Annotation of Mycobacteriophage JewelBug Genome

Barbara Ann McAnulty, Sarah McGinness, Kathryn Atherton, Madalyn Alm

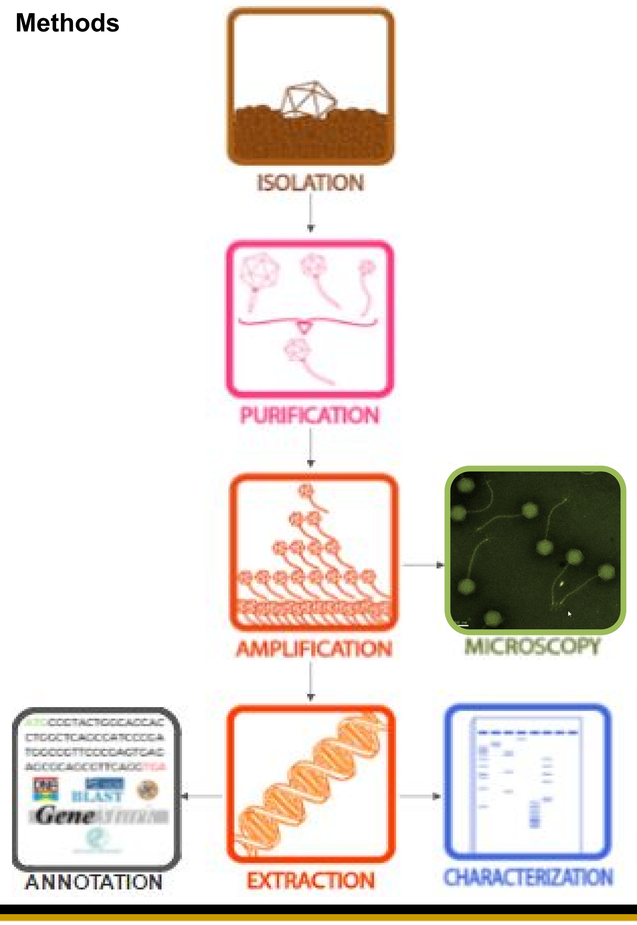
PURDUE UNIVERSITY

Agricultural & Biological ENGINEERING

Introduction

- This research aims to produce an annotated genome of JewelBug, a newly discovered species of mycobacteriophage, for the SEA-PHAGES project.
- Mycobacteriophage are virions that infect mycobacteria in order to replicate.
- The SEA-PHAGES project is an international effort of students characterizing mycobacteriophage found in the environment to discover new genes and their functions.
- In the Fall 2016 semester, various mycobacteriophage were isolated at Purdue using wet lab techniques.
- In the Spring 2017 semester, software and databases aided in annotating a JewelBug's genome.
- This annotation will be validated through wet lab testing to determine the type and function of the projected proteins.

Methods



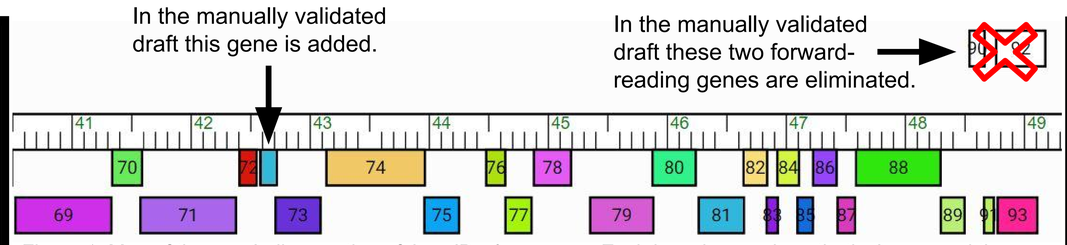


Figure 1: Map of the concluding section of JewelBug's genome. Each box shows a hypothetical gene and the measuring tape marks the gene's position in thousands of base pairs. Two major adjustments to the auto-annotated genome are shown, the addition of a gene following gene 72 and preceding gene 73 and the deletion of genes 90 and 92.

Results

In the final section of the genome (40900-50341 bp) there were originally 26 protein-encoding; but, after manual inspection, there are a total of 25 genes in this section because one gene was added and two genes were deleted.

The total length of the genome is 50,341 bp with 61.6% GC content and a 10 bp overhang sequence of CGGTCGGTTA.

The comparison of the auto-annotation of JewelBug to the manually-validated genome show a better adherence to the guiding principles of viral genome annotation.

- The validated genome has on average 5.7% longer open reading frames.
- The validated genome has 5.0% greater similarity, 4.8% greater alignment, and 4.7% greater identity percentage comparisons to proteins in the BLAST database.
- The average Shine Dalgarno scores, a measure of the ability of ribosomes to initiate protein synthesis, became 2.5% less negative, which aligns with the guiding principles.

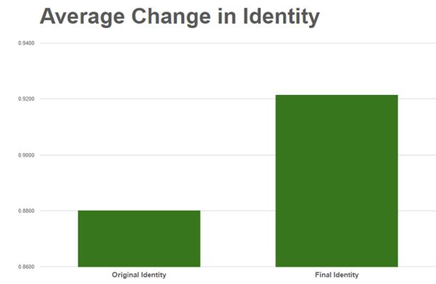


Figure 2: Bar graph representing average identity before manual annotation and identity after manual annotation. Identity refers to the extent to which two sequences have the same residues at the same positions.

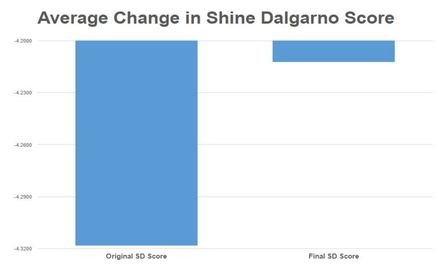


Figure 3: Bar graph representing average shine dalgarno (SD) score before and after manual annotation. The SD score evaluates the potential for different start codons. The smaller the score, the higher the potential.

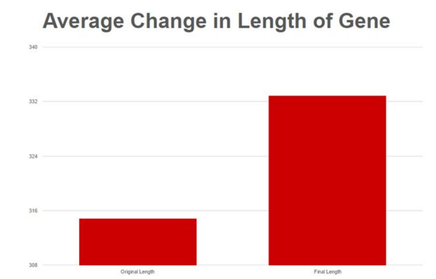


Figure 4: Bar graph representing the average gene length in base pairs before and after manual annotation. Lengthening genes increases gene density. Phages typically have high gene density.

Relevance

- Viruses are the most genetically diverse group of biological material as well as the least studied. Therefore, they contain the largest number of novel genes.
- Expanding the databases of viral genomes allows for more efficient and higher-quality future annotations.
- The databases provide the groundwork for research into the evolution and adaptation of viruses which also provides information about evolution, adaptation, and susceptibility of bacteria. Two paramount concerns for human medicine.
- Mycobacteriophages have specific applications in human medicine since their hosts include *M. tuberculosis*.

Conclusion

- The auto-annotation by DNAMaster utilizing GeneMark and Glimmer gene call algorithms required several revisions including two gene deletions, one gene addition, and many start adjustments on its 40900 - 50341 base pair section
- The final draft of JewelBug's genome represents the most plausible version given the information in current databases.
- As can be seen in the image to the right, there are turbid edges or halos around the plaques formed by JewelBug. This suggests that the virus follows a lysogenic life cycle. The genes found in this section of the genome are likely instrumental in the regulation of the JewelBug's life cycle as well as determining its host range.

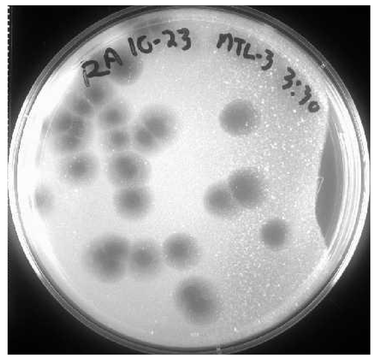


Figure 5: Plaques formed by Jewelbug on a lawn of *M. smegatis* bacteria

Future Work

- Send the completed annotation to University of Pittsburgh Howard Hughes Medical Institute quality control for wet lab verification of the gene calls
- Determine the function of the proteins that currently do not have a known function
- Determine the host range of JewelBug
- Compare to the rest of the A6 cluster to determine possible evolutionary path

Citations:

[Electron microscopy image of siphoviridae bacteriophage]. (n.d.). Retrieved April 5, 2017, from http://bacteriophages.igmors.u-psud.fr/media/images/FamilyPictures/Siphovirus.png

JewelBug PhagesDB Entry. (2012). Retrieved April 5, 2017, from http://phagesdb.org/phages/JewelBug/

Jacobs-Sera, D., Bowman, C. A., Pope, W. H., Russell, D. A., Cresawn, S. G., & Hatfull, G. F. (2015). Annotation and Bioinformatic Analysis of Bacteriophage Genomes A User Guide to DNA Master. Retrieved April 5, 2017, from http://phagesdb.org/media/docs/Annotation_Guide_12.01.2015.pdf

Poxleitner, M., Pope, W., Jacobs-Sera, D., Sivanathan, V., & Hatfull, G. (2016). Phage Discovery Guide. Retrieved April 5, 2017, from http://www.seaphages.org

Instructors:

Dr. Kari L. Clase[1, 2, 3, 4], Professor

Yi Li[1, 2, 3], Ph.D. candidate

[1]Agricultural and Biological Engineering, [2]Biotechnology Innovation and Regulatory Science Center, [3]Bindley Bioscience Center, [4]Technology Leadership and Innovation, Purdue University

Acknowledgements:

Howard Hughes Medical Institute, Science Education Alliance
Polytechnic Institute, Purdue University
Department of Agricultural and Biological Engineering, Purdue University
Department of Biological Sciences, Purdue University
Discovery Park, Purdue University
Genomics Core Facility, Purdue University

PURDUE AGRICULTURE
PURDUE UNIVERSITY

PURDUE ENGINEERING
Think impact.

Purdue University is an equal opportunity/equal access institution.