Actinobacteriophage Genome Annotation Submission Cover Sheet

This Cover Sheet will accompany each genome’s annotation file(s) submission and succinctly describe the work that your students and you have done. This document ensures that the work done was as complete and thorough as it could be. Most important to the QC reviewer, denote where the trouble spots were in your annotation and how they were resolved.

Phage Name. Lesiram

Your Name. Catalina Davila Aguer

Your Institution. Universidad Ana G. Mendez, Gurabo campus

Your email. Davilac2@uagm.edu

Additional emails. (for correspondence). frivera@uagm.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.

1. We found a large gap between genes 5 and 6. This gap was checked for possible coding sequences using Blast and entering the corresponding aminoacid sequence in HHPred. No function could be assigned to this region.
2. We found a large gap between genes 10 and 11. This gap was checked for possible coding sequences using Blast and entering the corresponding aminoacid sequence in HHPred. No function could be assigned to this region. We changed the start from 6629 (called by Glimmer) to 6569 (not called by Glimmer nor by GeneMark). Although this start has not the best Z Value and Final Score, it gives a longer ORF (not the longest) and reduces the gap from 283 to 223. Also, this change in the start improves the alignment from 85.8% to 100%. Finally, the % coverage in HHPRED is improved from 53.25% to 77.15%. The Staterator report is not informative for this start.
3. Only one tail assembly chaperone is found in this genome. Thus, a translational frameshift was not detected in this phage.
4. We found a large gap between genes 19 and 20. This gap was checked for possible coding sequences using Blast and entering the corresponding aminoacid sequence in HHPred. No function could be assigned to this region.
5. Gene 26 requires attention since it was only called by Glimmer and not by GeneMark.
6. DNA Master assigned the function “membrane protein” to genes 29, 30, and 66. None of the two non-draft members of cluster GF have the function "Membrane Protein" in these genes. We decided not to assign this function, since only one transmembrane domain was found when Deep TmHHm was run (the result is included in the Lesiram\_final submission file, MemProt\_Supporting info sheet).
7. One of the two non-draft members of cluster GF have the function "Holin" in gene 31. We decided to assign this function, based on Deep TmHHm results, since we found four transmembrane domains (results are included in the Lesiram\_final submission file, MemProt\_Supporting info sheet). Also, this gene has a Pfam hit to a holin with a high probability (98%).
8. Gene 43 was assigned the function SSB protein, according to what is requested in the SEA-PHAGES official function list.
9. DNA Master could not do the BLAST for gene 48. However, according to the BLAST in PECAAN, this gene is a membrane protein. None of the two non-draft members of cluster GF have the function "Membrane Protein" in this gene. We decided to assign this function, since two transmembrane domains were found when Deep TmHHm was run (results are included in the Lesiram\_final submission file, MemProt\_Supporting info sheet).
10. The Starterator report for gene 49 is not informative, since Lesiram is the only phage that has this start.
11. We found a large gap between genes 53 and 54. This gap was checked for possible coding sequences using Blast. No function could be assigned to this region. Gene 54 was not called by GeneMark.
12. Gene 55 is an orpham, no Starterator data available, no BLAST hits.
13. Gene 58 is an orpham, no Starterator data available.
14. We changed the start of gene 61 from 36202 (called by both Glimmer and GeneMark) to 36148. This start has the best Z Value and Final Score, and gives a longer ORF (not the longest), reducing the gap from 93 to 39 bp. The alignment % and similarity % in BLAST remained the same (100%), but the alignment changed from q1:s1 to q1:s19.

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):

yes 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.)?

yes 2. Are all the genes ‘Valid” when you click the [Validation button](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?

yes 3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.

yes 4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?

yes 5. Has the [documentation been recreated](https://seaphagesbioinformatics.helpdocsonline.com/article-86) from the Feature Table to match the latest file version?

N/A 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?

N/A 7. Has the [frameshift in the tail assembly chaperone](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?

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?

yes 9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?

yes 10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?

yes 11. Did you [delete the genes](https://seaphagesbioinformatics.helpdocsonline.com/article-65) that you meant to delete?

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)

yes 1. Have any duplicate genes been deleted?

yes 2. Has the Notes field been cleared (using the automated buttons)?

yes 3. Do the gene numbers and locus tags match?

yes 4. Are the correct Feature\_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?

yes 5. Do the function names in the Product field either match the official function list or say “Hypothetical Protein”?

yes 6. Has the Function field been cleared (using the automated buttons)?

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

      PECAAN output

      DNA Master shorthand (previously used format)

  X   Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.

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

      PECAAN output

      DNA Master shorthand (previously used format)

  X   Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.