Spilled 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. Spilled

Your Name. Ann Powell

Your Institution. University of Evansville

Your email. ap96@evansville.edu

Additional emails. (for correspondence). N/A

The Spilled genome was unclustered, but is very similar to the comparator phage IchabodCrane, MindFlayer, Wipeout, and Karimac (all cluster BE2 phage). The first part of the genome is duplicated and found at the end of the linear genome, such that SEA\_SPILLED 1-24 match with SEA\_SPILLED 261-284.

Many of the genes found in most phage genomes were located in our genome. Those not found include: Scaffolding protein (rank 2), minor capsid protein (rank 3), Head to tail adapter (rank3), Head to tail stopper (rank 3), and Tail terminator (rank 3). Siphoviridage phage proteins not found include: we only found one Tail assembly chaperones (rank 3) and Tail fiber or tailspike (rank 3).

We deleted thirteen draft genes and three tRNAS (highlighted in yellow and documented in notes).

Regions of concern in Spilled:

SEA\_SPILLED\_1: We called the Glimmer Start 1110, but start 1149 produces a shorter gap. Neither gap length is a violation.

SEA\_SPILLED\_9: Both starts 5336 and 5345 are possible starts. Start 5336 has the most Mas of starts available, but the gap is 132 bp.

SEA\_SPILLED\_14: Selected Genemark start 7360. Glimmer start, 7300 is the most annotated and both have 1:1 matches in comparator phage.

SEA\_SPILLED\_19: Glimmer predicted start 9639, start 6. It is the most annotated start site within this gene family and has 27 MAs (manual annotations). Start 3 at 9681 has 4 MAs. Start 9639 leaves a 170 bp gap.

SEA\_SPILLED\_44: We were unsure which function to call. One comparator phage, IchabodCrane\_43 called their gene LysM-like peptidoglycan binding protein and another comparative phage, MindFlayer\_44 called their gene endolysin.

SEA\_SPILLED\_67: We called start 51220, but 51229 is by far the most annotated start. Our start has a 26 bp overlap.

SEA\_SPILLED\_77: I am moderately confident of this gene call (function). We called DNA helicase. Both DNA Helicase and DNAB-LIKE Replicative Helicase are approved functions and are called in comparator phage.

SEA\_SPILLED\_138: I am moderately confident of this gene call (start). Although the start chosen was not the most annotated start, the most annotated start 85468 was not shown in PEACANN and it does not match comparator phage. Gap is 193 bp.

SEA\_SPILLED\_204: GeneMark and Glimmer both call 103265. We selected start 103177. Start 103264, start 5 has 14 MAs and is the most annotated start. We think Start 3, 103177, is a better start. Start 103376 has a 77 bp overlap.

SEA\_SPILLED\_205: Glimmer start 103376, start 3. Start 3 has 8 MAs and is the most annotated start. Start 4 has 7 MAs.

SEA\_SPILLED\_218: No function called, but there is a very large trans-membrane domain identified in the TMHMM analysis.

SEA\_SPILLED\_231: I am moderately confident of this gene call. Not sure which function to call, either phosphesterase or metallophosphoesterase. Both are called in comparator phage.

SEA\_SPILLED\_260: Not sure which function to call, Blast evidence supports endonuclease VII or HNH endonuclease. Both Endonuclease VII and HNH endonuclease are on the SEA PHAGE Official Function list.

We checked all gaps and no genes were added. All gaps were also present in the comparator phages IchabodCrane and Mindflayer.

Aragorn and tRNAScan SE identified a few tRNAs that we not in the autoannotation and we added these. They align with the tRNAs in the comparator phages IchabodCrane and/or Mindflayer.

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)

 N/A 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 FeatureTypes 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:

 X PECAAN output

      DNA Master shorthand (previously used format)

 X Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.

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.