

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. Wilca
Your Name. Jessica Rocheleau
Your Institution. University of Massachusetts, Amherst
Your email. jmrocheleau@umass.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.

Wilca is essentially identical to BirdInFrench (only one nucleotide difference). We had separate lab sections annotate these 2 genomes, and then did a final quality control check to make sure our final start and function calls matched. It probably makes sense to QC these genomes together.

Same note as BirdInFrench: **Gene 59 (41,149-40,379)**: It appears that *DNA recombinase* has been removed from the official function list. So, while BLAST results, HHpred results and synteny do show similarity of this gene with DNA recombinase, DNA repair protein and DNA recombinase respectively, we annotated this gene as no known function. This is because the only type of DNA recombinase in the official function list is RecA-like DNA recombinase. For a gene to have this function, it must have alignment with N-terminal domain preceding the core ATPase domain, a complete ATPase domain and alignment to C-terminal domain and Mg⁺⁺ binding domain. Or, if this gene lacks alignment to the N-terminal and C-terminal domain but still has a complete ATPase binding domain then this gene must be called a ASCE ATPase. But, this gene had neither the N-terminal and C-terminal domains nor a complete ATPase 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?
- 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

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.