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. Dashyla
Your Name. Nicholas Klotz
Your Institution. Webster University
Your email. nicholasklotz@webster.edu
Additional emails. (for correspondence). marypreuss34@webster.edu, andrewfazio@webster.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.

In the original auto-annotated version of Dashyla by PECAAN, there was a gene found in the forward direction from 2,788 to 3,135 that was deleted. The gene was listed as an orphan in PhagesDB and had no NCBI BLAST or PhagesDB BLAST results. Since there was no strong evidence supporting that it was a gene, it was deleted.

In gene 7, the WhiB family transcription factor, a -31bp gap was chosen over the other potential start sites. Originally we were going to select the next start site, which was a 29bp gap, however, in Starterator, this was not the most annotated start, and the selected start site of 5,531 had 147 manual annotations, which made it more compelling to pick. Also, in HHPRED, many of the hits had coverage beginning at the start of the protein when the -31bp start was selected, which indicated to us that the first part of the protein was necessary for its function. Please review.

In gene 8, the function call for RuvC-like resolvase may have been a stretch; there were some hits in both phagesDB and NCBI BLAST for that function, but most hits called hypothetical protein as the function. There were a couple HHPRED hits that indicated this function, however, the e-values were iffy (the two best hits had e-values of 0.093 and 0.13). We were unsure whether this was enough evidence to support this function call.

There was a very large gap (984 bp) between genes 30 and 29. The gap was BLASTed (BLASTx), and there were 2 results suggesting a protein call, however, when the nucleotide sequence was BLASTED, it had almost perfect identity with many more phages. 2 phages called a very tiny protein (found by GeneMark, but not Glimmer) 93bp in length. Those two phage calls were the only hits supporting a gene call here, and many other phages didn't call a protein for the same/similar sequence. It was concluded that this is likely too small to be a protein coding region, so the gene was deleted. Its position was from 16,646 to 16,738bp.

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?
N/A 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?
N/A 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
Yes Spreadsheet

What is the file type (sort) submitted for QC to document your gene calls? Choose only one.:
PECAAN output