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. Aubs

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.

I’m from an older cohort with older habits, switching back to actinobacteriophages this year. Sorry I don’t have notes from my students now requested by SEA PHAGES. The complete notes file is the complete cds export from pecan, without their “rationale” which is on their wiki pages.

Question about start choice:

**30585-30836**, called according to Glimmer/Genemark predictions, second highest RBS (but longer ORF), covers CP. There is a mixture of blastp 1:1 matches for this gene, and a mixture of starterator Mas, not sure what is best.

Gene additions:

None

Gene deletions:

**29493-28837** on reverse strand to make room for gp 35 on forward strand

**45677-46048** on reverse strand to make room for gp 69 on forward strand

**46062-46394** on forward strand, orpham directly overlaps 5662-46414 with 304 pham members

Deleted another gene that is gp70 in the phamerator map, with 2 pham members that directly overlaps the same gene above. Region around 70 is tricky.

Function:

**109-564**, gp 1, annotated as terminase small subunit by HHPred hit (not quite up to SEA PHAGES criteria I learned- 10-4 evalue and no publication), next to large subunit, and pham homology.

**3612-4367**, gp 4, annotated as Hypothetical Protein, others annotated as capsid maturation protease. We don’t havesupport for capsid maturation protease but maybe you know other cluster specific tips to call this? Hits this [pfam](http://pfam.xfam.org/family/PF04233.17).

**28859-29092**, gp 33, we annotated as Hypothetical Protein, others annotated as holin. Unsure from function guidelines if the criteria “AND” is required. No transmembrane domains, though adjacent to endolysin. Gp34 has a single transmembrane domain.

**41716-42693**, gp 63, we annotated as RecA-like DNA recombinase. Hits RecA [5JRJ](https://www.rcsb.org/structure/5JRJ) in HHPred, but other annotators have called this AAA family ATPase.

**45662-46414**, gp 69 we annotated as DNA methyltransferase. A more specific cytosine-5-methyltransferase is not in the function guidelines.

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

X DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

Word document (must be easily searchable)

Other: Describe.