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. Burley

Your Name. Marie Fogarty

Your Institution. Durham Tech

Your email. fogartym@durhamtech.edu

Additional emails. (for correspondence).

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.

**Gene 8** – Stop 4415 - We did not call a function for this gene. There is a HHPRED hit to head fiber protein, and DJ phage Nadmeg called this gene as head fiber protein. However, none of the other DJ phages (there are 28 total) are making this call.

**Gene 16** – stop 10,245. We called this as an RNA ligase based on the following, but we are not sure. 5 other DJ phages did call the function and there is a HHPRED hit (but only with 25 % coverage) Phagesdb blast: Nithya, RNA ligase,0 Blast NCBI function: RNA ligase, Afflac and Figkuar, E = 0

**Gene 28** stop 21,004 – We did not call a function, but there is some evidence for a minor tail protein = DJ phages, Tanis and Odesza, called minor tail protein but 27 other pham members did not. There is a HHPRED Match with > 90% probability: Hit with Fiber Upper, gp68; phage tail, tail tip, tape measure protein, VIRAL PROTEIN; 3.7A {Staphylococcus virus 80alpha}, P=95.9%

**Gene 31:** We did not call a function. Potential minor tail protein - Three ‘tail related hits’ HHPRED hits to a) minor capsid protein, 88% coverage, 95% probability, b) HK97-gp10\_like ;Bacteriophage HK97-gp10, putative tail-component c) Phage\_tail\_S ; Phage virion morphogenesis family.

Since this is normally where the tail assembly chaperones are called reluctant to call this as a minor tail protein.

**Gene 67:** We called the function as membrane protein. There is some evidence for PnuC-like Nicotinamide riboside transporter – HHPRED Match with > 90% probability: Nicotinamide riboside transporter PnuC, P = 99.1, 40 % coverage and also it is called in many C1 phages (44) but only 2 DJ phages.

**Gene 83**: The Glimmer and Genemark start sites were kept as this gives the longest coverage and longest ORF. We struggled a bit with this choice - It gives longest ORF, captures coding potential and genemark and glimmer agree. However there is no 1:1 blast match and starterator shows that a downstream start is the one most commonly called - (Start: 8 @53246 has 13 MA's). The site we went with leaves a 42 bp gap instead of a 180 bp gap.

All gaps were explored, and some genes were added. The most notable remaining gaps are 1) of 197 bp between gene 55 + 56 - There is a 193 bp gap, but no coding potential that warrants gene addition and the gap seems to be conserved in similar DJ phages including Crocheter and Nadmeg

2) 327 bp long between gene 85 and 86. We explored gene addition here due to the 371 bp gap. Although there is a small amount of coding potential in one frame, on adding the gene at stop 55368, there were no 1:1 matches. This gap is consistent in other phages so we decided not to add the gene

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?

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:

Yes - PECAAN output

      DNA Master shorthand (previously used format)

Yes - 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)

SS Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.