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. Altostratus

Your Name. Jonathan N. Lawson

Your Institution. Baylor University

Your email. Jonathan\_Lawson@Baylor.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.

Altostratus is a temperate phage placed in a new cluster, FS. Genome length is 39715 bp. With the lack of manually annotated genomes to compare to, we request additional review of this genome. Generally, there are many areas of weak CP, particularly in frames in the opposite orientation. Many called genes have “moderate” to weak CP and have been called to maintain orientation and fill gaps.

There are a couple of regions that we request detailed review of

* At the end of 20+ reverse genes between 21700-23000
* At the end of the genome 37900 to the end

We will describe genes that have been added and deleted from these regions.

Structural genes were able to be called with more detailed functions based on identifying HHPRED results required for Head-to-Tail adaptor and stopper.

The tyrosine integrase was identified, and we have called it an immunity repressor and Cro based on gene organization and HHPRED results.

The TAC was identified and FS called.



Isolated reverse gp26 well supported by CP. 42 bp gap for promoters.

Considered adding gene start 21949 stop 21764 to close gap, but weak CP and no blast or HHPRED hits kept us from adding it.

Glimmer and GM called gp28

Small 81bp ORF between gp28 and gp29 was not added due to lack of CP evidence and small size.

We added gp 29, gp30, gp31 to fill a large gap. All had weak CP but nothing more convincing on forward direction. Gp31 was short (90 bp) but fills a gap with a -8 bp overlap.

Gp32 called by glimmer and GM.

Gp33 only call by Glimmer but we added this fill a gap and had strong SD scores. Also had a BLAST hits and HHPRED hits in the 80%s.

END of GENOME

Unable to resolve gap between gp66 and gp67. We added gp67 to keep the strong HHPRED hits, but needed to delete a GM called reverse gene to do so. We considered added other ORFs on reverse strand here to fill gaps, but decided gp67 was a stronger candidate.

Kept gp68 with strong CP and HHPRED hits to PH domain proteins.

Deleted a gene start 38825 stop 38974 to make room for promoters needed for gp68 and gp69. CP and HHPRED hits in 80%s made this call difficult but we thought necessary.

Added gp70 to fill gap. Gene is small but with some CP and the fact it prevented a >100 bp gap we included it.

We worked with AlphaFold to support our annotations. Most of these ORFs have been folded as a monomer, dimer, trimer, with ATP, with NAD, with dsDNA, with ssRNA, and with adjacent genes in the genome. These can all be accessed here <https://baylor.box.com/s/52d0dsonbabe9vn16yoa78390s7vs0dp>

Students are currently working with other tools like FoldSeek in their independent projects. All are exploring questions related to how Alphafold/foldseek can be used to enhance annotation. I look forward to sharing insights at the faculty meeting in June.

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?

N/A 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:

Yes 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.:

Yes PECAAN output

      DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.