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. Rikishi
Your Name. Ann Powell
Your Institution. University of Evansville
Your email. ap96@evansville.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.

The Rikishi genome was very similar to the comparator phage Jollison, Spilled, and Spelly. It includes 244 genes and 42 tRNAs (highlighted in green in the notes file) that match with the tRNAs of Jollison and/or Spelly and one tmRNA.

Many of the genes found in Subcluster BE2 phage were also found in our phage genome in the same relative positions. These genes include: HNH Endonuclease (Rank 2), Terminase (Rank 1), Portal (Rank 1), capsid maturation protease (Rank 2), major capsid protein (Rank 1), head-to-tail adaptor (Rank 3), endolysin (Rank 1), holin (Rank 2). The genome also contained several genes found in Siphoviridae: major tail protein (Rank 1), tail assembly chaperone (Rank 3), tape measure protein (Rank 1), and minor tail proteins (Rank 1).

We deleted 13 draft genes from the auto-annotation and noted one gene that was in the phamerator map (highlighted in salmon in notes) but not in our draft annotation.

Regions of concern in Rikishi:

SEA_RIKISHI_14: we chose the Glimmer-selected start of 7151 as it had the most MAs. It is a very small and is in a small Pham. Check to make sure it is a gene.

SEA_RIKISHI_39: glimmer Calls 17588, genemark calls 17525. Start 17588 is supported. Some phage call endolysin and some phage call glycosyltransferase. We called endolysin based on "Two endolysin" comment for BE cluster phage in online annotation forum

SEA_RIKISHI_44: there was good (and equal) evidence for two approved functions, LysM-like peptidoglycan binding protein and endolysin. We selected LysM-like peptidoglycan binding protein.

SEA_RIKISHI_86: Glimmer and GeneMark call 63886. We selected 63889 as it is the second of two tandem starts.

SEA_RIKISHI_96: We called NKF, but some phage called terminase, small subunit.

SEA_RIKISHI_156: GeneMark called 88952, not called by Glimmer. Very small, only 108 bp. May not be a gene.

SEA_RIKISHI_205: GeneMark called 102982, not called by Glimmer. 77bp overlap (GM's predicted start at 102982); 59bp overlap (start at 103000). Start at 103000 has only 1 MA but may be the best start.

SEA_RIKISHI_223: Glimmer called 107447, GeneMark called 107405. Start 107447 has a 16 bp gap, Start 107405 has a 26 bp overlap. Start 107447 is the most annotated start with 38 MAs, while start 107405 has 20 MAs. We selected Start 107447.

SEA_RIKISHI_260: Glimmer called 119828, not called by GeneMark. May not be a gene, only 126 bp.

There is a large gap, 1163 bp between SEA_RIKISHI_260 and SEA_RIKISHI_261 (also in comparator phage). The is no coding potential in this area and we did not identify any ORFs with good support in this gap.

Aragorn and tRNAScan SE identified a few tRNAs that were not in the autoannotation and we added these. They align with the tRNAs in the comparator phage Jollison and/or Spelly. Aragorn also identified one tmRNA.

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?
NA 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:

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

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