Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name:	Asapag
Your Name:	Ann Koga
Your Institution:	The College of Idaho
Your email:	akoga@collegeofidaho.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.

1. Genes 15-16 (stops=8613 and 9014). This appears to be a translational frameshift. Unlike other tail assembly chaperone genes, the predicted genes are gp 15—large gap—gp16. In other words, there is not a predicted overlap like in other genomes. Cozz, Emalyn, Yoshi, DS6A and a few others seemed to have some similarities in sequence in this region, but the actual area of the frameshift doesn't quite match anything. The frameshift can be narrowed down to a fairly small region (~20 bp). In gp 15, there is a decent alignment with Cozz, which ends at amino acid 84, a serine (nucleotide 8584). Then there is similarity starting with the amino acid sequence G-I-L in the "intergene" region (nucleotide 8604). Note there is coding potential in this intergene region. We chose to annotate this as a -1 frameshift, though Blast data is better with a +2 frameshift that deletes the valine at position 89. We could not find any