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.

Bacteriophage: **Chako**

Susanne Pfeifer
Arizona State University
spfeife1@asu.edu
mmilhave@asu.edu

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.

DNA Master File:

6. There were no tRNAs found in this genome
7. The frameshift in EA1 phages is not commonly noted; other individuals in the forum have been told to skip the frameshift annotation, but still call the genes "tail assembly chaperones" (TAC).

Profile:

1. There were no duplicate genes

Points of interest:

1. Deleted gene 40 of the original annotation due to poor BLAST hit, it being a forward gene amongst reverse genes, poor RBS score, and no synteny
2. TAC frameshift was not annotated as a forum post mentioned that EA1 phages are particularly difficult to annotate in this regard. Nonetheless, genes 15 & 16 have been called as TAC.
3. Gene 11 BLASTs 1:1 to minor capsid protein, but everything else was uninformative. Related phages did not call this gene as functional.
4. Genes 20-22 were called minor tail proteins via synteny, due to being in between genes 18 and 23, which had strong support for minor tail proteins
5. Gene 47 BLASTed well to both AAA-ATPase and thymidylate kinase, but due to closely related phages calling it thymidylate kinase, we did as well.

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):

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? **Yes**

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? **N/A (no tRNAs were found)**
7. Has the frameshift in the tail assembly chaperone been annotated correctly (if applicable)? **No**
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? **Yes**

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

1. Have any duplicate genes been deleted? **N/A (no duplicated genes were found)**
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)? **Yes**

How are you documenting your gene calls in class? Choose any/all that apply:

PECAAN output
DNA Master shorthand (previously used format)
Spreadsheet **Yes**
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)
Spreadsheet **Yes**
Powerpoint
Word document (must be easily searchable)
Other: Describe.