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: **Micasa**
Your Name: **Nicholas Klotz**
Your Institution: **Webster University**
Your email. **nicholasklotz@webster.edu**
Additional emails. (for correspondence): **marypreuss34@webster.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.

Gene 1 was not found by PECAAN or GeneMark, but there is strong evidence for this gene on NCBI BLAST. An error pops up for this gene in DNA Master for validation, but because of its CTG start codon. Some other A5 phages have this protein listed with a start codon of CTG, including MarysWell and Zolita. It yields good hits for HNH endonuclease in NCBI BLAST with 1:1 hits. This gene also has synteny to most other A5 phages in Phamerator. *validation error in DNA Master, incorrect CTG start codon

Gene 9 I was unsure about the start site; the LORF was a –4 bp gap indicating an operon and there were many 1:1 hits in NCBI BLAST for that start site (6130), however, both Glimmer/GeneMark annotated the start as 6157. In Starterator, Micasa do not has the most annotated start, but there are manual annotations (MAs) for both 6130 (19 MAs) and 6157 (29 MAs). I ended up selected the suggested start 6157, however, was unsure if the –4bp gap was a better option.

Genes 20/21 represented the ribosomal frameshift. We found that there was a –1 bp shift back. We compared the gene from Micasa to Swirley's annotated gene, which had 100% coverage/synteny with Micasa's. Using ClustalO sequence alignment, we found the –1bp gap of a repeated g nucleotide. We then BLASTed the final product (BLASTn and BLASTp) and got many good hits with low e-vals to other A5 phages. *validation error in DNA Master (gene 20-21 share an upstream coordinate)

Gene 24 had a Topcons/DeepTMHMM hit, but the most likely function seemed to be minor tail protein. I was unsure if this should be annotated differently due to the liklihood of it being a membrane protein.

Gene 86 I was unsure about the start site. The LORF (49444) was exactly 120bp, and all other options were smaller than that. Glimmer/GeneMark did not agree, but neither selected the LORF. The next best option was 49438, which was GeneMark's suggested start, with a length of 114 bp. In Starterator, this option had the most MA's, and there were 1:1 hits in NCBI BLAST with good e-vals. Other shorter options also had 1:1 NCBI BLAST hits, but weren't manually annotated as frequently. Based off Starterator, I selected 49438 as the start.

GAPS: there were large gaps (>100bp) at genes 3, 7, 20, 30, 31, 54, 68, 82, and 86. All gaps were investigated in the GeneMark and in Phamerator to find potential synteny of an unannotated gene. The largest gap (gene 86, gap of 934) was BLASTed and no coding potential was found.

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.)?
NO 2. Are all the genes 'Valid" when you click the Validation button?
YES 3Are the genes (and matching LocusTag numbers) sequential, starting with #1, counting by 1s.
YES 4Are 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?
YES 7Has 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
Word documents

What is the file type (sort) submitted for QC to document your gene calls? Choose only one.:

PECAAN output (exported into DNA Master)