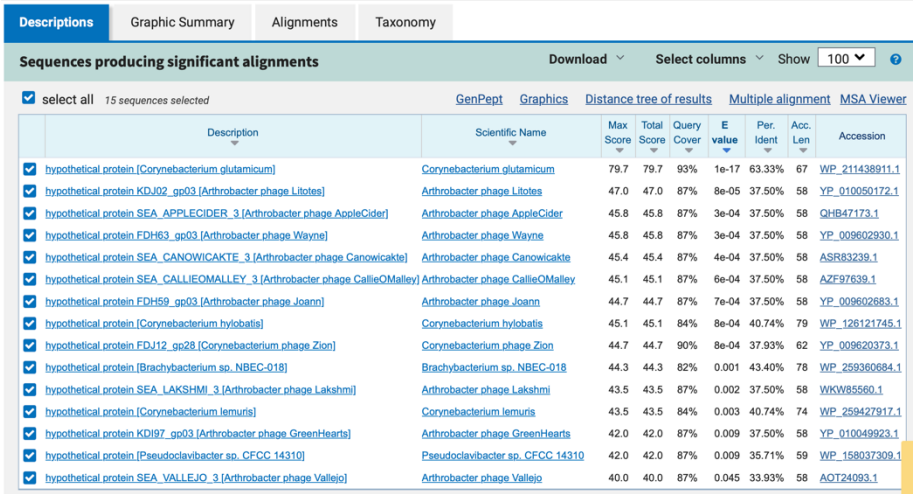
anoPS complete sequence, 14910 bp including 14-base 3' overhang (TCCCGCGCCAGCCC), Order 2, Window 96, Step 12, 7/9

Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 Sequences producing significant alignments: Score (bits) E 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 PotatoChip_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 hypothetical protein [Corynebacterium glutamicum] Corynebacterium glutamicum 79.7 79.7 93% 1e-17 63.33% 67 WP_211438911.1 hypothetical protein KDJ02_gp03 [Arthrobacter phage Litotes] Arthrobacter phage Litotes 47.0 47.0 87% 8e-05 37.50% 58 YP_010050172.1 hypothetical protein SEA_APPLECIDER_3 [Arthrobacter phage AppleCider] Arthrobacter phage AppleCider 45.8 45.8 87% 3e-04 37.50% 58 QHB47173.1 hypothetical protein FDH63_gp03 [Arthrobacter phage Wayne] Arthrobacter phage Wayne 45.8 45.8 87% 3e-04 37.50% 58 YP_009602930.1 hypothetical protein SEA_CANOWICAKTE_3 [Arthrobacter phage Canowicakte] Arthrobacter phage Canowicakte 45.4 45.4 87% 4e-04 37.50% 58 ASR83239.1 hypothetical protein SEA_CALLIEOMALLEY_3 [Arthrobacter phage CallieOMalley] Arthrobacter phage CallieOMalley 45.1 45.1 87% 6e-04 37.50% 58 AZF97639.1 hypothetical protein FDH59_gp03 [Arthrobacter phage Joann] Arthrobacter phage Joann 44.7 44.7 87% 7e-04 37.50% 58 YP_009602683.1 hypothetical protein [Corynebacterium hylobatis] Corynebacterium hylobatis 45.1 45.1 84% 8e-04 40.74% 79 WP_126121745.1 hypothetical protein FDJ12_gp28 [Corynebacterium phage Zion] Corynebacterium phage Zion 44.7 44.7 90% 8e-04 37.93% 62 YP_009620373.1 hypothetical protein [Brachybacterium sp. NBEC-018] Brachybacterium sp. NBEC-018 44.3 44.3 82% 0.001 43.40% 78 WP_259360684.1 hypothetical protein SEA_LAKSHMI_3 [Arthrobacter phage Lakshmi] Arthrobacter phage Lakshmi 43.5 43.5 87% 0.002 37.50% 58 WKW85560.1 hypothetical protein [Corynebacterium lemuris] Corynebacterium lemuris 43.5 43.5 84% 0.003 40.74% 74 WP_259427917.1 hypothetical protein KDI97_gp03 [Arthrobacter phage GreenHearts] Arthrobacter phage GreenHearts 42.0 42.0 87% 0.009 37.50% 58 YP_010049923.1 hypothetical protein [Pseudoclavibacter sp. CFCC 14310] Pseudoclavibacter sp. CFCC 14310 42.0 42.0 87% 0.009 35.71% 59 WP_158037309.1 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 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 ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	***Plus long ORF*** ***ATG*** ***Score faible : -2.4***
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:	**11343**

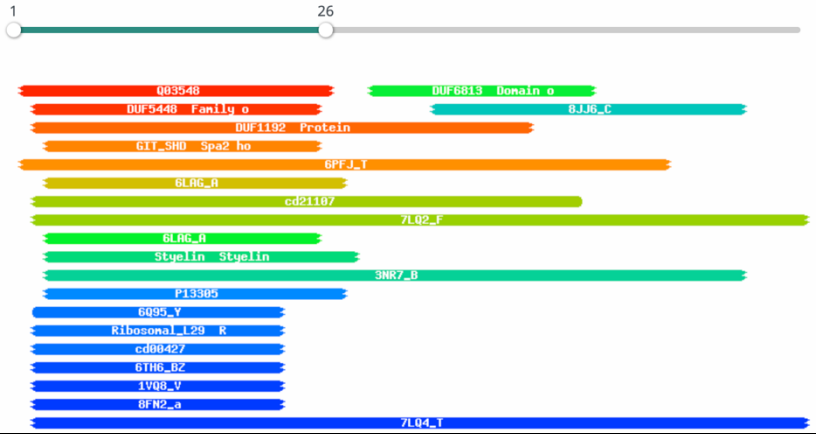
Décision #3 : Quelle est la function de la protéine putative ?

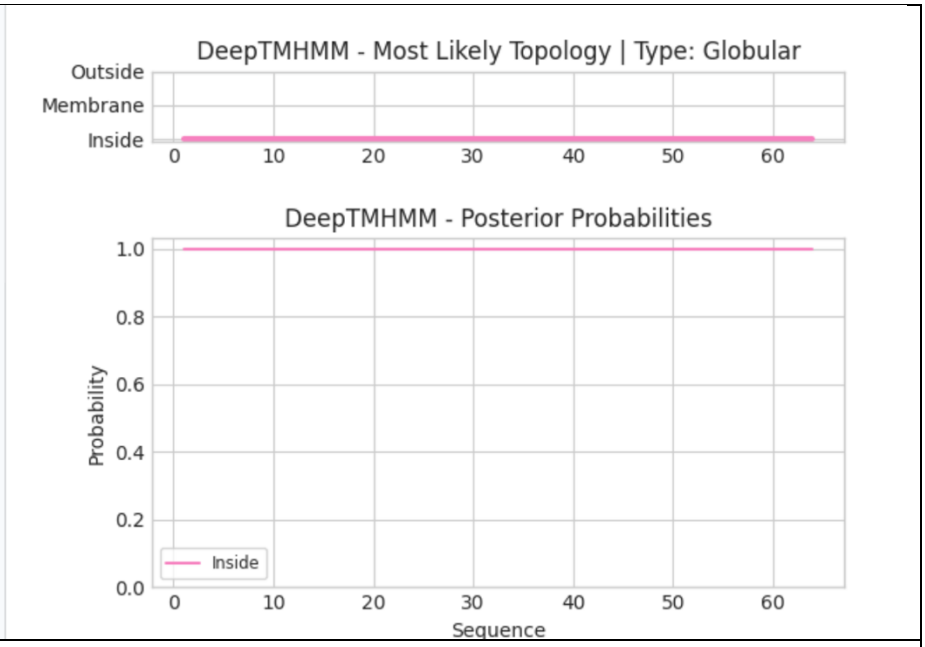
Collection des éléments de réponse	Rationnelle
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	***** PhagesDB :**  Sequences producing significant alignments: Score (bits) E 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 ***** nr :**   Description / Scientific Name / Max Score / Total Score / Query Cover / E value / Per. Ident / Acc. Len / Accession hypothetical protein [Corynebacterium glutamicum] Corynebacterium glutamicum 79.7 79.7 93% 1e-17 63.33% 67 WP_211438911.1 hypothetical protein KDJ02_gp03 [Arthrobacter phage Litotes] Arthrobacter phage Litotes 47.0 47.0 87% 8e-05 37.50% 58 YP_010050172.1 hypothetical protein SEA_APPLECIDER_3 [Arthrobacter phage AppleCider] Arthrobacter phage AppleCider 45.8 45.8 87% 3e-04 37.50% 58 QHB47173.1 hypothetical protein FDH63_gp03 [Arthrobacter phage Wayne] Arthrobacter phage Wayne 45.8 45.8 87% 3e-04 37.50% 58 YP_009602930.1 hypothetical protein SEA_CANOWICAKTE_3 [Arthrobacter phage Canowicakte] Arthrobacter phage Canowicakte 45.4 45.4 87% 4e-04 37.50% 58 ASR83239.1 hypothetical protein SEA_CALLIEOMALLEY_3 [Arthrobacter phage CallieOMalley] Arthrobacter phage CallieOMalley 45.1 45.1 87% 6e-04 37.50% 58 AZF97639.1 hypothetical protein FDH59_gp03 [Arthrobacter phage Joann] Arthrobacter phage Joann 44.7 44.7 87% 7e-04 37.50% 58 YP_009602683.1 hypothetical protein [Corynebacterium hylobatis] Corynebacterium hylobatis 45.1 45.1 84% 8e-04 40.74% 79 WP_126121745.1 hypothetical protein FDJ12_gp28 [Corynebacterium phage Zion] Corynebacterium phage Zion 44.7 44.7 90% 8e-04 37.93% 62 YP_009620373.1 hypothetical protein [Brachybacterium sp. NBEC-018] Brachybacterium sp. NBEC-018 44.3 44.3 82% 0.001 43.40% 78 WP_259360684.1 hypothetical protein SEA_LAKSHMI_3 [Arthrobacter phage Lakshmi] Arthrobacter phage Lakshmi 43.5 43.5 87% 0.002 37.50% 58 WKW85560.1 hypothetical protein [Corynebacterium lemuris] Corynebacterium lemuris 43.5 43.5 84% 0.003 40.74% 74 WP_259427917.1 hypothetical protein KDI97_gp03 [Arthrobacter phage GreenHearts] Arthrobacter phage GreenHearts 42.0 42.0 87% 0.009 37.50% 58 YP_010049923.1 hypothetical protein [Pseudoclavibacter sp. CFCC 14310] Pseudoclavibacter sp. CFCC 14310 42.0 42.0 87% 0.009 35.71% 59 WP_158037309.1 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 : VEAF_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**
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 ?**

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	Q03548	VEAF_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
2	PF17526.6	; DUF5448 ; Family of unknown function (DUF5448)	95.64	0.038	37.78	3.2	24	118
3	PF06698.15	; DUF1192 ; Protein of unknown function (DUF1192)	91.29	4.2	24.78	6.1	41	61
4	PF08518.15	; GIT_SHD ; Spa2 homology domain (SHD) of GIT	88.78	2.4	19.72	3.1	23	29
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
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
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



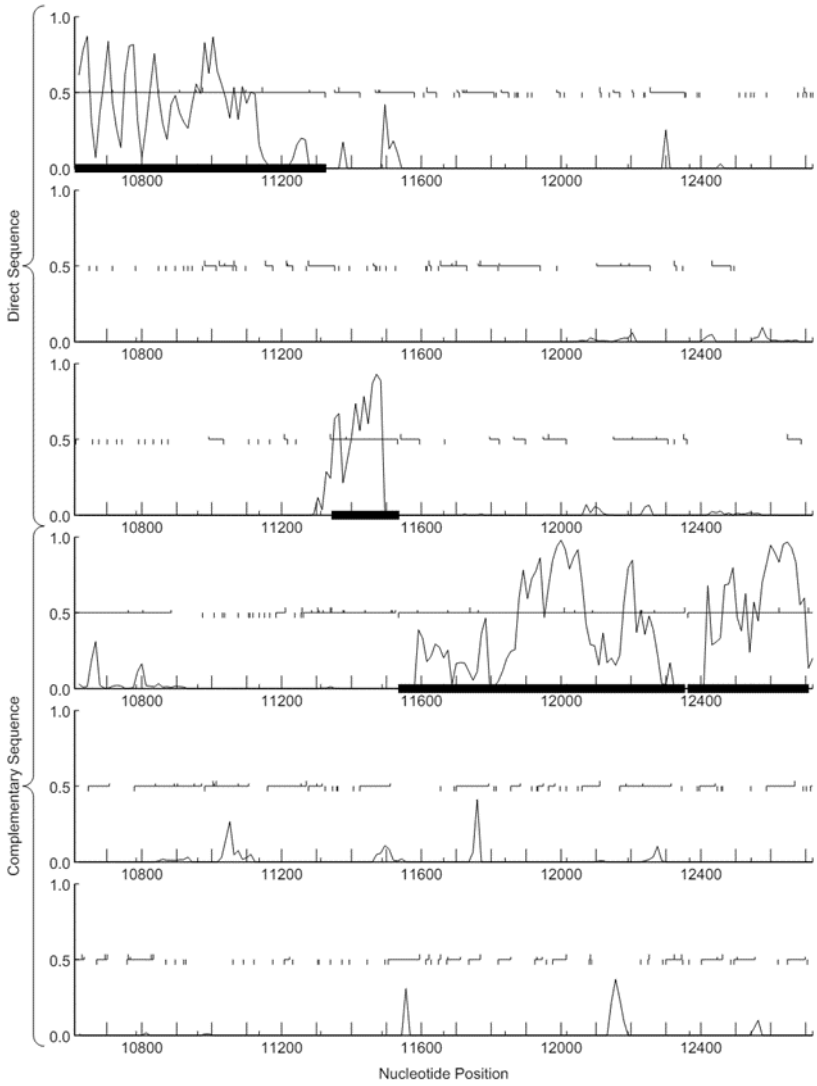
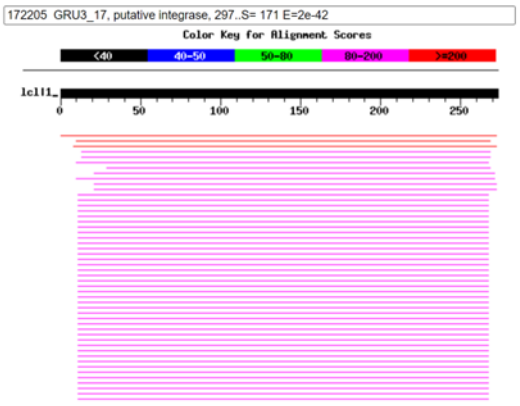
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Répondez Oui ou Non. Une fois que vous avez pris une décision quant-à-la fonction du gène, vérifiez la liste des fonctions officielles de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

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**

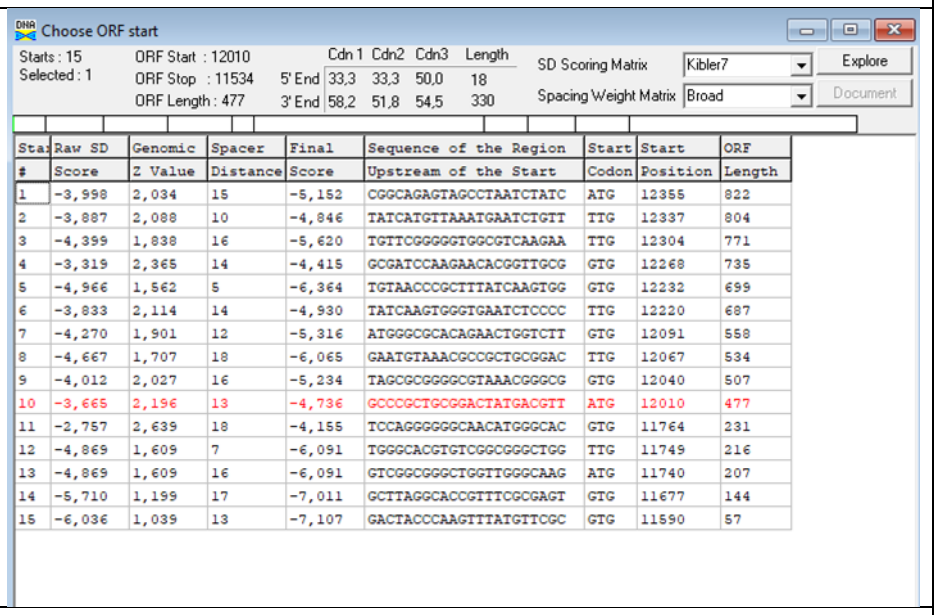
Décision #1 : Est-ce un gène ?

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 ?	anoPS 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: 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

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

Décision #2 : Quel est le start du gène ?

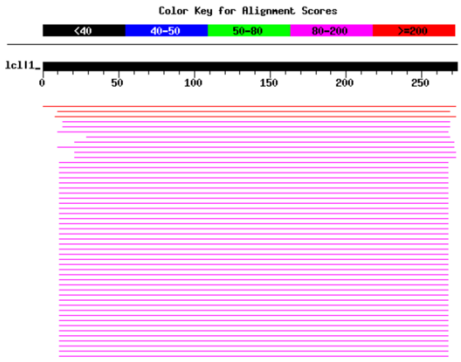
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 12355** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 12355**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
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**

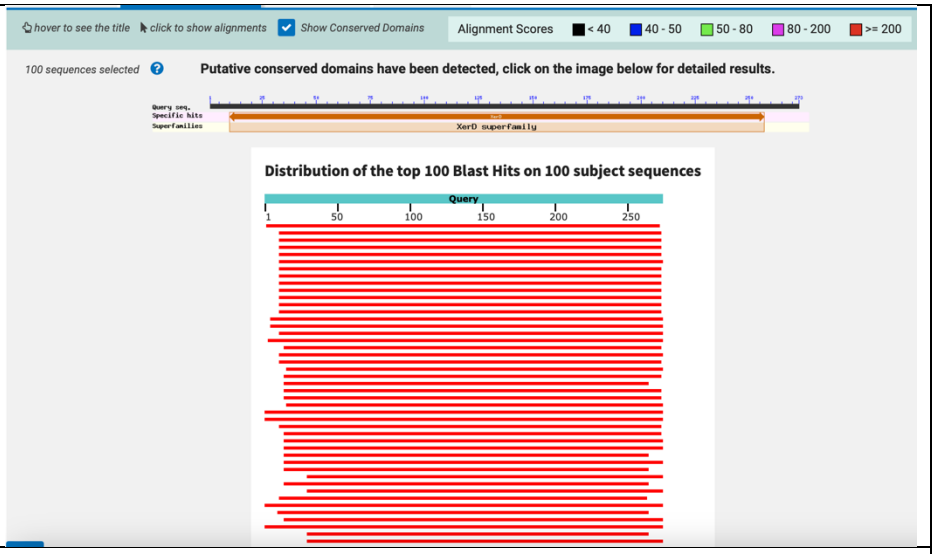
Choose ORF start

Starts : 15 Selected : 1 | ORF Start : 12010 | ORF Stop : 11534 | ORF Length : 477 | Cdn 1 | Cdn2 | Cdn3 | Length | 5' End 33,3 33,3 50,0 18 | 3' End 58,2 51,8 54,5 330 | SD Scoring Matrix: Kibler7 | Spacing Weight Matrix: Broad | Explore | Document

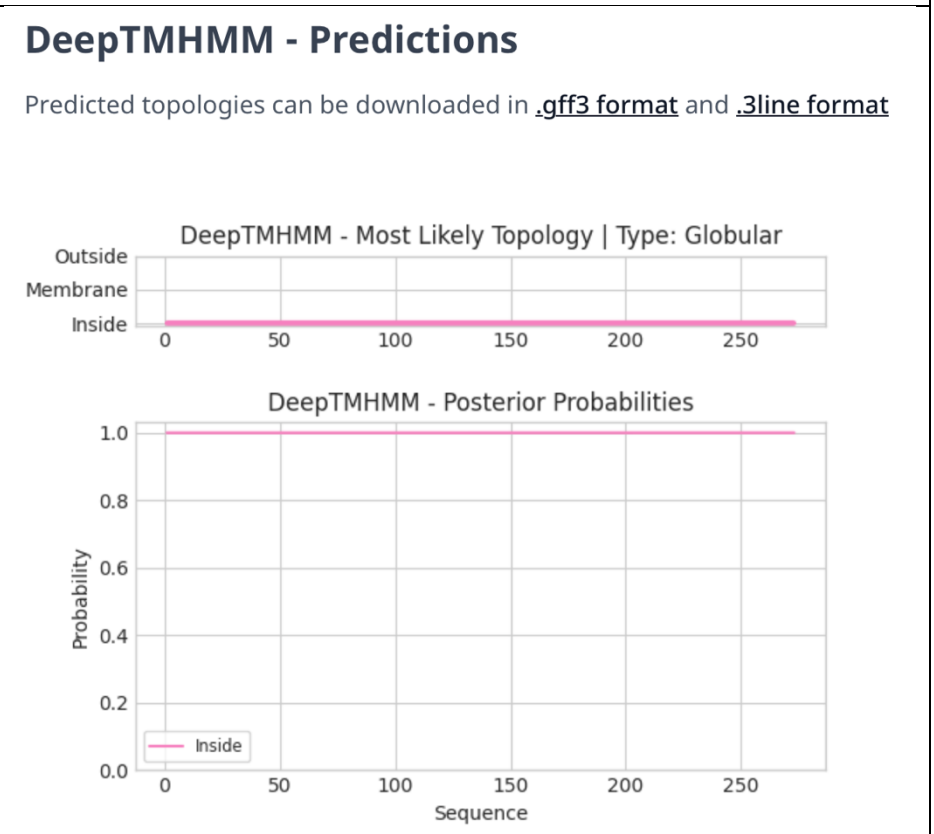
Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-3,998	2,034	15	-5,152	CGGCAGAGTAGCCTAATCTATC	ATG	12355	822
2	-3,887	2,088	10	-4,846	TATCATGTTAAATGAATCTGTT	TTG	12337	804
3	-4,399	1,838	16	-5,620	TGTTCGGGGGTGGCGTCAAGAA	TTG	12304	771
4	-3,319	2,365	14	-4,415	GCGATCCAAGAACACGGTTGCG	GTG	12268	735
5	-4,966	1,562	5	-6,364	TGTAACCCGCTTTATCAAGTGG	GTG	12232	699
6	-3,833	2,114	14	-4,930	TATCAAGTGGGTGAATCTCCCC	TTG	12220	687
7	-4,270	1,901	12	-5,316	ATGGGCGCACAGAACTGGTCTT	GTG	12091	558
8	-4,667	1,707	18	-6,065	GAATGTAAACGCCGCTGCGGAC	TTG	12067	534
9	-4,012	2,027	16	-5,234	TAGCGCGGGGCGTAAACGGGCG	GTG	12040	507
10	-3,665	2,196	13	-4,736	GCCCGCTGCGGACTATGACGTT	ATG	12010	477
11	-2,757	2,639	18	-4,155	TCCAGGGGGGCAACATGGGCAC	GTG	11764	231
12	-4,869	1,609	7	-6,091	TGGGCACGTGTCGGCGGGCTGG	TTG	11749	216
13	-4,869	1,609	16	-6,091	GTCGGCGGGCTGGTTGGGCAAG	ATG	11740	207
14	-5,710	1,199	17	-7,011	GCTTAGGCACCGTTTCGCGAGT	GTG	11677	144
15	-6,036	1,039	13	-7,107	GACTACCCAAGTTTATGTTCGC	GTG	11590	57

Décision #3 : Quelle est la function 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 :**  Sequences producing significant alignments: 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_18, integrase, 298 169 9e-42 Zombie_32, integrase, 398 167 2e-41 ***** nr :** 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 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 hadale] — Corynebacterium hadale — 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 >= 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**

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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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-trancription 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 »**

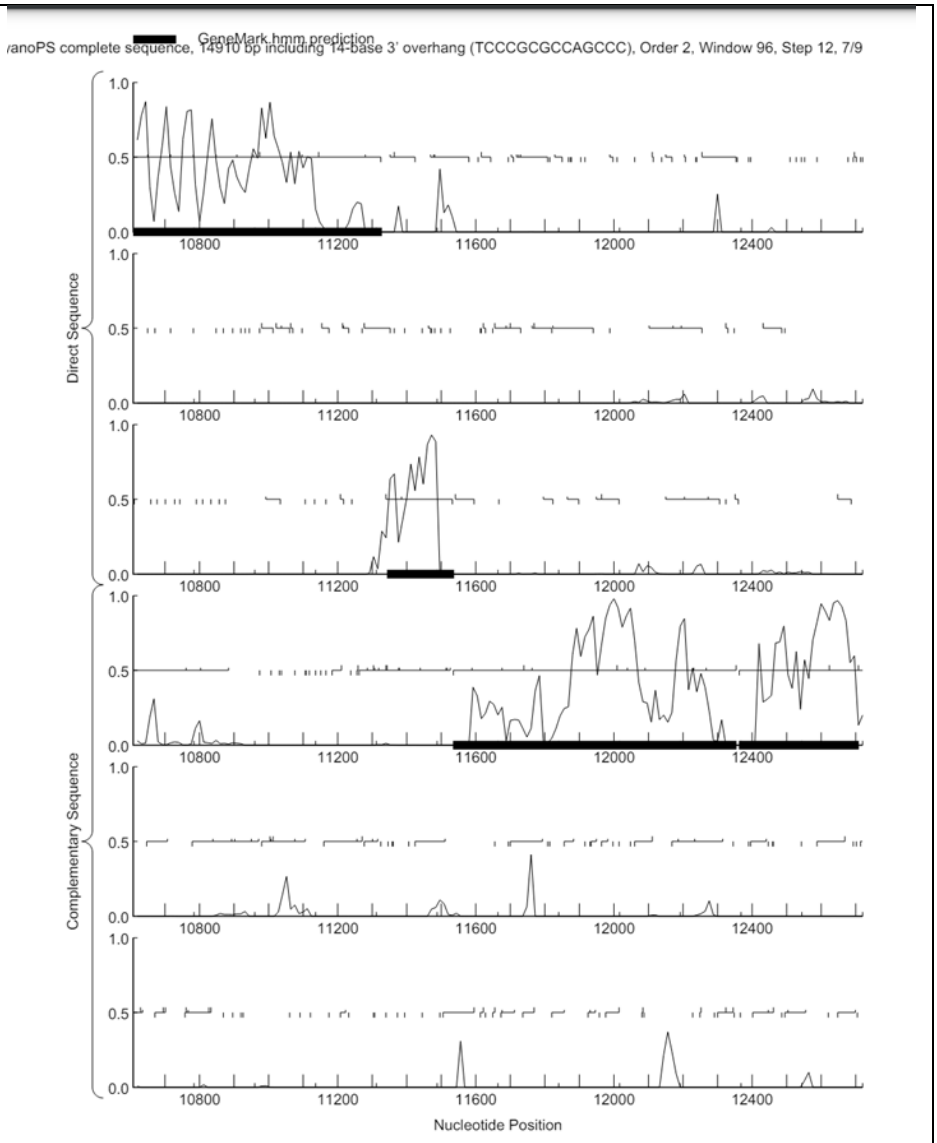
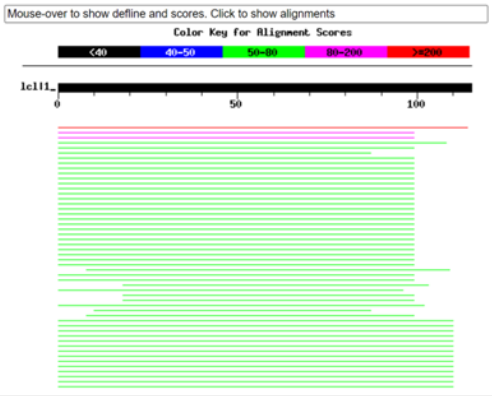
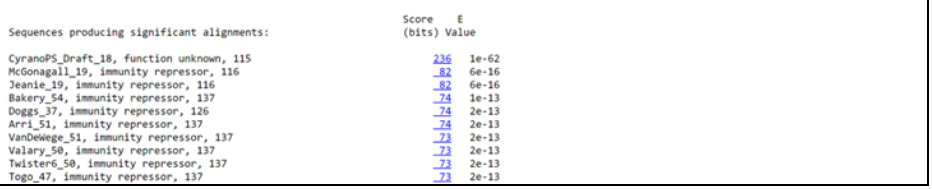
de fonctions approuvées par SEA-PHAGES ?	
DECISION:	**Tyrosine Integrase**

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**

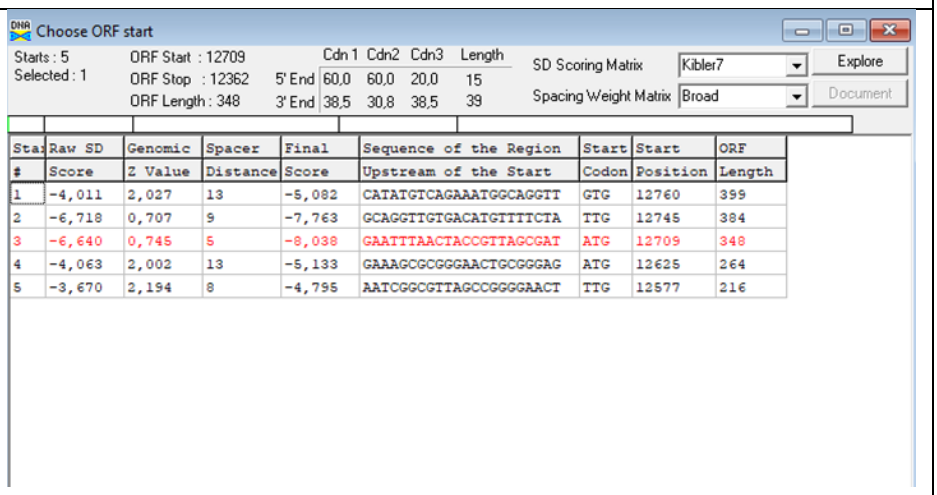
Y-a-t-il des éléments supportant un potentiel codant ?	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	 

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

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	123	123	85%	1e-33	67.68%	101	WP_235696498.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	121	121	85%	6e-33	66.67%	101	WP_226813673.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	121	121	85%	7e-33	66.67%	107	WP_310795193.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	119	119	81%	3e-32	68.42%	97	UJL54986.1
HTH family transcriptional regulator [Corynebacterium diphtheriae bv. intermedius str. NCTC 5011]	Corynebacterium diphtheri...	118	118	81%	1e-31	67.37%	97	EIK55653.1
helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	117	117	86%	3e-31	59.00%	102	QQU78348.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	116	116	76%	9e-31	70.79%	101	WP_339326148.1
helix-turn-helix domain-containing protein [Corynebacterium diphtheriae bv. gravis]	Corynebacterium diphtheri...	114	114	73%	3e-30	70.24%	88	UWF02311.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	113	113	86%	1e-29	57.28%	104	WP_205917829.1
transcriptional regulator [Corynebacterium diphtheriae bv. gravis]	Corynebacterium diphtheri...	112	112	73%	1e-29	71.76%	91	OWN57024.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	112	112	86%	2e-29	56.31%	104	WP_106361736.1
helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium heidelbe...	112	112	85%	2e-29	58.59%	101	WP_112768907.1
helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium accolens	112	112	85%	3e-29	59.60%	101	WP_284900651.1
helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium heidelbe...	111	111	85%	8e-29	58.59%	112	WCZ37641.1
helix-turn-helix transcriptional regulator [Corynebacterium freiburgense]	Corynebacterium freiburg...	110	110	85%	2e-28	59.60%	101	WP_051256139.1
helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	109	109	85%	3e-28	56.57%	101	WP_306592755.1
XRE family transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	108	108	71%	5e-28	71.08%	88	ARB87912.2
helix-turn-helix transcriptional regulator [Corynebacterium propinquum]	Corynebacterium propinqu...	108	108	86%	6e-28	57.00%	102	WP_284571910.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	108	108	86%	1e-27	53.40%	104	WP_088298544.1
helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	85%	2e-27	54.55%	101	WP_284790726.1
TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	86%	3e-27	55.00%	102	HAT66625290.1
helix-turn-helix transcriptional regulator [Corynebacterium propinquum]	Corynebacterium propinqu...	106	106	86%	8e-27	53.00%	102	WP_284594249.1
helix-turn-helix transcriptional regulator [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudod...	106	106	85%	8e-27	55.56%	101	WP_284587248.1

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	123	123	85%	1e-33	67.68%	101	WP_235696498.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	121	121	85%	6e-33	66.67%	101	WP_226813673.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	121	121	85%	7e-33	66.67%	107	WP_310795193.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	119	119	81%	3e-32	68.42%	97	UJL54986.1
HTH family transcriptional regulator [Corynebacterium diphtheriae bv. intermedius str. NCTC 5011]	Corynebacterium diphtheri...	118	118	81%	1e-31	67.37%	97	EIK55653.1
helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	117	117	86%	3e-31	59.00%	102	QQU78348.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	116	116	76%	9e-31	70.79%	101	WP_339326148.1
helix-turn-helix domain-containing protein [Corynebacterium diphtheriae bv. gravis]	Corynebacterium diphtheri...	114	114	73%	3e-30	70.24%	88	UWF02311.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	113	113	86%	1e-29	57.28%	104	WP_205917829.1
transcriptional regulator [Corynebacterium diphtheriae bv. gravis]	Corynebacterium diphtheri...	112	112	73%	1e-29	71.76%	91	OWN57024.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	112	112	86%	2e-29	56.31%	104	WP_106361736.1
helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium heidelbe...	112	112	85%	2e-29	58.59%	101	WP_112768907.1
helix-turn-helix transcriptional regulator [Corynebacterium accolens]	Corynebacterium accolens	112	112	85%	3e-29	59.60%	101	WP_284900651.1
helix-turn-helix transcriptional regulator [Corynebacterium heidelbergense]	Corynebacterium heidelbe...	111	111	85%	8e-29	58.59%	112	WCZ37641.1
helix-turn-helix transcriptional regulator [Corynebacterium freiburgense]	Corynebacterium freiburg...	110	110	85%	2e-28	59.60%	101	WP_051256139.1
helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	109	109	85%	3e-28	56.57%	101	WP_306592755.1
XRE family transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	108	108	71%	5e-28	71.08%	88	ARB87912.2
helix-turn-helix transcriptional regulator [Corynebacterium propinquum]	Corynebacterium propinqu...	108	108	86%	6e-28	57.00%	102	WP_284571910.1
helix-turn-helix transcriptional regulator [Corynebacterium diphtheriae]	Corynebacterium diphtheri...	108	108	86%	1e-27	53.40%	104	WP_088298544.1
helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	85%	2e-27	54.55%	101	WP_284790726.1
TPA: helix-turn-helix transcriptional regulator [Corynebacterium striatum]	Corynebacterium striatum	107	107	86%	3e-27	55.00%	102	HAT66625290.1
helix-turn-helix transcriptional regulator [Corynebacterium propinquum]	Corynebacterium propinqu...	106	106	86%	8e-27	53.00%	102	WP_284594249.1
helix-turn-helix transcriptional regulator [Corynebacterium pseudodiphtheriticum]	Corynebacterium pseudod...	106	106	85%	8e-27	55.56%	101	WP_284587248.1

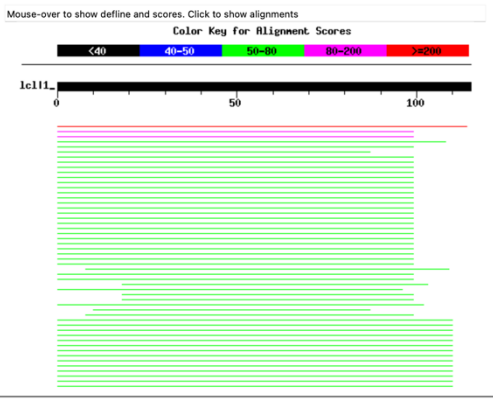
Décision #2 : Quel est le start du gène ?

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) :12709** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :12709**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	 Sta Raw SD Genomic Spacer Final Sequence of the Region Start Start ORF # Score Z Value Distance Score Upstream of the Start Codon Position Length 1 -4,011 2,027 13 -5,082 CATATGTCAGAAATGGCAGGTT GTG 12760 399 2 -6,718 0,707 9 -7,763 GCAGGTTGTGACATGTTTTCTA TTG 12745 384 3 -6,640 0,745 5 -8,038 GAATTTAACTACCGTTAGCGAT ATG 12709 348 4 -4,063 2,002 13 -5,133 GAAAGCGCGGGAACTGCGGGAG ATG 12625 264 5 -3,670 2,194 8 -4,795 AATCGGCGTTAGCCGGGGAACT TTG 12577 216
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 ?	***Troisième plus longue ORF avec un ATG*** ***score (-8)*** ***Mais plus long ORF avec un ATG***
Est-ce que le start est conservé chez les homologues voir Starterator ?	**–**

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	>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 MTSAFNQAGIIPEWTKQDRIRKARELREMKQSDLATAIGVSRGTLASIEQGVREPRRGEV 60 MT+A+ ++G +PE + + R+R ARE ++Q LA +GVSR T+++ E+G PR+ + Sbjct: 1 MTTAY-ESGRVPEISLRHRLRIAREEAGLEQQSLAERMGVSRNTVSAAEKGKNAPRKVVL 59 Query: 61 IAISFATGVSLDWLETGKTPADDNGGGEKWAHWDSNPRPA 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 MTSAFNQAGIIPEWTKQDRIRKARELREMKQSDLATAIGVSRGTLASIEQGVREPRRGEV 60 MT+A+ ++G +PE + + R+R ARE ++Q LA +GVSR T+++ E+G PR+ + Sbjct: 1 MTTAY-ESGRVPEISLRHRLRIAREEAGLEQQSLAERMGVSRNTVSAAEKGKNAPRKVVL 59 Query: 61 IAISFATGVSLDWLETGKTPADDNGGGEKWAHWDSNPRPA 100 A + ATG + WLETG P G + A DSNP+P+ Sbjct: 60 NAWALATGFDVKWLETGIAPQPGPEGDDGCARRDSNPKPS 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 MTSAFNQAGIIPEWTKQDRIRKARELREMKQSDLATAIGVSRGTLASIEQGVREPRRGEV 60 MT+A+ ++G +PE + + R+R ARE ++Q LA +GVSR T+++ E+G PR+ + Sbjct: 1 MTTAY-ESGRVPEISLRHRLRIAREEAGLEQQSLAERMGVSRNTVSAAEKGKNAPRKVVL 59
DECISION:	**12 709 ATG**

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

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



Sequences producing significant alignments:	Score (bits)	E Value
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 :**

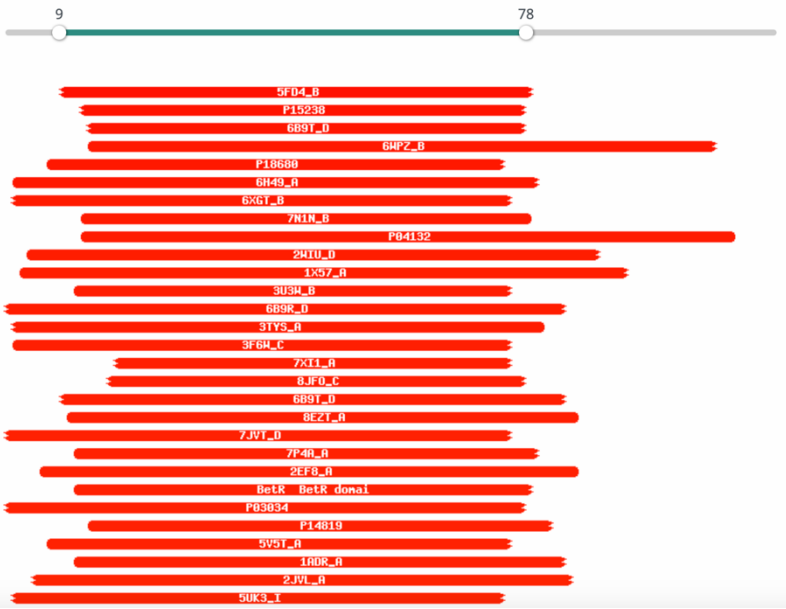
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée venant de la PDB ou autre base de données lors d'un crible HHPred avec une proba >= 90% et une couverture acceptable ?	**PFAM : BetR ; BetR domain** **CDD : DUF1870 ; Domain of unknown function (DUF1870)** **Cd : HTH_XRE; 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**

1	5FD4_B	ComR; Streptococcus, Competence, Quorum sensing, ComR, TRANSCRIPTION REGULATOR; 2.9A {Streptococcus suis (strain 05ZYH33	98.46	0.000022	45.11	9.5	69		324
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
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
4	6WPZ_B	Pf4r; DNA BINDING PROTEIN; 1.993A {Pseudomonas aeruginosa}	97.94	0.00057	32.12	9	90		95
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
6	6H49_A	Orf20; SaPI, Repressor, STRUCTURAL PROTEIN; HET: SO4; 1.8A {Staphylococcus aureus}	97.92	0.001	34.13	9.2	78		157
7	6XGT_B	Cyanate hydratase; cyanase, HYDROLASE; HET: MLI, FMT; 2.2A {Thermomyces lanuginosus}	97.78	0.0024	34.06	9.3	74		181
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
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
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
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
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
22	2EF8_A	Putative transcription factor; helix-turn-helix, DNA binding protein, TRANSCRIPTION REGULATOR; HET: CME; 1.95A {Enteroba	97.19	0.0088	26.72	6.8	80		84
23	PF08667.14	; BetR ; BetR domain	97.16	0.022	30.59	8.9	68		174
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
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

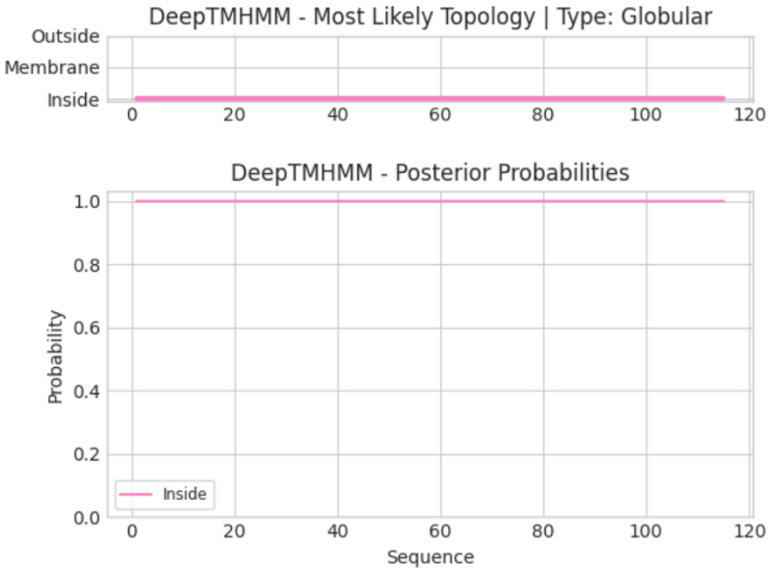
10	2WIU_D	HTH-TYPE TRANSCRIPTIONAL REGULATOR HIPB; TRANSFERASE TRANSCRIPTION COMPLEX, SERINE KINASE, DNA-BINDING, MERCURY DERIVATI	97.6	0.0021	28.52	9.2	85	88
11	1X57_A	Endothelial differentiation-related factor 1; EDF1, HMBF1alpha, helix-turn-helix, Structural Genomics, NPPSFA, National	97.4	0.0045	27.57	9.1	90	91
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
13	6B9R_D	Hydroxyethylphosphonate dioxygenase; Phosphonate, Hydroxymethylphosphonate, Iron, OXIDOREDUCTASE; HET: GOL, 2HE; 1.802A	97.37	0.013	35.89	8.3	83	450
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
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
16	7XI1_A	anti-CRISPR protein AcrIF24; Anti-CRISPR, Acr, Inhibitor of type I-F Cascade, IMMUNE SYSTEM; 2.53A {Pseudomonas aerugino	97.32	0.017	32.59	8.3	59	236
17	8JFO_C	AcrIIA15; IIA type anti-crispr protein, VIRAL PROTEIN; 2.3A {Staphylococcus delphini}	97.26	0.0072	31.82	5.9	61	171
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

Resubmit Section



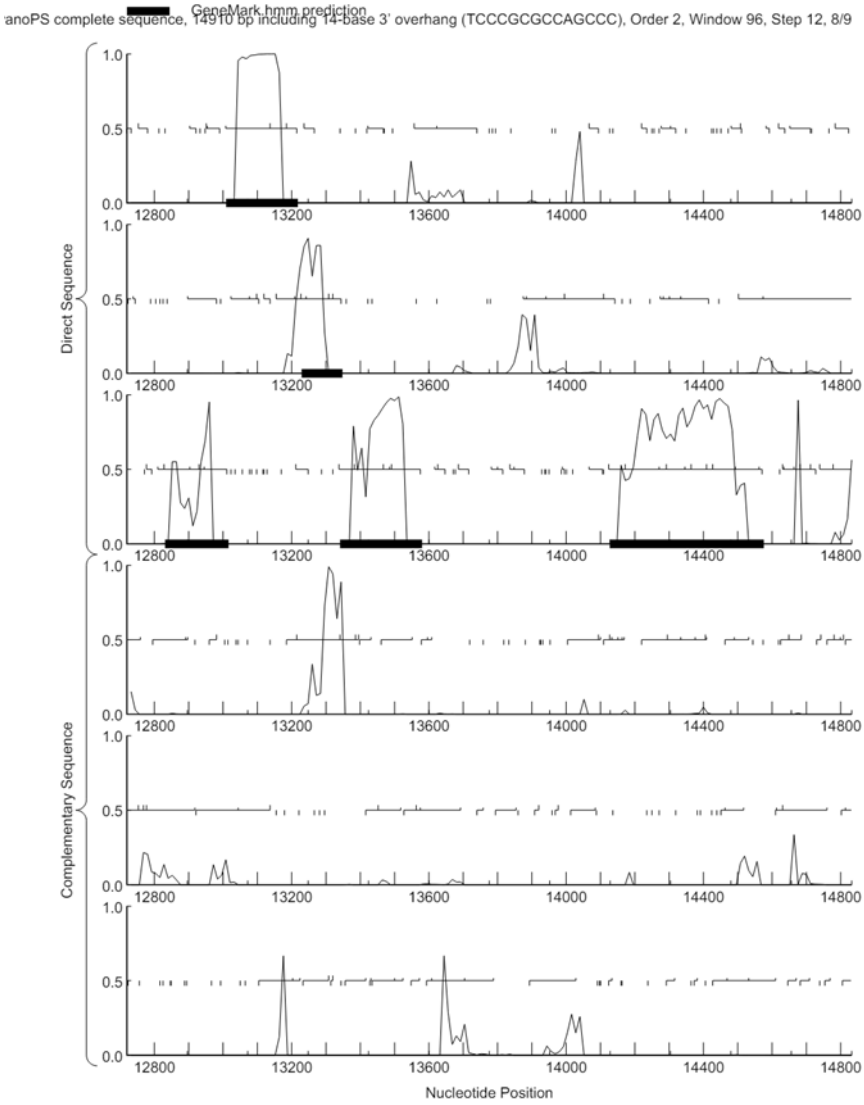
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	**Oui le gene avant = integrase et le gene apres helix turn helix**

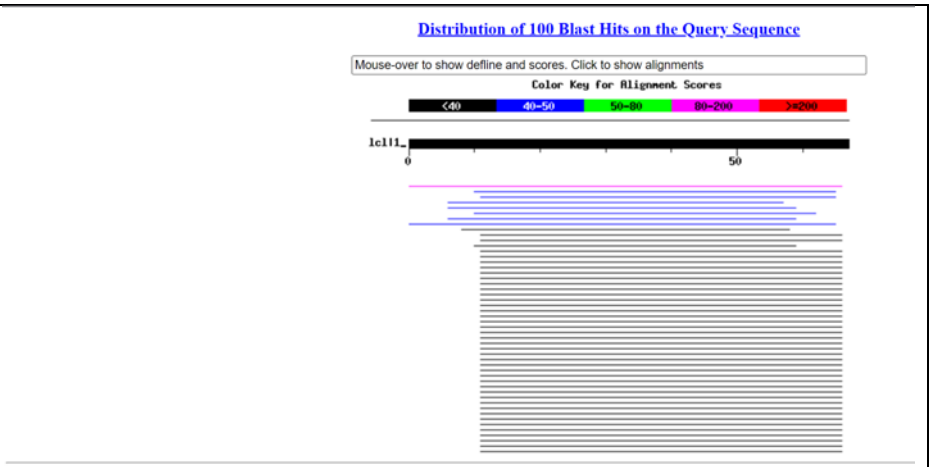
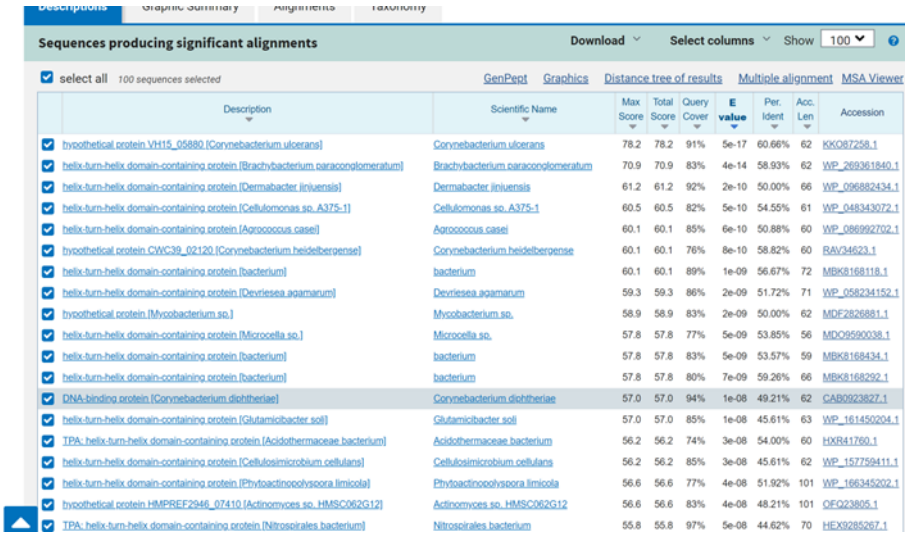
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format 
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Oui**
DECISION:	**Immunity repressor**

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**

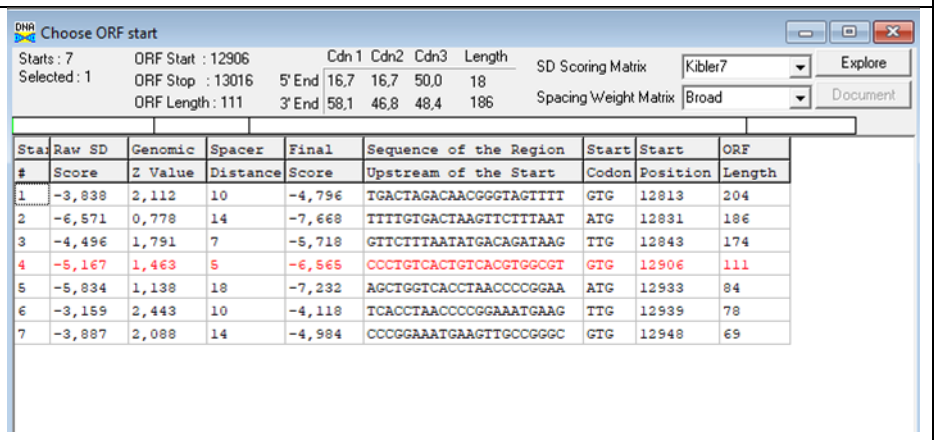
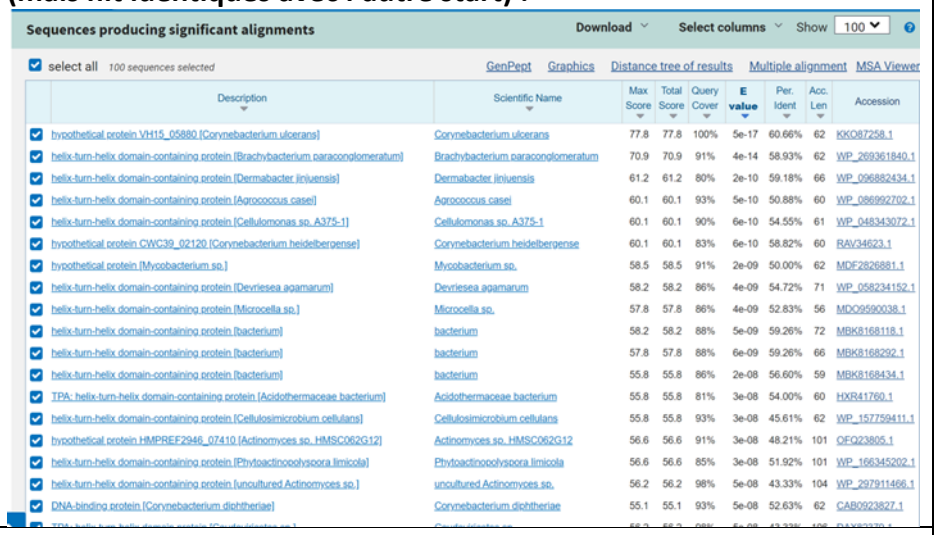
Décision #1 : Est-ce un gène ?

Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
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: Score (bits) E 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 
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) :12813** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) :12831**
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
Est-ce que le start prédit conduit au plus long ORF ? Sinon, est-ce que l'ORF le plus long conduit à un chevauchement excessif avec un autre ORF (>30bp) ? Si le plus long pas retenu, quel est l'écart intergène résultant ?	***Le start 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.*** **(mais hit identiques avec l'autre start) :** 
Est-ce que le start est conservé chez les homologues voir Starterator ?	–

Choose ORF start

Starts : 7 Selected : 1 | ORF Start : 12906 | ORF Stop : 13016 | ORF Length : 111 | Cdn 1 Cdn2 Cdn3 Length | 5' End 16,7 16,7 50,0 18 | 3' End 58,1 46,8 48,4 186 | SD Scoring Matrix: Kibler7 | Spacing Weight Matrix: Broad | Explore | Document

Sta #	Raw SD Score	Genomic Z Value	Spacer Distance	Final Score	Sequence of the Region Upstream of the Start	Start Codon	Start Position	ORF Length
1	-3,838	2,112	10	-4,796	TGACTAGACAACGGGTAGTTTT	GTG	12813	204
2	-6,571	0,778	14	-7,668	TTTTGTGACTAAGTTCTTTAAT	ATG	12831	186
3	-4,496	1,791	7	-5,718	GTTCTTTAATATGACAGATAAG	TTG	12843	174
4	-5,167	1,463	5	-6,565	CCCTGTCACTGTCACGTGGCGT	GTG	12906	111
5	-5,834	1,138	18	-7,232	AGCTGGTCACCTAACCCCGGAA	ATG	12933	84
6	-3,159	2,443	10	-4,118	TCACCTAACCCCGGAAATGAAG	TTG	12939	78
7	-3,887	2,088	14	-4,984	CCCGGAAATGAAGTTGCCGGGC	GTG	12948	69

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 VH15_05880 [Corynebacterium ulcerans]	Corynebacterium ulcerans	77.8	77.8	100%	5e-17	60.66%	62	KKO87258.1
helix-turn-helix domain-containing protein [Brachybacterium paraconglomeratum]	Brachybacterium paraconglomeratum	70.9	70.9	91%	4e-14	58.93%	62	WP_269361840.1
helix-turn-helix domain-containing protein [Dermabacter jinjuensis]	Dermabacter jinjuensis	61.2	61.2	80%	2e-10	59.18%	66	WP_096882434.1
helix-turn-helix domain-containing protein [Agrococcus casei]	Agrococcus casei	60.1	60.1	93%	5e-10	50.88%	60	WP_086992702.1
helix-turn-helix domain-containing protein [Cellulomonas sp. A375-1]	Cellulomonas sp. A375-1	60.1	60.1	90%	6e-10	54.55%	61	WP_048343072.1
hypothetical protein CWC39_02120 [Corynebacterium heidelbergense]	Corynebacterium heidelbergense	60.1	60.1	83%	6e-10	58.82%	60	RAV34623.1
hypothetical protein [Mycobacterium sp.]	Mycobacterium sp.	58.5	58.5	91%	2e-09	50.00%	62	MDF2826881.1
helix-turn-helix domain-containing protein [Devriesea agamarum]	Devriesea agamarum	58.2	58.2	86%	4e-09	54.72%	71	WP_058234152.1
helix-turn-helix domain-containing protein [Microcella sp.]	Microcella sp.	57.8	57.8	86%	4e-09	52.83%	56	MDO9590038.1
helix-turn-helix domain-containing protein [bacterium]	bacterium	58.2	58.2	88%	5e-09	59.26%	72	MBK8168118.1
helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	88%	6e-09	59.26%	66	MBK8168292.1
helix-turn-helix domain-containing protein [bacterium]	bacterium	55.8	55.8	86%	2e-08	56.60%	59	MBK8168434.1
TPA: helix-turn-helix domain-containing protein [Acidothermaceae bacterium]	Acidothermaceae bacterium	55.8	55.8	81%	3e-08	54.00%	60	HXR41760.1
helix-turn-helix domain-containing protein [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	55.8	55.8	93%	3e-08	45.61%	62	WP_157759411.1
hypothetical protein HMPREF2946_07410 [Actinomyces sp. HMSC062G12]	Actinomyces sp. HMSC062G12	56.6	56.6	91%	3e-08	48.21%	101	OFQ23805.1
helix-turn-helix domain-containing protein [Phytoactinopolyspora limicola]	Phytoactinopolyspora limicola	56.6	56.6	85%	3e-08	51.92%	101	WP_166345202.1
helix-turn-helix domain-containing protein [uncultured Actinomyces sp.]	uncultured Actinomyces sp.	56.2	56.2	98%	5e-08	43.33%	104	WP_297911466.1
DNA-binding protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	55.1	55.1	93%	5e-08	52.63%	62	CAB0923827.1

Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	Download GenPept Graphics ▼Next ▲Previous Descriptions **hypothetical protein VH15_05880 [Corynebacterium ulcerans]** Sequence ID: KKO87258.1 Length: 62 Number of Matches: 1 Range 1: 1 to 61 GenPept Graphics ▼Next Match ▲Previous Match Score 78.2 bits(191) Expect 5e-17 Method Compositional matrix adjust. Identities 37/61(61%) Positives 45/61(73%) Gaps 0/61(0%) Query 7 MTDKLLTTAEVAERLGISPVTVTWRVRAGHLTPEMKLPGVNGAYLFNPASIDQLTKETQS 66 M++KLLTTA VA+ LG+ TVTWR R G LTP MKLPG+NGAYLF+P I+QL + Sbjct 1 MSNKLLTTAAVAKILGLPHSTVTWRARTGRLTPAMKLPGINGAYLFDPREIEQLASKKGP 60 Query 67 A 67 A Sbjct 61 A 61 Download GenPept Graphics ▼Next ▲Previous Descriptions **helix-turn-helix domain-containing protein [Brachybacterium paraconglomeratum]** Sequence ID: WP_269361840.1 Length: 62 Number of Matches: 1 See 1 more title(s) See all Identical Proteins(IPG) Range 1: 3 to 58 GenPept Graphics ▼Next Match ▲Previous Match Score 70.9 bits(172) Expect 4e-14 Method Compositional matrix adjust. Identities 33/56(59%) Positives 39/56(69%) Gaps 0/56(0%) Query 8 TDKLLTTAEVAERLGISPVTVTWRVRAGHLTPEMKLPGVNGAYLFNPASIDQLTKE 63 + KLLT EVAERLGI+ TV+WRV G L P KLPG NGAYLF+ +D+L E Sbjct 3 SSKLLTNVEVAERLGIARQTVSWRVARGQLEPVQKLPGRNGAYLFDEEQVDRLAAE 58 Related Information Identical Proteins - Identical proteins to WP_269361840.1
DECISION:	**12813**

Décision #3 : Quelle est la function 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 defline and scores. Click to show alignments Color Key for Alignment Scores <40 40-50 50-80 80-200 >=200 lcl11_ 0 50 Sequences producing significant alignments: Score (bits) E 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 :**

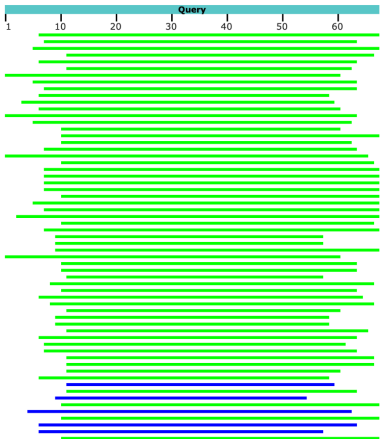
hypothetical protein VH15_05880 [Corynebacterium ulcerans]	Corynebacterium ulcerans	78.2	78.2	91%	5e-17	60.66%	62	KKO87258.1
helix-turn-helix domain-containing protein [Brachybacterium paraconglomeratum]	Brachybacterium paraconglomeratum	70.9	70.9	83%	4e-14	58.93%	62	WP_269361840.1
helix-turn-helix domain-containing protein [Dermabacter jinjuensis]	Dermabacter jinjuensis	61.2	61.2	92%	2e-10	50.00%	66	WP_096882434.1
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
helix-turn-helix domain-containing protein [Agrococcus casei]	Agrococcus casei	60.1	60.1	85%	6e-10	50.88%	60	WP_086992702.1
hypothetical protein CWC39_02120 [Corynebacterium heidelbergense]	Corynebacterium heidelbergense	60.1	60.1	76%	8e-10	58.82%	60	RAV34623.1
helix-turn-helix domain-containing protein [bacterium]	bacterium	60.1	60.1	89%	1e-09	56.67%	72	MBK8168118.1
helix-turn-helix domain-containing protein [Devriesea agamarum]	Devriesea agamarum	59.3	59.3	86%	2e-09	51.72%	71	WP_058234152.1
hypothetical protein [Mycobacterium sp.]	Mycobacterium sp.	58.9	58.9	83%	2e-09	50.00%	62	MDF2826881.1
helix-turn-helix domain-containing protein [Microcella sp.]	Microcella sp.	57.8	57.8	77%	5e-09	53.85%	56	MDO9590038.1
helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	83%	5e-09	53.57%	59	MBK8168434.1
helix-turn-helix domain-containing protein [bacterium]	bacterium	57.8	57.8	80%	7e-09	59.26%	66	MBK8168292.1
DNA-binding protein [Corynebacterium diphtheriae]	Corynebacterium diphtheriae	57.0	57.0	94%	1e-08	49.21%	62	CAB0923827.1
helix-turn-helix domain-containing protein [Glutamicibacter soli]	Glutamicibacter soli	57.0	57.0	85%	1e-08	45.61%	63	WP_161450204.1
TPA: helix-turn-helix domain-containing protein [Acidothermaceae bacterium]	Acidothermaceae bacterium	56.2	56.2	74%	3e-08	54.00%	60	HXR41760.1
helix-turn-helix domain-containing protein [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	56.2	56.2	85%	3e-08	45.61%	62	WP_157759411.1
helix-turn-helix domain-containing protein [Phytoactinopolyspora limicola]	Phytoactinopolyspora limicola	56.6	56.6	77%	4e-08	51.92%	101	WP_166345202.1
hypothetical protein HMPREF2946_07410 [Actinomyces sp. HMSC062G12]	Actinomyces sp. HMSC062G12	56.6	56.6	83%	4e-08	48.21%	101	OFQ23805.1
TPA: helix-turn-helix domain-containing protein [Nitrospirales bacterium]	Nitrospirales bacterium	55.8	55.8	97%	5e-08	44.62%	70	HEX9285267.1
hypothetical protein [Leifsonia sp.]	Leifsonia sp.	55.5	55.5	83%	5e-08	50.00%	59	MAT17174.1
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
TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	56.2	56.2	89%	7e-08	43.33%	106	DAX82379.1
helix-turn-helix domain-containing protein [Isoptericola variabilis]	Isoptericola variabilis	56.2	56.2	89%	7e-08	43.33%	104	MBF1252321.1
TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	56.2	56.2	89%	8e-08	43.33%	106	DAR87009.1
hypothetical protein CBR64_00030 [Cellulosimicrobium cellulans]	Cellulosimicrobium cellulans	55.8	55.8	85%	1e-07	45.61%	106	ARU50140.1
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
TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	55.1	55.1	89%	2e-07	41.67%	106	DAN08877.1
hypothetical protein A0J59_14960 [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	54.3	54.3	97%	2e-07	38.46%	73	KZM77960.1
TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	55.1	55.1	89%	2e-07	41.67%	106	DAN08877.1
hypothetical protein A0J59_14960 [Cellulosimicrobium sp. I38E]	Cellulosimicrobium sp. I38E	54.3	54.3	97%	2e-07	38.46%	73	KZM77960.1
hypothetical protein [Microbacterium lacticum]	Microbacterium lacticum	53.5	53.5	83%	3e-07	46.43%	61	WP_314429462.1
TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	54.7	54.7	89%	3e-07	41.67%	106	DAO74759.1
hypothetical protein GCM10025863_06820 [Microbacterium suwonense]	Microbacterium suwonense	53.5	53.5	71%	4e-07	58.33%	72	BDZ38068.1
helix-turn-helix domain-containing protein [Microbacterium suwonense]	Microbacterium suwonense	53.5	53.5	71%	4e-07	58.33%	70	WP_286303025.1
TPA: helix-turn-helix domain-containing protein [Rhodoglobus sp.]	Rhodoglobus sp.	53.1	53.1	86%	5e-07	49.18%	66	HOW00973.1
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
DNA-binding protein [Actinomycetota bacterium]	Actinomycetota bacterium	52.4	52.4	79%	1e-06	52.83%	69	RPI11310.1
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
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
TPA: helix-turn-helix domain protein [Caudoviricetes sp.]	Caudoviricetes sp.	51.6	51.6	86%	2e-06	43.10%	64	DAI53416.1
hypothetical protein [Streptosporangiaceae bacterium]	Streptosporangiaceae bacterium	51.2	51.2	79%	2e-06	49.06%	61	MCW2900027.1
helix-turn-helix domain-containing protein [Corynebacterium auriscanis]	Corynebacterium auriscanis	52.0	52.0	86%	2e-06	46.55%	80	WP_282939148.1
hypothetical protein N579_08860 [Corynebacterium pseudodiphtheriticum 090104]	Corynebacterium pseudodiphtheriticum 0...	51.2	51.2	86%	2e-06	43.10%	64	ERJ43622.1
helix-turn-helix domain-containing protein [unclassified Microbacterium]	unclassified Microbacterium	51.2	51.2	73%	3e-06	53.06%	59	WP_143016360.1
helix-turn-helix domain-containing protein [Dermabacter sp. HSID17554]	Dermabacter sp. HSID17554	51.2	51.2	73%	3e-06	51.02%	64	RUP86599.1
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
helix-turn-helix domain-containing protein [Betaproteobacteria bacterium]	Betaproteobacteria bacterium	50.8	50.8	80%	3e-06	49.09%	59	MBK6602370.1
hypothetical protein [Microbacterium sp. GCS4]	Microbacterium sp. GCS4	50.4	50.4	85%	4e-06	45.61%	59	WP_152926074.1
helix-turn-helix domain-containing protein [Gulosibacter bifidus]	Gulosibacter bifidus	50.4	50.4	80%	4e-06	40.74%	66	WP_110477041.1
helix-turn-helix domain-containing protein [Acidipropionibacterium timonense]	Acidipropionibacterium timonense	50.8	50.8	83%	5e-06	46.43%	70	WP_130864622.1
helix-turn-helix domain-containing protein [Gordonia westfalica]	Gordonia westfalica	50.4	50.4	82%	5e-06	47.27%	65	WP_159441558.1
DNA binding domain-containing protein, excisionase family [Gordonia westfalica]	Gordonia westfalica	50.4	50.4	82%	5e-06	47.27%	74	SDU50629.1
hypothetical protein DI609_07700 [Corynebacterium urealyticum]	Corynebacterium urealyticum	50.4	50.4	73%	6e-06	51.02%	77	PZO99721.1
excisionase and transcriptional regulator [Gordonia phage SallySpecial]	Gordonia phage SallySpecial	50.1	50.1	77%	7e-06	50.00%	63	YP_010674640.1
TPA: helix-turn-helix domain-containing protein [Propionibacteriaceae bacterium]	Propionibacteriaceae bacterium	49.7	49.7	71%	8e-06	47.92%	63	HEY5482789.1
helix-turn-helix domain-containing protein [Metallococcus carri]	Metallococcus carri	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.

Distribution of the top 100 Blast Hits on 100 subject sequences



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

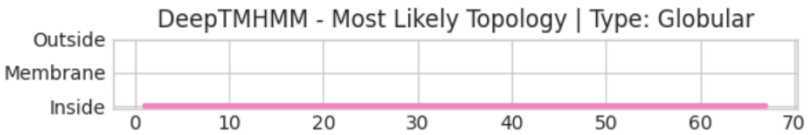
Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	5FD4_B	ComR; Streptococcus, Competence, Quorum sensing, ComR, TRANSCRIPTION REGULATOR; 2.9A {Streptococcus suis (strain 05ZYH33	98.46	0.000022	45.11	9.5	69	324
2	P15238	RPC_BP163 Repressor protein C OS=Rhizobium phage 16-3 OX=10704 GN=C PE=1 SV=4	98.39	0.000034	43.09	8.9	65	263
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
4	6WPZ_B	Pf4r; DNA BINDING PROTEIN; 1.993A {Pseudomonas aeruginosa}	97.94	0.00057	32.12	9	90	95
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
6	6H49_A	Orf20; SaPI, Repressor, STRUCTURAL PROTEIN; HET: SO4; 1.8A {Staphylococcus aureus}	97.92	0.001	34.13	9.2	78	157
7	6XGT_B	Cyanate hydratase; cyanase, HYDROLASE; HET: MLI, FMT; 2.2A {Thermomyces lanuginosus}	97.78	0.0024	34.06	9.3	74	181

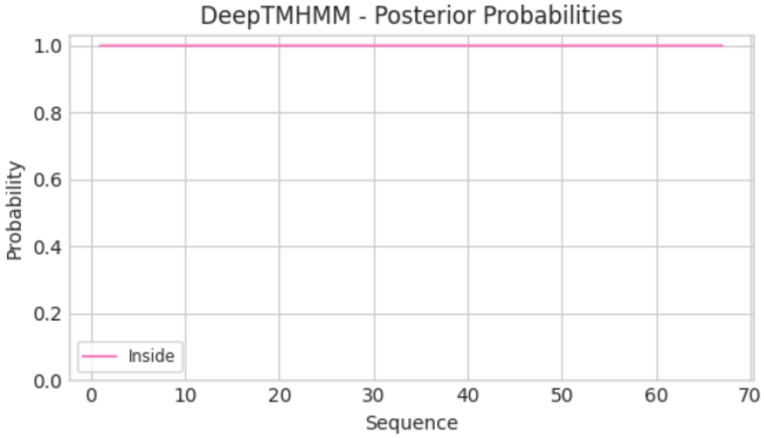
8	6G1T_A	AM32; Repressor, Protein-DNA complex, Type IV secretion system, DNA BINDING PROTEIN; HET: EPE; 1.93A {Enterococcus faeca	98.94	1.9e-8	50	6.1	62	122
9	PF06806.16	; DUF1233 ; Putative excisionase (DUF1233)	98.93	3.1e-8	43.81	6.1	61	70
10	2KVV_A	Putative excisionase; excisionase, DNA binding, wHTH motif, Structural Genomics, NORTHEAST STRUCTURAL GENOMICS CONSORTIU	98.91	4.7e-8	45.1	6.4	66	78
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
12	PF04936.16	; DUF658 ; Protein of unknown function (DUF658)	98.88	5.8e-8	51.11	6.9	56	186
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
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
15	PF11112.12	; PyocinActivator ; Pyocin activator protein PrtN	98.81	1.4e-7	43	6.2	59	73
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
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

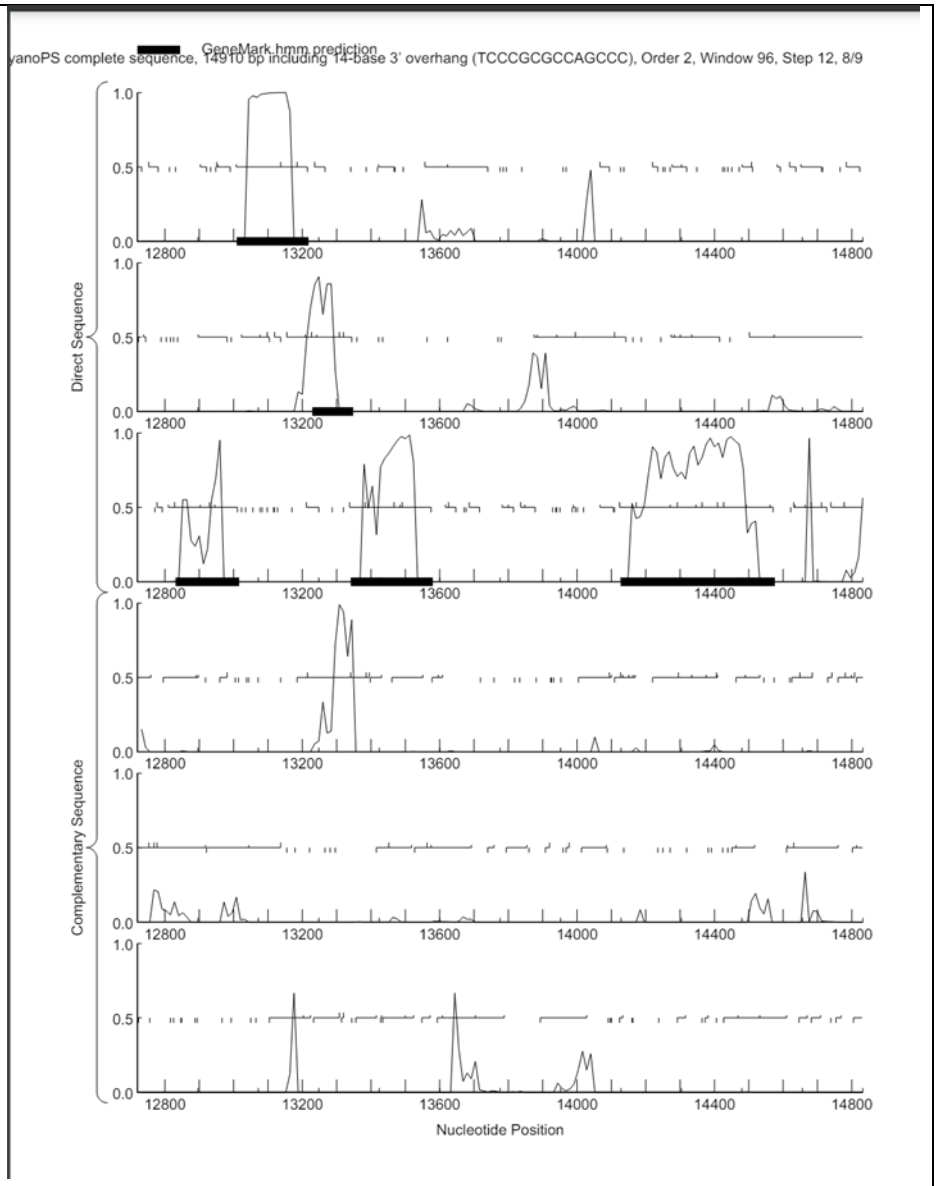
excise

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

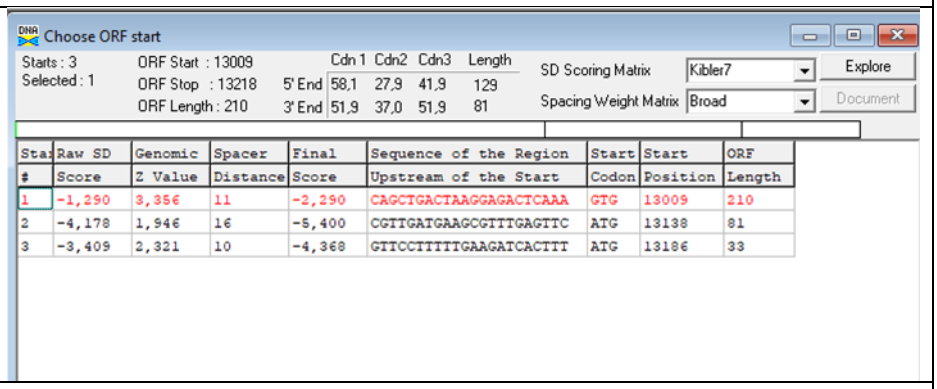
Décision #1 : Est-ce un gène ?

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** Mouse-over to show defline and scores. Click to show alignments Color Key for Alignment Scores <40 40-50 50-80 80-200 >=200 lcl|1_ 0 50 Sequences producing significant alignments: Score (bits) E Value CyranoPS_Draft_20, function unknown, 69 140 9e-34 >CyranoPS_Draft_20, function unknown, 69 Length = 69 Score = 140 bits (354), Expect = 9e-34 Identities = 69/69 (100%), Positives = 69/69 (100%) Query: 1 MLKEDRDWAIKLTASILEGSINQREIENSKYDLEAIVDEAFEFMKETGTYGLFLFEDHFM 60 MLKEDRDWAIKLTASILEGSINQREIENSKYDLEAIVDEAFEFMKETGTYGLFLFEDHFM Sbjct: 1 MLKEDRDWAIKLTASILEGSINQREIENSKYDLEAIVDEAFEFMKETGTYGLFLFEDHFM 60 Query: 61 EIARKHRRA 69 EIARKHRRA Sbjct: 61 EIARKHRRA 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 ?

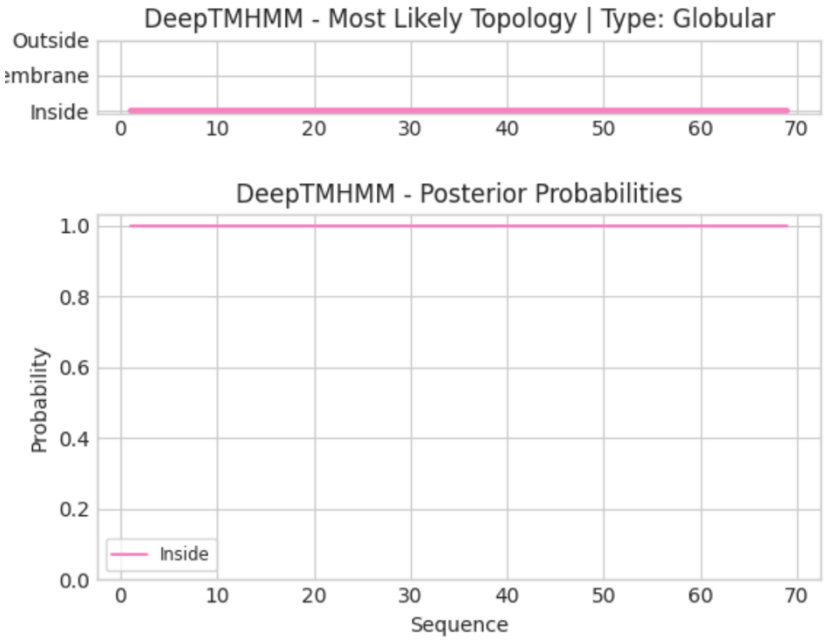
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : 13009 Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : 13009
Est-ce que le start est associé à un RBS (Ribosome Binding Site) de bon score ?	
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

Sta #	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	-1,290	3,356	11	-2,290	CAGCTGACTAAGGAGACTCAAA	GTG	13009	210
2	-4,178	1,946	16	-5,400	CGTTGATGAAGCGTTTGAGTTC	ATG	13138	81
3	-3,409	2,321	10	-4,368	GTTCCTTTTTGAAGATCACTTT	ATG	13186	33

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

Collection des éléments de réponse	Rationnelle

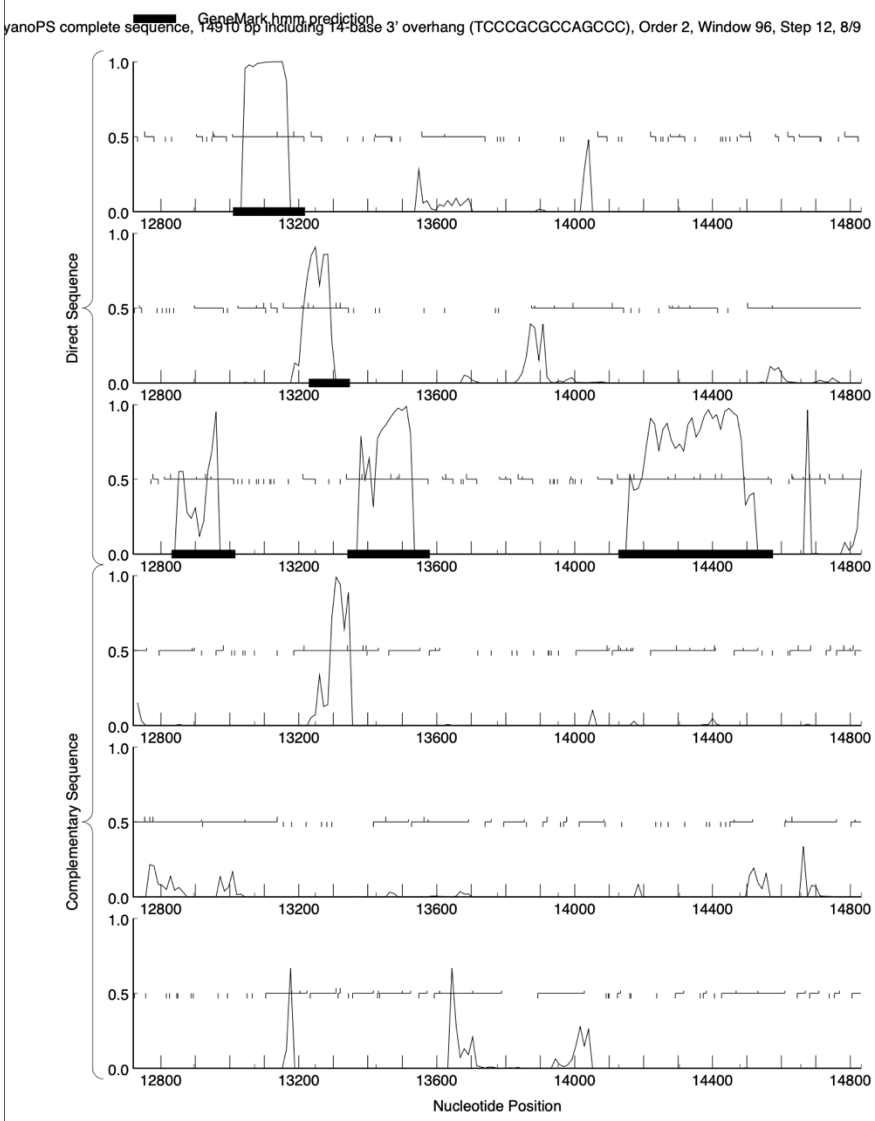
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	**Listez le meilleur hit Blastp pour chaque source :** ***** PhagesDB :** **(q#: s#) : [alignment] ; e-value :** **Rien** ***** 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 ?	
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**

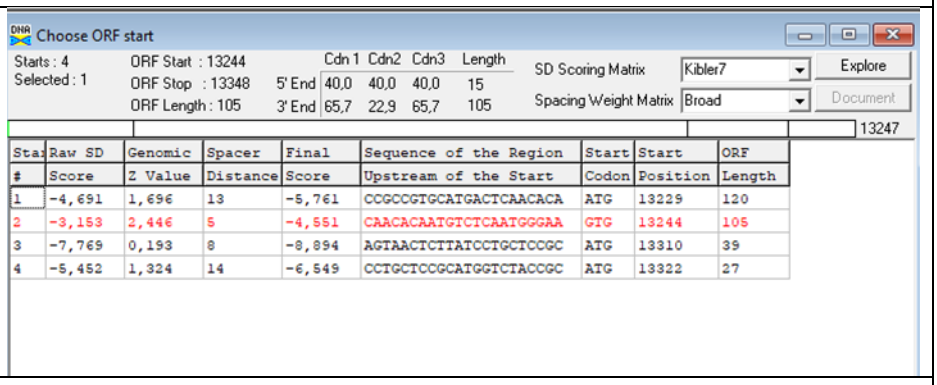
Student Gene Annotation Worksheet

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 ?

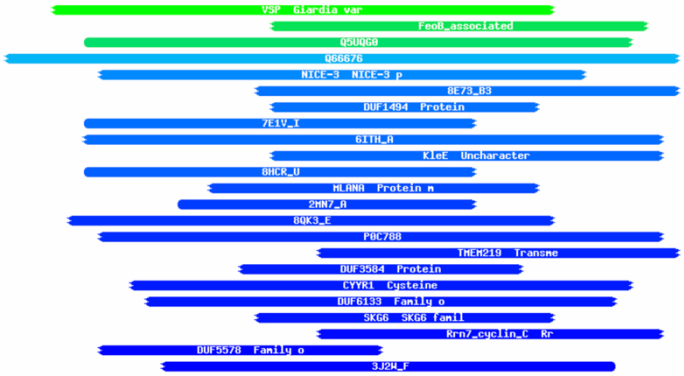
Collection des éléments de réponse	Rationnelle
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**YES BOTH**
Y-a-t-il des éléments supportant un potentiel codant ?	**GeneMarkS** 
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**non**
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**non**
DECISION:	**OUI**

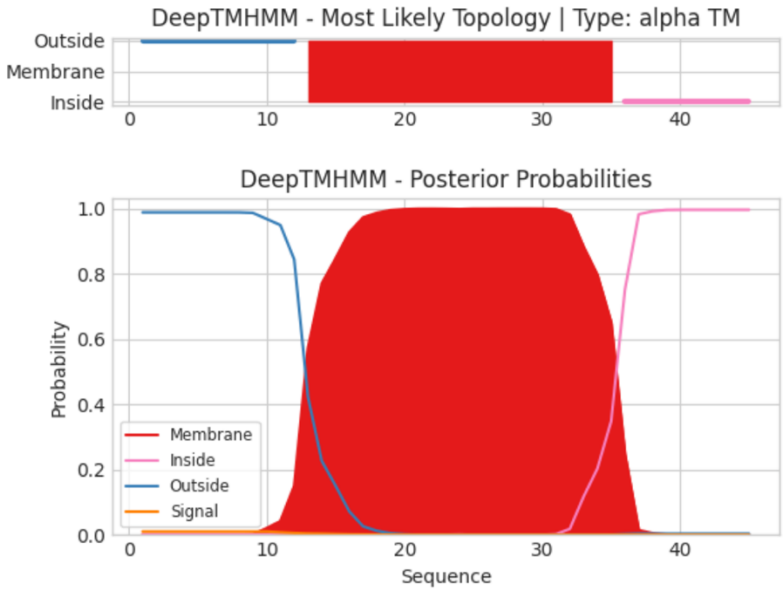
Décision #2 : Quel est le start du gène ?

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 ?	
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 function de la protéine putative ?

Collection des éléments de réponse	Rationnelle

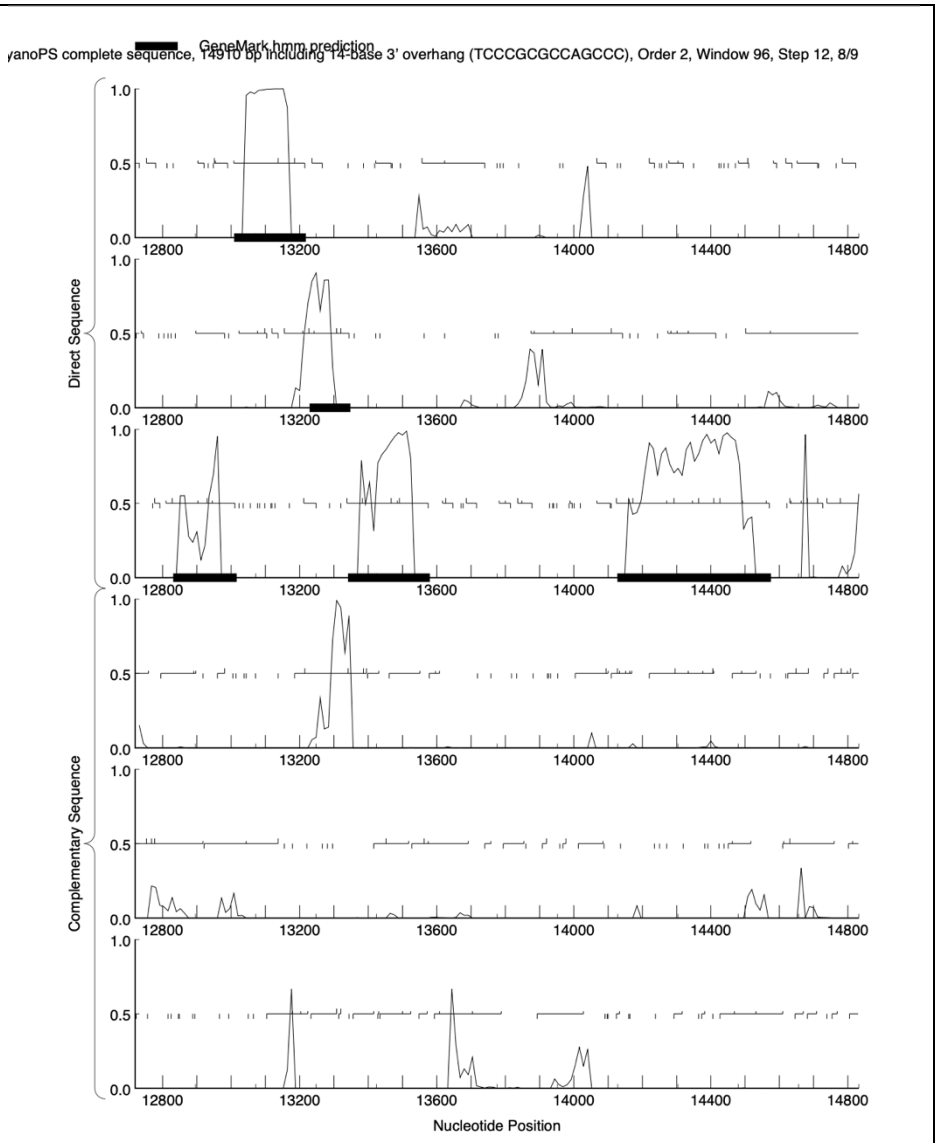
Est-ce que la séquence protéique s'aligne avec une protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	**Listez le meilleur hit Blastp pour chaque source :** ***** PhagesDB :** **rien** ***** nr :** **rien**
Est-ce que la séquence 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 ?	**PFAM proba de 70% : VSP ; Giardia variant-specific surface protein**  | Nr | Hit | Name | Probability | E-value | Score | SS | Aligned cols | Target Length | |---|---|---|---|---|---|---|---|---| | 1 | PF03302.17 | ; VSP ; Giardia variant-specific surface protein | 70.05 | 27 | 21.55 | 3.6 | 33 | 403 | | 2 | PF12669.11 | ; FeoB_associated ; FeoB-associated Cys-rich membrane protein | 64.92 | 33 | 17.18 | 3.8 | 25 | 43 | | 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 | | 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 | | 5 | PF07406.15 | ; NICE-3 ; NICE-3 protein | 51.64 | 90 | 19.5 | 3.7 | 31 | 178 | | 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 |
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	**"Aucune synténie observée".**

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**OUI.** **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 ?	**non**
DECISION:	**Membrane protein**

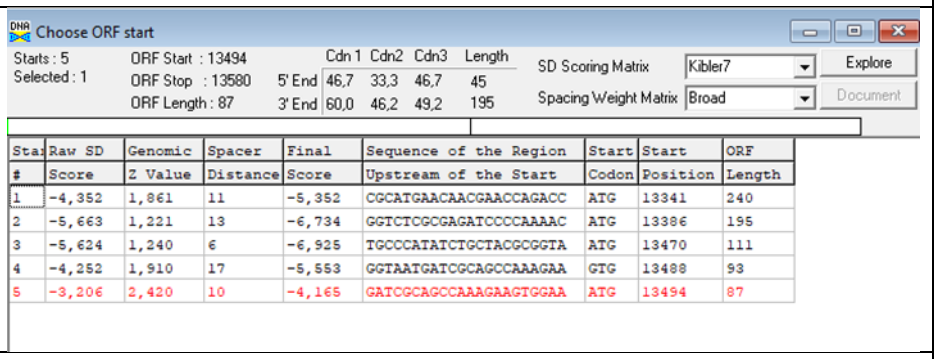
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**22**
Coordonnées du Stop	**13580**
Direction (For/Rev)	**for**
Gap/chevauchement avec un autre gène	**Ove 8**
Coordonnées du Start retenu	**13 341**
Fonction prédite	**NFK**

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 ?	**GeneMarkS :**	

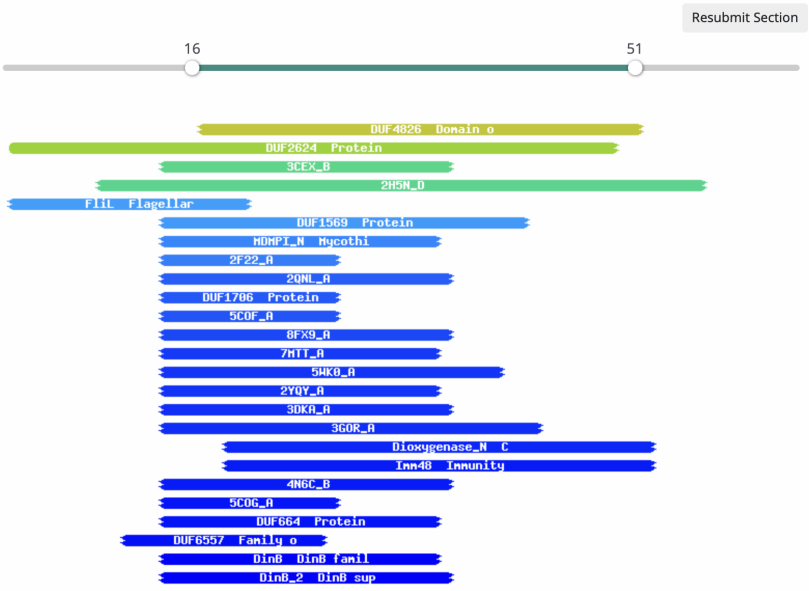
	
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**rien**
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**Pas d'homologues**
DECISION:	**OUI**

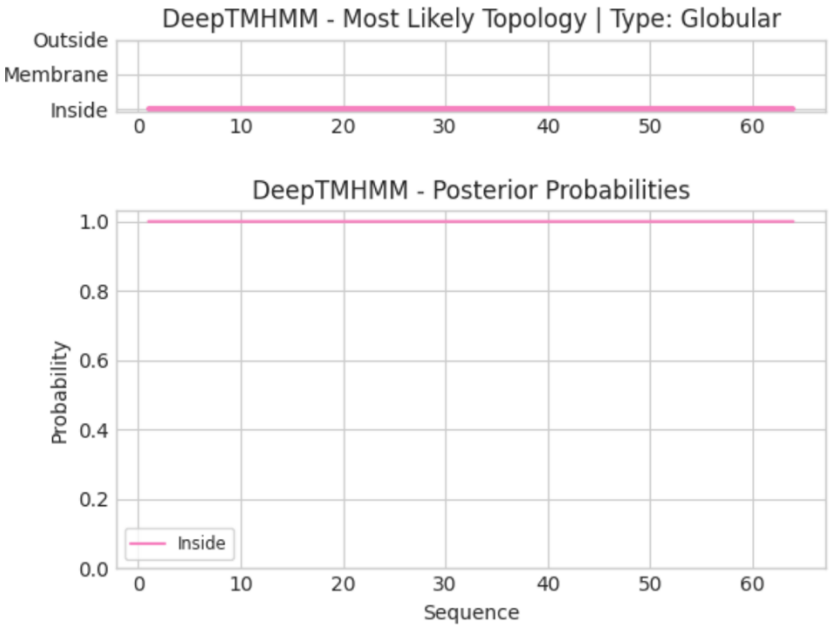
Décision #2 : Quel est le start du gène ?

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) : 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 ?	 Sta# / 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 / GGTCTCGCGAGATCCCCAAAAC / ATG / 13386 / 195 3 / -5,624 / 1,240 / 6 / -6,925 / TGCCCATATCTGCTACGCGGTA / 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 ⚠ No significant similarity found. For reasons why,click here
DECISION:	13 341

Décision #3 : Quelle est la function 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 :

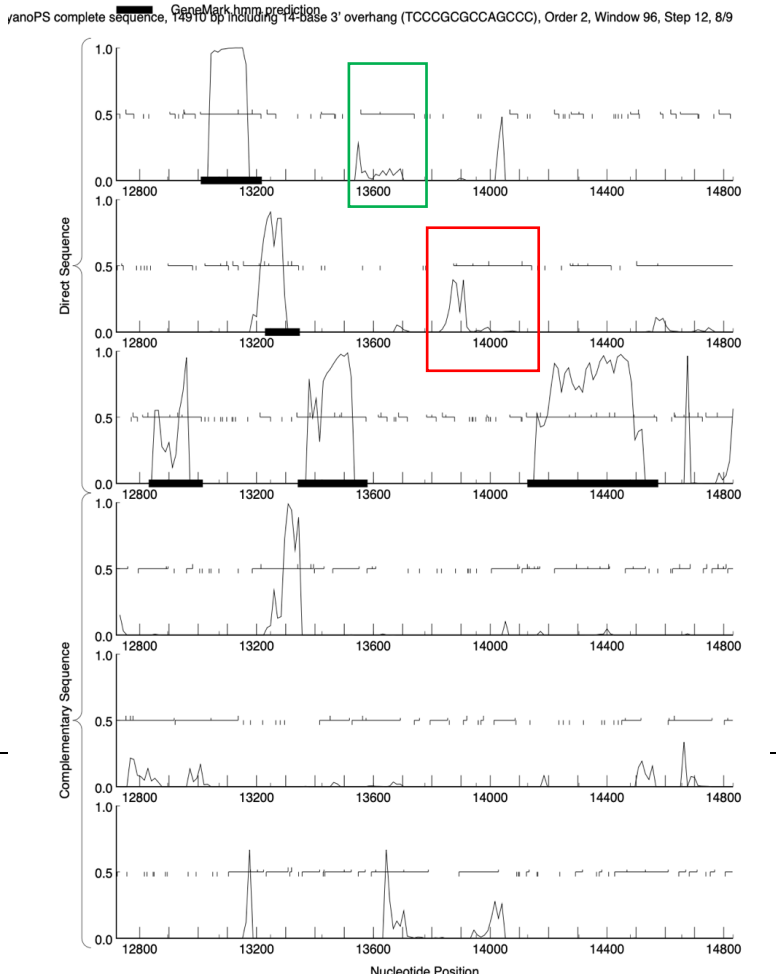
protéine de fonction annotée lors d'un BlastP (phagesDB et/ou nr) avec une e-value < 10^{-4} et une couverture acceptable ?	***** PhagesDB :** **rien** ***** nr :** **rien** ⚠ No significant similarity found. For reasons why,click here
Est-ce que la séquence protéique s'aligne avec une protéine de fonction 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)** 1 PF16108.9 ; DUF4826 ; Domain of unknown function (DUF4826) 81.31 13 24.54 4.6 36 126 2 PF11116.12 ; DUF2624 ; Protein of unknown function (DUF2624) 78.5 25 21.65 5 43 83 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 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 5 PF03748.18 ; FliL ; Flagellar basal body-associated protein FliL 53.35 31 19.41 1.8 20 92 6 PF07606.15 ; DUF1569 ; Protein of unknown function (DUF1569) 53.06 80 18.36 3.4 30 151 7 PF11716.12 ; MDMPI_N ; Mycothiol maleylpyruvate isomerase N-terminal domain 51.35 68 16.56 2.8 23 135 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 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 transcirptor**

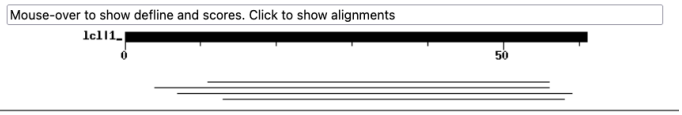
dans une région du génome qui montre une forte conservation de l'ordre des gènes ?	
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	
Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	
DECISION:	**NKF**

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 ?	**Green frame below** 

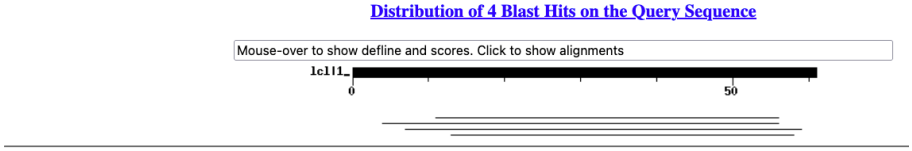
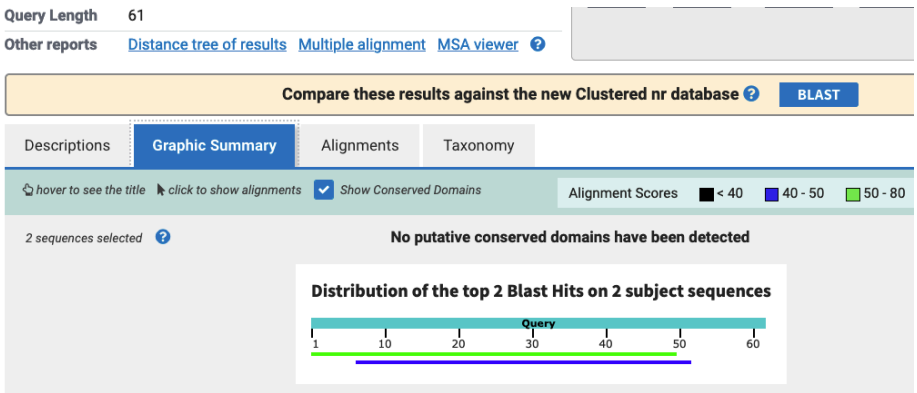
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**Query=** (61 letters) **Distribution of 4 Blast Hits on the Query Sequence**  Sequences producing significant alignments: Score (bits) E Value Schiebs_21, function unknown, 62 32 0.74 EpicDab_22, function unknown, 106 30 2.2 Mbo4_11, minor capsid protein, 156 29 4.8 Donovan_48, function unknown, 130 29 4.8 >Schiebs_21, function unknown, 62 Length = 62 Score = 31.6 bits (70), Expect = 0.74 Identities = 20/47 (42%), Positives = 24/47 (51%), Gaps = 6/47 (12%) Query: 12 PSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCK-HRPALNPP 57 P + H +C CRR FAA VP DG IC DC+ + A PP Sbjct: 11 PDRASAH--RCECCRRKFAAGHVPT---DGLLICRDCRAEQEAAAPP 52 >EpicDab_22, function unknown, 106 Length = 106 Score = 30.0 bits (66), Expect = 2.2 Identities = 19/55 (34%), Positives = 25/55 (45%), Gaps = 6/55 (10%) Query: 5 PHTNEFRPSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCKHRPAL--NPP 57 P T E R + +CV C R+FA P D+ +C DC+ AL PP Sbjct: 48 PSTTEQRRAALLAAKPRCVDCERLFAPRVEPPRDL----LCRDCRTARALAQEPP 98
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**This candidate gene was revealed when we analyzed large gaps in the annotation (gap 546 bp)**
DECISION:	**OUI**

Décision #2 : Quel est le start du gène ?

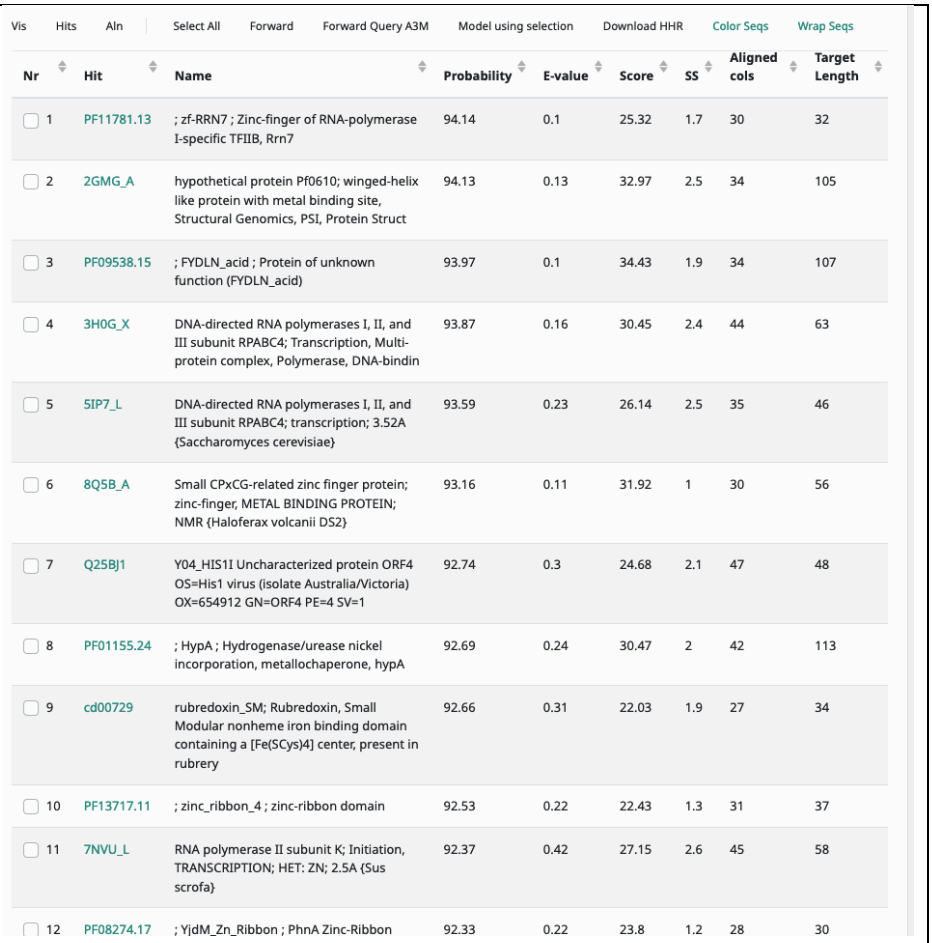
Collection des éléments de réponse	Rationnelle
Quel start Glimmer et GeneMark suggèrent-ils ?	**Coordonnées du start données par Glimmer (mettre NA si ne donne pas de résultats) : NA** **Coordonnées du start données par GeneMark (mettre NA si ne donne pas de résultats) : NA**
Est-ce que le start est 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 ?	**NON** **Query=** (61 letters) Distribution of 4 Blast Hits on the Query Sequence Mouse-over to show defline and scores. Click to show alignments lcl|1_ 0 50 Sequences producing significant alignments: 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 >Schiebs_21, function unknown, 62 Length = 62 Score = 31.6 bits (70), Expect = 0.74 Identities = 20/47 (42%), Positives = 24/47 (51%), Gaps = 6/47 (12%) Query: 12 PSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCK-HRPALNPP 57 P + H +C CRR FAA VP DG IC DC+ + A PP Sbjct: 11 PDRASAH--RCECCRRKFAAGHVPT---DGLLICRDCRAEQEAAAPP 52 >EpicDab_22, function unknown, 106 Length = 106 Score = 30.0 bits (66), Expect = 2.2 Identities = 19/55 (34%), Positives = 25/55 (45%), Gaps = 6/55 (10%) Query: 5 PHTNEFRPSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCKHRPAL--NPP 57 P T E R + +CV C R+FA P D+ +C DC+ AL PP Sbjct: 48 PSTTEQRRAALLAAKPRCVDCERLFAPRVEPPRDL----LCRDCRTARALAQEPP 98
Est-ce que le start est conservé chez d'autres homologues retrouvés par Blastp ?	

DECISION:	

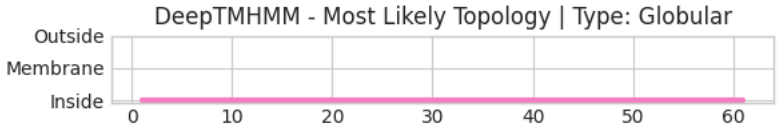
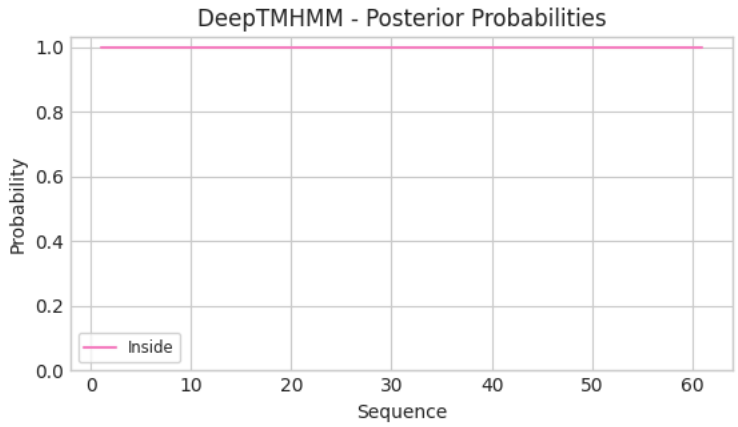
Décision #3 : Quelle est la function de la protéine putative ?

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 :** **Query=** (61 letters)  Sequences producing significant alignments: 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 >Schiebs_21, function unknown, 62 Length = 62 Score = 31.6 bits (70), Expect = 0.74 Identities = 20/47 (42%), Positives = 24/47 (51%), Gaps = 6/47 (12%) Query: 12 PSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCK-HRPALNPP 57 P + H +C CRR FAA VP DG IC DC+ + A PP Sbjct: 11 PDRASAH--RCECCRRKFAAGHVPT---DGLLICRDCRAEQEAAAPP 52 >EpicDab_22, function unknown, 106 Length = 106 Score = 30.0 bits (66), Expect = 2.2 Identities = 19/55 (34%), Positives = 25/55 (45%), Gaps = 6/55 (10%) Query: 5 PHTNEFRPSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCKHRPAL--NPP 57 P T E R + +CV C R+FA P D+ +C DC+ AL PP Sbjct: 48 PSTTEQRRAALLAAKPRCVDCERLFAPRVEPPRDL----LCRDCRTARALAQEPP 98 ***** nr :** 

	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 62.8 bits(151) Expect 1e-10 Method Compositional matrix adjust. Identities 27/49(55%) Positives 35/49(71%) Gaps 1/49(2%) Query 1 MPIDPHTNEFRPSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCK 49 MP D + E+RP +RDP GR+C C R+FAA VP P +DG R+CTDC+ Sbjct 40 MPWD–ESGEYRPYRRDPEGRRCNHCNRVFAAGCVPVPYLDGTRLCTDCR 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 45.8 bits(107) Expect 2e-04 Method Compositional matrix adjust. Identities 21/45(47%) Positives 27/45(60%) Gaps 0/45(0%) Query 7 TNEFRPSQRDPHGRKCVKCRRIFAADAVPAPDIDGNRICTDCKHR 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 >= 90% et une couverture acceptable ?	Number of Hits: **250** Query MSA diversity (Neff): **3.76964** Visualization Resubmit Section 16 51 zf-RRN7 Zinc-fi 2GMG_A FYDLN_acid Prot 3H0G_X 5IP7_L 8Q5B_A Q256J1 HypA Hydrogenas cd00729 zinc_ribbon_4 z 7NVU_L YjdM_Zn_Ribbon 2KDX_A 8HIM_L 1SPW_A 5K2M_N 7ZOK_A zinc_ribbon_5 z Cys_rich_KTR Cy

	Vis Hits Aln | Select All Forward Forward Query A3M Model using selection Download HHR Color Seqs Wrap Seqs  **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".**

Nr	Hit	Name	Probability	E-value	Score	SS	Aligned cols	Target Length
1	PF11781.13	; zf-RRN7 ; Zinc-finger of RNA-polymerase I-specific TFIIB, Rrn7	94.14	0.1	25.32	1.7	30	32
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
3	PF09538.15	; FYDLN_acid ; Protein of unknown function (FYDLN_acid)	93.97	0.1	34.43	1.9	34	107
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
5	5IP7_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
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
7	Q25BJ1	Y04_HIS1I Uncharacterized protein ORF4 OS=His1 virus (isolate Australia/Victoria) OX=654912 GN=ORF4 PE=4 SV=1	92.74	0.3	24.68	2.1	47	48
8	PF01155.24	; HypA ; Hydrogenase/urease nickel incorporation, metallochaperone, hypA	92.69	0.24	30.47	2	42	113
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
10	PF13717.11	; zinc_ribbon_4 ; zinc-ribbon domain	92.53	0.22	22.43	1.3	31	37
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
12	PF08274.17	; YjdM_Zn_Ribbon ; PhnA Zinc-Ribbon	92.33	0.22	23.8	1.2	28	30

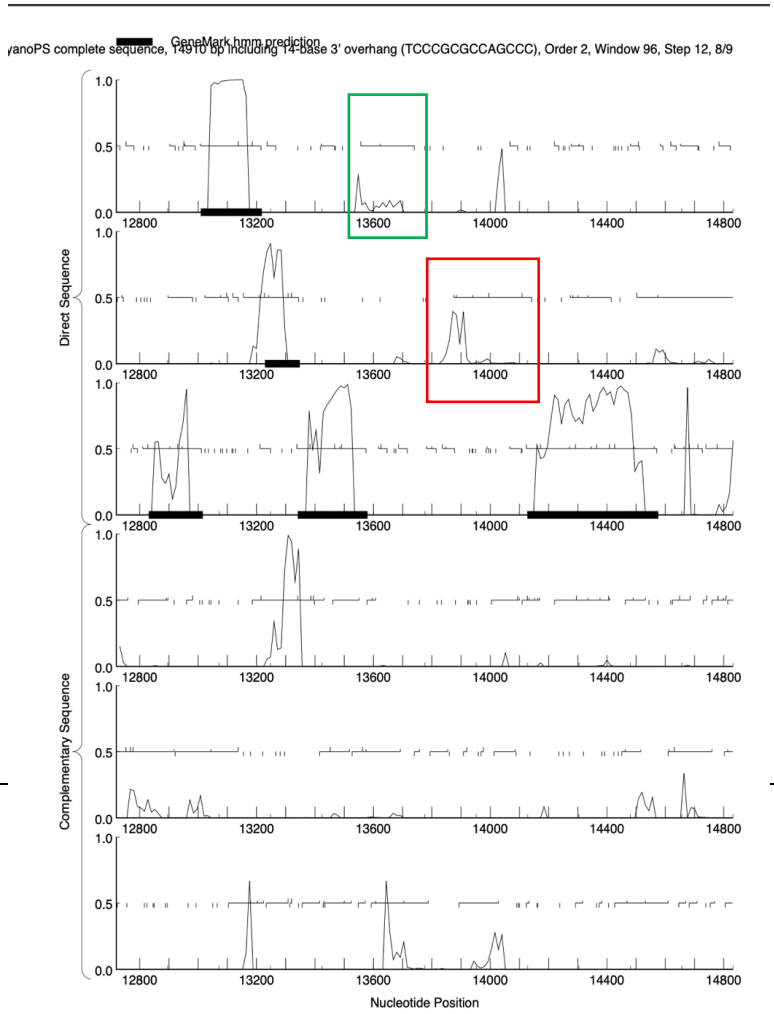
Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	**DeepTMHMM - Predictions** Predicted topologies can be downloaded in .gff3 format and .3line format   You can download the probabilities used to generate this plot here **Predicted Topologies** >Unnamed | GLOB
Est-ce que la fonction proposée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

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 ?	**Red frame below** 

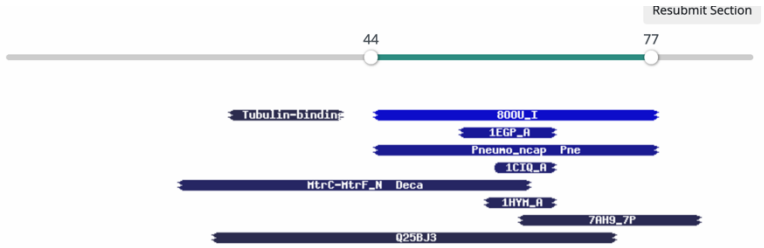
Est-ce que le candidat est retrouvé chez d'autres génomes annotés ?	**Non** Query Length 89 Other reports No significant similarity found. For reasons why,click here
Est-ce que le candidat est en contradiction avec les principes d'annotation ?	**We are filling a gap (546 bp)**
DECISION:	**OUI**

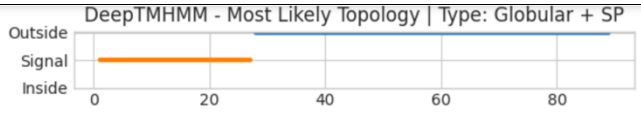
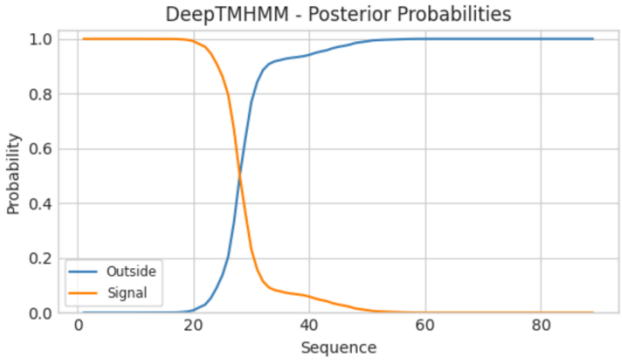
Décision #2 : Quel est le start du gène ?

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) : 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 function 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 1 | 8OOU_I | Nucleoprotein; D10-symmetry, N-RNA, nucleoprotein, RSV, VIRAL PROTEIN; 2.9A {Trichoplusia ni} | 36.17 | 68 | 29.25 | 2.1 | 34 | 391 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 3 | PF03246.18 | ; Pneumo_ncap ; Pneumovirus nucleocapsid protein | 30.92 | 71 | 29.17 | 1.4 | 34 | 392 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 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".**

Est-ce que ce gène code pour une protéine transmembranaire (TM) ?	  You can download the probabilities used to generate this plot here **Predicted Topologies** >Unnamed | SP MLDVASQTTSFNIPALLRLIPVVTFLVSSRPNEPYRPIGGMTLTFLTSVGDTRALPPSRRSWPECGAILPLTHNHYQLMIGTIHEHTNP SSSS

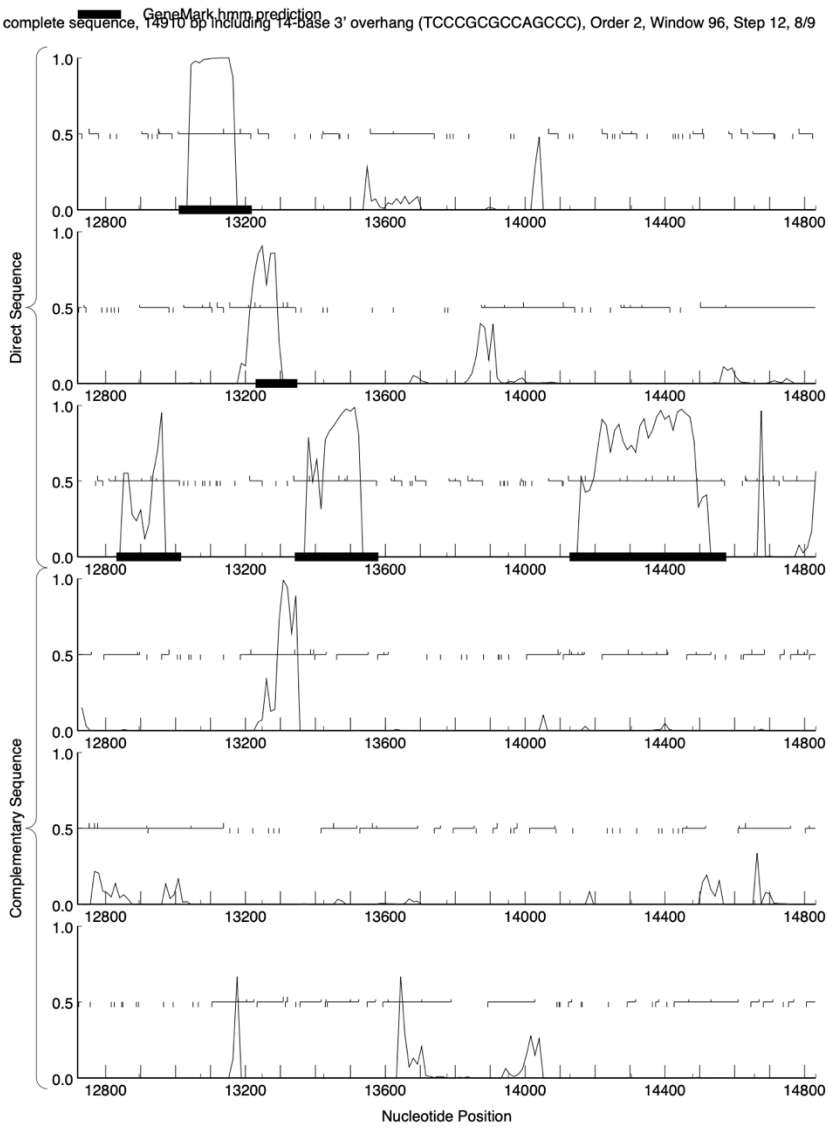
Student Gene Annotation Worksheet

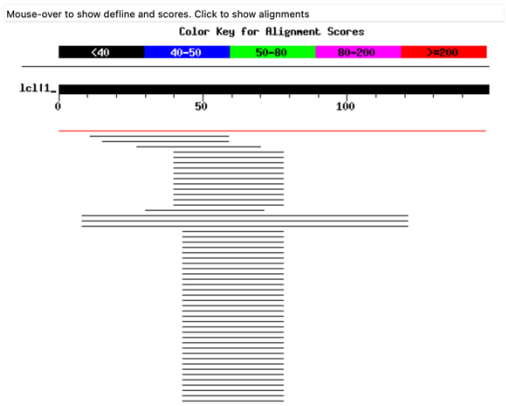
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**25**
Coordonnées du Stop	**14576**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	**Gap 546**
Coordonnées du Start retenu	**14 127**
Fonction prédite	**WhiB family transcription factor**

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 ?	

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

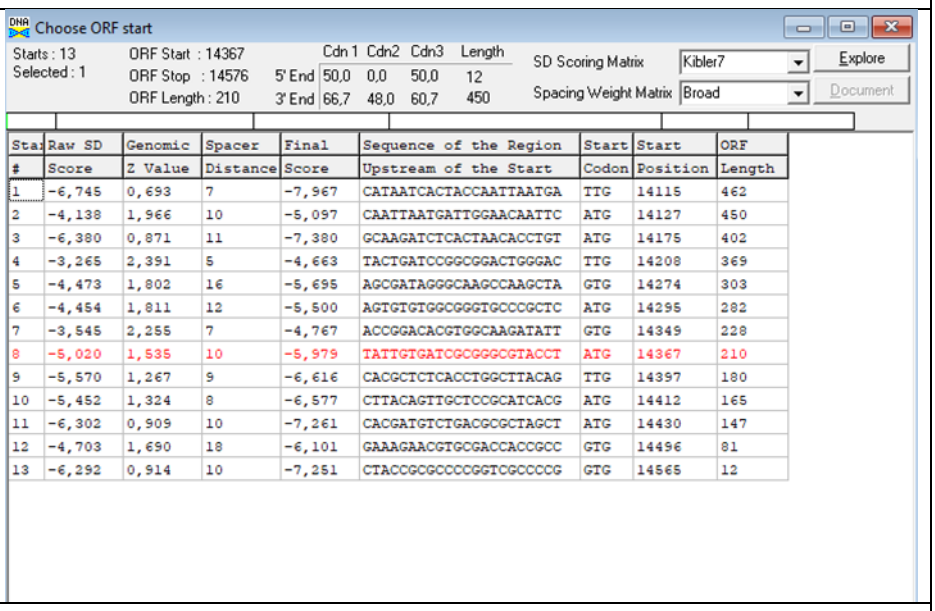
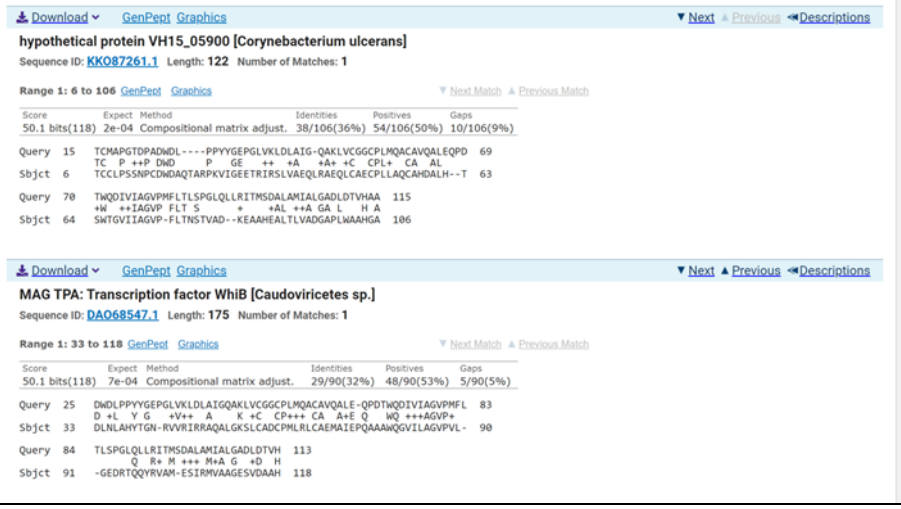




Sequences producing significant alignments:	Score (bits)	E Value
CyranoPS_Draft_23, function unknown, 149	308	3e-84
Trax_62, WhiB family transcription factor, 251	36	0.039
Neville_63, WhiB family transcription factor, 251	35	0.051
Rimmer_91, WhiB family transcription factor, 86	34	0.15
Petra64142_Draft_98, function unknown, 86	34	0.15
Pat3_90, function unknown, 86	34	0.15
mu16_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
Quel start Glimmer et GeneMark suggèrent-ils ?	**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**
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 ?	**ORF la plus longue → 462pb TTg 14115, gap 534pb** **ORF de 450 → ATG 14127 → 546pb**
Est-ce que le start est conservé chez les homologues voir Starterator ?	

Choose ORF start

Starts : 13 Selected : 1 | ORF Start : 14367 | ORF Stop : 14576 | ORF Length : 210 | Cdn 1 Cdn2 Cdn3 Length | 5' End 50,0 0,0 50,0 12 | 3' End 66,7 48,0 60,7 450 | SD Scoring Matrix: Kibler7 | Spacing Weight Matrix: Broad | Explore | Document

Sta#	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	-6,745	0,693	7	-7,967	CATAATCACTACCAATTAATGA	TTG	14115	462
2	-4,138	1,966	10	-5,097	CAATTAATGATTGGAACAATTC	ATG	14127	450
3	-6,380	0,871	11	-7,380	GCAAGATCTCACTAACACCTGT	ATG	14175	402
4	-3,265	2,391	5	-4,663	TACTGATCCGGCGGACTGGGAC	TTG	14208	369
5	-4,473	1,802	16	-5,695	AGCGATAGGGCAAGCCAAGCTA	GTG	14274	303
6	-4,454	1,811	12	-5,500	AGTGTGTGGCGGGTGCCCGCTC	ATG	14295	282
7	-3,545	2,255	7	-4,767	ACCGGACACGTGGCAAGATATT	GTG	14349	228
8	-5,020	1,535	10	-5,979	TATTGTGATCGCGGGGCGTACCT	ATG	14367	210
9	-5,570	1,267	9	-6,616	CACGCTCTCACCTGGCTTACAG	TTG	14397	180
10	-5,452	1,324	8	-6,577	CTTACAGTTGCTCCGCATCACG	ATG	14412	165
11	-6,302	0,909	10	-7,261	CACGATGTCTGACGCGCTAGCT	ATG	14430	147
12	-4,703	1,690	18	-6,101	GAAAGAACGTGCGACCACCGCC	GTG	14496	81
13	-6,292	0,914	10	-7,251	CTACCGCGCCCCGGTCGCCCCG	GTG	14565	12

Download GenPept Graphics ▼Next ▲Previous ◀◀Descriptions

hypothetical protein VH15_05900 [Corynebacterium ulcerans]

Sequence ID: KKO87261.1 Length: 122 Number of Matches: 1

Range 1: 6 to 106 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
50.1 bits(118)	2e-04	Compositional matrix adjust.	38/106(36%)	54/106(50%)	10/106(9%)

```
Query  15   TCMAPGTDPADWDL----PPYYGEPGLVKLDLAIG-QAKLVCGGCPLMQACAVQALEQPD  69
            TC  P ++P DWD     P   GE    ++  +A    +A+ +C  CPL+  CA  AL
Sbjct  6    TCCLPSSNPCDWDAQTARPKVIGEETRIRSLVAEQLRAEQLCAECPLLAQCAHDALH--T  63

Query  70   TWQDIVIAGVPMFLTLSPGLQLLRITMSDALAMIALGADLDTVHAA  115
            +W  ++IAGVP FLT S      +    +AL ++A GA L   H A
Sbjct  64   SWTGVIIAGVP-FLTNSTVAD--KEAAHEALTLVADGAPLWAAHGA  106
```

Download GenPept Graphics ▼Next ▲Previous ◀◀Descriptions

MAG TPA: Transcription factor WhiB [Caudoviricetes sp.]

Sequence ID: DAO68547.1 Length: 175 Number of Matches: 1

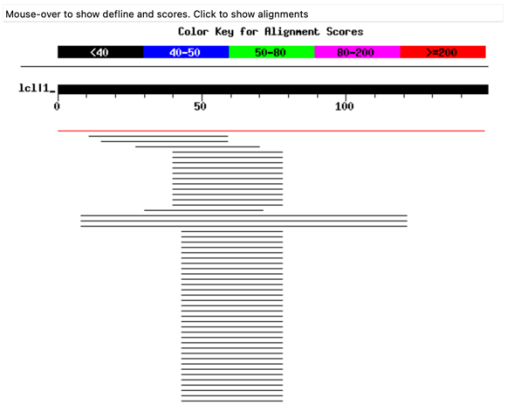
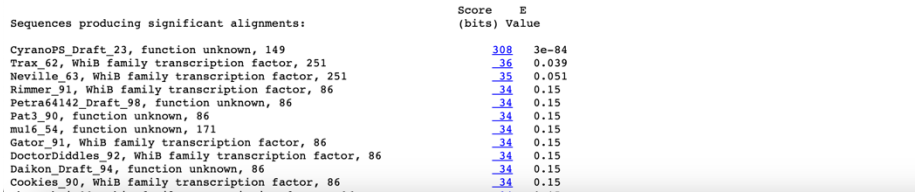
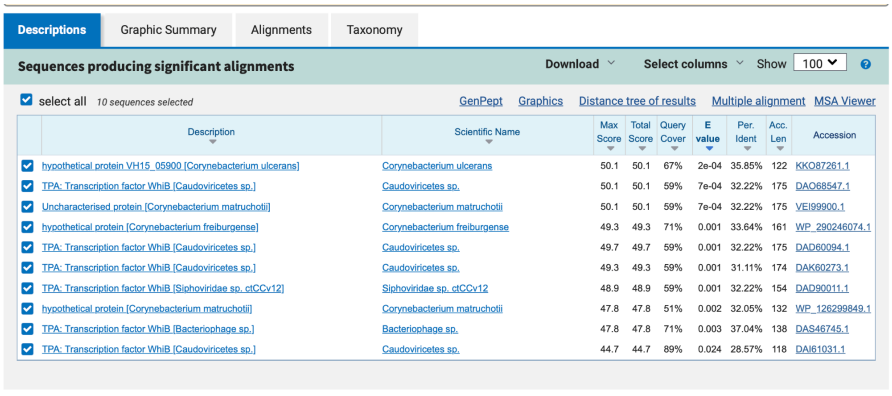
Range 1: 33 to 118 GenPept Graphics ▼Next Match ▲Previous Match

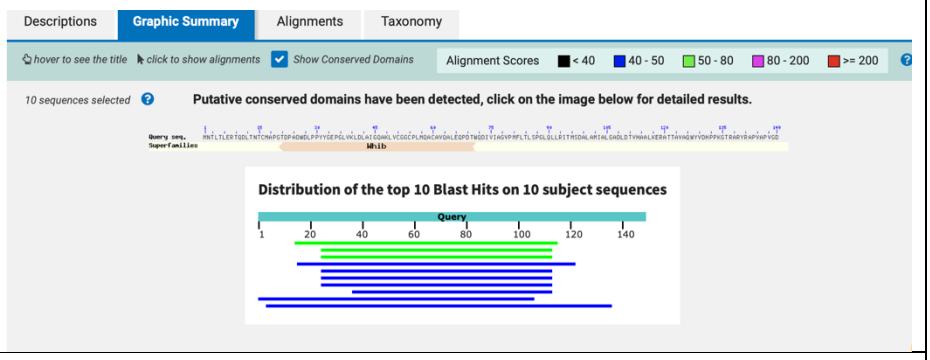
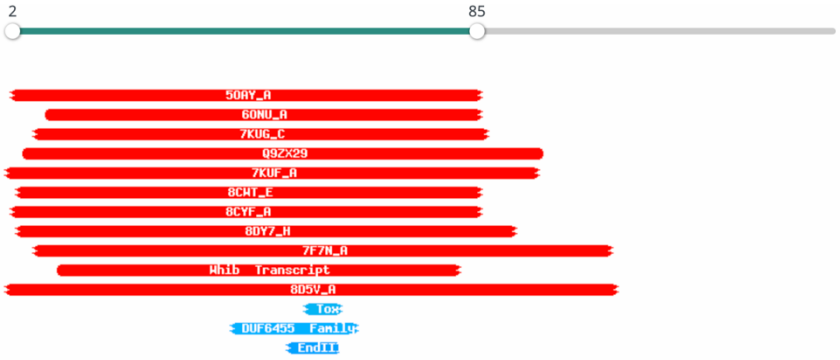
Score	Expect	Method	Identities	Positives	Gaps
50.1 bits(118)	7e-04	Compositional matrix adjust.	29/90(32%)	48/90(53%)	5/90(5%)

```
Query  25   DWDLPPYYGEPGLVKLDLAIGQAKLVCGGCPLMQACAVQALE-QPDTWQDIVIAGVPMFL  83
            D +L  Y G   +V++  A    K +C  CP+++ CA  A+E Q   WQ +++AGVP+
Sbjct  33   DLNLAHYTGN-RVVRIRRAQALGKSLCADCPMLRLCAEMAIEPQAAAWQGVILAGVPVL-  90

Query  84   TLSPGLQLLRITMSDALAMIALGADLDTVH  113
              Q  R+ M +++ M+A G  +D  H
Sbjct  91   -GEDRTQQYRVAM-ESIRMVAAGESVDAAH  118
```

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 function 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 :**   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 ***** nr :**  Descriptions Graphic Summary Alignments Taxonomy Sequences producing significant alignments Download Select columns Show 100 select all 10 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 VH15_05900 [Corynebacterium ulcerans] Corynebacterium ulcerans 50.1 50.1 67% 2e-04 35.85% 122 KKO87261.1 TPA: Transcription factor WhiB [Caudoviricetes sp.] Caudoviricetes sp. 50.1 50.1 59% 7e-04 32.22% 175 DAO68547.1 Uncharacterised protein [Corynebacterium matruchotii] Corynebacterium matruchotii 50.1 50.1 59% 7e-04 32.22% 175 VEI99900.1 hypothetical protein [Corynebacterium freiburgense] Corynebacterium freiburgense 49.3 49.3 71% 0.001 33.64% 161 WP_290246074.1 TPA: Transcription factor WhiB [Caudoviricetes sp.] Caudoviricetes sp. 49.7 49.7 59% 0.001 32.22% 175 DAD60094.1 TPA: Transcription factor WhiB [Caudoviricetes sp.] Caudoviricetes sp. 49.3 49.3 59% 0.001 31.11% 174 DAK60273.1 TPA: Transcription factor WhiB [Siphoviridae sp. ctCCv12] Siphoviridae sp. ctCCv12 48.9 48.9 59% 0.001 32.22% 154 DAD90011.1 hypothetical protein [Corynebacterium matruchotii] Corynebacterium matruchotii 47.8 47.8 51% 0.002 32.05% 132 WP_126299849.1 TPA: Transcription factor WhiB [Bacteriophage sp.] Bacteriophage sp. 47.8 47.8 71% 0.003 37.04% 138 DAS46745.1 TPA: Transcription factor WhiB [Caudoviricetes sp.] Caudoviricetes sp. 44.7 44.7 89% 0.024 28.57% 118 DAI61031.1

	
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 : 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** 

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
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
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
4	Q9ZX29	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
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
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
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
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
9	7F7N_A	Transcriptional regulator WhiB4; apo-WhiB4, helix-turn-helix domain, intrinsically	99.57	8.7e-15	106.93	5.8	88	126

10	PF02467.20	; Whib ; Transcription factor WhiB	99.56	2.2e-14	92.75	6.4	64	65
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
12	PF08097.15	; Toxin_26 ; Conotoxin T-superfamily	55.61	7.1	17.02	-0.4	7	11
13	PF20056.3	; DUF6455 ; Family of unknown function (DUF6455)	54.95	45	23.17	3.2	23	85
14	PF10576.13	; EndIII_4Fe-2S ; Iron-sulfur binding domain of endonuclease III	53.7	7.6	19.28	-0.5	10	17
15	Q70LD4	Y223_AFV1Y Uncharacterized protein ORF223 OS=Acidianus filamentous virus 1 (isolate United States/Yellowstone) OX=654909	27.05	78	24.01	0.9	13	223
16	cd09637	Cas4_I-A_I-B_I-C_I-D_II-B; CRISPR/Cas system-associated protein Cas4. CRISPR (Clustered Regularly Interspaced Short Pali	26.79	42	23.77	-0.5	10	170
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
18	PF01930.21	; Cas_Cas4 ; Domain of unknown function DUF83	23.56	89	22.35	0.7	10	162
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
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

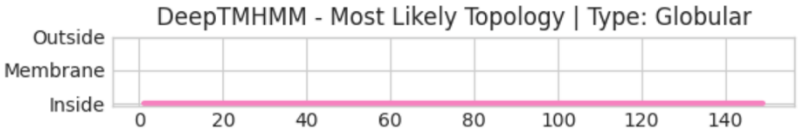
Ce gène est-il situé à côté de gènes de fonction connue et dans une région du génome qui montre une forte conservation de l'ordre des gènes ?

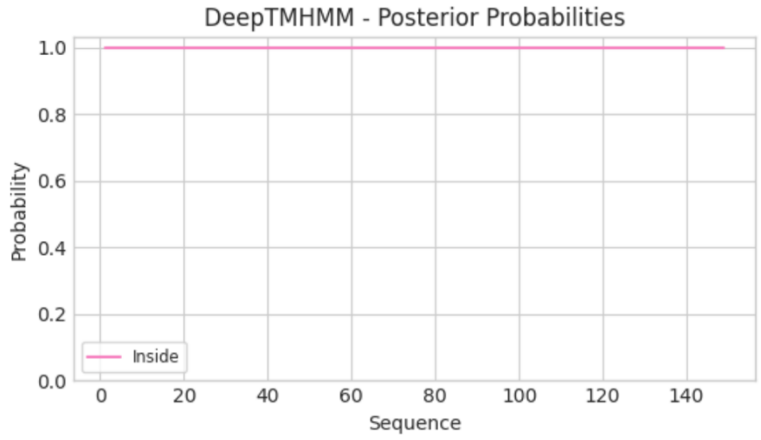
"Aucune synténie observée".

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

DeepTMHMM - Predictions

Predicted topologies can be downloaded in .gff3 format and .3line format



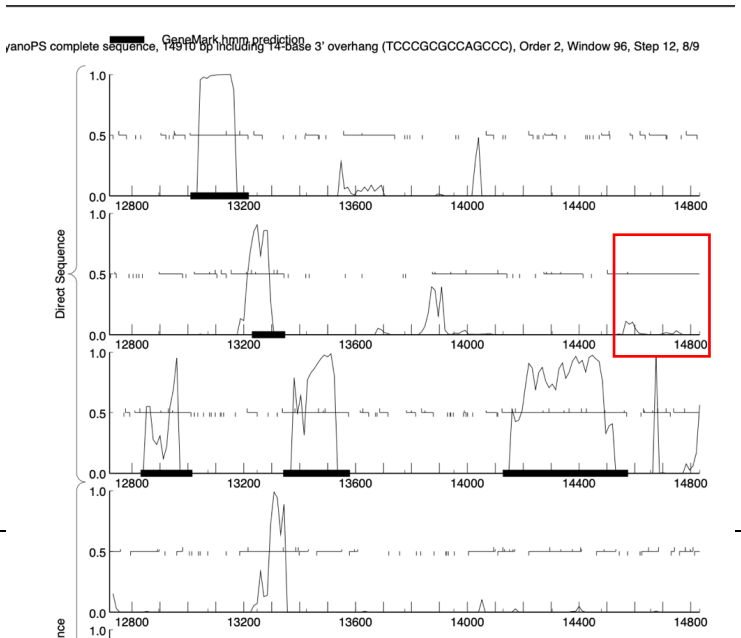


Est-ce que la fonction proposée fait partie de liste de fonctions approuvées par SEA-PHAGES ?	**Oui**
DECISION:	**WhiB family transcription factor**

Student Gene Annotation Worksheet

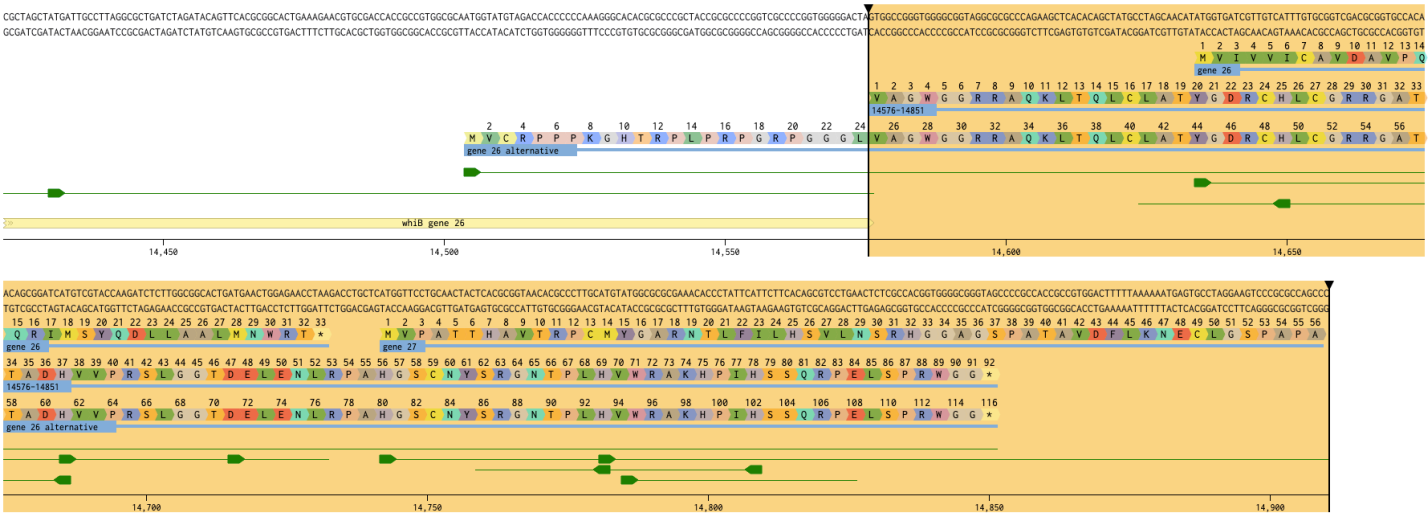
Basic Phage Information	
Nom du Phage	**CyranoPS**
Gène #	**26**
Coordonnées du Stop	**14851**
Direction (For/Rev)	**For**
Gap/chevauchement avec un autre gène	
Coordonnées du Start retenu	**14576**
Fonction prédite	**HNH endonuclease**

Décision #1 : Est-ce un gène ?

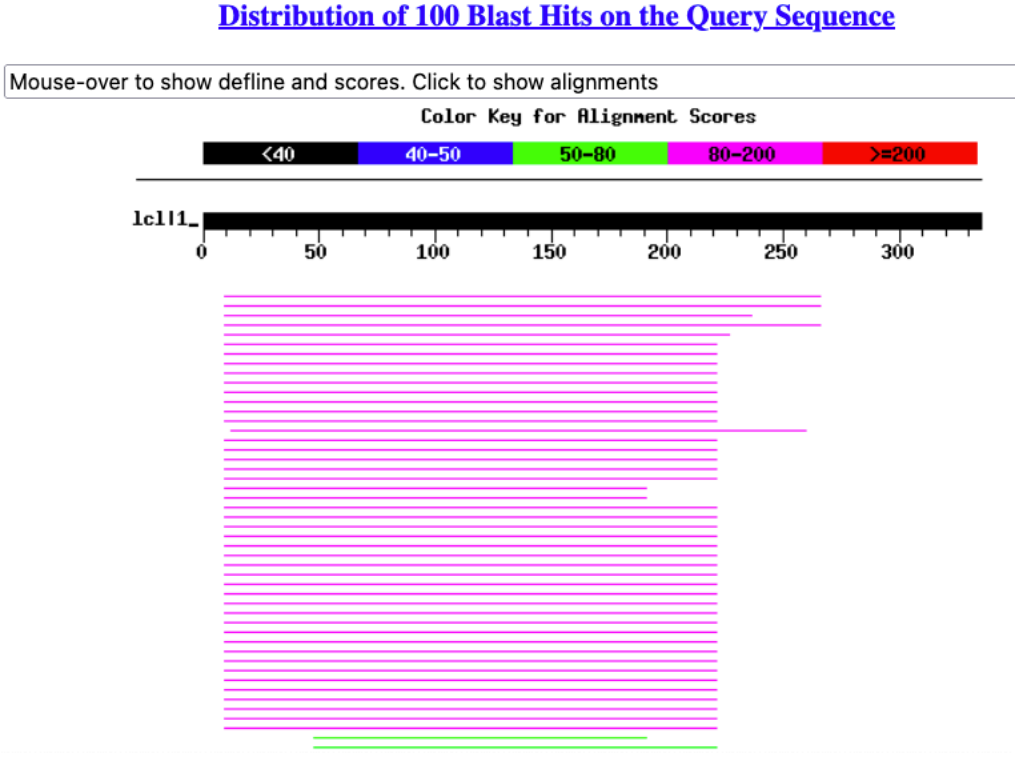
Collection des éléments de réponse	**Rationnelle**
Est-ce que le candidat a été trouvé par un pg d'auto-annotation (Glimmer, GeneMark)?	**NO**
Y-a-t-il des éléments supportant un potentiel codant ?	**Red frame below** 

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)



```
>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  WGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 189
           WGGR AQ LT L L TYGD CHLCGR GATTADH+VPRS GG D L NLRPAHG CN +R
Sbjct: 4   WGGRAAQALTALVLITYGDVCHLCGRPGATTADHLVPRSRGGDDSLGNLRPAHGPCNSAR 63

Query: 190 GNTPLHVWRAKHPIHSSQRPELSPRW 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  WGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 189
           W GR AQ+LT L LATYGD CHLCGR GATTADHVVPR  GG + ++NLRPAH SCN  R
Sbjct: 8   WSGRVAQRLTALTLATYGDVCHLCGRAGATTADHVVPRVQGGDNSIDNLRPAHLSCNSRR 67

Query: 190 GNTPLHVWRAKHPIHSSQRPELSPRW 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
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:	

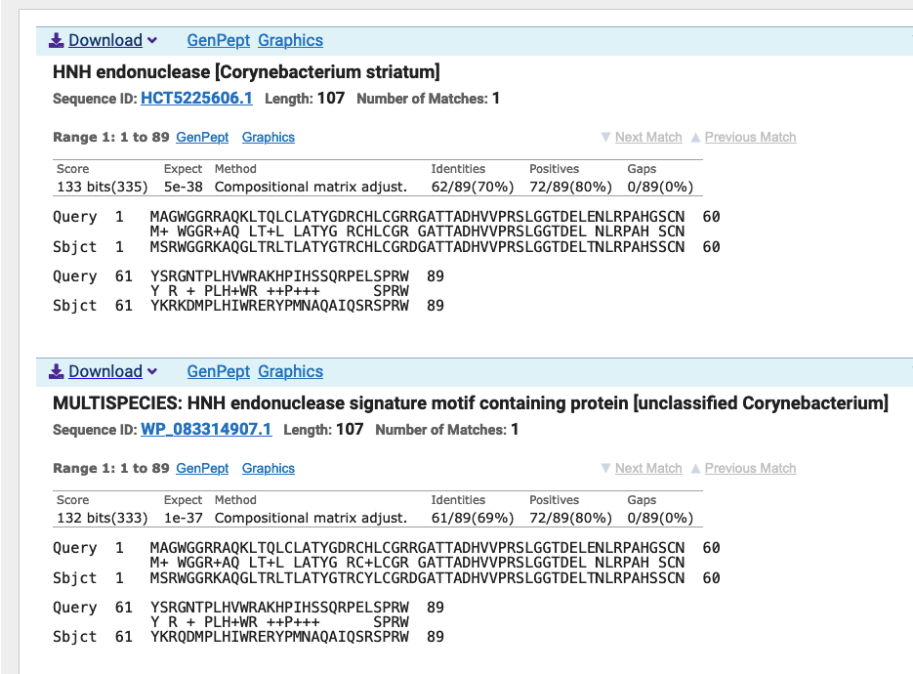
Décision #3 : Quelle est la function de la protéine putative ?

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 :**

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  MAGWGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCN 60
          M+ WGGR AQ LT L L TYGD CHLCGR GATTADH+VPRS GG D L NLRPAHG CN
Sbjct: 1  MSRWGGRAAQALTALVLITYGDVCHLCGRPGATTADHLVPRSRGGDDSLGNLRPAHGPCN 60

Query: 61 YSRGNTPLHVWRAKHPIHSSQRPELSPRW 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  WGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 63
          W GR AQ+LT L LATYGD CHLCGR GATTADHVVPR  GG + ++NLRPAH SCN  R
Sbjct: 8  WSGRVAQRLTALTLATYGDVCHLCGRAGATTADHVVPRVQGGDNSIDNLRPAHLSCNSRR 67

Query: 64 GNTPLHVWRAKHPIHSSQRPELSPRW 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  WGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCNYSR 63
          WGG+ A+ LT L LATYG  CHLCGR GA+TADHV+PRS GG D L+NLRPAH SCN SR
Sbjct: 7  WGGQVAKTLTSLTLATYGRVCHLCGRPGASTADHVIPRSRGGRDALDNLRPAHLSCNQSR 66

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

***** nr :**



Download GenPept Graphics

HNH endonuclease [Corynebacterium striatum]

Sequence ID: HCT5225606.1 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   MAGWGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCN  60
           M+ WGGR+AQ LT+L LATYG RCHLCGR GATTADHVVPRSLGGTDEL NLRPAH SCN
Sbjct  1   MSRWGGRKAQGLTRLTLATYGTRCHLCGRDGATTADHVVPRSLGGTDELTNLRPAHSSCN  60

Query  61  YSRGNTPLHVWRAKHPIHSSQRPELSPRW  89
           Y R + PLH+WR ++P++         SPRW
Sbjct  61  YKRKDMPLHIWRERYPMNAQAIQSRSPRW  89
```

Download GenPept Graphics

MULTISPECIES: HNH endonuclease signature motif containing protein [unclassified Corynebacterium]

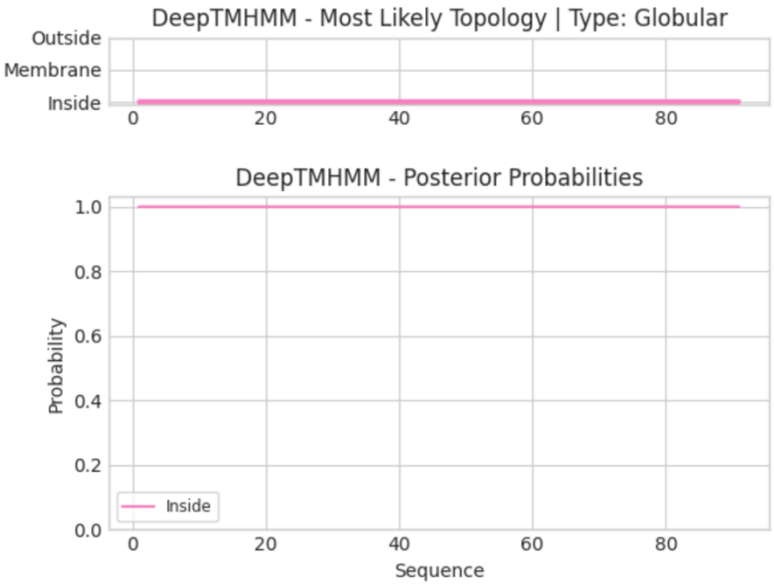
Sequence ID: WP_083314907.1 Length: 107 Number of Matches: 1

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   MAGWGGRRAQKLTQLCLATYGDRCHLCGRRGATTADHVVPRSLGGTDELENLRPAHGSCN  60
           M+ WGGR+AQ LT+L LATYG RC+LCGR GATTADHVVPRSLGGTDEL NLRPAH SCN
Sbjct  1   MSRWGGRKAQGLTRLTLATYGTRCYLCGRDGATTADHVVPRSLGGTDELTNLRPAHSSCN  60

Query  61  YSRGNTPLHVWRAKHPIHSSQRPELSPRW  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 >= 90% et une couverture acceptable ?	Number of Hits: **75** Query MSA diversity (Neff): **12.7961** Visualization **Nr** | **Hit** | **Name** | **Probability** | **E-value** | **Score** | **SS** | **Aligned cols** | **Target Length** 1 | 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 2 | 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 3 | 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 4 | 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 5 | PF09665.15 | ; RE_Alw26IDE ; Type II restriction endonuclease (RE_Alw26IDE) | 97.52 | 0.001 | 39.62 | 4.9 | 79 | 511 6 | 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 7 | 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 8 | 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 9 | P13340 | END7_BPT4 Recombination endonuclease VII OS=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)**