In Silico annotation of the genome of novel bacteriophage Kady

John Sherwood[1], Jasmina Cunmulaj[1], Dr. Joshua Thomson[2], Dr. Jacob D. Kagey[1]
1.Department of Biology, College of Engineering and Science, University of Detroit Mercy
2. University of Detroit Mercy School of Dentistry

Kady is a novel bacteriophage isolated from soil samples as part of the first half of the HHMI SEA--Phages project. The research that is part of this project is targeted to help understand the genetic makeup of the gene sequences within the bacteriophages and add more annotated genomes to the growing databases, as this will help with the growing number of bacterial strains that cannot be controlled with common antibacterial chemicals anymore. The research discussed in this poster was the annotation of the genome of novel bacteriophage Kady, using *In Silico*, or an entirely computer based method. To do this the isolated bacteriophage had its DNA sequenced and then the auto-annotation systems, GLIMMER and GeneMark, were used to predict genes in this novel DNA sequence. These gene predictions were then annotated using *In Silico* methods. In total, three tRNA's and 47 genes were predicted, including two that were not called by auto annotation, as well as not including the two predicted to be false positives. As a result, this annotation will help future researchers better understand bacteriophage genomes, it will be especially helpful when trying to understand bacteriophages that can infect *Mycobacterium Tuberculosis*, as that bacteria is in the same family as *Mycobacterium Smegmatis*, the bacterial strain Kady was isolated with.

Background

- 1st half of HHMI SEA-phages project
  - Kady was a bacteriophage collected from a soil sample
  - Isolated using Mycobacterium smegmatis
  - Sent in for DNA sequencing

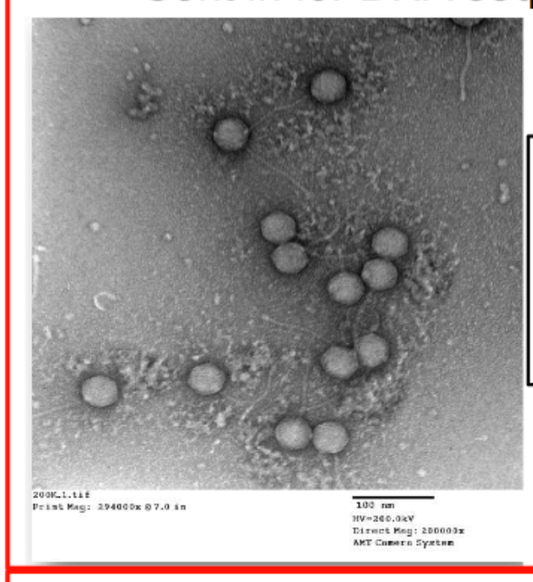
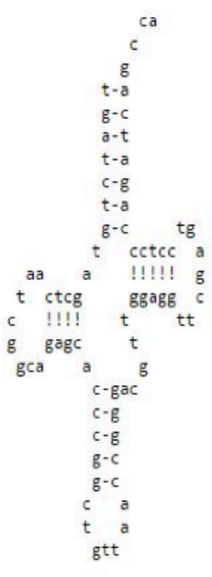


Figure 1. Electron microscope image of bacteriophage KADY, which classifies it into the Siphoviridae family.

tRNA



- In addition to 47 Genes, 3 tRNA were predicted
- Predicted between basepairs 3922-4221
- Predicted using Aragorn and Mobyle Pasteur software

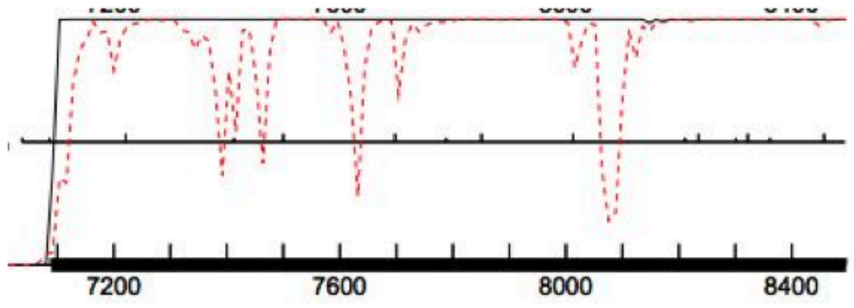
Figure 6. This is a picture of the predicted shape of tRNA 1 as predicted by Aragorn software.

Final Gene Predictions

- In total, 47 genes predicted from base pairs 1-12,871 and 39,630-50,900
  - 2 deleted from original auto-annotation (ORF 2, ORF 89)
  - 2 called by annotator, not found by auto-annotation (ORF 2.5, ORF 60.5)

In Silico Gene Annotation

Figure 2. This figure shows the coding potential as predicted by GLIMMER.



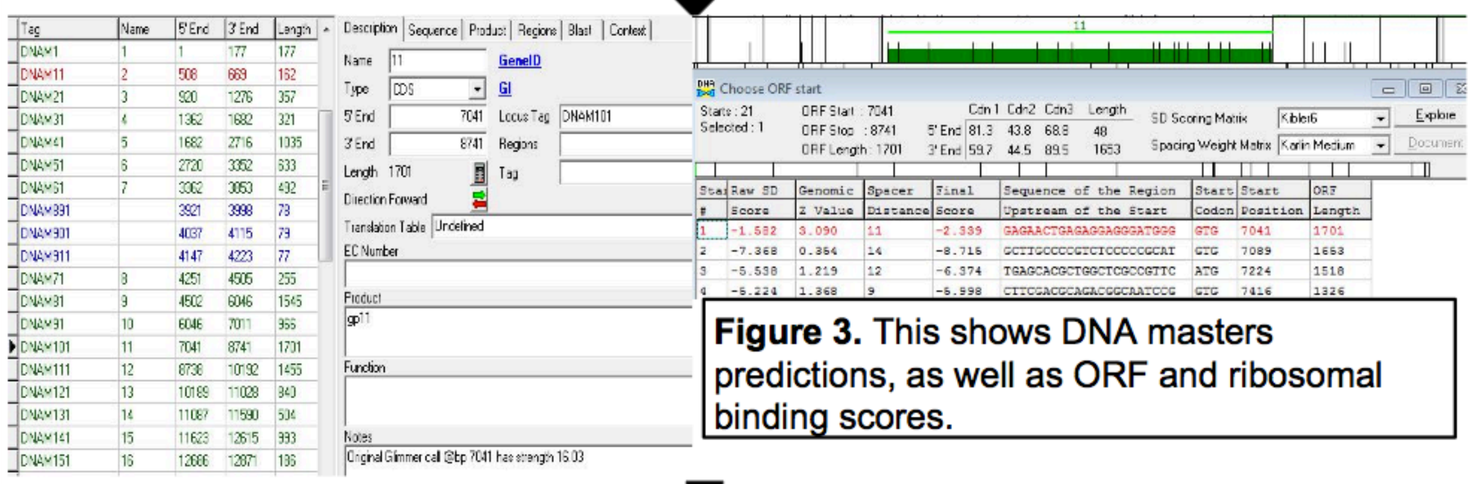


Figure 3. This shows DNA masters predictions, as well as ORF and ribosomal binding scores.

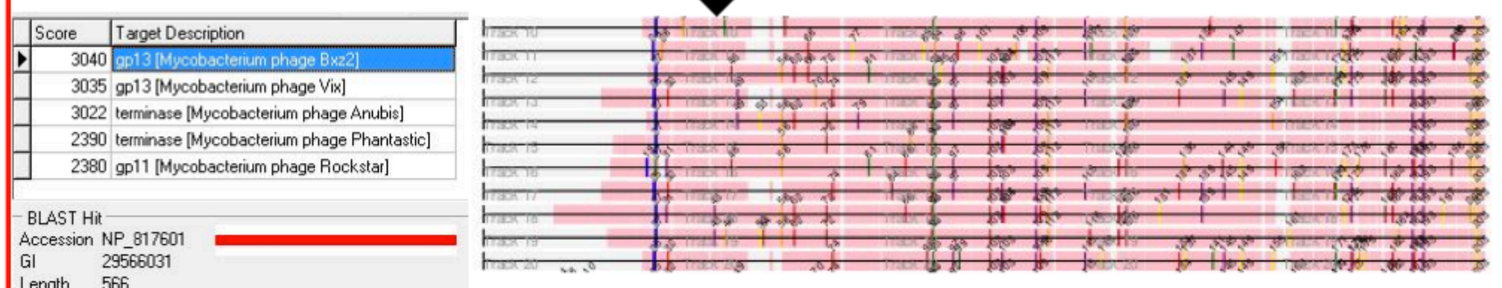


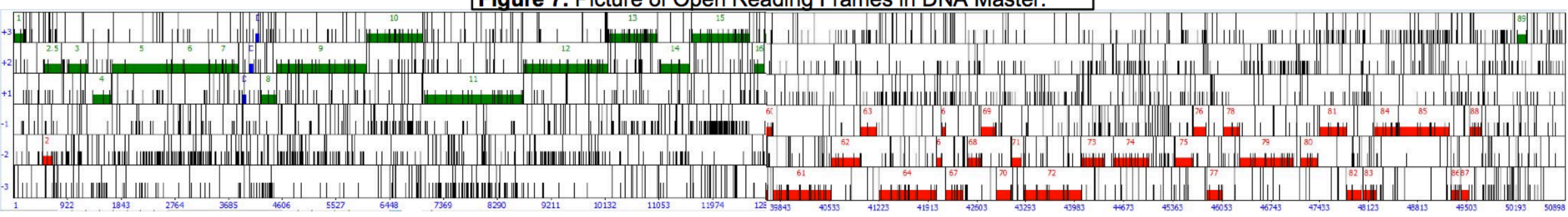
Figure 4. This is an image of a DNA Master BLAST (left), and a Starterator image, (right).

Figure 5. This is an example of all data compiled and considered in final gene predictions.

Gene Prediction 11

SSC: Start: 7041 Stop: 8741
CP: Yes, all coding potential is contained
SD: Best possible score
Final Score: -2.339
Z-Score: 3.090
SCS: Both GeneMark and GLIMMER call 7041
LO: Yes, longest open reading frame
Gap: 30 base pair gap
BLAST: 1:1 alignment with gp13 of Mycobacterium phage Bxz2
ST: Starterator agrees and calls for 7041

Figure 7. Picture of Open Reading Frames in DNA Master.



Acknowledgments

This work was supported by NIH Rebuild grant and the HHMI SEA Phage program. Thank you to Stephanie Conant, for guidance in the isolation process, and Wayne State University for the Electron Microscopic Image.

Howard Hughes Medical Institute

University of Detroit Mercy