**Genome Annotation Submission Cover Sheet**

**Pre-QC Phage Genome Annotation Checklist**

 **Phage Name:** Kenna

**Names:** Leticia Candra, Amanda Makara, Amro Nasser, Ilan Schwell, Evelyn Okorie, Carly Grossman, Yash Agarwal, Chris Houseworth, Haley Aull

 **Institution:** University of Pittsburgh

**E-mails:** LEC117@pitt.edu, ARM196@pitt.edu, AMN84@pitt.edu, IYS1@pitt.edu, ECO10@pitt.edu , CAG160@pitt.edu, YAA30@pitt.edu, CDH53@pitt.edu, HAA85@pitt.edu

 **Additional E-mails:** Instructor - Rebecca Bortz (RLB6@pitt.edu)

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 Kenna\_FinalCompleteNotes.dnam5 complete notes file:

* Have any duplicate genes (or any with the same stop coordinate?) been removed?
* Do the functions in the Notes match the official function list?
* Is the function field EMPTY for all features?
* Do the notes contain the Initial Glimmer/GeneMark data from the auto- annotation?

 For the Kenna minimalistic file:

* Have any duplicate genes (or any with the same stop coordinate?) been removed?
* Is the notes field empty for all the features with no known function?
* Do the function names in the Notes match the official function list, when applicable?
* 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 34 (26,501-27,1095bp) had HHpred results with probability >90%, cover=25%, e=0.1 that suggest **toxin** in *Yersinia*, but there was insufficient evidence to call this function more specifically to comply with function list.
* **Deleted** original gene 42 (29,921-30,043 bp) - No coding potential, no room for promoters to bind
* Original gene 56 (36059-36730 bp Rev) - HHPRED suggests function as **HTH DNA binding and Repressor** with probability >90%, but contradictory evidence suggested this was not a strong call.
* Original gene 85 (47,387-48,184 bp) - Hit to **replisome organizer** in HHpred with probability >90%, but not on functional calls list
* **Deleted** original gene 86 (47,804-47,950 bp Rev)- No CP. Complete overlap with forward gene 85
* **Deleted** original gene 101 (52,769-52,461 bp Rev) - Overlaps with genes 99:100. Reverse gene in the middle of forward genes with less CP.
* Gene 89 (49,368-49,775bp) had HHpred results with probability >90%, cover=74%, e=0.0011 that suggest **EAD-22**. This function is also called by *Gordonia* phage BritBrat.
* Found ORFs within gaps at 31,998-32159bp and 36,909-37, 130bp but they did not show GeneMark coding potential so no genes were added