Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name: Purky

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

- Which gene is the excisionase?
    - Gene 37 (Start:28450) HHPRED has a good match to excisionase over its full length. Purky has an integrase and immunity repressor. This region of the genome does not match any other phage in the database, so there is no data to draw from other phage annotations. We called it an excisionase based on the HHRPED data and its location near the integrase.
    - Gene 31 (Start:26,584) and Gene 40 (Start - 30,247) also come back as having matches to excisionase, but we called the function as helix-turn-helix DNA binding domain

- 64 - (start:41,545) both glimmer and Genemark called, no matches to anything in NCBI blast - OrPham?

- 54 - (start: 41855) possible MmcB-like endonuclease. Good HHPRED hit with 99% coverage. No other phage call it with a function - all label it as hypothetical.

Added Gene

- Gene 39 - Start: 30118, Stop: 29705. Immunity Repressor. Not called by GeneMark of Glimmer. Strong matches to other phages in the cluster when this ORF was added and blasted.

Deleted Genes

- Forward Gene - 26747(start) to 27016 (stop) - only called by GeneMark. The start of this gene overlaps with gene 32 (phamerator) that is going in the reverse direction. Both Glimmer and GM call the reverse gene

tRNAs

- Deleted tRNA - Start:29787; only called by tRNA Scan SE with a COVE score of 1.54

White Space Checks

- Region 29500 to 30200 - this region was examined and the immunity repressor was identified here. GM had coding potential, but did not call the gene. We added the gene - see above.