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. Kharcho and Ottawa

Your Name. Adam Rudner

Your Institution. University of Ottawa

Your email. arudner@uottawa.ca

Additional emails. (for correspondence). edoddmoh@uottawa.ca

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.

**Gp7: Fills in gap, low coding potential**

**Gp14: new gene added, large overlap with previous and low coding potential gene but fills in gap**

**Gp16: added gene to fill gap, low coding potential**

**Gp17: No coding potential, fills in large gap**

**Gp45 : minor tail protein based on Blast but question about the carbohydrate binding domain**

**Gp91 : Potential reverse strand gene to fill in gap, low coding potential**

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):

Y 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.)?

Y 2. Are all the genes ‘Valid” when you click the [Validation button](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?

Y 3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.

Y 4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?

Y 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?

Y 7. Has the [frameshift in the tail assembly chaperone](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?

Y 8. Have you cleared your Draft\_Blast data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?

Y 9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?

Y 10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?

Y 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)

Y 1. Have any duplicate genes been deleted?

Y 2. Has the Notes field been cleared (using the automated buttons)?

Y 3. Do the gene numbers and locus tags match?

Y 4. Are the correct Feature\_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?

Y 5. Do the function names in the Product field either match the official function list or say “Hypothetical Protein”?

Y 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:

Y PECAAN output

Y DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

Y Other: Describe. Word/Google Docs

What is the file type (sort) submitted for QC to document your gene calls? Choose only one.:

Y PECAAN output

Y DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe. Notes were recorded in PECAAN, and imported into DNA Master.

**Gene 7 (Low coding potential):**

Start Sites:

Graphical user interface, text, application, email

Description automatically generated

GenemarkS/Host Genemark:

Diagram

Description automatically generated A picture containing chart

Description automatically generated

Pham Maps:

Chart

Description automatically generated

**Gene 14 (Low coding potential):**

Start sites:

Graphical user interface, text, application, email

Description automatically generated

GenemarkS/Host Genemark:

Diagram, box and whisker chart

Description automatically generatedBox and whisker chart

Description automatically generated with medium confidence

c

c

Pham Map:

Diagram

Description automatically generated

**Gene 16 (Low coding potential)**

Start sites:

Graphical user interface, application

Description automatically generated

GenemarkS/Host Genemark:

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

c

Pham Map:

A picture containing diagram

Description automatically generated

**Gene 17 (No coding potential):**

Start sites:

Graphical user interface, application

Description automatically generated

GenemarkS/Host Genemark:

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

Pham Map:

Chart

Description automatically generated

**Gene 91 (Reverse gene. Fills gap. Minimal coding potential in GenemarkS)**

Start sites:

Graphical user interface

Description automatically generated with low confidence

Graphical user interface

Description automatically generated with medium confidence

GenemarkS/Host Genemark:

Diagram, timeline

Description automatically generatedDiagram

Description automatically generated

Pham Map:

Timeline

Description automatically generated

**Gene 45:**

HHPred: high probability to carbohydrate binding domain. Would this be the reason as to why all the other genes in the pham are called as a minor tail protein?