Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name: JuJu

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".

- [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?
- [x] 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number?
- [x] 4. Have all old BLAST hits been cleared, and all gene features reBLASTed?
- [x] 5. Are the locus tags the"SEA_ PHAGENAME"?
- [x] 6. Has the Documentation been recreated to match the information in the feature table?
- [x] 7. Have tRNA ends been adjusted with web-based Aragorn and/or tRNAscan SE?
- [x] 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:

- [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
- [x] c. Do the functions in the Notes match the official function list?
- [x] d. Are all three lines of functional evidence described for EVERY gene?
- [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 (including hidden marks?)
- [x] c. Do the function names in the Product field either match the official function list or say "Hypothetical Protein"?
- [x] d. Is the Function field empty (including hidden marks?)

- [x] 10. Did you use PECAAN to annotate your phage?

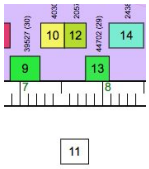
If, so please describe how in the text field after question 11.

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

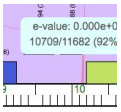
We used PECAAN in combination with PhagesDB, Phamerator, and the HHRPED and BLAST sites. Students worked with PECANN and Phamerator maps to annotate the genome - going to HHPRED and BLAST to look more closely at certain genes. They inserted notes similar to the DNA Master Notes field in the PECAAN notes 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.

Overall we feel pretty confident with the gene calls and functional identifications in the annotation

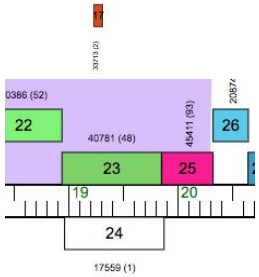
Deleted Genes and Justification (# that is in Phamerator_Draft Map)



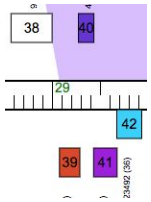
- Deletion of gene 11 - Gene 10 and 12 in the Phamerator map are forward and 11 is listed as an OrPham and overlapping in the reverse direction.



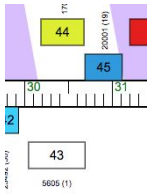
- Deletion of Gene 17 - This gene is in the reverse orientation in the middle of the tail assembly chaperone gene that is in the forward direction.



- Deletion of Gene 24. Not present in GeneMark. It is a called gene by Glimmer in the reverse orientation overlapping with Gene 23 and 25, which are in the forward orientation.



- Deletion of Gene 40: The ORF is only 81 bps and its orientation is in reverse-forward-reverse for the surrounding genes before and after it. Doesn't have enough room for a promoter and it was deleted. It does not appear in closely related phages



- Deletion of Gene 43: The gene is an OrPham and is in the reverse orientation while overlapping with a forward orientation gene.

Empty White Space Checks

35,600 to 35,800 - nothing in GeneMark with any coding potential
53,300 to 53,600 - nothing in GeneMark with any coding potential