**When you are finished, please save it as "PhageName\_coversheet.pdf".**

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. Eugenia

Your Name. Vipaporn Phuntumart

Your Institution. Bowling Green State University.

Your email. vphuntu@bgsu.edu

Additional emails. (for correspondence). [vphuntu@bgsu.edu](mailto:vphuntu@bgsu.edu)

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.

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?

Y 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:

x PECAAN output

x DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

Word document (must be easily searchable)

x Other: Describe. We combined our word doc. Into single PDF

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.

**For SMART member**

There was NO tRNA present.

There was NO frameshift in the tail assembly chaperone.

Gene **3082** - **2399** (Reverse) has been **deleted,** based on

1. Phamarator: PhageDB indicated that it belongs to **Pham 6989** with 7 Members, however they all are drafted genomes
2. No Synteny with other B1 phages in genome map
3. NCBI BLAST showed no hit

A screenshot of a computer

Description automatically generated

Gene **8944 -9378** (Forward) has been **added,** is annotated as HNH endonuclease, based on

1. excessive gap on Phamarator (shown below as gene 10)
2. there is an ORF with LORF,
3. Length= 435 bp, Gap=9, Spacer=13, Z-score=1.392
4. BLAST-Start: [HNH endonuclease [Mycobacterium phage JacAttac] ],,NCBI, q1:s1 100.0%
5. RBS: Kibler 6, Karlin Medium, 1.392, -6.354315264641582, 5.33247E-99
6. SIF-BLAST: ,,[HNH endonuclease [Mycobacterium phage JacAttac] ],,YP\_009018323,100.0,5.33247E-99 SIF-HHPRED: HNH\_3 ; HNH endonuclease,,,PF13392.10,31.25,99.1 This gene was added based on Phamerator, ORF and BLAST
7. According to Phamerator, it belongs to Pham 150005

A graph with numbers and lines

Description automatically generated with medium confidence 

1. Gene Within the gene candidate portion of PECAAN, the gene selected above has a smaller gap and spacer than the gene that starts at 34,040. Also, it is the longer open reading frame with a better final score.

Gene **13714** to **14013** (Forward): [**Pham 152825 report**](http://phages.wustl.edu/starterator/Pham152825Report.pdf)**: Potential start site are 13663 and 13714, I called 13714 (not LORF) because of:**

1. **Coding potential starts at a second ATG not the firs GTG**

A graph of a graph

Description automatically generated

1. **A better Z-core and final scoreA screenshot of a score

   Description automatically generated**
2. **Blast results**

**A screenshot of a computer

Description automatically generated**

**Gene 33,965-34,351:** In PECAAN, Glimmer and GeneMark did not agree on a start site.

Glimmer said that the start site was 34,043 while GeneMark said 33,965. I chose the GeneMark start 33,965 for two reasons:

1. coding potential starts at approximately 33,900 as shown.

A graph of a number

Description automatically generated with medium confidence

1. Within the gene candidate portion of PECAAN, the gene selected above has a smaller gap and spacer than the gene that starts at 34,040. Also, it is the longer open reading frame with a better final score.

A screenshot of a computer

Description automatically generated

Gene **16974** to **16711** (Reverse): Not LORF

Gene **18355** to **18684** (Forward): Not LORF