Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name:	Arudococcus
Your Name:	A. Javier Lopez
Your Institution:	Carnegie Mellon University
Your email:	jlaa@andrew.cmu.edu
Additional emails: (For correspondence)	swm3@andrew.cmu.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.

1. New gene added with start at 879. Not called by Glimmer or GeneMark. Glimmer called an overlapping gene on the opposite strand @bp 690 with strength 4.69 ** not called by GeneMark. We deleted that gene because there was No coding potential and no BLAST hits. Instead we called this gene on the bottom strand based on significant coding potential and protein BLAST hits against other phages in GenBank.

2. Programmed frameshift identified between gene19 (start 11349) and gene20 (stop 12124)

3. Genome contains multiple significant gaps between genes. Note large gaps on both sides of gene 68 (start 45856 TTG). Although this gene appears somewhat dubious (TTG start, not called by GeneMark, 51 codons), we agreed with Glimmer because the entire ORF and its start are conserved in multiple other phages