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. Gilberta

Your Name. Frederick Baliraine

Your Institution. LeTourneau University

Your email. FredBaliraine@letu.edu

Additional emails. N/A

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. Investigated gap downstream of feature 36 at 27166-27261 bp, but even though there is a significant coding potential plateau in the area (between 27000-27150 bp), there is no start in phagesDB to accommodate it and there is even no CP line. Besides there are no hits to anything in phagesDB. As such, this gap was ignored.
2. Function for feature 58 (37505-37777 rev). There is a "mixed bag" of "specific" functions for this gene, with many hits to NrdH-like glutaredoxin and glutaredoxin in phagesDB, with a few hits to Thioredoxin for this Gilberta sequence. There are however more than 22 hits in HHpred to thioredoxin. I would go for thioredoxin based on the many hits to it, but the forum post of May 8, 2022 (https://seaphages.org/forums/topic/5376/?page=1#post-9554) suggests that the more general term "oxidoreductase" would do. I note though that in PDB, the reference sequence phage Onyinye gene78 for "oxidoreductase" is almost thrice as long 792 bp vs 273 bp of Gilberta) and mostly hits "Polyketide oxygenase PgaE," chain A, with ref Sequence in pfam hitting the "FAD folding domain" but I do not see any hits in Gilberta to any "FAD folding domain." Instead, in PDB Gilberta hits the "Molecule" THIOREDOXIN chains A & B, with the pfam ref hitting "Glutaredoxin or Glutaredoxin-like NRDH-redoxin, THIOREDOXIN; OXIDOREDUCTASE, GLUTAREDOXIN." On the other hand, the ref sequence for Thioredoxin, phage Cjw1 gp 37 (246 bp) has a comparable length to Gilberta's 273 bp and they both have several hits the molecule Thioredoxin chain A, but not the "Polyketide oxygenase PgaE," or "FAD domain."
3. The inserted gene at 46691-46747 rev (feature 83) is only 57 bp long, but it is part of an operon with a 4 bp overlap with the downstream gene and its start has a strong RBS score. Coding potential is significant but only seen in GeneMarkS. It has hits in phagesDb with q1: s1,100% to 6 phages, but all of them with e-value of 0.36.
4. No BLAST data for feature 83 (46691-46747 rev) in NCBI upon BLASTing in DNA Master, but has data in phagesDb as stated above.
5. Membrane protein: Gilberta gp 37 (27268-27423 rev); Gilberta gp 80 (45416-45712 rev); and Gilberta gp 86 (46976-47173 rev)? For these three genes, SOSUI and TMHMM show they are membrane proteins with 1 transmembrane domain detected by each of these two software, but DeepTMHMM (which the website claims that it is more accurate than TMHMM) and TOPCONs shows no transmembrane domain, but rather, signal peptides. We have called them membrane proteins, nevertheless, based on the SOSUI and TMHMM 2.0 calls as per the current (May 13, 2022) Official SEA-PHAGES Functions List.

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?

Yes 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

Yes DNA Master shorthand (previously used format)

Yes 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)

Yes Spreadsheet

      PowerPoint

      Word document (must be easily searchable)

      Other: Describe.