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. WaddleDee
Your Name. Olga Y. Lubman
Your Institution. Maryville University
Your email. olubman@maryville.edu
Additional emails. (for correspondence). Olgalubman2018@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.

Formatting of Annotation Notes in Excel Spreadsheet:

The first column contains locus tags, which will number the genes as documented in the DNA Master.

1. Genes that were automatically called and subsequently deleted have been denoted as "deleted" in the spreadsheet.
2. Each gene is indicated to be read forward or reverse.
3. The original start and gene stop columns lists the start and stop codons automatically called by Glimmer and/or GeneMark.
4. "Start" column indicates the selected start site for the gene.
5. "Was start changed? "column indicates whether this start site is different or the same as what was automatically called, and updates automatically as start sites are changed.
6. The "length" column calculates the length of the gene automatically from the called start site and stop site. Genes less than 200 bp will be automatically highlighted in orange and genes less than 120 bp will be automatically highlighted in red tRNA will not be highlighted as long as their bp length falls within 70 to 85 bp.
7. The "Gene gaps", automatically calculates gaps based on the start site of the selected gene and the stop site of the previous gene, ignoring genes denoted as "deleted." It does not ignore these genes automatically, but the equations were adjusted to reflect the deletions. Gaps larger than 50 bp are automatically highlighted in yellow. Gaps of 0 bp or less highlight in orange. Gaps of -30 bp or less highlight in red. These are all true with the exception of gaps of -1 bp and -4 bp in length, which are excluded from being highlighted as those are acceptable overlaps.
8. Column J denotes whether the assigned open reading frame is the longest possible choice or not (*true* means yes for longest ORF, and *false* is NOT the longest ORF).
9. Functions are assigned based on the approved protein function list. The "Annotation Notes" column lists fold based probability and % sequence coverage based on HH pred
10. Highlighted in Neon Blue are ORFs that we propose to add to the function list. Column titled " new function support" has a link to a supporting document google doc file.

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.)?
    a. Yes
2. Are all the genes 'Valid" when you click the Validation button?
    a. Yes
3. Are the genes (and matching LocusTag numbers) sequential, starting with #1, counting by 1s.

    a. Yes
4. Are the Locus Tags the "SEA_PHAGE NAME" format?
    a. Yes
5. Has the documentation been recreated from the Feature Table to match the latest file version?
    a. 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?
    a. Yes
7. Has the frameshift in the tail assembly chaperone been annotated correctly (if applicable)?
    a. No
8. Have you cleared your Draft_Blast data and have you re-Blasted the submitted DNA Master file?
    a. Yes
9. Has every gene been described and supported in your Supporting Data file?
    a. Yes
10. Did you investigate 'gaps'?
    a. Yes
11. Did you delete the genes that you meant to delete?
    a. Yes

Genes too short :
WaddleDee_231

Genes that overlap:
284 overlap with 285

Places where we attempted to insert genes:
Gene 294; 1000 bases on each side of it. There is no coding capacity and there is a change in the direction of the transcription

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?
    a. Yes
2. Has the Notes field been cleared (using the automated buttons)?
    a. Yes
3. Do the gene numbers and locus tags match?
    a. Yes
4. Are the correct Feature_Types correctly selected (most will be ORFs, but check that tRNAs and tmRNAs are correctly labeled)?
    a. Yes
5. Do the function names in the Product field either match the official function list or say "Hypothetical Protein"?
    a. Yes
6. Has the Function field been cleared (using the automated buttons)?
    a. Yes

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

1. PECAAN output
2. DNA Master shorthand
3. Spreadsheet
4. Word document

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