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. Melc17

Your Name. Jillian Nissen

Your Institution. SUNY College at Old Westbury

Your email. nissenj@oldwestbury.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.

Gene 9: Assigned function was commonly represented across pham members but was not listed as an accepted function on PECAAN

Gene 20: Assigned function was commonly represented across pham members but was not listed as an accepted function on PECAAN

Gene 26: We attempted to investigate a possible frameshift mutation in this gene, but the pham number did not match between PECAAN, Phamerator, and PhagesDB. Based on closest related phages we concluded that a frameshift was unlikely.

Gene 36/37: Both genes are from the same pham.

Gene 39: There is support for using the start at 36231, but closest related phages use the start at 36237 so that was selected.

Gene 40: 36605 start was chosen based on Starterator, GeneMark and Glimmer evidence, but the start at 36602 uses a -4 gap so has potential to be the start as well.

Gene 44: Start site is debatable. Three potential starts are identified: 38991: supported by GeneMark, includes all GeneMark coding capacity; 38979: Most MAs, has a -4 gap, chosen by closest related phages, does not include all GM coding capacity; 38937: Selected start in starterator, chosen by Glimmer. Start at 38979 was selected.

Gene 46: Start site was chosen to match the suggested start in GeneMark as the starterator suggested start does not include all coding capacity. The closest related phages use this start as well. GeneMark did not agree with Glimmer and Starterator

Gene 47: Start site was hard to decide as Glimmer and Starterator agreed on 40360, however GeneMark chose 40276. Additionally, 40360 has a better gap, yet 40276 has a better Z-score (although they are both above 2). Ultimately, 40360 was assigned.

Gene 48: Start site was difficult as Glimmer and Starterator agreed on 40384, yet GeneMark chose 40396. Even though 40384 has a worse Z-score, it has a better gap and includes all coding capacity and was decided in the end.

Gene 51: Various enzyme functions in Phages DB for Pham 38387 (exonuclease, DNA helicase, and helicase subunit), although exonuclease was most prevalent which was the reasoning for assigning this function. Based on HHPred results and coverage it is possible that this protein has helicase and nuclease functions.

Gene 58: Function of DNA primase/polymerase also listed among pham members and some strong HHPred results indicate this function

Gene 64: Start site on this gene is debatable. Three potential starts: 56647: Closest related phages use this start, smallest gap of possible choices, most MAs; 56668: Glimmer and Starterator chosen start 56680: GeneMark chosen start, best z-score. 56647 was chosen.

Gene 65: This gene is in a different pham than the corresponding gene in the closest related phage (Weher20\_65), possibly due to start site location

Gene 70: Start site not clear. 59260: Includes all GM coding capacity, higher z-score, but has no MAs; 59677: selected by Glimmer, GeneMark, and Starterator, has more MAs, but does not include all GM coding capacity. Start at 59677 was chosen.

Gene 82: High homology to other phages identifying this gene as HNH endonuclease, but no high probability HHPred evidence

Gene 84: Difficult start site choice. Three potential starts: 64019: chosen by Glimmer and Starterator, slightly higher z-score; 64076: chosen by GeneMark; 64046: -4 gap, highest z-score, most MAs. 64046 was chosen.

Gene 86: Glimmer and Starterator agreed on start 64461, however, GeneMark picked 64458 which has a -4 gap which is why it was assigned. Another note is that both sites did not have total coding capacity.

Gene 10 was added, and two genes (Melc17Draft\_21 and Melc17Draft\_59) were deleted.

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?

N/A 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 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:

Yes 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.