Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name: Shaobing

Your Name: Stephanie Voshell

Your Institution: Virginia Tech

Your email: svoshell@vt.edu

Additional emails: (For correspondence) stephanie.voshell.87@gmail.com

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.

Shaobing differed from related K1 phage that have been annotated in a few regions, most notably at the very beginning and towards the end of the genome. This did not present any issues, but the alignment was not perfect for a few genes due to slight differences in the N-terminus amino acids. We feel that we have made reasonable start site calls in these regions. We annotated a -1 frameshift in the tail assembly chaperone. Throughout the genome we noticed that there are often larger gaps between genes than we have seen in other genomes, but we did not see evidence for adding more genes or lengthening others.