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. Glaske16

Your Name. Frederick Baliraine

Your Institution. LeTourneau University

Your email. FredBaliraine@letu.edu

Additional emails. JoshMcLoud@letu.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. There has been an update on how to call membrane proteins, but this has not yet been updated in the Official Functions List (see <https://seaphages.org/forums/topic/5165/?page=2#post-10138>).
2. Feature 19 (13215-13607). Though hits “tail completion protein” in NCBI, all phagesDB hits are NKF. Moreover, the “tail completion protein function” is not on the Official Functions List.
3. Features 27 (21273-21719 bp) & 28 (21728-22174 bp): We contend that these two are not minor tail proteins. According to the Resource Guide, minor tail proteins genes are the big genes down the tape measure and usually not more than 5. Whereas there are several BLASTp hits to minor tail protein, probably based on synteny calls by previous annotators, this is a small gene that does not warrant being called a minor tail protein in our view. The following forum posts suggest not calling small proteins minor tail proteins (<https://seaphages.org/forums/topic/4464/>; <https://seaphages.org/forums/topic/4546/>; and the instructions for using synteny <https://seaphagesbioinformatics.helpdocsonline.com/article-90>). To use synteny, ALL the following three conditions must be met: (1) of the correct size , (2) adjacent to other structural genes of known, verifiable function and (3) the only possible option for that function in the genome. “You can (use synteny to) call minor tail proteins for the 'big' genes downstream of the tape measure protein.” The Forum post by Debbie on May 29, 2023 too seems to concur with this determination (<https://seaphages.org/forums/topic/5511/>).
4. Feature 33 (26409-26837 bp). Though hits excisionase (aka excise) in NCBI, the top hit in HHPred is Gene 34.1 protein of the prototypic phage D29, which is NKF. Besides, all hits in phagesDB are NKF.
5. Feature 48 (32530-33069 Rev) hits “ribonucleoside reductase class II” function in NCBI is not on the Official SEA Functions List. All hits in phagesDB are NKF. Moreover, in HHPred it hits gp 49 of phages L5 & D29, with 100% probability, but both are NKF. So we settled for NKF.
6. Feature 51 (36026-36793 rev) is indeed a metallophosphoesterase, because it contains the HEXXH motif (HEGNH) that coordinates the metal ion, unlike the subcluster M1 phages (<https://seaphages.org/forums/topic/5557/?page=1#post-10335>).

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?

**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?

**Yes** 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

**Yes** DNA Master shorthand (previously used format)

**No** 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

**Yes** DNA Master shorthand (previously used format)

Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.