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. Starcevich
Your Name. Marisa Pedulla
Your Institution. Montana Technological University
Your email. mpedulla@mtech.edu
Additional emails. (for correspondence).

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.

Please investigate our proposed attP site between genes 43 and 44 The top hit, bases 4,847981-4,847934 covers the tRNA-Lys gene in the M. smegmatis genome. This is presumably the attB

```
gene            4847908..4847983
                     /locus_tag="MSMEG_RS22905"
                     /old_locus_tag="MSMEG_4746"
     tRNA            4847908..4847983
                     /locus_tag="MSMEG_RS22905"
                     /old_locus_tag="MSMEG_4746"
                     /product="tRNA-Lys"
```

```
>gi|118467340|ref|NC_008596.1| Mycobacterium smegmatis str. MC2 155,
           complete genome
         Length = 6988209

 Score = 79.8 bits (40), Expect = 3e-15
 Identities = 46/48 (95%)
 Strand = Plus / Minus


Query: 219     gtgcgccgtgagggtttcgaacccccgacccgctgattaagagtcagc 266
               ||||||||| |||||||||||||| ||||||||||||||||||||||
Sbjct: 4847981 gtgcgccgtcagggtttcgaacccccggacccgctgattaagagtcagc 4847934
```

Please look at the lysinA region. Gene 35 is called as encoding lysinA; however, gp 33 has blast hits with n-termini of other lysinA genes. Gene 34, an pourported HNH endonuclease may have been added to the genome interrupting an ancestral lysinA gene. We called gene 33 as endolysin.

Deletions: We deleted our auto-annotated gene 50 (36950-37042), 70 (45970-45725, reverse), 71 (45960-46334), 72 (46720-46947, reverse).

We added genes
30 which was phagesdb draft gene 34 (26056-26196, changing start from GenemMarkS 26077),
31 phagesdb draft gene 35 (26193-26324, called by GeneMarkS),
43 which is phagesdb draft gene 47(32008-32208, reverse, GeneMark called),
69 whch is phagesdb draft gene 75 (45704-46807, not called by GeneMarkS),
70 which is phagesdb draft gene 79 (46804-46965, not called by GeneMarkS),
71 not in phagesdb draft (46946-47122, not called by GeneMarkS),

73 which is phagesdb draft gene 81 (47241-47459, not called by GeneMarkS),
91 not called in phagesdb draft (52521-52643, not called by GeneMarkS),
92 not called in phagesdb draft (52640-52795, not called by GeneMarkS),
96 not in pahgesdb draft (53988-54083, not called by GeneMarkS),
100 phagesdb draft gene 106 (54674-54871, called by GeneMarkS),
103 phagesdb draft gene 109(55513-55674, not called by GeneMarkS).

We **changed starts** for our auto-annotated genes: 12, 14 (frameshift), 18, 30, 41, 42, 61, 76, 86 93)

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?
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!)

N/A 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
DNA Master shorthand (previously used format)
Spreadsheet
Powerpoint
XXXX 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
Powerpoint
XXXX Word document (must be easily searchable)
Other: Describe.