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. **RomansRevenge**
Your Name. Amanda Freise
Your Institution. UCLA
Your email. afreise@ucla.edu
Additional emails. (for correspondence). tbouklas@mednet.ucla.edu

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.
—-----------------------------------

Useful genomes for comparison: Cluster AP

Stop @ 61794 (reverse): Possible HTH DNA binding protein, but the spacer between the first and second alpha helix is 5 amino acids long instead of 3-4. Called it a HTH. Also selected a reasonable start site (though not the best one) that produced a gene the same length as the last gene in the genome, since RomansRevenge has DTRs.

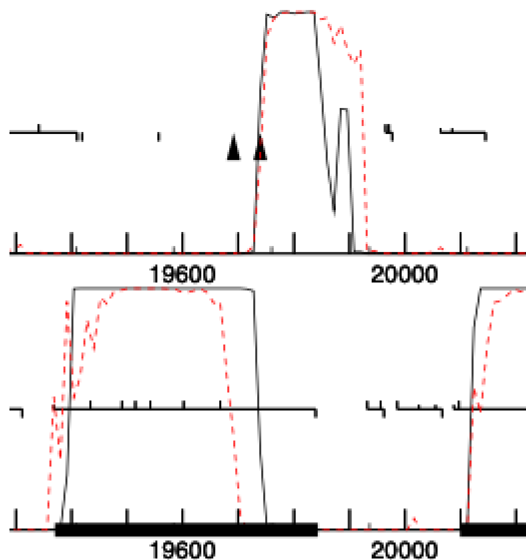
Stop @ 4486: Somewhat tricky start. best call based on stats is start 4085, but produces -17 gap/overlap. RomansRevenge doesn't have the most commonly called start site (7); choices are start sites 9 (4085) or 11 (4103). Other phages that have this gene also overlap with the preceding gene stop site, but not necessarily by 17 base pairs.

For function call, the qualification for a ribbon-helix-helix based on this forum (https://seaphages.org/forums/topic/5437/) was met in this gene. The gene has a B-sheet of 6-10 amino acids (indicated by Ee) followed by a short spacer and two alpha helices (indicated by Hh).

Stop @ 8228: Start: 41 @6378 has 11 MA's, so went with this one instead of LORF (0 MA's).

Stop @ 13249: There is a large gap before the gene but there is no coding potential according to host-trained GeneMark. Synteny with MellowYellow. While there are many hits to the major capsid hexamer protein in NCBI BLAST and phagesDB BLAST, there are no hits on HHPred. However, this call is consistent with the observed major capsid hexamer protein in Cluster DC, which has been experimentally proven to have a hexamer structure in its capsid protein.
https://seaphages.org/forums/topic/5430/

Tail assembly chaperone fusion protein (or lack thereof):
Relatives SilentRX and BruhMoment have fusion proteins and canonical slippage, but neither of their sequences matched RomanRevenge's closely. Debbie and I could not identify slippage in RomansRevenge, despite the presence of coding potential in another frame (below).
https://seaphages.org/forums/topic/4665/



Stop @ 31559: Potential gene here, but not sure. Only atypical CP present on self-trained genemark. Produces gene that has -28 overlap with downstream gene (which DOES have other members in its pham), if this one is included. No hits of any kind in either BLAST or HHpred. Decided to remove.

Stop @ 42971: Sufficient evidence as supported by NCBI Blast, Phages DB, and HHpred for a HicA-like toxin since there are multiple calls with sufficient e-values for that function. A SEA-PHAGES forum (https://seaphages.org/forums/topic/5342/) indicates that it is possible for HicA-like toxin to be present without the antitoxin (which we did not identify).

Stop @ 43480: deepTMHMM called 1 TMD with a 20 AA length. Because this is longer than SEA-PHAGES guidance, left call as NKF.

Stop @ 51584: Possible HTH DNA binding protein based on evidence hits, but the spacer between the first and second alpha helix is 8 amino acids long instead of 3-4. Called it a HTH.

—------------------------------------

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.