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. HubbaBubba
Your Name. Dane Bowder and Erin Doyle
Your Institution. Doane University
Your email. dane.bowder@doane.edu; erin.doyle@doane.edu
Additional emails. (for correspondence). NA

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 a number of interesting genes in HubbaBubba that we spent some extra time on. Most of our issues were with functions- described below:

- Gene 24- We have a hit to MuF capsid decoration protein, but that is no longer an approved function, so we ultimately called it an NKF. We did have hits to the capsid maturation protease from D29, but after consulting with Debbie on forums, we decided to go with NKF because of poor HHPRED hits
- Gene 25- Based on synteny, if Gene 24 was the capsid maturation protease, we might call this scaffolding protein, but because we didn't make that call and low e value hits to scaffold protein in HHPRED we went with NKF. We realize this may be unconventional and get changed during QC, but wanted to be more conservative and careful than to call it "because everyone else did"
- Gene 30- we called this a head to tail stopper due to HHPRED hits to a phage head completion protein, and because many other phages have assigned this function. Curiously, the other CR4s have not, but we felt the evidence was strong enough.
- Gene 42- we were unsure whether this should be considered a minor tail protein. It is a big gene following the tape measure (it would be the 5th of them), but other CR4s were split on whether they called the minor tail function or not. Ultimately, we decided not to, but are open to it being called.
- Gene 45- we called the lysin A, glycosyl hydrolase domain function because of (poor quality) HHPRED hits to other hydrolases (chitinase), and other phages called this, but we seriously considered calling an NKF. We were not sure which way to go on this.
- Gene 60- we called ASCE-ATPase function based on the tutorial provided by Sally M at last years faculty meeting. We considered RecA and believe we followed the rules properly on this, but it is a new function call for us.

We also had a start we were unsure on, but made the best choice based on what we thought:

- Gene 63- we ultimately went with start 48828, but were pretty torn due to very poor RBS values. This gene is not present in other CR4 phages, and starterator was uhelpful- but we chose the start because of the 8 bp overlap to reduce the gap substantially.
- We considered adding a gene between gene 71 and 72, but ultimately decided we wouldn't because we couldn't get a start to work. Another CR4 phage, IDyn, has a gene here, and we noticed some minor alternative coding potential, but we just couldn't find a good start to utilize, so we opted to drop it.

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?
5. Has the documentation been recreated from the Feature Table to match the latest file version?
NA 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)?
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
- X 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.:

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