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. **DejaVu**

Your Name. **Robert Logan**

Your Institution. **Eastern Nazarene College**

Your email. **robert.logan@enc.edu**

Additional emails. **roblogan6@gmail.com**

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.

1. **According to DNA Master, tRNA Scan SE, and Phamerator-based synteny, a tRNA sits in the middle of gene 69. Both Glimmer and GeneMark call gene 69.**
	* **Gene 69 starts at 40778 and ends at 40915. It is a forward gene that is surrounded reverse genes.**
	* **The tRNA starts at 40849 and ends at 40783. It is in the reverse direction.**
	* **Gene 70 starts at 41118 and ends at 40897. There is less of a gap between the tRNA call and gene 70 than between gene 69 and gene 70.**
	* **Gene 68 starts at 40756 and ends at 38297. There is less of a gap between gene 68 and 69 than between 68 and the tRNA.**
	* **Aragorn v 1.2.41 did not detect a tRNA in the DejaVu genome with either the Standard genetic code or the Bacteria and Plant Plastid Code.**
	* **Six ED1 cluster genomes had synteny evidence of a tRNA. Six ED1 cluster genomes had no tRNA reported. Dejavu seems to have an orphan coding sequence at the loci in leu of the tRNA sequence. Based on this synteny, however, I would be inclined to call the tRNA. Yet, I am inclined to not trust DNA Master's judgment in this case. See the following point for more details on this.**
2. **Alarmingly, DNA Master doesn’t recognize the PECAAN reported gene calls as being valid, either in Standard genetic code or the Bacteria and Plant Plastid Code. This is bewildering since the gene calls have been confirmed outside of DNA Master (i.e., Glimmer, GeneMark, Starterator, syntenty, etc.).**
3. **The start positions for genes 14, 15, and 117 (called 117 in DNA Master, 116 in the notes because the tRNA wasn’t assigned a gene ID number) have been changed and should be confirmed.**
4. **There are also two orphans. One at gene 32 and another at gene 88/89. The one at 88/89 is another lone forward gene among reverse genes, which makes it a little suspect. Gene 69 is also considered an orphan, but that might be because it is being contaminated by a junk gene call or a junk tRNA call.**

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):

**No!** 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.)?

**No!** 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?

**No!** 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?

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

**No** 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?

**No!** 4. Are the correct Feature\_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?

* **DNA Master was throwing a "range" related error, which prevented me from opening the archived .dnam5 file to go back and correctly annotate the two back-to-back tail assembly chaperon genes (PECAAN genes #51 and #52).**
* **The predicted ribosomal slippage genomic loci is CDS join (24677. .25339;25339. .26558)**

**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.