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. Atlantica

Your Name. Nic Vega

Your Institution. Emory University

Your email. nic.vega@emory.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.

SEA\_ATLANTICA\_230 (CDS 17622 – 17945) annotated as holin based on synteny, length, and number and structure of TM domains (2-3 TMDs separated by linker, with short N-terminal domain and longer proline-rich C-terminal domain). SEA\_ATLANTICA\_200 (4 TMDs) appears to be part of the lysis cassette and is a possible additional call for holin, but the structure is somewhat less typical for phage-associated holins, albeit within theoretical limits for these proteins.

SEA\_ATLANTICA\_250 (CDS complement (18206 - 18520)) is annotated as hypothetical; we have requested permission to annotate as “Pleckstrin homology (PH)-like protein”. See notes.

SEA\_ATLANTICA\_280 (CDS complement (19926 - 20165)) is not very convincing? There was enough evidence in favor that the gene was retained, but we are not happy with it.

SEA\_ ATLANTICA\_370 (CDS complement (23907 - 24284)) and SEA\_ ATLANTICA\_380 (CDS 24504 - 24773) are annotated as immunity repressor/Cro respectively. Immunity repressor has been called here several times within cluster AS, but Cro is not a common call. See notes for rationale.

SEA\_ATLANTICA\_460 (CDS 28808 – 29587) called as "RepA-like replication initiator" (e.g. lambda O-like) based on HHPred results and AlphaFold, partly on synteny. See notes for rationale.

SEA\_ATLANTICA\_480 (CDS 30984 - 31448) called as “helicase loader” (e.g. lambda P-like) using similar process as for gp46. Given lambda-like structure, helicase not expected on genome; we do not find a helicase. See notes.

Putative tRNA (not included; originally auto-annotated at 19756- 19821) is non-functional based on folded structure and anticodon-acceptor mis-match, but is homologous to 3’ 38 bases of one copy of *Arthrobacter* tRNA-fMet and to previously annotated *attP* in *Arthrobacter* phage Galaxy (AS1). The region is very highly conserved in nucleotide identity across cluster AS phage and may represent a conserved *attP* site. Hoping to get wet lab data.

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 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

      DNA Master shorthand (previously used format)

x Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

x Other: Describe. Shared OneNote (for full documentation including screenshots of results; summaries in PECAAN)

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)

x Other: Describe. Text file