Actinobacteriophage Genome Annotation Submission Cover Sheet

Phage Name. Wrackline

Your Name. Kit Williams

Your Institution. Southern Maine Community College

Your email. kmwprints@gmail.com

Additional emails. (for correspondence). esavage@mainecc.edu, btarbox@mainecc.edu

Describe any issues or specific genes that you would like to highlight for the QC reviewer. This includes any genes that you had questions about or received help with or that warrant further inspection in the QC review process. Include those genes that you deliberated on and/or want to strongly advocate for. If you contacted SMART, workshop facilitator, or a buddy school for help, please document.

* See attached notes at bottom

Please record yes/no for each of the questions below. If further explanation is needed, please add this item to the above box.

In the submitted DNA Master file (Yes/No):

1. Does the genome sequence in your submitted DNA Master file match the nucleotide fasta file posted on phagesDB (same number of bases, no N bases, etc.)?
	1. Yes
2. Are all the genes ‘Valid” when you click the [Validation button](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?
	1. Yes
3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.
	1. Yes, however the second part of my frame shift had to be added manually and when DNA Master organizes by index, this gene appears out of order. When organized by start, it sits in the correct location and is numbered correctly regardless. (SEA\_Wrackline\_25)
4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?
	1. Yes
5. Has the [documentation been recreated](https://seaphagesbioinformatics.helpdocsonline.com/article-86) from the Feature Table to match the latest file version?
	1. Yes
6. Have tRNAs followed the [tRNA protocol](https://seaphagesbioinformatics.helpdocsonline.com/undefined), **COPYING** tRNA-AMINOACID type (DNA equivalent of the anti-codon) from Aragorn output - ﻿tRNA-Gln(ctg) - AND the ends been adjusted to match the Aragorn output?
	1. N/A
7. Has the [frameshift in the tail assembly chaperone](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?
	1. Yes
8. Have you cleared your Draft\_Blast data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?
	1. Yes
9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?
	1. Yes
10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?
	1. Yes
11. Did you [delete the genes](https://seaphagesbioinformatics.helpdocsonline.com/article-65) that you meant to delete?
	1. Yes

Now, [make a profile of the file](https://seaphagesbioinformatics.helpdocsonline.com/article-64) you plan to send. (And you can save this file for [Review to Improve!)](https://seaphagesbioinformatics.helpdocsonline.com/untitled-18)

1. Have any duplicate genes been deleted?
	1. Yes
2. Has the Notes field been cleared (using the automated buttons)?
	1. Yes
3. Do the gene numbers and locus tags match?
	1. Yes
4. Are the correct Feature\_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?
	1. Yes
5. Do the function names in the Product field either match the official function list or say “Hypothetical Protein”?
	1. Yes
6. Has the Function field been cleared (using the automated buttons)?
	1. Yes

How are you documenting your gene calls in class? Choose any/all that apply:

* Spreadsheet

What is the file type (sort) submitted for QC to document your gene calls? Choose only one.:

* Spreadsheet

### Gene stop 458 (Gene 2)

* Orpham, no starterator data
* GeneMark and Glimmer disagree
	+ GM start is longest ORF, 4 bp overlap with previous gene
	+ No strong evidence for glimmer start

### Gene stop 1061 (Gene 5)

* Orpham, no starterator data
* GeneMark and Glimmer disagree
* GM start has the longest ORF and a 4 bp overlap with previous gene

### Gene stop 2439 (REVERSE, Gene 11)

* Called only by glimmer
* Short gene, 123 bp
* There is coding potential
* Orpham, no starterator data
* There is not enough space between the start of this gene and the start of the next downstream gene which is in the forward direction (only a gap of 18)
* Lone reverse gene in this area
* No compelling hits on hhpred or NCBI and no hits at all on phagesdb

Due to the above reasons, I chose to remove this gene from the annotation, as I do not believe there is sufficient evidence for its existence.

### Gene stop 12664 (New proposed gene 25)

* Added due to coding potential present on GM print out
* Second part of the frame shift, directly upstream of the tape measure gene

FRAME SHIFT:

25 sequence ATGCCATACGTTGACCTCTCCGAGTTCCTGACCGAGAATGACCTGGTGATCCAGGGCCTCGGCCCCCGTGACTACACCGTGCCCGCGCCGGACGTGGACACCGGCCTCCGGTACTCCGCCCTCTCCGGCATCGCCATCAAGGTGCAGAACGGTGAGGCCGTGTCGCCCGAGGAACTGGCCAGCCTGAAACTGGAGGGCGCGGAGGAGCGCACGTTCGTGCATCAGGTCCTCTCGCACGGCGTGGTGGAGGAGATGCAGTCGGACGGCCTGAAGTGGCCCGTCGTCGTGCGCGCCGCGAACTACGCGTTCACCCACTTTGCCGTCTCCCCGGAGGCCGCCAAGAAGGCGTTCGAGGCGGGTGCGTTCTCGGGAAAAGCACAGGCTCCGACGAATCGCGCGGCACGTCGGACAACCCCATCGGCCCCGCGGGCCTCGCGCGGTTCGAAGAAGCCCCAGTAG

GP 25:

MPYVDLSEFLTENDLVIQGLGPRDYTVPAPDVDTGLRYSALSGIAIKVQNGEAVSPEELASLKLEGAEERTFVHQVLSHGVVEEMQSDGLKWPVVVRAANYAFTHFAVSPEAAKKAFEAGAFSGKAQAPTNRAARRTTPSAPRASRGSKKPQZ

### Genes 35, 36, and 37

* These three genes are all membrane proteins directly downstream of the endolysin, with two of them hitting to holin-like proteins on hhpred. I called them membrane proteins but I found this arrangement interesting and would love more eyes on it.

### Gene stop 34966

* Starterator has all other GF phages calling start 2 (33989)
* Scores are identical, both start with ATG
* St. 1 has spacer of 6, st. 2 has spacer of 12
* Option one is the longest ORF and overlap of 4 with the previous gene
	+ I went with option 1 for the start (GM and Glimmer call)
* 16:15 alignment with other GF phages
	+ DelGarza called it a RecA-like DNA recombinase
	+ Other GF phages called it DNA binding protein
* HHpred alignments with RAD52 homologs, 99.9% probability with e value of 1.9x10^-22, 99.89% w/ e value of 2.8x10^-22

### Gene stop 23033 (REVERSE)

* Called HNH Endoneuclease, evidence is below

>Wrackline\_Draft gp39 (Minor variations such as HNK, HKH, HNN, and HNNH are allowed)

MRAVCGATHPRRRVACVLGTSHTGVHTSATGITWDRETKHRRGATNSNDRGNAAERRKRKCWALAWWGDGISCMCFRCGRVLLYSTLELDKIIPGVYGGTYARGNIRPACGECNREMGHALRALIRAKTPKRTILRMNRNGEL

### Gene stop 29275 (Gene 54)

* Called hth DNA binding protein but interesting evidence of possible repressor protein found on hhpred, including good alignment
	+ See also gene 59 for a hth DNA binding protein with solid hits to repressor protein on hhpred

### Gene stop 37759 (Gene 70)

* HNH Endoneuclease, see evidence below

>Glaske16\_gp70\_(44853-45341 bp) EXAMPLE FROM FORUM MPDGNQPACKYGACNDPVLARGFCKLHYYRNRDGKPMDGPRRSYSTGPRAWTYERLASVPITSTGAHQRVRRLWGSASLYPCATCGGPAKDWAYDGTDPTHYYEQGRKAWSHFSRWPEFYMPMCKPCHSNHDRRAAADELREYRQWKMRNPGKTLEDLEGVAZ

>Wrackline\_Draft gp70 (Minor variations such as HNK, HKH, HNN, and HNNH are allowed)

MTAKLTVARVAPRDGRACAHCGATDGLSVQHRANKGMGGSKDAERMSNGIILCATYNALLEQSAAEAERAIRMGWKCSKYGRAEVGNHTTQVPYWHYPTQGWRLAFDDGTWRPATVDEVKAHLARVTAAA

### Gene stop 38481 (Gene 71)

* Called ParB N-terminal-like domain, but slightly unsure about this call, would love extra eyes on this
* Also gene w/ stop 40128 (gene 73), called the same product