

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. Schomber
Your Name. Bonnie Madden
Your Institution. George Mason University
Your email. Bmadden1@gmu.edu
Additional emails. (for correspondence). Aschere2@gmu.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.

GP_13

SSC: 5411 – 5695 F: oxidoreductase SIF-BLAST: NKF SIF-HHPred: Oxidoreductase, PDB_mmCIF70_12Aug, 7071_W, Yarrowia liopytica, 92.9% probability, 79.7872% coverage The official function on list is oxidoreductase. I think it codes for a subunit of the pump that drives the terminase.

GP_17

SSC: 7542 – 7811 CP: F: membrane protein SIF-BLAST: Gray_18, NCBI, AZS07873, 100% identity, 100% aligned, 1:1, 4.40468e-55 SIF-Syn: membrane protein same as Chidiebere and Gray. Upstream gene has NKF same as Chidiebere and Gray. Downstream gene is an orpham. **TmHhm: 2**

GP_27

SSC: 16794 – 17450 SIF-BLAST: NKF SIF-HHPred: Phage portal protein, PDB_mmCIF70_12Aug, 6TE9_A, Rhodobacter capsulatus, 98.4% probability, 37.156% coverage Did not call the function. Possible portal protein based on HHPred High probability but low coverage.

GP_40

SSC: 29577 – 30032 SIF-HHPred: Phage_tail_S, Pfam-A_v35, Phage virion morphogenesis family, PF05069.16, 99.8% Probability, 89.404% Coverage. Function is possible minor tail protein. It is in the region of other tail proteins.

GP_55

SSC: 42984 – 44156 F: baseplate wedge protein SIF-HHPred: baseplate wedge protein, Myoviridae, Phage, Vibrio phage XM1, PDB_mmCIF70_31_Jul, 7KH1_F2, 99.3% Probability 53.3333% coverage

GP_59

SSC: 49686 - 50225 F: membrane protein SIF-BLAST: membrane protein, Gray_61, NCBI, AZS07915, 100% Identity, 100% Aligned, 1:1, 6.76891e-125 SIF-HHPred: SMODS and SLOG-associating 2TM effector domain 2, Pfam-A_v35, PF18183.4, 95.1% Probability, 83.7989% Coverage SIF-Syn membrane protein, same as ChisanaKitsune and Gray. **TmHhm: 2**

GP_61

SSC: 50620 - 51114 CP: F: membrane protein SIF-BLAST: membrane protein, Gray_63, NCBI, AZS07917.1, 100%, 9.47181e-114 SIF-Syn membrane protein same as in Gray and Chidiebere, gene is located downstream two genes from a membrane protein. **TmHhm: 4**

GP_62

SSC: 51114 – 51452 SIF-HHPred: Cell division protein FtsB, PDB_mmCIF70_31_Jul, 6H9N_B, Escherichia coli K-12, 95.9% Probability, 89.2857% Coverage - Nothing on the official function list so gene function was not called.

GP_74

SSC: 60492 - 60896 F: helix-turn-helix DNA binding domain, MerR-like SIF-HHPred: Transcriptional regulator, MerR family, PDB_mmCIF70_31_Jul, 3GPV_B, Bacillus thuringiensis serovar konkukian, 97.5% Probability, 69.403% Coverage – protein has HTH conformational folding

GP_91

SSC: 68534 – 68752 F: membrane protein SIF-BLAST: Gray_93, NCBI, AZS07975, 94.444% identity, 98.6111% aligned, 1:1, 5.97148e-35 SIF-Syn: membrane protein, same as in Chidiebere and Gray. Gene is located downstream two genes from a lipoprotein and upstream a gene with NKF same as Chidiebere and Hanem. **TmHmm: 2**

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](#)?
- Yes 3. Are the genes (and matching LocusTag numbers) [sequential](#), starting with #1, counting by 1s.
- Yes 4. Are the Locus Tags the “[SEA PHAGE NAME](#)” format?
- Yes 5. Has the [documentation been recreated](#) from the Feature Table to match the latest file version?
- Yes 6. Have tRNAs followed the [tRNA protocol](#), **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](#) been annotated correctly (if applicable)?
- Yes 8. Have you [cleared your Draft Blast](#) data and have you [re-Blasted](#) the submitted DNA Master file?
- Yes 9. Has every gene been [described and supported in your Supporting Data file](#)?
- Yes 10. Did you investigate ‘[gaps](#)’?
- Yes 11. Did you [delete the genes](#) that you meant to delete?

Now, [make a profile of the file](#) you plan to send. (And you can save this file for [Review to Improve!](#))

- 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
- Yes 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.