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. Diminimus

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 (to just use DeepTMHMM), but this has not yet been updated in the Official Functions List (see <https://seaphages.org/forums/topic/5165/?page=2#post-10138>)
2. Some features would not get BLAST results during Re-Blasting in DNA Master, despite having BLAST results in phagesDB. This was particularly so for the tRNAs.
3. Feature 22 (11870-12274 bp). Called NKF despite >90% hit to minor capsid protein in HHPred because this function was not backed up with extensive proof. Neck protein is not a function on the official functions list and phamerator shows that PegLeg gp21 as well as gp21 of the prototypic phage D29 hit in HHPred (hit # O64215) too does not actually have the minor capsid protein function (it is NKF in phagesDB gene list), and all hits in phagesDB are NKF.
4. Feature 30 (24157-24615 bp): We contend that this is NKF rather than a minor tail protein. 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. Whereas HHPred hits for minor tail protein in UniProt database are labelled, "Evidence at protein level", this particular hit is simply listed as, "Predicted," implying that there is no experimental data verifying it. 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/>)
5. Feature 35 (28791-29789 bp) is a minor tail protein despite various NKF hits in phagesDB. We sought clarification on the Forums and Debbie concurs that this indeed is a minor tail protein (<https://seaphages.org/forums/topic/5511/>)
6. Feature 57 (39487 – 40347 bp) matches the Cas4 exonuclease function rather than simply, “exonuclease” as explained in the notes, as well as the criteria in the Official Functions list. I contained Debbie Jacobs-Sera and she too concurs that she too would be inclined to call it Cas4 exonuclease.
7. Feature 107 (58239-58496 bp) overlaps the tRNA downstream by 9 bp. In Pope et al (2014); DOI: <https://doi.org/10.1128/jvi.03363-13>), I note that whereas the tRNA in Bongo gp 105 at 58492-558566 bp has a 9 bp overlap with gp 105 (58244-85801 bp, pham 4696), it was called in the paper (Fig 12A & table 3). The overlap is not obvious from the figures and the paper since it was not discussed, but I when I downloaded the curated minimal file of Bongo from phagesDB, I was able to calculate and determine that indeed there is an overlap. It is more obvious in phamerator. Other examples of the genes in this same pham 4696 showing the tRNA overlap are Bricole (58392-58649 bp), Skinny (59486-59743 bp), SlimeJimmy (59697-59954), & LilhomieP (59134-59391 bp). From this, and even after consulting with Debbie Jacobs-Sera, I was able to determine that both the gene and tRNA in question in other M1 phages we are annotating are legitimate based on this paper and the current knowledge, unless experimental data ends up proving otherwise.
8. Feature 126 (66621-66180 bp). Despite many BLASTp hits to metallophosphoesterase in phagesDB, this gene does not meet the criterion for this function since it lacks the HEXXH motif required to coordinate the metal ion. The correct function is phosphoesterase (<https://seaphages.org/forums/topic/5557/>).

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.