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. ShaiHulud

Your Name. Rivka Glaser

Your Institution. Stevenson University.edu

Your email. rglaser@stevenson.edu

Additional emails. (for correspondence). mparker11@stevenson.edu

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. There were a few start sites that we would like you to review (these are highlighted in yellow on the spreadsheet):
   1. Gp16, for the original start site of 9118, the query to subject alignment are mostly not 1:1, yet there was only MAs for 9188. However, we extended start site to 8966 (as seen in GeneMarkS), and saw 1:1 alignment with other genomes. Please verify.
   2. Gp61, there were alot of MAs for 40037 start site but had 1:17 with genomes. However, the Z and RBS scores were better for start site 39989. Should MAs have higher priority over Z and RBS scores?
   3. Gp63, more MAs for 40623 vs 40572 start site and longer length w/ coding potential at 40623, but 40572 did have better Z and RBS scores. Should MAs have higher priority over Z and RBS scores?
   4. Gp64, start site was moved up to 40987 from 41185 due to coding potential and BLAST alignment (further details below). Please verify.
2. There were a few functional calls that we would like you to review (these are highlighted in yellow on the spreadsheet):
   1. Gp11, 12, 19, 25, 26, 30, 43, and 46 all had high matches, but we were unsure whether the length of the area of alignment was long enough to be considered a good quality match for their designated function.
   2. According to phagesdb, gp15 and gp16 could be tail assembly chaperones, but there is not enough evidence or resources available to us to determine whether this is true.
   3. Gp52 had good evidence on HHPred of being either thymidylate synthase or methyltransferase, but we did not know what the definite function was. There were high quality matches to various phages for each of these functional calls. We ended up calling it thymidylate synthase, based on the frequency at which we saw this function in the hits.
3. Initially, we believed that there was a programmed frameshift between the gp15 and gp16. However, after exploring the forums, we found that other EA phages have frameshifts, however, EA1 phages, including ours, have been found to not have the programmed frameshift.
4. The original DNA master file had a gene at gp41, but its short length (less than 120 bps) and lack of BLAST data to other annotated phages led us to believe that it is not a gene. Therefore, we deleted gp41 from the genome.
5. Upon investigation of the gap between final genes gp63 and gp64, we determined that the start site gp64 should be adjusted to 40987 bps from 41185 bps. This conclusion was drawn due to the presence of coding potential on GeneMarkS and 1:1 alignment of with other annotated phages with which we saw high homology (namely, Vispistious and Winzigespinne).
6. In using numerous programs, we found that there are no tRNA’s in this genome.

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?

N/A 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)

N/A 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)

X 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)

X Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

      Other: Describe.