

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. **SophKB**

Your Name. Charlotte Berkes

Your Institution. Merrimack College

Your email. berkesc@merrimack.edu

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.

1. Five genes were ADDED due to finding gaps with atypical coding potential: gp96, gp115, gp123, gp126, gp140 (highlighted in yellow in annotation notes spreadsheet)
2. Changed start sites of gp50, 51, 93, 95, 102, 105, 119, 124, 125, 134, 135. (highlighted in blue in annotation notes spreadsheet)
3. I called three HNH endonucleases in this genome: gp96 (59438-59899), gp102 (61980-62270), and gp140 (75004-75351). I added gp96 and gp140. I found the "HNH within a 30 aa span" in all three of these ORFs. I am very confident in calling gp96 as an HNH endonuclease (CDD search finds an "HNHc superfamily" domain). CDD search didn't find the HNHc superfamily domain in either gp102 or gp140, but gp102 contains a NUMOD4 domain (NUMOD4 is a putative DNA-binding motif found in homing endonucleases and related proteins). gp140 has HHPred high probability low e value hits to HNHs. More notes in the 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?

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?

N/A 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!](#))

No 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:

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