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. Brunswick
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.

Very tricky genome! Multiple difficult functional calls. Most of the below is explanations for decisions made. I have colored purple the calls that I was really not sure about.

Stop @ 11324: Function call notes: https://seaphages.org/forums/topic/4633/

Stop @ 11943: Possible tail terminator function. 5A21_G is a hit in HHpred (98% prob, 98% coverage, e-value: 0.000068). 6TE9_F is also a hit with an even better e-value. Pham has mixed function calls (NKF vs tail terminator); many do call tail terminator. Approved function list note from Aug 2023 seems to suggest this is all acceptable evidence to make function call.

AU phages have two major tail protein genes: https://seaphages.org/forums/topic/4631/

Stop @ 17782: Called NKF, but suspect possible minor tail due to synteny, gene size (819bp), and HHpred hit 7EY9_d (T7 tail fiber protein) (97% prob, 59% coverage, evalue 0.008). FIve other AU1 phages call this function. Many pham nembers do not call a function at all.

Stop @ 18757: Called NKF, but suspect possible minor tail due to synteny, gene size (948bp), and CDD hit for collagen triple helix repeat (though the ID (35%) and coverage (15%) are low, and e-value is 0.0017). There are also some HHpred hits with to receptor binding proteins domains with high probability (97%), okay coverage (28-32%), and poor e-value hits (0.0052 and 0.019). A couple other phages in pham call minor tail protein.

Stop @ 19596: Duplication of preceding gene. Called NKF, but suspect possible minor tail due to synteny, gene size (828bp), and CDD hit for collagen triple helix repeat (though the ID (35%) and coverage (15%) are low, and e-value is 0.0017). There are also some HHpred hits with to receptor binding proteins domains with high probability (97%), okay coverage (33%), and poor e-value hits (0.00077, 0.0037). A couple other phages in pham call minor tail protein.

Stop @ 29247: TMHMM and SOSUI each called one TMD, but DeepTMHMM called none. Based on latest guidance, leaving as NKF.

Stop @ 29407: Added gene. No other AU1 phages call a gene in this location. Good atypical CP on GM-Self. Small, but fills gap. Very similar sequence to: Giantsbane_36, Ingrid_35, LilHuddy_36, Loretta_35. No TMDs detected. This gene does interrupt several other membrane proteins in a row.

Stop @ 29990: TMHMM and SOSUI each called one TMD, but DeepTMHMM called none. Based on latest guidance, leaving as NKF.

Stop @ 30556: Added gene. Good CP on GM-self only. Fills otherwise large gap. Phage Tenno has gene in this location.

Stop @ 31495: TMHMM predicts 2 TMDS. TOPCONS Suite also predicts 2 TMDs. SOSUI predicts none, DeepTMHMM predicts none. Leaving as NKF according to new TMD rules.

Stop @ 31977: One great HHpred hit to hydrolase but that's it, so leaving as NKF for now. No other genes in pham call a function.

Stop @ 33175: Added gene. VERY low atypical CP on GM-self only. CastorTray, ElephantMan, Niktson call gene in this location. Pham of their gene has 18 members, many from other clusters.

Stop @ 34714: Added gene. Good atypical CP on GM-self only. Found in phages ScienceWizSam, CapnMurica, DevitoJr

Gap between 36799-37140: No coding potential present.

Stop @ 43823: Multiple good hits in HHpred to SSBs, along most of the query protein length.

Stop @ 46180: Added gene. Typical and atypical CP on GM-self only. Found in phages ElephantMan, Niktson, Teacup, Tenno, Nightmare, and others

Stop @ 46211: Ultimately opted not to add a gene in this region at stop 46211, even though there was an ORF with coding potential. It would have overlapped the next gene completely, and more genes called the larger previous gene at stop 46180. That gene also filled the gap better than the smaller one that would result from the ORF with stop 46211.

Stop @ 46417: Added gene. Some atypical CP on GM-self only. Found in phages ElephantMan, Niktson, Teacup, DevitoJr, and others

Stop @ 46563: Added gene. Some atypical CP on GM-self only. Found in phages ElephantMan, Niktson, CapnMurica, DevitoJr, and others

Stop @ 47825: Ultimately opted not to add a gene in this region at stop 47825, even though there was an ORF with coding potential. It would have overlapped the downstream gene substantially,. IN comparison to the genes on either end which had high typical CP, this ORF had very low atypical coding potential only. Only 3 other AU1 phages called this gene; other calls in pham were from AU2 or AU3 phages.

Stop @ 50659: Possible "atypical" endonuclease with an HNK motif near the end (per this forum post: https://seaphages.org/forums/topic/5505/). has good hits to HNH proteins in HHpred. I have bolded the residues that would work for this call:
MSEWRTLPEFPDYEITSDGDVRNKETFYVLKEIQNKNTGAWSYSLRRPDGRATQRNFWSLIYSAWPELKPAEDEPQ
DVRRSPARQYAERGRWKAIPGFPNYQA**H**PEGLVRYIKTRKPRKMKYEKRGEQYFRLY**N**EYGDYSDV**K**LSVILDR
TFQKVSA

Stop @ 50868: NKF. Good HHpred hits to HNH endonuclease, but doesn't have the requisite motif. Also, the HHPred alignments don't align with the HNH parts of the PDB hits.

Stop @ 51726: Added gene. Good atypical CP on GM-self. THis region is somewhat variable aomong phages in the cluster. Some phages have the upstream and downstream gene immediately adjacent to each other, with no gap (CapnMurica, CastorTray, DevitoJr). This genome has a gap that can be filled.

Stop @ 53279: not sure if this gene is real. Very small, some genomes do NOT add a gene here. Good CP on GM-self. ElephantMan, Niktson have a gene of the same size, pham 28967 CapnMurica, DevitoJr, Gordon have a gene in this gap, but it's longer.

Region between 53610-54283 is tricky to call. There are several ORFs in different frames with good CP. I called genes to match CapnMurica, ElephantMan, Niktson. Exception is gene with stop at 53848: this gene had good CP in Brunswick GM-self, but was not found in other non-draft AU phages.

stop @ 55126: tricky call; diverse pham. HHpred hits that included both nucleases and helicases (UvrD) were present so went with RecB functional call

—----------

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.