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. SJReid
Your Name. Brenna Fox
Your Institution. McGill University
Your email. Brenna.fox@mail.mcgill.ca
Additional emails. (for correspondence). Patrick.lypaczewski@mcgill.ca

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.

There are still numerous large gaps with minimal coding potential, no functional assignment, and no similarities to previously annotated genes. Without strong evidence after analysis, these gaps were mostly left alone. The largest gap is 809bp right before gene 243 (158434 start site). The second largest gap is 501bp right before gene 287 (176485).

There were some genes added that were not called by Glimmer or GeneMark, but rather called by PhagesDB or Pharokka. Most of these contained decent coding potential, good alignment with genes in other FC cluster phages, or a functional assignment. However, gene 47.5 (called by PhagesDB) was added and may not have enough evidence for its addition. Other gaps had coding potential but no known function or alignment with other genomes, so they were not added.

tRNA 8 (7.5 in notes) was only predicted by tRNAScan and was predicted as a tRNA over 100bp long. There was difficulty trimming it properly, as trimming leads to a poor infernal score and an altered tRNA shape. It was not called but could be legitimate.

It was difficult for me to determine the aligned domains for the DNA Primase (172 in notebook) and DNA primase/helicase (176 in notebook). 172 was initially thought to be a DNA Primase/Polymerase, but I could not find enough alignment with the pol domain. 176 was called as a DNA primase/helicase but may need further confirmation that both domains are well aligned.

A tail assembly chaperone was identified, but no frameshift was found.

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?
Yes 3. Are the genes (and matching LocusTag numbers) sequential, starting with #1, counting by 1s.
Yes 4. Are the Locus Tags the "SEA_PHAGE NAME" format?
Yes 5. Has the documentation been recreated from the Feature Table to match the latest file version?
Yes 6. Have tRNAs followed the tRNA protocol, **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?
No 7. Has the frameshift in the tail assembly chaperone been annotated correctly (if applicable)?

Yes 8. Have you cleared your Draft_Blast data and have you re-Blasted the submitted DNA Master file?
Yes 9. Has every gene been described and supported in your Supporting Data file?
Yes 10. Did you investigate 'gaps'?
Yes 11. Did you delete the genes that you meant to delete?

Now, make a profile of the file you plan to send. (And you can save this file for Review to Improve!)

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)
Spreadsheet
Powerpoint
Word document (must be easily searchable)
Yes Other: Describe. Microsoft OneNote

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)
Spreadsheet
Powerpoint
Word document (must be easily searchable)
Yes Other: Describe. Microsoft OneNote