

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. BigShaq

Your Name. Erin Doyle

Your Institution. Doane University

Your email. erin.doyle@doane.edu

Additional emails. (for correspondence). dane.bowder@doane.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.

We had no major issues with this genome. In general the CS2s seem pretty well annotated so that made our job easier.

We added gene 52 (47565-47720) to capture a peak of coding potential. This did create a slightly larger overlap (47bp) with the previous gene, but other phages do have this gene added (for example, Breezic_53).

Genes 45 (43867 – 44046) and 47 (44745 – 45131) we made start calls that disagreed with starterator and some other CS2 phages. Our reasoning is explained more fully in the notes, but we considered coding potential coverage, BLAST hits, and RBS scores, and the size of the CS2 subcluster and relationships between those phages.

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)?
- 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)
- Spreadsheet
- Powerpoint
- Word document (must be easily searchable)
- Other: Describe. Students worked in DNA Master but recorded their thought processes in a Google Slides presentation. Final starts and functions were put in a Google sheet for a quick overview. While grading annotations, we entered the data into PECAAN, which was used to generate the final files.

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)
- Spreadsheet
- Powerpoint
- Word document (must be easily searchable)
- Other: Describe.