**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. Psullivan

Your Name. Vipaporn Phuntumart

Your Institution. Bowling Green State University.

Your email. vphuntu@bgsu.edu

Additional emails. (for correspondence). 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

 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.:

      PECAAN output

X DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.

**For SMART members**

Gene 16021-16398 (Forward),There is a **frameshift translation** in the tail chaperone assembly which will need to be corrected for gene 16021 forward (16021 - 16398 ; 16021 - 16820). The sequence for the frameshift is 16370-16371 GGGG**G**AAG. The **G** in bold is the slippery base. The shift is a -1 frameshift which will look like this:



Delete; Gene 27860 to 28012 (Forward), Length 153 base pairs, 50 amino acids.

We believe this is not a gene. We delete it because it is a single member, colored white. Starterator also called it Orpham, No data.

Gene **6491-7459** (Forward), Length: 969bp, 322aa. We believe that this is the more accurate start site (not 6545) due to it being the LORF as well as genes with high synteny being 322aa long as well. GeneMark called this 6491 start, while DNA Master & Glimmer did not (it calls 6545). Based on PECAAN, LORF and BLAST compared to annotated genome, **we believe 6491 is the correct start site.** Lysin B is highly conserved (100% identity with the three most similar phages to PSullivan).







Gene **45597-45776** (Reverse) Length: 180bp, 59aa. We believe that this is a more accurate start site due to the fact that GeneMark called it (45776) as the start site, and the sequence length between this gene and other genes in the pham with synteny and a high identity being the same. The alternative start site which Glimmer called does not have the same sequence length between it and other genes it has synteny with. Neither of these predicted start sites is the LORF, however the selected start site has the shortest gap.





Gene **45773 to 46042** (Reverse) Length: 270bp, 89aa. We believe this is a more accurate start site because GeneMark called it as the start site, it is the LORF, and it is the only predicted start stie with a short gap. In addition, the gene has the same sequence length as other genes in the pham with synteny and a high percent identity.





Gene **46767 to 46919** (Reverse) Length: 153bp, 50aa. We believe this is a more accurate start site because it is the site with the LORF, and it has the same sequence length as other genes in the pham with high synteny.





Delete; Gene **49702** to **49544** (Reverse), **159** base pairs
**52** amino acids. We delete it because it is a single member, colored white. Starterator also called it Orpham, No data.

\***Note: Bold numbers are our calls**