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. Koreni

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 (using only 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. The function of the feature gp 70 at 434443-43841 bp has been confirmed as HNH endonuclease. Has over 20 hits to HNH endonuclease in phagesDb, with high probability in HHPred but low e-values (see <https://seaphages.org/forums/topic/5505/>)
3. The feature gp 69 at 53720-53881 bp was inserted, though coding potential not significant. It has an operon with a 4 bp ATGA overlap with the upstream gene and a 1 bp overlap with the downstream gene, despite not having significant CP. We do not see a reason why the ribosome would skip this overlap and go over a gap, when this operon is long, starting several genes upstream. Moreover, there were significant BLASTp hits in phagesDb.
4. Feature 108 (58718-58975 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.
5. I posted a question about the gp117 at position 60940-61320 bp on the Forum (<https://seaphages.org/forums/topic/5503/>), because it has no significant CP, but has more than 70 hits to the HNH endonuclease in phagesDB. Debbie stated that she would “avoid calling this gene. The instances where I know there is a tRNA in a predicted protein do not look like this one. However, when I saw the hit to an HNH that was confounding. However, I don't think it is an HNH. The H-N-H seems missing to me. I would call the protein but include the tRNAs. I would also draw attention to this in your cover sheet. The QCer will review all of this to confirm.” Christopher Shaffer thinks it is HNH endonuclease, and the hit matches HNH endonuclease (<https://seaphages.org/forums/topic/5505/>) see the hit at <https://www.rcsb.org/structure/5H0M> which matches that of the HNH refence sequences provided in the Official Function List. Later personal communication with Debbie was consistent with calling it HNH. So we are leaving this gene and predicting it as HNH endonuclease. Notably, this gene is present in M1 phages Skinny, Pegleg, & LilhomieP, but absent in M1 phages Bricole, IPhnae7, Reindeer, & TyDawg.
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. Functions of gp 98 (56773-57150 bp) and gp117 have been confirmed as HNH endonuclease after inquiring from the forum (see Forum post <https://seaphages.org/forums/topic/5505/>).
8. Feature 129 (66640-67299). 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/>).
9. Feature 89 (53720-53881); despite not showing significant CP, this gene was inserted, as it has an operon with a 4 bp ATGA overlap with the upstream gene and a 1 bp overlap with the downstream gene. We do not see a reason why the ribosome would skip this overlap and go over a large gap, when this operon is long, starting several genes upstream. Moreover, there were significant BLASTp hits in phagesDb.

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.