Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

Phage Name:	Joselito
Your Name:	Bernadette Connors
Your Institution:	Dominican College of Blauvelt
Your email:	bernadette.connors@dc.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*
- [x] 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:

- [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?
- [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.

1.SD scoring mstrix Kibler6 and Spacing weight matrix Karlin medium were used
2.Five genes deleted: (1) 40305-40385 and 43077-43322 because both were forward amidst reverse, (2) 47569-47625 (57nt) and 47622-47720 (99nt) because they were short and had no significant similarity with anything when searched against multiple databases, and (3) 1560-31 because it had no similarity to anything (please investigate as this seems where the genome circularizes and I am not sure if I can delete it)
3.This phage crashed Starterator. As a result, we searched phams on http://phages.wustl.edu/starterator/. We do have "percentage support" for each feature, but did not include those values in this submission.

Joselito

4. Joselito_26: 6.5% of phage call the start at 17613. This start is a candidate start in Joselito, and gives a 1:1 with other tape measure proteins. Neither Glimmer or Genemark called it at this start site.

5. Joselito_ 45 matches what is called DNA polymerase 1 with good identity, but has similarity to DnaQ-like (or DEDD) 3'-5' exonuclease domain superfamily. Unsure to call it DnaQ exonuclease (DNA polymerase III subunit) or DNA polymerase 1.

6. Joselito_4 not called initially, so it was inserted. Starterator supports this start site for other A11s (such as Jabith and Mulciber)

7. Regions 25791-26120 and 45102-45300 were investigated because of the gap size, but nothing of note was discovered.