

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. RavenCo17

Your Name. Amanda Freise

Your Institution. UCLA

Your email. afreise@ucla.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.

Stop **4705**: Location call a little tricky. See PECAAN notes for more info.

Stop sites **9959**, **10368**: These two genes are the two tail assembly chaperones. Many CZ phages call the usual programmed translational frameshift. However, these genes in RavenCo17 appear more similar to those of Bjanex7 and Faith8x8, both of which do not call a frameshift.

Genes 20, 21, 22: All have 4 TMDs. Gene 19 is lysin A, so one of these is possibly holin. Some genomes call gene 20 (stop **19447**) the holin gene, but more members of that pham do NOT call a function. Gene 20 has some HHpred hits for holin, but they are all very poor (I suspect these hits are what led other people to call this gene as holin). Calling "membrane protein" for now on all three, since it is not clear which one is the holin.

Genes 30-33 (starting with stop site **28459**): 4 orphans in a row! Please review in particular for accuracy.

Stop **30167**: conflicting evidence for which start site to choose. Starterator heavily suggests 30337, but stats are better for 30415 + two other CZ phages called this start site.

Stop **33175**: There was a small (140bp) auto-annotated forward orphan here, which we deleted due to low coding potential on both GM host and self. However, this leaves a large gap.

Stop **35661**: tricky start site call. Pham is diverse, and Starterator only somewhat helpful.

Region between genes 40-44: Lots of big gaps, changes in direction. I reviewed coding potential and did not find any gaps that could be filled, but second look would be appreciated!

Stop **47035**: not sure about functional call. Limited evidence to support HTH DNA binding domain.

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)?

NO 8. Have you [cleared your Draft_Blast](#) data and have you [re-Blasted](#) the submitted DNA Master file?

I was unable to re-BLAST within DNA master and have yet to figure out what the issue is. Debbie said it would be fine for now to note this here and submit.

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

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

DNA Master shorthand (previously used format)

Spreadsheet

Powerpoint

Word document (must be easily searchable)

Other: Describe.