Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name:	StevieRay
Your Name:	Michael Kart
Your Institution:	St. Edward's University
Your email:	michaelkart@stedwards.edu
Additional emails: (For correspondence)	charlesh@stedwards.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) We suspect that the genome may have been misassembled: (a) Sequencing Notes: "A small percentage of reads (~7%) in this project matched a different phage, a Cluster A. It's assumed this was a minor contaminant." (b) The regions in StevieRay's genome from (38344, 38430) and (41562, 41648) match exactly and are 87 bases long. (c) In particular, we suspect the region around (36095, 36625) might be misassembled since we get a lot of Strep and other bacterial stuff. Solid matches don't show up again until the (37992, 38351) region. (2) Called (40026, 40466) a hypothetical protein (matches BOBI_76) as opposed to calling it a DNA methylase (matches DNA methylase in Malithi) since HHPred does not support a DNA methylase domain. (3) In gap (47010, 47219) there is homology to Jebeks hypothetical protein with Jebeks' coordinates (43733, 43927). No coding potential in GeneMark (Matrix : smeg)