Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name: Finnry

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. If you are not sure how to do something, please see the Online Bioinformatics manual page "How to Pass Preliminary Review".

- ☑ 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?
- ☑ 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number?
- ☑ 4. Have all old BLAST hits been cleared, and all gene features reBLASTed?
- ☑ 5. Are the locus tags the"SEA_ PHAGENAME"?
- ☑ 6. Has the Documentation been recreated to match the information in the feature table?
- ☑ 7. Have tRNA ends been adjusted with web-based Aragorn and/or tRNAscan SE?
- ☑ 8. Has the frameshift in the tail assembly chaperone been annotated (where applicable?)

9. For the items below, generate a genome profile, and review the following. 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
- ☑ c. Do the functions in the Notes match the official function list?
- ☑ d. Are all three lines of functional evidence described for EVERY gene?
- ☑ 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 (including hidden marks?)
- ☑ c. Do the function names in the Product field either match the official function list or say "Hypothetical Protein"?
- ☑ d. Is the Function field empty (including hidden marks?)

☑ 10. Did you use PECAAN to annotate your phage?

If, so please describe how in the text field after question 11.

11. Describe any issues or specific genes that you were unable to satisfactorily resolve, and warrant further inspection in the Quality Control review.

We used PECAAN in combination with PhagesDB, Phamerator, and the HHRPED and BLAST sites. Students worked with PECANN and Phamerator maps to annotate the genome - going to HHPRED and BLAST to look more closely at certain genes. They inserted notes similar to the DNA Master Notes field in the PECAAN notes section. They also used a two page form for each gene in which they hand wrote the data and justified the choices that they made. These were handed in and used to provide feedback on their annotations- they also serve as a permanent written research notebook for their annotation. The data was then exported and put into DNA master for the final finishing and to annotate the ribosomal frame shift.

Genes to Check

- Gene 81 - HHPRED suggests a good match over the whole length of the protein for a toxin in toxin/antitoxin system, HicA-like. Other phage in the cluster call this as hypothetical protein, but this seems like a clear function for this ORF.

Added Gene

- Gene 54 - Start 42629 to Stop 42844. Picks up coding potential in Gene Mark that was not called by either glimmer or GM. Also called in Samty.

Deleted Genes

- Gene 130 (Phamerator #) - This gene is in the forward orientation in a stretch of reverse genes. Also, the start of Gene 130 is within the coding sequence of the previous gene 129, going in the reverse direction. Keeping 130 would mean the promoter for this gene would be in the coding region of 129 and the promoter for 129 would be within the coding region of gene 130. Gene 130 is not present in the highly similar phage Samty.

tRNAs

- Deleted tRNA - Start - 39829; in the middle of the coding sequence of gene 49 (start 39675 - stop 40340).
- Deleted tRNA - start 63568 - stop 63640; overlaps the start of gene 102 (start 63590).
- Deleted tRNA - start 65781 - stop 65873; tRNA (stop), overlaps with another tRNA - Asn - start 65802; which is also called in Samty

White Space Checks

- Region 31400 to 31900 - no coding potential in this region. Nothing called in similar phages in the cluster
- Region 36000 to 36500 - no coding potential in this region. Nothing called in similar phages in the cluster
- Region 63800 to 64200 - no coding potential in this region. Nothing called in similar phages in the cluster
- Region 69900 to 70100 had some coding potential in GeneMark. This is just before gene 125, which is the only forward gene on the far left arm. This gene is called in Samty. The region just prior to gene 125 had an ORF with coding potential that wasn't called. There might be enough room to place a gene upstream of 125 so that a promoter could go in both directions for this new gene and the previous reverse gene 124. Blast analysis did not reveal any homologous protein for this region. Our conclusion was to not call a gene in this space.