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. AdaS
Your Name. Julia Simler
Your Institution. Purdue University
Your email. jsimler@purdue.edu
Additional emails. (for correspondence). klclase@purdue.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.

Draft gene 18 from 10531-10770 bp was a small reverse gene in between the tail assembly chaperone genes and tape measure protein gene, which is not typical of most phage genomes. However, this was consistent with the AY cluster report.

Draft gene 32 from 22784-22918 bp was excluded from the final annotation due to multiple factors that violated major guiding principles of gene calling. It was only predicted by Glimmer, with no supporting coding potential observed, and no evidence from GeneMark or other algorithms. Its reverse orientation and overlap of 147 bp with adjacent forward genes also raised concerns about biological plausibility. Additionally, the gene had no BLAST or HHPred hits, did not align with any known proteins, and lacked conserved synteny across other phage genomes. As an apparent orpham with no functional clues or homologs, Gene 32 was determined not to be a real gene and was therefore deleted from the annotation

Draft gene 48 presented a complex case. The original Glimmer-predicted forward gene (29470–29592) was poorly supported by BLAST and lacked conservation or functional alignment, prompting deletion. Instead, a GeneMark-called reverse gene was annotated beginning at 29721. This new gene had a high-quality alignment to a hypothetical protein in Arthrobacter phage Seahorse and showed improved RBS and ORF characteristics. Despite a low-confidence HHPred result and lack of synteny, the switch to the reverse-strand gene at 29721 was supported by both sequence identity and gene prediction logic, though the function remained unknown. This special case highights the difference between the Glimmer and GeneMark algorithms.

Draft genes 73 (41320-42669 bp) and 74 (41633-42409 bp) were completely overlapping. Draft gene 73 was the predicted gene call by Glimmer while draft gene 74 was the predicted gene call by GeneMark. Draft gene 74 was deleted because it isn't supported by Starterator and is not conserved in other phage genomes. It also would have created large gaps with its upstream and downstream genes. This special case once again highlights the difference between the Glimmer and GeneMark algorithms.

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?
Yes 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)
- 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)
- Spreadsheet
- PowerPoint
- Word document (must be easily searchable)
- Other: Describe.