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. BaBullseye

Your Name. Allison Johnson

Your Institution. Virginia Commonwealth University

Your email. aajohnson@vcu.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.

I’m from an older cohort with older habits, switching back to actinobacteriophages this year. Sorry I don’t have notes from my students now requested by SEA PHAGES. The complete notes file is the complete cds export from pecaan. My students document annotation on a wiki and I’ve compiled there notes [here](https://wiki.vcu.edu/display/phagelab/BaBullseye%2Bannotation%2Bdocumentation), but another SMART team member didn’t really want that format.

**24354-24548**, went with start with highest RBS and most MAs in starterator/most 1:1 blastp matches, covers all the coding potential.It’s not the longest and there is some disagreement in starterator.

**26694-25152 reverse gene.** I don’t understand the starterator report for this pham, which would support a start position at the Genemark/Glimmer predicted 26460 position rather than the full length 26694. Blastp, CDD and HHPred support annotation of the full-length gene, starting at 26694. RBS is fairly similar for all starts, 26460 and 26694 cover all the coding potential. I suspect the pham switches once the full length gene is annotated, and this is supported by 99% query coverage to D29 homolog, which is in a different pham.



**26460-26694**, predicted by both tools, has 4 MAs/most 1:1 blastp matches, covers all coding potential. It’s soooo short.

**30456-29878**, predicted by both tools (though black bar on Genemark map starts more internal), has most MAs, 1:1 blastp matches, covers all coding potential. There is some disagreement in starterator with starts at 30471.

**Three tRNAs** called according to Aragorn coordinates, confirmed infernal scores > 35

5402-5474 ends with first C of CCA, (following sequence is TC)

5479-5552 ends with first C of CCA (following sequence is AA)

5594-5666 ends with first C of CCA (following sequence is AA)

Gene additions:

**2143-2439 on forward strand**, to cover obvious strong coding potential in a genome region where genes are transcribed in forward direction. Function: HNH endonuclease

Gene deletions:

**2307-2158** on reverse strand to make room for HNH endonuclease, gene 4 on forward strand

**5554-5435** on reverse strand to make room for three tRNAs on forward strand

Function:

**3037-3999, 3999-4664** annotated as minor tail protein based on Welkin’s comments in cluster A Annotation Tips 🡪 gene length, position. This annotation isn’t consistent throughout the pham for 3037-3999. More confident in 3999-4664 because of CDD to collagen triple helix repeat.

**10812-11618**- capsid maturation protease by synteny? There is no HHPred or CDD support, and no homology to Langerak\_gp4, but pham seems to be annotated with this function.

**13475-13852** hits [5A21](https://www.rcsb.org/structure/5A21), I believe chain C- please check function for head-to-tail adaptor

**22579-24000** we annotated as hypothetical protein. Please check functional annotation of minor tail protein? There is no HHPred or CDD support, just not sure why other people have called this.

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?

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

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

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)🡪 only for frameshift

      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

      DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

Word document (must be easily searchable)

X Other: Describe. Web link to our wiki-base electronic notebook. This should be publicly viewable.