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. **Shrubaron**

Your Name. **Dr. Hari Kotturi**

Your Institution. **University of Central Oklahoma (old)/West Virginia School of Osteopathic Medicine**

Your email. **hkotturi@osteo.wvsom.edu**

Additional emails. (for correspondence). **Harikotturi@gmail.com**

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.

We investigated the 111 bp gap between 5775-5887 (gene 13 and 14) and did not find any matches. There is no evidence of it being a gene.

Deleted Gene 19 – (7730-7560-reverse). DNAMaster thinks this is a gene. There is no evidence of it being a gene. We checked Phamerator, PhagesDB, Starterator, coding potential map, NCBI BLASTN and BLASTP, and HHPred. Also, when compared to other annotated members of this cluster there is no reverse gene in this region. Based on all the evidence, we deleted this gene.

Added Gene 19 – (7527-7757-forward). DNAMaster missed this gene. There is enough evidence of it being a gene. We checked Phamerator, PhagesDB, coding potential map, NCBI BLASTN and BLASTP, and HHPred. Also, when compared to other annotated members of this cluster, there is a forward gene in this region. Based on all the evidence, we added this gene.

We investigated the gap (7757-7982) right before the portal protein (gene 20). There is no evidence of a gene in that region.

Deleted Gene - (10252-10133-reverse). DNAMaster thinks this is a gene. There is no evidence of it being a gene. We checked Phamerator, PhagesDB, Starterator, coding potential map, NCBI BLASTN and BLASTP, and HHPred. Also, when compared to other annotated members of this cluster there is no reverse gene in this region. Based on lack of no evidence, we deleted this gene.

Added Gene 22 – (10112-10201). DNAMaster missed this gene. There is enough evidence of it being a gene. We checked Phamerator, PhagesDB, coding potential map, NCBI BLASTN and BLASTP, and HHPred. Also, when compared to other annotated members of this cluster, there is a forward gene in this region. Based on all the evidence, we added this gene. SMART members can check it.

We investigated the gap between 13196-13384 (gene 26 and 27). There is no evidence of a gene.

Gene 28 – (15441-16253) – **SMART members can look at it**. some annotated members of this EC cluster call this gene “major capsid pentamer protein.”we did not find any evidence of this other than some NCBI BLASTP and PhagesDB. There is no evidence in HHPRED. As per Debbie's posting on cluster-specific annotation tips, the cluster EC has two major capsid proteins. We think it should be called NKF.

Genes 34 (19527-19970) and 35 (20036-20260). This cluster does not have a frameshift. None of the annotated members call a frameshift. SEAPHAGES faculty forum also confirms no frameshift.

Added Gene 57 – (37378-37581). Glimmer and GeneMark did not call this gene. We found this gene by investigating the space between 37192-37542. We also looked at the coding potential in GeneMarkS. This gene is also found in other annotated members of this cluster. It also has a good Z score (2.56) and final score (-4.37). We added the gene

Added Gene 62 – (39385-39558). Glimmer and GeneMark did not call this gene. We found this gene by investigating the space between 39385-39555. We also looked at the coding potential in GeneMarkS. This gene is also found in other annotated members of this cluster. It also has a good Z score (1.04) and final score (-6.62). We added the gene.

Gene 65 – (40390-41022). **SMART members should look at this gene**. We get hits for DNA ligase in PhagesDB and NCBI. HHPRED hits do not show any evidence of DNA ligase. We think it should be called NKF.

**SMART members should look at this space**. We investigated the gap between 42226-42441 (gene 68 and 69). Some members in this cluster have a gene in this gap, and some do not have any genes. There is no evidence of a gene in this region in this phage genome. There is no PhagesDB, NCBI BLASTP match. We did not add this gene.

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? **We checked and no tRNA in this genome**

**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? **We were having trouble with blasting**.

**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)

      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.:

**X** PECAAN output

      DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.