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. Benllo
Your Name. Suparna Bhalla
Your Institution. Mount Saint Mary College
Your email. Suparna.bhalla@msmc.edu
Additional emails. (for correspondence). Suparna.bhalla@msmc.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.

Gene 17 – needs to be looked at. We annotated it as an ORPHAM. In this region of the phage, several members have two minor tail proteins. Gene 17, has no sequence similarity to other phages in this cluster but some similarity to phages in other clusters calling it a "tail spike protein". We left it as a protein with no function

Genes 47 and 48 on PECAAN (45 and 48 on phamerator) were not included as the ORF's were predicted on the reverse strand and did not fit the profile of a valid genes. Additionally they overlapped with two forward genes)

GENE 64 – 41775 – 41995 Forward – Keeping this gene and a start site at 41755
Only Glimmer calls start at 41775. Region encompasses all coding potential. Phages with high alignment such as Circum and KeaneyLin with e-value of 5e-40. There is no significant overlap with other genes. Starterator supports start at 41775 which is a shorter ORF, with an ATG start and good scores, but it also equally supports start at 41730 which is a longer ORF, but a TTG start and poorer score. Both give equal number of 1:1 alignment with 5 similar phages in the cluster each. Will stay with the ATG at 41775.
Evidence for a function:
PhagesDB, NCBI BLAST and HHPRED all call this gene to be a hypothetical protein with unknown function.

Gene 65 - 41806 – 41988 Forward – Not included
Overlaps with gene 64 which is a longer ORF and will not include in the final annotation

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?

NA 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)?
YES 8. Have you cleared your Draft Blast data and have you re-Blasted the submitted DNA Master file?
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:

- Yes PECAAN output
  - DNA Master shorthand (previously used format)
  - Spreadsheet
  - Powerpoint
- Yes 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.:

- Yes PECAAN output
  - DNA Master shorthand (previously used format)
  - Spreadsheet
  - Powerpoint
  - Word document (must be easily searchable)
  - Other: Describe.