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. CandC
Your Name. Alyssa Gleichsner
Your Institution. SUNY Plattsburgh
Your email. Aglei002@plattsburgh.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.

1. At the faculty meeting this year I worked with a group on EG cluster annotations and we discovered that the PTFS is not on the official list of sequences verified with bench evidence. It looks like half of the annotations call the frameshift and half do not. The slippery sequence seems pretty clear and those of us with EG phage experience all agreed on the sequence and slip – so I am calling it in this submission (but could use some scrutiny I think). In this genome it includes genes with stop coordinates 21,392 and 21,577.

2. Gene with stop 20,527 had a lot of debate in class about the correct start. The class voted to keep the predicted start (20,030), but one group argued hard for a change to 20,015. They argued that the change was supported by the genemark self potential and the z score – going against starterator and blast with the point that past annotations may have just gone along with previous choices. I included their evidence slide for the gene (with start of 20,015) in the ppt document, because I think it's a fair argument, even though the final annotation submitted does use the 20,030 start (which is better supported by starterator, and the host genemark file). It was enough of a heated argument that I wanted to raise it with whoever is QCing though!

3. Gene with stop 59,452 is one that everyone opted to delete – but I wanted to bring it to the attention of QC. It has strong coding potential in the host trained genemark – and has one significant hit on NCBI's database to a bacteria – could be a host gene that was picked up?

4. Last – We called a major capsid hexamer protein for this genome (stop coordinate: 14,376) because we think the evidence supports it – but did not call a pentamer because we lacked support to do so. I wasn't sure if we should be calling a hexamer without also calling a pentamer, so wanted to bring that up as well.

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 – N/A 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:

Some PECAAN output
Some DNA Master shorthand (previously used format)
Spreadsheet
All 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
This Powerpoint
Word document (must be easily searchable)
Other: Describe.