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. GalacticEye
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 @ **13151** - Possible more specific functional call. At minimum, minor tail protein (based on PhagesDB BLAST), May also be a glycoside hydrolase (based on some HHpred hits + BLAST). Multiple HHpred hits with >95% probability refer to hydrolase or glycoside hydrolase, but query coverage is low. However, only 1/44 members of pham call this as glycoside hydrolase, vast majority of others call it a minor tail protein.
stop @ **13527** - Tricky start site. Only 2 other non-draft members in this pham. They both call start sites that result in small gene of 135bp, possibly because most typical CP is towards end of gene, but there is still some coding potential in first half, so we called this this as longer gene w/ start site (13222) with third-best final score and Z-score. Earlier start sites (13189, 13195) close gap but are TTG codons and have much worse final scores.
stop @ **18037 -** Tricky start site. Overwhelming (58/60 non-draft calls) starterator data for start @ 17543, but why?? Upstream gene is head-to-tail adaptor; this gene has strong HHpred hit for "PF05069.16 Phage_tail_S ; Phage virion morphogenesis family". Next start site of 17552 has gap of -4, suggesting operon, which negates poorer RBS score. Going with start of 17552; unsure as to why so many people chose first one.
stop @ **21050** - Not sure if real due to overlap and interruption of synteny. There is atypical CP on GM-self, none on GM-host. Hail2Pitt, Luker, and Newt all have it in the same location, but are the only three members of this pham.
stop @ **37891** - Possible more specific functional call. Mostly NKF in pham, but Newt calls as minor tail protein, perhaps based on synteny - is that still appropriate? Called here as NKF but could be updated.
Stop @ **42942** - very tricky start site call. Based on the evidence in our annotation notes, chose start site @ 42394. Summary: The location call for this gene is difficult to make because there is evidence that could support several start sites, including 42394, 42403, and 42499. Start 42394 provides the longest ORF among reasonable calls at 549 bp,and the most favorable z-score and final score of the reasonable calls. It also has a more likely start codon, ATG, than the other two. The 13 bp overlap is a little longer than typical but shows synteny with several other phages within this cluster. This start site also has the most manual annotations for this Pham, although it is just 2 more calls than 42403.
Note: Two DNA helicase genes called with high confidence (genes 2, 56). Acceptable?

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.