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. PSonyx

Your Name. Ombeline Rossier

Your Institution. University Paris-Saclay

Your email. ombeline.rossier@universite-paris-saclay.fr

Additional emails. (for correspondence). christophe.regeard@universite-paris-saclay.fr

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.

**Structural annotation**

* Gene 50 (stop 34995) is a small gene that overlaps (33nt) with gene 51. We debated back and forth to shorten it (but it results in a 66 bp gene) or delete altogether (but this creates a 86 bp gap). As Genemark indicates a strong coding potential, we decided to keep it.
* Gaps that were not filled: between gene 57 and 58; and between gene 82 and gene 83

**Functional annotation**

* Gene 37 (stop 30970) was annotated “DnaQ-like (DNA polymerase III subunit)”, based on alignments found on HHPRED. However, other HHPRED hits with even better scores aligned gp37 with known structures of exoribonuclease (for which there is no official name suggested in the “official gene list”). We are wondering whether we should expect to have other subunits of the DNA polymerase encoded in the PSonyx genome to carry on with the decided annotation.
* Gene 77 (stop 43855) was annotated as “HNH endonuclease” as HHPRED aligns with PacI. There is HNH within 30 aa but it is not in the region with homology with PacI. Not sure about decision.
* Gene 83 (stop 46032) was annotated as “hypothetical protein”. BlastP aligned with proteins annotated as “HNH endonuclease” or “endonuclease VII”. We are not inclined to use these annotations because: 1) the residues HNH were in a span wider than 40 aa and 2) HHPRED finds two restriction endonucleases with evalue score not great (prob 93% evalue=0.11) and gp83 is shorter than the hits, so we are not sure that it has all the features to function as a nuclease.
* Gene109 (stop 57961) encodes a big protein and was annotated AAA-ATPase (first hit with BlastP) but with HHPRED we find proteins that use ATP for recombination (RecF or RecN) or for chromosomal partition or maintenance (Smc). In Actinophagesdb, BlastP shows that prior annotations are “all over the place”. No sign of consensus on the functional annotation.
* Gene 117 (stop 62309) was annotated DNA binding protein, but with HHPRED we find alignments with proteins whose structure were resolved and which are RNA polymerase sigma factors. The SEA-PHAGES official function list recommends the term DNA binding protein however so we stick to that.

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 individual genes were re-blasted but not within DNA Master (we get a “could not load SSL library” error message when attempting to use Blast) 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:

      PECAAN output

      DNA Master shorthand (previously used format)

x Spreadsheet

x Powerpoint

x 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)

      Spreadsheet

      Powerpoint

      Word document (must be easily searchable)

x Other: Describe. Zipped folder with Excel spreadsheet and Word document