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. **Nicole21**
Your Name. **Amanda Raimer**
Your Institution. **Radford University**
Your email. **araimer@radford.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.

There were 2 genes that had functions that appeared in blastp hits, HHpred hits, and matched similar genes in related phages but were not in the approved function list (highlighted in notes):

- **Gene 9 (Stop 8931), Function: "capsid maturation protease and MuF-like fusion protein", related phages: LeeLot, Giraffe, JhangoPhett. The closest on the approved list is "capsid maturation protease".**
- **Gene 20 (Stop 15857) – Function: "queuine tRNA-ribosyltransferase", related phages: LeeLot, Giraffe, JhangoPhett.**

Start of Gene 16 (Stop 13992) – two potential starts (13642, 13693), more confident in 13642 but 13693 also has strengths and was the original called start

Start of Gene 21 (Stop 16688) – two potential starts (16987, 16951), more confident in 16951 but not 100% confident because there is a longer ORF that is called in many phages (Starterator).

Start of Gene 59 (Stop 48959) – called 51766 as start because it is longest ORF and shows 1:1 ratio with similar proteins, but 51706 is a possibility because it is more commonly annotated and has higher RBS scores.

Start of Gene 66 (Stop 56227) – Starterator, Phamerator, and blastp show many possible start sites with very similar likelihoods of being called on Starterator. We called the most annotated on Starterator (56099), but the chance of calling when present was very similar to other starts we considered.

There are a number of genes near the end of the genome where the starts didn't produce majority 1:1 starts in blastp but were the preferred start in Starterator by a fairly large margin from the other potential start. I listed the genes here and also highlighted them in yellow:

- **Gene 83 (Stop 62426), Gene 96 (Stop 65802), Gene 99 (Stop 67226), Gene 102 (Stop 68322)**

We considered adding a gene between Genes 91 (Stop 64305) and 92 (Stop 64647) because there was a potential ORF that would reduce the gap and there was a gene in a related phage (Giraffe). However, the gene would be under 100bp and still leave some gap, there were only 2 significant hits in blastp, and only 10 genes in the pham of the Giraffe gene. In the end we didn't end up adding a gene there.

Genes 36 (Stop 34988) and 37 at first look like duplicates, but we see this duplication in other related phages (LeeLot, Giraffe, JhangoPhett) and comparing the amino acid sequence is an e-value $<10^{-10}$

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