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. GigiOuiOui
Your Name. Cade O'Neill
Your Institution. Montana Technological University
Your email. Coneill2@mtech.edu
Additional emails. (for correspondence). Coneill1287@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.

Gene 64, Gene 65, Gene 66.

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?
Yes 3. Are the genes (and matching LocusTag numbers) sequential, starting with #1, counting by 1s.
Yes 4. Are the Locus Tags the "SEA_PHAGE NAME" format?
Yes 5. Has the documentation been recreated from the Feature Table to match the latest file version?
No tRNAs in genome, so N/A. 6. Have tRNAs followed the tRNA protocol, **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 been annotated correctly (if applicable)?
Yes 8. Have you cleared your Draft Blast data and have you re-Blasted the submitted DNA Master file?
Yes 9. Has every gene been described and supported in your Supporting Data file?
Yes 10. Did you investigate 'gaps'?
Yes 11. Did you delete the genes that you meant to delete?

Now, make a profile of the file you plan to send. (And you can save this file for Review to Improve!)

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
- [X] DNA Master shorthand (previously used format)
- [ ] 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
- [X] DNA Master shorthand (previously used format)
- [ ] Spreadsheet
- [ ] Powerpoint
- [ ] Word document (must be easily searchable)
- [ ] Other: Describe.

The region between Genes 39 (Reverse, hypothetical protein) and 40 (Forward, tyrosine integrase) contains the attB/attP recombination region of GigiOuiOui's genome. This is the reason for the 394 bp gap between the two genes.

Genes 64, 65, and 66 are a significant region that needs to be investigated further. We hypothesize that due to the coding potential shown in the GeneMark output, there are multiple possibilities: 1) the three ORFs that comprise these three genes undergo two programmed ribosomal frameshifts to produce one functional protein (DNA methyltransferase), 2), the resulting protein is a defunct pseudogene that serves no role and has been mutated over time from a functional ancestral protein, or 3) the genome was incorrectly sequenced. We called all three ORFs to bring this to your attention. When comparing the combined sequence spanning all three ORFs, there was an extreme similarity between the combined sequence and a bacterial gene product. This is why all three ORFs were called, and also why there is significant overlap in this region.

Inserted genes (not originally called in the auto-annotation) include: Gene 37, Gene 48, Gene 56, Gene 60, Gene 64, Gene 65, Gene 78, Gene 80, Gene 85,

Deleted genes that were removed from the original auto-annotation include: Gene 26, Gene 54

The programmed ribosomal frameshift gene occurs in Genes 12 and 13.