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: Frankenweenie

Your Name: Kathleen Cornely

Your Institution: Providence College

Your email: kcornely@providence.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.

When we had questions about the gene annotation, we posted our questions to the cluster-specific forum on the SEAPHAGES website. When we had questions about where to call a start, we indicated our thought process in the Notes field in the complete DNA Master file. We found this genome to be particularly challenging to annotate, not just because of its length, but because there was often conflicting information regarding the start site. Many times Starterator called a start that was not the most annotated and we found that the most annotated start provided better BLAST data. Other times we were confused about whether to prioritize a most annotated start, better BLAST data, a -4 overlap versus a -1 overlap, production of a protein that had two initial Met residues or that was similar in length to the other BM cluster phages. Having only five BM cluster phages also made this annotation challenging, as Frankenweenie sometimes followed the pattern of the other BM cluster phages, but other times did not. We also noted that Frankenweenie has a lot of orphams, which also made the annotation challenging. We also investigated the gaps, but many times we couldn’t find a viable gene to insert in these gaps, either because of lack of coding potential or lack of BLAST hits to phage or host genes.

The annotation was carried out by a group of four students as a summer research project in the Summer of 2022. Students used DNA Master, phagesdb, Phamerator, HHpred, TMHMM, TOPCONS, SOSUI, JPred, and coding potential maps. The students were not given access to PECAAN for this genome. Only the instructor used PECAAN to double-check gaps and other data that students posted in the Notes.

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](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?

yes 3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.

yes 4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?

yes 5. Has the [documentation been recreated](https://seaphagesbioinformatics.helpdocsonline.com/article-86) from the Feature Table to match the latest file version?

yes 6. Have tRNAs followed the [tRNA protocol](https://seaphagesbioinformatics.helpdocsonline.com/undefined), **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](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?

yes 8. Have you cleared your Draft BLAST data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?

yes 9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?

yes 10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?

yes 11. Did you [delete the genes](https://seaphagesbioinformatics.helpdocsonline.com/article-65) that you meant to delete?

Now, [make a profile of the file](https://seaphagesbioinformatics.helpdocsonline.com/article-64) you plan to send. (And you can save this file for [Review to Improve!)](https://seaphagesbioinformatics.helpdocsonline.com/untitled-18)

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.

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.