

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. **GoldDust**

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 GoldDust does not exhibit any particularly remarkable genomic features relative to other phages characterized to date. Comparative analysis revealed that the majority of its predicted open reading frames (ORFs) are highly conserved with homologous sequences in other bacteriophage genomes archived in PhagesDB.

Nevertheless, closer examination of the GoldDust genome identified several notable intergenic regions with relatively large gaps between adjacent genes. Specifically, gaps exceeding 100 base pairs but not exceeding 401 base pairs were observed between genes 3–4 (401 bp), 6–7 (129 bp), 16–17 (197 bp), 17–18 (208 bp), 22–23 (79 bp), 30–31 (207 bp), 37–38 (115 bp), 40–41 (123 bp), 58–59 (390 bp), and 70–71 (133 bp).

Despite the extended length of these intergenic regions, no significant coding potential was detected based on Glimmer or GeneMark predictions. Furthermore, BLAST analyses yielded no meaningful alignments corresponding to these genomic regions. These findings suggest that the large gaps are unlikely to harbor cryptic protein-coding genes and may instead represent non-coding regions or structural elements specific to GoldDust's genome organization.

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.