Actinobacteriophage Genome Annotation Submission Cover Sheet

This Cover Sheet will accompany each genome's annotation file submission and will 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: Kraw
Your Name: Mason Krawczynski
Your Institution: Providence College
Your email: mkrawczy@friars.providence.edu
Additional emails (for correspondence): kcornely@providence.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.

Start #	Description
21958	Believe the gene is too small to be called a minor tail protein. BLAST data is about half hypothetical and half minor tail protein function calls. HHPred has no relevant results; went with the hypothetical call.
24509	Believe the gene is also too small to be called a minor tail protein. BLAST data is about half hypothetical and half minor tail protein function calls. HHPred also has decent results. Went with the hypothetical call but a bit less confident than gp27.
*39728-39802	Investigated adding a gene with a start-stop of 39728-39802, but ultimately did not add it. The only convincing evidence was the start has an RBS (used ApE to investigate), though the score is poor. Also, gene with start 39799 would have a -4 bp gap with it added. Otherwise, there were no BLAST hits and no relevant HHPred hits. The annotation for it was below. SSC: Start 39728 Stop 39802; CP: NA; SCS: NA; BLAST-Start: none; Gap: 25 bp; LO: yes; RBS: Kibler 6 Karlin medium, raw score = -6.627, spacer = 13, Z score = 0.746, final score = -7.672, is best score; F: SIF-BLAST: none; SIF-HHPred: no hits greater than 90%; SIF-Syn: none; RBS score not great. Added this gene based on this forum from stukey: https://seaphages.org/forums/topic/5769/?page=2#post-11554, particularly the second paragraph. ApE shows the proposed gene start site has an RBS.
43630	Called for DNA methyltransferase. There are conflicting function calls in the BLAST results (methyltransferase and DNA methyltransferase). Called for DNA methyltransferase because of HHPred results. We called the function a DNA methyltransferase because we were able to locate DNA binding domains.
45107	Called function for DNA primase. There was at least one hit on HHPred for a DNA helicase. Some of the BLAST results have DNA primase/polymerase. One BLAST hit (Juliette) has primase/helicase. We investigated HHPred hits to crystallized proteins on the PDB, and found hits to primases that aligned with the protein expressed by this gene, but no convincing hits for helicases or polymerases, so we decided to call the function a primase only.
57199	Called for methyltransferase, but some of the BLAST hits have DNA methyltransferase. Felt more confident with methyltransferase based on the HHPred hits. We didn't locate DNA binding domains, so we decided not to call the function a DNA methyltransferase.

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 Locus Tag numbers) sequential, starting with #1, counting by ones?
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 been documented following 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!)

NA 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:

NO: PECAAN output
YES: DNA Master shorthand (previously used format)
YES: Spreadsheet
NO: PowerPoint
NO: Word document (must be easily searchable)
NO: Other: NA

What is the file type (sort) submitted for QC to document your gene calls? Choose only one:

NO: PECAAN output
YES: DNA Master shorthand (previously used format)
NO: Spreadsheet
NO: PowerPoint
NO: Word document (must be easily searchable)
NO: Other: NA