Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name: SuperSulley

Your Name: Chris Korey

Your Institution: College of Charleston

Your email: koreyc@cofc.edu

Additional emails:
 (For correspondence)

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

[✓] 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?

[✓] 2. Are all the genes "valid" when you click the "validate" button? *Section 9.3.2*

[✓] 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number? *Section 9.3.3*

[✓] 4. Have all old BLAST hits been cleared, and all gene features reBLASTed? *Section 9.3.4*

[✓] 5. Are the locus tags the phage name? *Section 9.3.3*

[✓] 6. Has the Documentation been recreated to match the information in the feature table? *Section 7.4*

[✓] 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:

[✓] a. Have any duplicate genes (or any with the same stop coordinate?) been removed?

[✓] b. Does every gene have **one and only one** complete set of Notes (see fig 12.2 in the Annotation Guide)?

[✓] c. Do the functions in the Notes match the official function list?

[✓] d. Is the function field EMPTY for all features?

[✓] e. Do the notes contain the initial Glimmer/GeneMark data from the autoannotation?

For the YourPhageName .dnam5 file:

[✓] a. Have any duplicate genes (or any with the same stop coordinate?) been removed?

[✓] b. Is the Notes field empty for all the features with no known function?

[✓] c. Do the function names in the Notes match the official function list, when applicable?

[✓] 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.

SuperSulley Genome Annotation Notes
Lead Annotators: Jabbarrius Ervin and Tommi Nagumo

Genes that were deleted and justification (Using the original gene numbers)

- Gene 4: This gene is not called by GeneMark and it is a reverse frame in the middle of consecutive forward frames. It does not fit reasonably between gene 3 and 5 which has an acceptable 4bp gap.
- Gene 57: This gene is not called by Genemark and it is a forward frame in the middle of consecutive reverse frames. It does not fit reasonably between gene 56 and 58 which has a 43bp overlap.
- Gene 82: We were planning to delete this gene since it was not called by GeneMark; however, deleting it would create a 559 bp gap between Gene 81 and 83, which is a transitional region from reverse to forward frame. Also, Gene 82 has a blast match so more reason to keep it.

Genes that were added and justification

- Gene 73: Not called by GeneMark but called by Glimmer. Keeping the gene since it closes the gap between Gene 71 and Gene 72 appropriately with a 3bp overlap at the start start.

Empty Regions - Justification for no gene calls
- Between Gene 1 and Gene 2: 747 bp gap. There is no gene potential shown on GeneMark for any possible frames.
- Between Gene 39 and Gene 40: 210 bp gap. No gene potential seen on GeneMark in any possible frames. Although it looks like the start site could be extended forward, DNA master does not give earlier start sites as a possibility.