Actinobacteriophage Genome Annotation Submission Cover Sheet

This Cover Sheet will accompany each genome’s annotation file(s) submission and succinctly describe the work that your students and you have done. This document ensures that the work done was as complete and thorough as it could be. Most important to the QC reviewer, denote where the trouble spots were in your annotation and how they were resolved.

Phage Name. KentuckyRacer

Your Name. Ann Powell

Your Institution. University of Evansville

Your email. ap96@evansville.edu

Additional emails. (for correspondence). jm757@evansville.edu

Describe any issues or specific genes that you would like to highlight for the QC reviewer. This includes any genes that you had questions about or received help with or that warrant further inspection in the QC review process. Include those genes that you deliberated on and/or want to strongly advocate for. If you contacted SMART, workshop facilitator, or a buddy school for help, please document.

The KentuckyRacer genome was very similar to the comparator phage IchabodCrane, Mindflayer, Spilled, and Spelly. It includes 43 tRNAs (highlighted in blue in the notes file) that match with the tRNAs of IchabodCrane and/or Spilled and one tmRNA (highlighted in salmon in the notes file).

Many of the genes found in Subcluster BE2 phage were also found in our phage genome in the same relative positions. These genes include: HNH Endonuclease (Rank 2), Terminase (Rank 1), Portal (Rank 1), capsid maturation protease (Rank 2), major capsid protein (Rank 1), head-to-tail adaptor (Rank 3), endolysin (Rank 1), holin (Rank 2). The genome also contained several genes found in Siphoviridae: major tail protein (Rank 1), tail assembly chaperone (Rank 3), tape measure protein (Rank 1), and minor tail proteins (Rank 1).

We deleted seven draft genes from the auto-annotation and noted one gene that was in the phamerator map (highlighted in yellow in notes) but not in our draft annotation.

We added two genes SEA\_KENTUCKYRACER\_37 and SEA\_KENTUCKYRACER\_189

Regions of concern in KentuckyRacer:

SEA\_KENTUCKYRACER\_2: we chose the Glimmer-selected start of 2218 as it had the most MAs, but the Genemark-selected start, 2263, would give a longer reading frame. Both starts are supported by data.

SEA\_KENTUCKYRACER\_27: a 317 bp gap 5’ to the selected start. There are additional starts that would increase the length of the ORF and decrease the size of the gap, but they were not supported by Starterator or RBS scores (change in direction here from reverse to forward).

SEA\_KENTUCKYRACER\_46: there was good (and equal) evidence for two approved functions, LysM-like peptidoglycan binding protein and endolysin, N-acetylmuramoyl-L-alanine amidase domain. We selected endolysin, N-acetylmuramoyl-L-alanine amidase domain in a 5-4 class vote.

SEA\_KENTUCKYRACER\_109: 632 bp gap (also in comparator phage). Did not find any ORFs supported by Starterator or coding potential in this gap.

SEA\_KENTUCKYRACER\_189: not called by Glimmer or Genemark. Two possible overlapping ORFs that fill this gap. Selected start 97971 as it gives a 1 bp gap while the other ORF would produce a 16 bp overlap. 97971 also has a few MAs on Starterator.

There is a large gap, 1154 bp between SEA\_KENTUCKYRACER\_263 and SEA\_KENTUCKYRACER\_264 (also in comparator phage). The is no coding potential in this area and we did not identify any ORFs with good support in this gap.

SEA\_KENTUCKYRACER\_290: a 317 bp gap 5’ to the selected start. There are additional starts that would increase the length of the ORF and decrease the size of the gap, but they were not supported by Starterator or RBS scores (change in direction here from reverse to forward).

Aragorn and tRNAScan SE identified a few tRNAs that were not in the autoannotation and we added these. They align with the tRNAs in the comparator phage IchabodCrane and/or Spilled. Aragorn also identified one tmRNA.

Please record yes/no for each of the questions below. If further explanation is needed, please add this item to the above box.

In the submitted DNA Master file (Yes/No):

Yes 1. Does the genome sequence in your submitted DNA Master file match the nucleotide fasta file posted on phagesDB (same number of bases, no N bases, etc.)?

Yes 2. Are all the genes ‘Valid” when you click the [Validation button](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?

Yes 3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.

Yes 4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?

Yes 5. Has the [documentation been recreated](https://seaphagesbioinformatics.helpdocsonline.com/article-86) from the Feature Table to match the latest file version?

Yes 6. Have tRNAs followed the [tRNA protocol](https://seaphagesbioinformatics.helpdocsonline.com/undefined), **COPYING** tRNA-AMINOACID type (DNA equivalent of the anti-codon) from Aragorn output - ﻿tRNA-Gln(ctg) - AND the ends been adjusted to match the Aragorn output?

N/A 7. Has the [frameshift in the tail assembly chaperone](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?

Yes 8. Have you cleared your Draft\_Blast data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?

Yes 9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?

Yes 10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?

Yes 11. Did you [delete the genes](https://seaphagesbioinformatics.helpdocsonline.com/article-65) that you meant to delete?

Now, [make a profile of the file](https://seaphagesbioinformatics.helpdocsonline.com/article-64) you plan to send. (And you can save this file for [Review to Improve!)](https://seaphagesbioinformatics.helpdocsonline.com/untitled-18)

No 1. Have any duplicate genes been deleted?

Yes 2. Has the Notes field been cleared (using the automated buttons)?

Yes 3. Do the gene numbers and locus tags match?

Yes 4. Are the correct Feature\_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?

Yes 5. Do the function names in the Product field either match the official function list or say “Hypothetical Protein”?

Yes 6. Has the Function field been cleared (using the automated buttons)?

How are you documenting your gene calls in class? Choose any/all that apply:

      PECAAN output

X DNA Master shorthand

X Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

X Other: Describe. OneNote Notebook

What is the file type (sort) submitted for QC to document your gene calls? Choose only one.:

      PECAAN output

      DNA Master shorthand (previously used format)

X Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.