Genome Annotation Submission Cover Sheet

Pre-QC Phage Genome Annotation Checklist

**Phage Name**: PhrostedPhlake

**Name**: Deborah Lee, Brianna Austin, Rakesh Chatrath, Dylan Falk, Shayna Dehmler, Logan Hellinger,

Deesha Desai, Cameron Hallihan, Swathi Thia Paruchuru

**Institution**: University of Pittsburgh

**email**: del52@pitt.edu, dbf14@pitt.edu, rchatwrath7@pitt.edu, deesha.desai@pitt.edu, logan.hellinger@pitt.edu, spd47@pitt.edu, clh176@pitt.edu, bca15@pitt.edu

**Additional emails**: 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 Section 1.4 recreated to match the information in the feature table?

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 complete notes file - 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 minimalistic file - 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.

**Deleted Genes:**

Original Gene 4 (Rev 2230-1994): Deleted due to having no similarity to any previous genes, orpham, no coding potential, and it's a reverse gene in the middle of many forward genes. Extending start on next gene gives 1:1 alignment and leaves 12 base-pair gap.

Original Gene 18 (Rev 10185bp-10364): Deleted because it’s a reverse gene in between two forward genes. Length of gene was 180bp. No significant BLAST and HHPred results. Also overlaps with gene#17 and 19 which are frameshift proteins.

Original Gene 39 (29304-29528): Deleted due to having no room for a ribosome binding site, no coding potential, and large overlap with gene 40.

Original Gene 90: (51427-52011) Deleted due to lack of coding potential and massive overlap with gene 91, which has a function called by BLAST and HHpred. Gene 90 has no significant HHpred hits and no BLAST results.

**Potential functional calls:**

Gene #38 (Rev 29676-29990) and #39 (Rev 29987-30253) antitoxin-toxin system is referred to as BrnA-BrnT from HHPRED and BLAST using Eyre, but this system is not an approved functional call.

**ADDED Gene #44** (32413-32850 Rev) 2nd LO that filled large gap. New gene has 1:1 alignment with several bacterial genes with XRE transcriptional regulator as functional call through NCBI BLAST. HHpred best hit 98%probability to SOS transcriptional repressor and many hits > 90%prob to transcriptional regulators.