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. HairyMan

Your Name. Alyssa Gleichsner

Your Institution. SUNY Plattsburgh

Your email. alyssagleichsner@gmail.com

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 some tough start calls with this genome, which contains many orphams and frequently did not have shared starts to other phages even when assigned a pham.
	+ We did occasionally go against the starterator suggested start for some genes as it appeared to not include all of the coding potential and changing the start reduced the gap between genes. This includes the genes with stop coordinates of 28738, 35837, and 36458.
	+ The start call for gene 61 (in Pecaan – stop 40,244) was particularly challenging. We voted to go with the glimmer start @40510 because it has 1:1 blast hits and the highest z score as well as an HHPred hit with good coverage, but the genemark start includes all of the GM coding potential, is the LORF, and reduces the gap so it was a strong consideration.
* There were two genes that have good coding potential, were called by both glimmer and genemark, but do not have significant blast hits to either database. We erred on the side of including these genes, as excluding them would create a large gap in the genome. These genes have stop coordinates of 20,303 and 30,602 bp.
* We identified and called a PTFS for this genome, despite the lack of other EA10 genomes having a documented frameshift. We feel the evidence for the frameshift is strong enough to call it. This includes the genes with stop coordinates at 8754 and 9152 bp.
* Gene 29 (on pecaan – stop @ 21242) is assigned the function of DNA primase/helicase instead of RecA DNA Recombinase. This is because it doesn’t have a clearly identifiable hydrolytic residue and KNK motif so we felt lacked the full evidence outlined via the function list linked slides. It does have the Walker A/Walker B motifs and a partial initial alpha helix (no predicted transmembrane domain) and the ATP-binding loop. We would appreciate a second look to see if someone else can identify the ‘missing’ components.
* Likewise, Gene 30 (stop 22879) was assigned the function nuclease, rather than the more specific functions of VRR-NUC domain protein or Holliday Junction resolvase because it is the most general term and we did not think we had sufficient evidence to be confident making a more specific function call.

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](https://seaphagesbioinformatics.helpdocsonline.com/article-84)?

Yes 3. Are the genes (and matching LocusTag numbers) [sequential](https://seaphagesbioinformatics.helpdocsonline.com/article-77), starting with #1, counting by 1s.

 Yes   4. Are the Locus Tags the “[SEA\_PHAGE NAME](https://seaphagesbioinformatics.helpdocsonline.com/article-77)” format?

 Yes  5. Has the [documentation been recreated](https://seaphagesbioinformatics.helpdocsonline.com/article-86) from the Feature Table to match the latest file version?

N/A 6. Have tRNAs followed the [tRNA protocol](https://seaphagesbioinformatics.helpdocsonline.com/undefined), **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](https://seaphagesbioinformatics.helpdocsonline.com/article-54) been annotated correctly (if applicable)?

Yes – except for the frameshift which was added after. 8. Have you cleared your Draft\_Blast data and have you [re-Blasted](https://seaphagesbioinformatics.helpdocsonline.com/article-57) the submitted DNA Master file?

Yes 9. Has every gene been [described and supported in your Supporting Data file](https://seaphagesbioinformatics.helpdocsonline.com/article-44)?

Yes 10. Did you investigate ‘[gaps](https://seaphagesbioinformatics.helpdocsonline.com/article-31)’?

Yes 11. Did you [delete the genes](https://seaphagesbioinformatics.helpdocsonline.com/article-65) that you meant to delete?

Now, [make a profile of the file](https://seaphagesbioinformatics.helpdocsonline.com/article-64) you plan to send. (And you can save this file for [Review to Improve!)](https://seaphagesbioinformatics.helpdocsonline.com/untitled-18)

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

Powerpoint

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

Powerpoint