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. ShaboiShabazz

Your Name. Allison Johnson

Your Institution. Virginia Commonwealth University

Your email. aajohnson@vcu.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.

Functional annotation:

Gene 9 annotated as head-to-tail adaptor based on synteny (right before head-to-tail stopper) compared to G1 phages (no HHPred or CDD support).

Gene 20, 22 annotated as hypothetical protein with no HHPred or CDD support for function. I’m of a generation that was trained with those as requirements for functional annotation, so I may be out of date. These genes appear to be a minor tail protein for most of the other G1 phages in phagesdb blast and NCBI blast, with 100% sequence identity.

Gene 51 may be a transcription initiation factor, through [HHPred hit with 2021 publication](https://www.rcsb.org/structure/7NVU) (probability of 98%, coverage of 94% and evalue 10-9. I did not post this to the forum for approval, it’s not on the approved functions list and I suspect you prefer I submit this annotation so have annotated as hypothetical protein.

Gene additions:

Gene 46 (SSC: 35517-35906, no evidence of coding potential, excellent RBS score) were inserted due to large gaps as well as synteny in G1 cluster phages where these genes have been previously called.

Gene 55 (SSC: 38643-39014), to cover coding potential in a gap and 100% alignment with 98 other cluster G1 phages. Annotated function as MPME1 based on the 100% identity to the group.

Gene 62 (SSC: 41545-41766 ) was inserted to cover coding potential and align with other cluster G1 phages (100% identity to 30 others). My students didn’t catch this one!

Gene deletions:

Deleted gene 55 from reverse strand (no coding potential), extended start position of gene 56 on forward strand over this region to cover very significant coding potential and align with cluster G1 genomes and synteny.

Deleted gene 60 from the reverse strand (no coding potential) at the end of the genome and directly overlaps gene 61.

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?

Yes 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:

X PECAAN output

X DNA Master shorthand (previously used format)🡪 only for frameshift

      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)

X Other: Describe. Web link to our wiki-base electronic notebook. This should be publicly viewable.