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. JanetJ

Your Name. Christa Bancroft

Your Institution. University of Southern California

Your email. cbancrof@usc.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.

**gp21** was explored which regard to its carbohydrate-binding domain alignment in HHPred. We believe this is a minor tail protein because it align closely to other proteins in their carbohydrate-binding domains, even though the overall coverage is small.

Contact Karen Klyczek for help about **gp44** (stop 33687) about the -77 overlap. These were comments, so we kept the -77 start site overlap:

77 bp is a large overlap - not impossible, but not common. The first question I would ask is whether this overlap is necessary to capture all of the coding potential for this ORF. See the GeneMark graph attached. The 32617 start would cut off some coding potential, whereas the 32563 start includes it all. Also, in Starterator, the 32653 start is well conserved in this pham and is the only start shared with Aoka. For these reasons I would choose the longest start even with the overlap. I would not use RBS scores alone to make that call; with an overlap the gene may not need a good RBS to be translated.

I also checked out the upstream gene that is overlapped (stop 32639), since it is an orpham in a region with a few orphams in both directions that completely overlap each other, and maybe it could be deleted? But this forward gene has good coding potential, better than the reverse genes, so I would not delete this gene.

We also checked whether **gp44** (stop 36237) is a HNH endonuclease as in Aoka gp46. However, it doesn’t have a HNH sequence within 30 amino acids as required in the function list, so we kept it as NFK.

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?

yes 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 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, we checked all gaps and in particular evaluated closely the following potential ORFs

Reverse:

26077- 25855 (no BLAST or HHPred matches, no Coding potential)

26057-25800 (no BLAST or HHPred matches, no Coding potential)

25819-25403 (no BLAST or HHPred matches, no Coding potential)

Forward

25271-25489 (no BLAST or HHPred matches, no Coding potential)

25613-25711 (no BLAST or HHPred matches, no Coding potential)

25712-25870 (no BLAST or HHPred matches, no Coding potential)

25891-25986

25996-26055 (no BLAST or HHPred matches, no Coding potential)

25459-25563 (no BLAST or HHPred matches, no Coding potential)

25704-25823 (no BLAST or HHPred matches, no Coding potential)

25983-26183 (no BLAST or HHPred matches, no Coding potential)

25407-25652 (no BLAST or HHPred matches, no Coding potential)

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:

X PECAAN output

X DNA Master shorthand (previously used format)

X Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

X Other: Describe. Google sheet

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

      PECAAN output

X DNA Master shorthand (previously used format)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.