Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name: Delian

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.

Questions about the functional assignments of two genes:

- SEA_DELIAN_35 - It is called as anti-toxin RelB-like in Frokostdame and just anti-toxin (not an approved function) in CaptainKirk2. We called it like Frokostdame, but the Blast and HHPRED information matches YefM family of anti-toxins, which isn't an approved choice.
- SEA_DELIAN_30 - called as NKF. Looks like it is in the set of Minor Tail proteins. The HHPRED data suggests a good match to an esterase, but that seems out of context for this region. So, we left it as NKF.

Deleted Genes and Justification

- Gene 79 (# based on Phamerator)--Has no supporting data from NCBI or HHPRED and is a reverse sequence in the middle of a forward sequenced group. There is no forward oriented gene in this region. Additionally, Captainkirk2 and Obliviate left out gene this gene.
- tRNA at Start - 31989 and Stop - 32054 was deleted. It was found by only tRNA Scan SE (Cove Score 9.33) and it is found within the putative immunity repressor coding sequence. Immunity Repressor is in the reverse orientation and the tRNA is in the forward orientation.

Empty White Space Checks

- No empty space in the genome had any regions that suggested the presence of a gene that wasn't called by Glimmer or GeneMark