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. Albanese

Your Name. Tammy Adair

Your Institution. Baylor University

Your email. Tamarah\_Adair@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.

CDS 77 - 1699

 /note=Minor tail protein seen on a portion of the genes for AL and AU cluster but not strong evidence on HHPred, NCBI, PhagesDB blast. Baseplate or receptor binding protein

CDS 1709 - 2587

 /note=Minor tail protein called on 3 AU phage. Same HHPRED hits as Gene 1.

CDS 2774 - 3199

 /note=Small amount of coding potential is not covered according to the GenMark Report with the start at 2774, however, moving the start back means increasing the overlap by 44bp.

CDS 3132 - 4697

 /note=(Start: 67 @3132 has 72 MA`s), (Start: 168 @3306 has 4 MA`s), (Start: 190 @3342 has 1 MA`s), Other comparable genomes have moved the start even though it is a 68 bp overlap.

CDS 6754 - 7527

 /note=(Start: 22 @6688 has 3 MA`s), (Start: 44 @6754 has 88 MA`s)

 /note=The 47 bp overlap from the initially called start, the poor RBS scores, and its uncommon called start in starterator made it optimal to push the start forward to get a more plausible gap, better RBS scores, and the second most common start in starterator.

CDS 10902 - 11810

 /note=No matches on HHPred with an acceptable e-value, but other two pieces of evidence support major tail protein function.

CDS 19329 - 19769

 /note=Confirmed with TMhMM and SOSUI prediction

CDS 21431 - 21856

 /note=Starterator not available for gene 27, LORF smallest gap; matches Oxynfrius; CDD evidence is low 0.00126484

CDS 21849 - 22175

 /note=Original Glimmer call @bp 21933 has strength 3.94; GeneMark calls start at 21834

 /note=Changed to limit overlap

 /note=The most common length (309) is not available. 342 is too much overlap and -5 vs -4 final score.

CDS 22144 - 23535

 /note=Start Changed. 1:1 (Delete overlapping gene 30)

CDS 23651 - 24061

 /note=BLAST- Aligned with Greenhearts-Helix-Turn-Helix DNA-binding domain protein

 /note=HHPred- NKF

 /note=Phagesdb- Aligned with Oxynfrius 30 HTH All pham members call HTH

 /note=CDD e-value is not great.

CDS 25146 - 25703

 /note=Unsure whether the function was a RuvC-like resolvase or Holliday junction, so Holliday junction was selected because it was inclusive and the highest frequency in the pham. This group did some exploring and suggest Swiss Model indicates RuvC-like.

CDS 27314 - 27745

 /note=membrane protein confirmed with SOSUI 1 5-24 LWVAIGASCLVAGAIGAAAG Primary and TMHMM

CDS 27745 - 28020

 /note=TMHMM and SOSUI suggest membrane protein

 /note=1 4-25 PDSILLAFALIFFFAALFLWGR Primary

CDS 28164 - 29111

 /note=The gene has coding potential and was identified as a Cas4 family nuclease by Blastp, Phamerator. However, HHPred identified it as a DNA binding protein. An exploration using Swiss Model and FatCat indicates this is closer to 6PPU\_A (a double-stranded DNA break resection protein) than 4r5q (Cas4). From April 14, 2022.

CDS complement (39887 - 40120)

 /note=Potential helix-turn-helix DNA binding domain; No significant evidence on HHPRED or CDD but called by 1 Pham member (EastWest\_39) and several hits in NCBI BLAST from 2020/2021.

CDS complement (40245 - 40808)

 /note=ORF not found in gap.

CDS 43774 - 44034

 /note=HHPRED evalue 0.000003; CDD evalue 0.00003; low identity and coverage

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?

NA 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)?

NO 8. Have you cleared your Draft\_Blast data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?

I am working on a remote desktop server. It times out before the BLAST is complete. All genes have been BLASTED and the starts confirmed. This info is on Pecaan and in the complete notes 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)

no 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

yes DNA Master shorthand (previously used format)

yes Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

yes Other: Describe.

Procedure for Annotation:

* Students were given instruction and practice on Arthrobacter Phage Elesar using DNA Master and the template and the bioinformatics guide.
* Each student then annotated 3 or more genes for Phage Albanese and checked 3 or more annotations.
* The students paired up and presented their annotations and received feedback.
* The students submitted corrected annotations in their notebooks.
* Following review they entered their annotations using Pecaan. This is the first time I have submitted Pecaan notes as final annotations.
* I reviewed the PECAAN annotations and made some corrections.
* This AK genome is very similar to Oxynfrius.

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.