Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 6-28-2018

Phage Name:	Kalah2
Your Name:	Victoria Brown-Kennerly
Your Institution:	Webster University
Your email:	vbrownkennerly64@webster.edu
Additional emails: (For correspondence)	marypreuss34@webster.edu, schroeds@webster.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".

- ☑ 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.

- 6245-6433: This ORF has a significant overlap with the previous ORF (71bp).
- 11101-11610: Glimmer calls 11116 as the start, and it is the most annotated according to Starterator. However 11101 creates a longer ORF with a 4bp overlap, and it covers all the CP unlike 11116. Wanda is the only other phage that calls this 11101 start.
- 15583-16332: Klein and Minerva call 15583, but 1:1 alignment with hypothetical protein. Other phages call15604 and have a 1:1 alignment with a glycosyltransferase, but not all the CP is covered?
- 18367-18633 GAP: large gap, but no significant similarity by BLASTp on phagesdb and BLASTx on NCBI
- 25754-25936 GAP: ORF in this gap shows similarity to gp35 from Klein (32:8 alignment) and gp32 from BAKA (97:33 alignment). Both phages have this ORF in the same location/context in the genome, however no coding potential is seen in this region for Kalah2.
- 48357-48428: very short ORF (53bp); PHAM 5971; is called in two other draft genomes and 1 annotated genome (Klein)
- 53822-53782: deleted this ORF; has very little CP, no significant BLAST results, and other phages have a similar gap in this area.
- 108257-108391: without inserted ORF, there would be a 177bp gap. Inserted ORF 224.5, very short (134bp), but also found in LittleE and Omega (PHAM 38426) in same genomic context. Thin spike of coding potential in this region, but these coordinates do not cover all of it. Lengthening the ORF creates significant overlap with previous ORF.
- 110371-110535: end of the genome; BLASTed entire region and found similarity to gp 238 from Omega/gp218 from Thibault (PHAM 5634). Very little atypical coding potential in this region. Inserted this feature.

Genome Annotation Submission Cover Sheet

Preliminary Annotation Review Checklist 6-28-2018 PAGE 2: IMAGE

Phage Name:	Kalah2
Your Name:	Victoria Brown-Kennerly
Your Institution:	Webster University
Your email:	vbrownkennerly64@webster.edu
Additional emails:	marypreuss34@webster.edu, schroeds@webster.edu

Below is the image of a 2-ORF frameshift for the tail assembly chaperone, involving original ORFs 32-33 (in Phamerator.org). This is for PHAM 48329.

The frame shifts from +1 to +2. Nucleotide number 20512, an "A base" at the end of a small repeated poly(A) region, gets skipped.

In the sequence: 5'- CCCCAAAA it's the last A-base shown (at genome position 20512) gets skipped. ORF 32 should span 19990 to 20511 then frameshift skips the "A" to pick up at ORF 33 at 20513 - 20785. In the .dnam5 file, the ORF 33 shows a span of: 19990 to 20785.

Ribosomal frameshift:
19990..20511; 20513..20785. Below image: highlighted in dark yellow / orange is the shifted region.