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. **Emotion**
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 @ 19074: Functional call question**.** HHpred has a 99.5% probability hit for "N-acetylmuramoyl-L-alanine amidase; amidase" (PDB: 6SSC). (Emotion_21 HHpred results: https://toolkit.tuebingen.mpg.de/jobs/Emotion_21). I would like a second look in case I have misinterpreted it. The other AZ phages all have a different pham, which will be designated "endolysin" going forward per the AZ function harmonization.
stop @ 19998: Start site call is a little tricky. Orpham. Smallest possible ORF is also by far best RBS score (Z score is 3+)
stop @ 20762: Start site call is a little tricky. Orpham. Smallest possible ORF is also by far best RBS score (Z score is 3+)
stop @ 23344: Possible membrane protein. TMHMM called one TMD; TOPCONS only detected a signal peptide. I tried to run it on SOSUI for a tiebreaker, but the SOSUI server has been down for days. Not sure I can call this a membrane protein; needs confirmation.
stop @ 27236: Strong CP in reverse direction. Reverse gene found in several other phages. However, note that there are multiple areas of good CP in reverse frames in this general region, which overlap with some forward genes. Not sure if this gene is real.
stop @ 33724: Tricky. Closest start site in location compared to other genes in pham would be 33410 (called by Genemark). But 33359 (Glimmer-called) has much better score.
stop @ 34743: Tricky - doesn't have most annotated start site. Start #27 looks good. but in downstream gene (stop 35778) which also has 27, the first start site (#21) seems better. Should they be the same?
stop @ 35778: Tricky: see above
stop @ 38826: Added gene based on coding potential in self-trained GM. Small, may not be real. possible DNA-binding function? Some good hits in HHpred.
Large gap around ~39000: Looked at both Genemark outputs; did not see any coding potential in this region.
stop @ 41998: Tricky. Start site chosen has best RBS stats, but could be made larger to fill gap
stop @ 43170: Tricky. Start site chosen has best RBS stats, but could be made larger to fill gap

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 re-BLAST 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.