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. **Soondubu**
Your Name. Amanda Freise
Your Institution. UCLA
Your email. afreise@ucla.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.
—----------

Stop @ 35605: Hits to HNH endonucleases, but I did not see the requisite H-N-H in a 20-30 aa span. HHpred hits also not the nest, mostly below 30% coverage. Orpham except for one other Left as NKF.

Stop @ 39171: Tricky start. Glimmer/genemark predict 37732 (LORF) and 37720 respectively. Both have okay final scores. But start site 37756 has a less negative final score and larger z-score than the sites called by glimmer and genemark. This is currently the only AZ phage in the pham. I picked 37756 because the length of the gene matched the length of some of the other genes in the pham, and that start site still captured the majority of the coding potential in the the ORF. It also leaves a larger gap for the upstream reverse gene.

Stop @ 43805: Decided not to keep this auto-annotated gene. There was basically no CP in either genemark. It did fill the gap, but overlapped both the upstream and downstream genes. There was one hit in NCBI BLAST to a bacterial protein but the e-value was only 10^-3.

Checked the region at the end of the genome and confirmed there was not coding potential/options for ORFs in the gaps.

—----------

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)?
No 8. Have you cleared your Draft_Blast data and have you re-Blasted the submitted DNA Master file?
I was unable to save my re-BLAST results within DNA master and have yet to figure out what the issue is. Debbie said it would be fine for now to note this here and submit.
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:

- X PECAAN output
- DNA Master shorthand (previously used format)
- Spreadsheet
- X 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.:

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