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: Halloweekend
Your Name. Marcelo Guerrero
Your Institution. Purdue University
Your email. guerre52@purdue.edu
Additional emails. (for correspondence). klclase@purdue.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.

The bacteriophage *Halloweekend*, a member of Cluster FF, exhibits distinctive genomic characteristics that both reinforce and expand upon the known traits of this phage cluster. Consistent with other FF phages, *Halloweekend* possesses a high GC content (~65%) and contains a predicted tRNA-Arg gene at position 26966, aligning with the cluster-wide trend of at least one tRNA per genome—most commonly tRNA-Arg .

A total of 64 genes were called in *Halloweekend*, which falls within the typical range (61–70 genes) reported for Cluster FF. Moreover, *Halloweekend* contains all 34 phams that are fully conserved across the cluster. These include genes involved in structural assembly (e.g., terminase, portal protein, capsid maturation protease, scaffolding protein, major capsid and tail proteins, tail assembly chaperone, tape measure protein, minor tail protein) and lysogeny-related functions such as tyrosine integrase. Additionally, *Halloweekend* contains key phams noted for variable conservation across the cluster, including genes encoding a ParB-like nuclease domain protein, an HNH endonuclease, and helix-turn-helix DNA-binding proteins—further supporting its integration within the genetic landscape of Cluster FF.

A distinctive feature of *Halloweekend* is the presence of frameshifts in its genome. While frameshift events are not commonly noted in the cluster report, their occurrence here may suggest a unique or underrecognized genomic mechanism or regulatory strategy.

Halloweekend also displays several notable intergenic gaps, particularly between genes 1–2 (88 bp), 33–34 (551 bp), and 35–36 (43 bp). These regions show no coding potential according to Glimmer or GeneMark predictions and yield no significant homology in BLAST searches. The size and lack of identifiable content in these gaps may suggest novel regulatory elements or reflect unique aspects of genome organization not broadly seen in other FF phages.

Finally, the genomic architecture of *Halloweekend* mirrors the transcriptional layout typical of Cluster FF phages: structural genes are located on the left arm and transcribed forward, followed by a central region of reverse-oriented genes, and concluding with forward-transcribed genes on the right arm. This conserved organization—combined with the full set of conserved phams, presence of key variable-function genes, 64-gene count, and characteristic tRNA content—firmly situates *Halloweekend* within Cluster FF, while its atypical intergenic spacing and frameshifts highlight its potential to reveal new insights into phage genome evolution and regulatory architecture.

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:

PECAAN output
Yes DNA Master shorthand (previously used format)
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
Yes DNA Master shorthand (previously used format)
Spreadsheet
Powerpoint
Word document (must be easily searchable)
Other: Describe.