Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name: Cerulean

Your Name: Kathleen Cornely

Your Institution: Providence College

Your email: kcornely@providence.edu

Additional emails:
(For correspondence)

Please check each box indicating completion of each task. Annotation Guide section #'s indicated

- ☑ 1. Does the genome sequence in your final contain the same number of bases and is it the same as the posted sequence on phagesdb.org?
- ☑ 2. Are all the genes "valid" when you click the "validate" button? *Section 9.3.2*
- ☑ 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number? *Section 9.3.3*
- ☑ 4. Have all old BLAST hits been cleared, and all gene features reBLASTed? *Section 9.3.4*
- ☑ 5. Are the locus tags the phage name? *Section 9.3.3*
- ☑ 6. Has the Documentation been recreated to match the information in the feature table? *Section 1.4*
- ☑ 7. Have tRNA ends been adjusted with web-based Aragorn and/or tRNAscan SE? *Section 9.5.3-4*

8. For the items below, generate a genome profile, and review the following. *Section 11.3*

For the YourPhageName_CompleteNotes.dnam5 file:

- ☑ a. Have any duplicate genes (or any with the same stop coordinate?) been removed?
- ☑ b. Does every gene have **one and only one** complete set of Notes (see fig 12.2 in the Annotation Guide)?
- ☑ c. Do the functions in the Notes match the official function list?
- ☑ d. Is the function field EMPTY for all features?
- ☑ e. Do the notes contain the initial Glimmer/GeneMark data from the autoannotation?

For the YourPhageName .dnam5 file:

- ☑ a. Have any duplicate genes (or any with the same stop coordinate?) been removed?
- ☑ b. Is the Notes field empty for all the features with no known function?
- ☑ c. Do the function names in the Notes match the official function list, when applicable?
- ☑ d. Is the function field EMPTY for all features?

9. Describe any issues or specific genes that you were unable to satisfactorily resolve, and warrant further inspection in the Quality Control review.

Gene 1: could probably be deleted; there are no BLAST results; although we did not delete the gene in our submitted DNA Master file.

Gene 6 SSC: Start 3367 Stop: 3759 (FWD). SCS: Agrees with both Glimmer and GeneMark. Original Glimmer call @bp 3367 has strength 9.49. CP: ORF includes all coding potential shown on GeneMark output. SD: -3.812, best score. Gap: 22 bp overlap. BLAST: hypothetical protein gp6 of Obama12, q7:s1; gp5 of Arturo, gp6 of Wile and hypothetical protein gp8 of Nyxis, q8:s1 for all three; hypothetical protein gp6 of Gadost, q1:s1. LO: 393 bp, not the longest possible ORF. ST: Suggested start coordinate agrees with prediction. F: Hypothetical protein