Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 5-15-2018

Phage Name:	Zepp
Your Name:	Hui-Min Chung
Your Institution:	University of West Florida
Your email:	hchung@uwf.edu
Additional emails: (For correspondence)	hchung@uwf.edu

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.

- Delete gene #17 of the original draft, and add one tRNA gene (current gene#38)

- Genes of special interests:

Gene #6 encodes MuF-like minor capsid protein. This product seems to be a truncated version of Neferthena gp5 (capsid maturation protease and MuF-like fusion protein), of

which the similarity is in first 206 amino acids.in a location similar to that in phage Neferthena gp5. It differs from Neferthena gp5 as it lacks the C-terminal half of Neferthena gp5. Other EA5 phages seem to share similarity with Neferthena on this gene homolog, Zepp seems unique on this regard.

Gene #12 This protein is similar to phage Neferthena gp12, whose function was determined as NKF. Both has Bacteriophage HK97-gp10, putative tail-component. It has 98.4% probability being minor capsid protein as suggested by HHPred. This is the main reason we decided to call its function as minor capsid protein

Gene #33 This product might have something to do with ribosome assembly according to the HHPred hits.

Gene #38 tRNA-Ala(GCA) tRNA gene, was not predicted by DNA Master

Gene #42 This product might be associated with cell adhesion mechanisms, according to HHpred hits. I could not find a functional term on the SEA-PHAGES list

Gene #45 This protein is homologous to AAA-ATPase [Microbacterium phage Neferthena]. According to HHpred, however, it the hits go to thymidine kinase with 95% probability. It is a bit confusing why in several other phages, the AAA-ATPase function seems more preferably assigned.

Gene #47 This product, glycosyltransferase, does not exist in the similar area of EA5 phage Neferthena. This product might be derived from gene duplication of upstream gene gp46

Gene #51 This product might have toxin/antitoxin function; according to HHpred hit, it is homologous to VapB_antitoxin, but cannot find this term on the SEA-PHAGES functional assignment list.

Gene #54 This product has strong homology to Alpha-glutamyl/putrescinyl thymine pyrophosphorylase clades 1, 2, and 3, according to HHpred, but I could not find a function term in the SEA-PHAGES list.

Gene #59 This product for BLAST has no matches with e values below 10-7. It has hits in HHPred with 99% probability similar to the Streptomyces temperate phage serine recombinase, fC31 integrase. (INTEGRASE; HYDROLASE, SERINE RECOMBINASE, UNIDIRECTIONAL, SITE-SPECIFIC; 2.15A {STREPTOMYCES PHAGE PHIC31})

For question 10:

We use PECAAN extensively. We use the DNA Master in parallel with PECAAN. Practically, since PECAAN has almost all the essential aspects of DNA Master (except the BLAST alignments, which is easier to see on DNA Master), PECAAN was the primary tool for students to do genome annotation, for the sake of convenience and speed.

STEP 1: compare annotations between PECAAN and DNA Master. The discrepancy between the predictions of PECAAN and DNA Master or genes of debate were further assessed and finalized on PECAAN first, and then redo the coordination on DNA Master. All the evidences for the gene function obtained from HHPred and NCBI-BLAST, Conserved Domain Database andTmHmm were saved on PECAAN, as well as synteny explanation and other notes.

STEP2: export the annotation record from PECAAN (it has “Export (New SEA Format) CDS Full Annotation” choice). The export text file made the documenting the function and notes in DNA Master quite easy.

The annotation work was then finalized in DNA Master.