JPandJE 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. JPandJE

Your Name. Ann Powell

Your Institution. University of Evansville

Your email. ap96@evansville.edu

Additional emails. (for correspondence). N/A

The JPandJE genome was very similar to the comparator phage Immanuel3, ToriToki, RetrieverFever, and Manuel. It includes a cassette of 22 tRNAs that matches with the tRNAs of Immanuel3.

Many of the genes found in most phage genomes were not located in our genome, these include: HNH endonuclease (rank2) but we have endonuclease 7 (gene 2), Capsid maturation protease (rank 2), Scaffolding protein (rank 2), No minor capsid protein (rank 3), Head to tail adapter (rank3), Head to tail stopper (rank 3), Tail terminator (rank 3), Major tail protein (rank 1) (siphoviridae), Tail assembly chaperones (rank 3) (sipoviridae), Tape measure protein (rank 1) (siphoviridae), Tail fiber or tailspike (rank 3) (siphoviridae), Tail tube protein (rank 1) (podoviridae), and Lower collar protein(podoviridae). Based on phage morphology and cluster (BF), we believe the phage type for this phage may be podoviridae.

We deleted two draft genes: JPandJE\_15 and \_16 (documented in notes)

We identified a split DNA primase (different reading frames with a large overlap) that was originally noted in Cluster A and BD phages, per Chris Schaeffer’s response to our question in the Annotation Forum.

Regions of concern in JPandJE:

SEA\_JPandJE\_11: We called the Glimmer Start 11783 to keep the overlap under 30 bp, but there is a longer start with Mas that produces a 30 bp overlap.

SEA\_JPandJE\_37: Glimmer predicted start 18570, start 12 is the most annotated start site within this gene family and has 11 MAs. Genemark predicted start, Start 10, 18534, has 3 MAs. Start 18570 has a gap of 115 bp, 18534 has a gap of 79 bp. Neither is a violation. We selected the Glimmer start 18570, but it could be either one.

SEA\_JPandJE\_71: Glimmer calls 37568 and Genemark calls 37553. Glimmer predicted start 37568 is not the longest ORF. Both starts encompass all the CP. Glimmer start 37568 has more MAs when compared to Genemark start 37553. 51 bp gap, 26 bp overlap.

SEA\_JPandJE\_85: Glimmer and Genemark predicted start 44959, start 2 is the most annotated start site within this pham and has 12 MAs. No other start has any MAs. Large gap to end of genome. Violation.

We checked all gaps and the only gene added was SEA\_JPandJE\_59, the second half of the split DNA Primase (see above). These gaps were also present in the comparator phages Immanuel3 and/or RetrieverFever.

Aragorn and tRNAScan SE identified several TRNAs that we not in the autoannotation and we added these. They align with the tRNAs in the comparator phage Immanuel3.

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)

Yes. 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

      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.