Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name:	Gompeii16
Your Name:	JoAnn Whitefleet-Smith for Daniel T.
Your Institution:	Worcester Polytechnic Institute
Your email:	jaws@wpi.edu
Additional emails: (For correspondence)	

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.

Note, Bircsak and Gompeii16 have identical sequences except for the addition of a single "C" at nt position 4004 in Gompeii16 (between 4003 and 4004 in Bircsak). Initial QC was done side-by-side, with one decision (see gene73 below) where there is a good argument for either choice.

Gene 2: NNF or terminase?

Gene 15: One excellent blast to head-to-tail connector protein of TheloniusMonk phage, assign this function?

Gene 36: Additional upsteam start sites that would increase number of aa aligning with top blast hits.

Gene 38: Use upstream start site to get better alignment with top blast hit? (note no GeneMark