

Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name: Roy17
Your Name: Mary Ann Smith
Your Institution: Marywood University
Your email: masmith@marywood.edu
Additional emails:
(For correspondence)

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

- 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? *Section 9.3.2*
- 3. Have the genes been renumbered such that they go sequentially from 1 to the highest number? *Section 9.3.3*
- 4. Have all old BLAST hits been cleared, and all gene features reBLASTed? *Section 9.3.4*
- 5. Are the locus tags the phage name? *Section 9.3.3*
- 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*

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

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 (see fig 12.2 in the Annotation Guide)?
- c. Do the functions in the Notes match the official function list?
- d. Is the function field EMPTY for all features?
- 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 for all the features with no known function?
- c. Do the function names in the Notes match the official function list, when applicable?
- 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 10 - addition, not called by Glimmer or GeneMark; Gene 66- addition, not called by Glimmer or GeneMark; Gene75- translation initiation/elongation factor, most blast data from phagesdb and ncbi say nkf, but hhpred has very strong evidence for this call (Prann on ncbi also does); Gene 50 - we listed nkf, but hhpred comes back with ATP-dependent helicase or ATP-dependenet hydrolase, but no one else calls it - I want to, but am nervous about it - it seems something is there, but do we just go helicase; Gene 61 - has a large gap before it, deleted gene in order to remove 22bp overlap with "removed gene" - seems split on issue many annotated B1 do, many do not; we have nkf, but found RDF (reverse directionality function) Nacho and Vivaldi use it, but not sure if it could be in B1; Gene 89 - transcription termination/antitermination protein - not phagesdb. 1 (brann) ncbi but also hhpred