

## 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. Pembroke

Your Name. Marly Cassford

Your Institution. Providence College

Your email. mcassfor@friars.providence.edu

Additional emails. (for correspondence). kcornely@providence.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.

One of the greatest concerns in the genome fall within the first few genes. In Pembroke, the initial gene 1 that was called by Glimmer and GeneMark was a reverse gene that has since been deleted. In its place, two genes have been added, however two very similar A3 cluster phages have an additional gene at the beginning of the genome. It was decided that due to the lack of coding potential on the coding potential maps, along with the lack of BLAST data, only two genes would be added to the start of the genome. Gene 3 in Pembroke was assigned the function HNH endonuclease, in contrast to other A3 cluster phages that were assigned the function of minor tail protein. HHPred data had multiple strong hits to HNH endonuclease, providing motive to call this as the function, rather than minor tail protein. Genes 5, 6, and 7 were assigned the function of minor tail protein. There was slight doubt about having 3 minor tail proteins at the beginning of the genome, however someone on the SEA PHAGES forums had posted that many A3 cluster phages have a series of 3 minor tail proteins at the beginning of the genome. Gene 57 had hits to proteins that had been assigned the function phosphofructokinase, however this function was ultimately not called in Pembroke because phosphofructokinase is not a function listed on the official SEA PHAGES function list.

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?
- Yes 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!](#))

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

Yes 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

Yes DNA Master shorthand (previously used format)

Spreadsheet

Powerpoint

Word document (must be easily searchable)

Other: Describe.