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. Clarkson
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 @ 495: Tricky start site. Disagreed with other only 3 pham members who called the site corresponding to 274bp in Clarkson. Coding potential does extend slightly past start site of 271/274, but not all the way back to 121. There are only 3 non-draft members in pham, and 3 of the 3 non-draft members call start site 4, which correlates to a start site of 274 bp for Clarkson. However, site 3 (271) is just one codon earlier, GTG, and has a better Z/RBS score. Calling that one.
Gap after stop @ 3115: In Clarkson, there is an ORF with some SP o GM-Self only. But the CP in that ORG overlaps completely with gene 9 (stop 3115). The remaining area in the ORF had no coding potential at all.
Stop @ 43844: Unsure if gene is real (also see below note). Added this gene based on CP in GM-self; presence of gene in other phages Beelzebub, Huphlepuff. Other phages (Gattaca) have an overlappeing gene in this location too but with start site at 43701, resulting in gap of -74, but appropriate gene length of 144. However, overlap is so large @ start 43701, that I am choosing 43767 and favoring smaller gene and better RBS score. Note that many phages choose one or othe other genes here and don't include both. not sure which is real: this gene, the upstream gene, or both. Both have smilar coding potential on both Genemark maps.
Region between genes 78-84: variable and hard to tell which genes are real (for current genes 79, 80, 82).
Stop @ 58091: Some HHpred evidence in support of "RNA binding protein" as function. What do you think?
stop @ 61205: tricky start site call. Small pham (only 2 other members). Chose 60927 based on better RBS scores and ATG start codon and starterator, but leaves large gap.
stop @ 64094: Unsure if correct gene called. Final gene in genome is in a pham with only one other draft gene. All other S phages (except draft LilBit) have a gene in this exact location but they are all in a different pham. In this location, the selected ORF is the only one with coding potential in any of the forward frames. Unclear why this version is in a different pham. However, this has also happened at several other places in the genome (e.g. Clarkson_34, Clarkson_59, Clarkson_65) so may not be an issue here either.

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?
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)?
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.