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.  Tongui

Your Name.  Catalina Dávila Aguer

Your Institution.  Universidad Ana G. Méndez, Gurabo campus

Your email.  davilac2@uagm.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.

1. The starterator report for gene 1 is not informative, since Tongui is the only phage having this start. Starterator reports were also NI for genes 3, 6, 11, 63, and 83.
2. Gene 11 requires further inspection. Its Starterator report is not informative, since the start has only been chosen 28.6% of time when present. Glimmer did not call this gene. There is a large gap (129 bp) between genes 10 and 11. However, no BLAST hits could be found in this section.
3. Legitimate large gaps were found at different locations within the genome (i.e. between genes 10-11, 15-16, 38-39, 42-43, 43-44, 47-48, 63-64, 65-66, 74-75). We researched all these gaps and did not find blast hits.
4. Gene 20 have been called head-to-tail stopper in similar phages within the cluster, such as Hager. Although we found an HHPRED alignment to SPP1, we didn't assign this function since the HHPRED hit was not to the 5A21 chain E or F in the macromolecular complex.
5. We changed the start in genes 21, 52, and 54, in order to improve the alignment and have better Z scores and Final scores. We also took into account the Starterator reports for these genes.
6. We identified a programmed translational frameshift in the two tail assembly chaperone genes (27 and 28). This is a +1 frameshift with the slippery sequence GGGGAAG, which changed GE to GK. Specifically, the four G was the nucleotide skipped. The same is observed in other phages of cluster EF, such as Rie18.
7. Although DNA Master and PECAAN identified gene 34 as Lysin A, we did not add the domain and just call it Lysin, since we only found one gene in the entire genome to call a lysin.
8. We identified more than one transmembrane domain in genes 35, 36, and 65 when Deep TmHHm was run but did not call them membrane proteins. Our decision was based on comparison with all the phages from the cluster. Images form the TmHHm results were included in the spreadsheet named Tongui\_MemProt\_Supporting info.
9. We assigned the function holin to gene 37, based on the TmHHm and HHPred results. We found four transmembrane domains when Deep TmHHm was run and a PFam hit to a holin in HHPred, with a probability of 86.3%. Images form the TmHHm results were included in the spreadsheet named Tongui\_MemProt\_Supporting info.
10. Gene 66 is an Orpham, no Starterator report available. Genemark shows coding potential but it doesn't have BLAST hits.
11. After investigating a large gap after gene 78, we decide to add gene 79, based on results from BLAST. This gene was not called by Glimmer or GeneMark.
12. tRNAs and tmRNAs were not detected in Tongui’s 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?

 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)

 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.