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. BotCity
Your Name. Erin Doyle
Your Institution. Doane University
Your email. erin.doyle@doane.edu
Additional emails. (for correspondence). dane.bowder@doane.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.

Overall, we had fun annotating this genome. The DNs are a weird cluster, but we had previous experience with a DN phage from a couple of years ago (BearBQ). Genes and areas that could use some additional scrutiny are as follows:

- Genes 15 (8333-8611) and 16 (8791-9012) are the Tail Assembly Chaperones. We identified the most likely frameshift as a -1 frameshift at position 8602. However, the slippery sequence CCCGGAA is not canonical, so we did not call it.

- Reverse gene 39 (Pecaan) (42 in Phamerator) located at 29,825-29,682 vs the forward gene located at 29,824-29,967 (Gene 43 in Phamerator). This region of the genome has a lot of mosaicism. Only one of these genes can be kept because keeping both does not allow for a large enough gap to switch from reverse to forward transcription. The forward gene (Phamerator 43) is found in 52 other phages; the reverse gene (Pecaan 39, phamerator 42) is an orpham. However, the orpham has stronger coding potential and is called by Glimmer and Genemark and creates a 4 base overlap with the previous gene (Pecaan 38, phamerator 41). Based on the CPs and the overlaps, we have chosen to keep the orpham. We carefully examined CPs of other phages and the gaps between genes in other phages to make this call. However, other phages that keep the forward gene don't have the orpham and have a larger gap that makes the transition to forward strand transcription feasible. No other phage exactly matches BotCity in this region.

- We added 3 genes that filled relatively small gaps, but were supported by CP, were present in other DN phages, and created favorable overlaps. These genes are:
    - Gene 67 (Pecaan): 41,572--41,745
    - Gene 71 (Pecaan): 43,242-43,379
    - Gene 75 (Pecaan): 44,935-45,111
    - The first two create 4bp overlaps with the genes on either side. The third creates a 14bp overlap with the previous gene and a 4bp overlap with the gene following. Other phages that have this gene (based on BLAST hits) have the same overlaps.

- We called the function of Gene 74 (Pecaan) 44,601-44,948 as HNH endonuclease based on the presence of an HNN motif which is described in the forums.

- We called the function for Gene 53 (Pecaan) Cro because we were able to find sufficient evidence for this because of an identified integrase and immunity repressor, and it is in the appropriate syntenic region for this.

- For gene 49 (Pecaan), we ultimately chose to go with IrrE-like protein because we found the specific HEXXH motif that was required along with an HTH pattern. This aligns with the rules and with what is found in discussions in the forums.

- Regarding holin- we have identified what we believe to be the holin at gene 24 (Pecaan) and have found two additional membrane proteins in the area (Pecaan 25 and 27). We have seriously debated calling gene 24 a membrane protein as well, however, there are actual HHPRED hits to holins and most of the DN phages have called gene 24 a holin. We realize that the rules state that in this situation it is typical to call them all membrane proteins, but based on the stronger evidence for holin at gene 24, we kept the call.

- Start for Gene 43 in PECAAN was called because we are the only phage that has this longer start site, it has a slightly better (if not insignificant) RBS score and better accommodates the coding potential for this gene. It does extend the overlap a few bases to a 13 bp overlap rather than an 8 bp overlap, but we felt that it was important to call the longer start because of the coding potential.

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. Yes 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.)?
2. Yes Are all the genes 'Valid" when you click the Validation button?
3. Yes Are the genes (and matching LocusTag numbers) sequential, starting with #1, counting by 1s.
4. Yes Are the Locus Tags the "SEA_PHAGE NAME" format?
5. Yes Has the documentation been recreated from the Feature Table to match the latest file version?
6. Yes 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?
7. Yes Has the frameshift in the tail assembly chaperone been annotated correctly (if applicable)?
8. Yes Have you cleared your Draft_Blast data and have you re-Blasted the submitted DNA Master file?
9. Yes Has every gene been described and supported in your Supporting Data file?
10. Yes Did you investigate 'gaps'?
11. Yes 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!)

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

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

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

Students completed their annotations using DNA Master and documented their thought processes in an online Google-drive based lab notebook. Final starts and functions were put in a Google sheet for a quick overview. While grading annotations, we entered the data into PECAAN, which was used to generate the final files.

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.