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. Survivors
Your Name. Pam Connerly & Beth Rueschhoff
Your Institution. Indiana University Southeast
Your email. pconnerl@ius.edu
Additional emails. (for correspondence). erueschh@ius.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.

1. The start site for gene 66 was controversial. We called start 44109, which gives an overlap of 121 with gene 65, but includes all coding potential. The alternative start we considered at 44226 has an overlap of 4 with gene 65, but misses about 50 bp of coding potential (strong in self, weak in G terrae). The RBS for both are similar, and both good, 44226 slightly better. Including all coding potential was the priority that swayed the majority of annotators. The structure of gene 66 and gene 65 suggest the possibility of a programed translational frameshift. We know that's only called for the tail assembly chaperone, but thought that by calling the upstream start with a large overlap with 65, the annotation for 66 might draw attention of researchers looking for genes like this.
2. The start site call for gene 34 was nearly a toss-up. We went with 26318 based on better BLAST hit alignment and including all coding potential. It does have 2 manual annotations in Starterator, but the RBS numbers are not good. The alternative start considered was 26351, which had more Starterator MAs, and the best RBS, but misses some coding potential and does not align as well with BLAST hits. Again, we favored including coding potential.
3. 54 was added as a new gene after auto-annotation. This small gene aligns with the 5' end of the gene 55 homolog in Cleo. Because the gap had some coding potential, and some BLAST alignment, we added the gene.
4. We called gene 53 as RecA-like DNA recombinase. However, HHpred data shows it has an N-terminal primase domain and a C-terminal helicase domain, so it could also be called a primase-helicase. HHpred alignment with RecA homologs appears to include both domains, so our call is based on thinking that RecA-like DNA recombinase is the more specific call, though literature research left us not 100% certain. The call for gene 38 as a RecA-like DNA recombinase was easier because the HHpred data showed a single domain.
5. Identified a couple of genes with multiple pieces of evidence favoring the called start sites, but did notice another start site option with a 4 bp overlap. Bulk of data favors our calls, but wanted to mention the observation – features 25, 48.
6. Please also note that: (a) No tRNA genes were found during auto-annotation or subsequent searching with Aragorn and tRNA-scanSE. (b) feature start sites that were changed after auto-annotation are highlighted in yellow in the Survivors Annotation Notes spreadsheet.

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, no duplicate genes were present 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)
X 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)
X Spreadsheet
Powerpoint
Word document (must be easily searchable)
Other: Describe.