Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name: Sheila

Your Name: Jordan Moberg Parker

Your Institution: UCLA

Your email: jparker@microbio.ucla.edu

Additional emails: (For correspondence)

Please check each box indicating completion of each task. Annotation Guide section #'s indicated

- [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? *Section 9.3.2*
- [x] 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number? *Section 9.3.3*
- [x] 4. Have all old BLAST hits been cleared, and all gene features reBLASTed? *Section 9.3.4*
- [x] 5. Are the locus tags the phage name? *Section 9.3.3*
- [x] 6. Has the Documentation been recreated to match the information in the feature table? *Section 1.4*
- [ ] 7. Have tRNA ends been adjusted with web-based Aragorn and/or tRNAscan SE? *Section 9.5.3-4*

No tRNAs

8. For the items below, generate a genome profile, and review the following. *Section 11.3*

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 (see fig 12.2 in the Annotation Guide)?
- [x] c. Do the functions in the Notes match the official function list? See gp20
- [x] d. Is the function field EMPTY for all features?
- [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 for all the features with no known function?
- [x] c. Do the function names in the Notes match the official function list, when applicable?
- [x] d. Is the function field EMPTY for all features?

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

Gene 8: The function of gene may need to be reconsidered. The gene has HHPred hits that suggest it
may be a portal protein. However, there are only two of such HHPred hits. The NCBI BLASTP and PhagesDB BLASTP do not suggest a function. However, gene 9 has BLASTP hits from both PhagesDB and NCBI that suggest it is a portal protein. So gene 8 may not be a portal protein.
Gene 18: The function of gene may need to be reconsidered. There is PhagesDB BLASTP and NCBI BLASTP hits that suggest it may be a major tail protein. Both have good e-values. However, there was no HHPred and CDD hits.

Gene 20: We have evidence from PhagesDB and NCBI BLASTP, as well as HHPred that this gene may be a queuine tRNA-ribosyltransferase, but this function is not listed on the function list provided. This function could potentially be added to the list of functions.
Gene 25: The function of this gene may need to be reconsidered. PhagesDB BLASTP suggests it may be a tail assembly chaperon with good e-value. However, HHPred shows functional support for tail component protein. It is very likely that this gene is involved in tail structure, but we are not sure what the exact function may be.
Gene 31: This gene's function should be reconsidered because even though it has only few hits for similarity to minor tail protein, those hits have a very low e-value (0.0). In additions, its placement in the genome, surrounding other tail proteins, is also evidence for it potentially being a tail protein.
Gene 69: This gene should be reconsidered because although it has good evidence for a PLA2, phospholipase function according to HHPred, NCBI and PhagesDB BLASTP, as well as CDD, it is a plant-specific protein.
Explanation of gaps in the genome
Between Gene 56 and 57: We believe the 137 gap between gene 56 and 57 is not a real gene. Although this region is called by other phages in the cluster in the same transcription direction as gene 56 and 57, such as Eremos, Sheila does not have coding potential in this region in the reverse direction. Other phages in the same cluster have called it to be a forward gene in between an array of reverse genes. However, for the reasons: 1) it is unlike for a single gene to change transcription direction 2) If a gene does change direction, it requires about 50 bp minimal for transcription machinery binding 3) the gap is only 137, which is smaller than the normal 200 bp requirement for a gene, 4) the possible start codon for the gap in the forward direction is mostly TTG, which is rare, it seems unlikely that this is a real gene.
Between 62 and 63: We believe that there is no gene in this gap because this is a place where the transcription changes orientation and a gap in the genome is usually required for the binding of transcription machinery. Additionally, this region does not have coding potential.