Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name:	Jumbo
Your Name:	Karen Klyczek
Your Institution:	University of Wisconsin-River Falls
Your email:	karen.k.klyczek@uwrf.edu
Additional emails: (For correspondence)	j.alfred.bonilla@uwrf.edu

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

- [x] 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?
- [x] 2. Are all the genes "valid" when you click the "validate" button? *Section 9.3.2*
- [x] 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number? *Section 9.3.3*
- [x] 4. Have all old BLAST hits been cleared, and all gene features reBLASTed? *Section 9.3.4*
- [x] 5. Are the locus tags the phage name? *Section 9.3.3*
- [x] 6. Has the Documentation been recreated to match the information in the feature table? *Section 1.4*
- [x] 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:

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

For the YourPhageName .dnam5 file:

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

We used both GenemarkS and the coding potential in Godornia bronchialis (the only Gordonia choice in Genemark.hmm); the latter did not have strong CP throughout the genome. We erred on the side of keeping genes even if they did not have good coding potential since CP may be under-representing genes.

For function analysis we used HHPred with the current pdb70 and Pfam_A; tgr_pham was not listed. In addition, we used the Phamerator maps of GMA3 and Gmala1 (though with caution, since these are not SEA-PHAGES genomes, but are the closed matches to Jumbo).

We did not use Starterator since Jumbo is a singleton and had few matches to other phams