BIOS 252 Science Research

Annotation Template

Panagakis 4/20/21

List most closely related phages to JalFarm20 by BLAST in phagesdb.org:

Annotation for Gene # \_\_\_36 in DNAMaster\_\_\_\_\_

Glimmer predicted nucleotide start 29,586-29,822 (reverse); DNAMaster, phagesdb, Phamerator. BUT this is #37 in PECAAN.

Glimmer predicted nucleotide stop \_\_29,822\_\_\_\_\_

**1. Is the ORF a protein coding gene?**

a. Do the Glimmer and GeneMark gene calls in the auto-annotation agree? \_\_\_\_NO; not called by GeneMark.\_\_\_

b. Does the ORF have high coding potential based on the GeneMark output? Mostly atypical.



Lysin B

28,603-29,601

#35: 28,603-29,601 +3 (Lysin B)

#36: 29,822-29,586 -1 (reverse)

#37: 29,611-29,844 +3 (Holin)

#38: 29,841-30,215 +2



c. Do we see this ORF in phage of the same subcluster using a Phamerator gene map? \_\_\_Just drafts\_\_\_\_ If so, what is the Pham number? 



d. Does a protein BLAST (from either phagesdb.org BLAST, or from NCBI BLAST) produce any matched alignments to proteins with well-known functions? NO

Conclusion: Is the ORF a gene? NO?

Rationale for the decision:

1. GeneMark did not call this as a gene

2. Coding potential is mostly atypical

3. Starterator was not informative; only draft genomes in Pham #6182

4. In DNAMaster: Forward genes 35 and 37; this gene, 36, is a reverse gene that overlaps: 29,822-29,586 (reverse)

5. No synteny evident

 **2.** **If it is a gene, where does it start?**

a. Is this start conserved among all members of the pham in Starterator? \_\_\_No\_\_\_\_

(phagesdb.org/Phage/Phams/enter phamily number/get starterator report) to evaluate gene starts.



b. Does the currently predicted start site include all of the coding potential in the GeneMark output? \_\_\_Yes; BUT atypical\_\_\_\_

c. Did Glimmer and GeneMark agree on the start? NO

d. Is the predicted start codon the longest possible for the ORF without causing excessive overlaps? Check DNA Master DNA/Frames. Yes

e. For genes with functional predictions from BLAST, does the start choice include the full-length protein? \_\_\_N/A\_

f. Does the start site match other starts for similar genes in NCBI BLAST and the phagesdb.org BLAST? \_\_\_Yes, but only drafts

g. Does the predicted start have an associated ribosome binding site (RBS, i.e. Shine-Dalgarno sequence) with a high score or recognizable sequence? \_\_\_\_

Use RBS Sequence Finder in DNA Master.

Conclusion: At which nucleotide does the gene start? \_\_\_\_

Rationale for the decision:

**3. What is the gene's function?**

a. BLASTP: Does this protein align with a protein with a functional assignment in NCBI or phagesdb.org? \_\_\_\_Only small portion-minor tail protein\_

b. HHPred: Does this protein align with a protein with a functional assignment in the PDB? \_\_\_

Use <https://toolkit.tuebingen.mpg.de/#/tools/hhpred> and paste protein sequence into the box.

c. Synteny: Is this gene located adjacent to other genes of known function, in a region of the genome that shows high conservation of gene order? \_\_\_\_\_\_NO

d. Official function from the current SEA-PHAGES FUNCTIONAL ASSIGNMENTS list found at https://seaphagesbioinformatics.helpdocsonline.com/article-96.

Conclusion: What is the gene function? \_

Rationale for the decision:

**4. Annotation Notes.**

This is the short-hand code list in the Features/Notes section of DNA Master for each gene.
SSC: CP: SCS: ST: BLAST-Start: Gap: 40 LO: RBS: F: SIF-BLAST: SIF-HHPred: SIF-Syn

The instructions how to record the notes are found at https://seaphagesbioinformatics.helpdocsonline.com/officialdocumentation