Comparative Genomic Analysis of Mycobacteriophage JewelBug:   
40900 - 50341 base pairs

KATHRYN ATHERTON

*IT 22700, Purdue University, West Lafayette, Indiana, United States of America*

*Howard Hughes Medical Institute Science Education Alliance – Phage Hunters Advancing Genomics and Evolutionary Science Program*

3 May 2017

**JewelBug is a phage of Mycobacterium smegmatis that was isolated in 2012 by Robert Agee at Purdue University. It has a siphoviridae morphology and forms plaques with halos of turbid edges, suggesting that it is a lysogenic phage. This section of the genome contained 23 genes in the final annotation; two genes were deleted from the original auto-annotation and one was added. All features were read in the backward direction on the complementary strand and had little to no evidence supporting functions for most of the genes; wet lab testing is necessary to verify the final annotation and find functions for those genes which have no known functions.**

**INTRODUCTION**

Mycobacteriophage are the group of bacteriophage that infect mycobacterial cells to utilize the bacteria’s reproductive mechanisms. Phages have the largest set of unexplored DNA sequences in nature.5 Determining the function of these genes could give scientists insight into medical problems posed by such bacterial ailments as Tuberculosis, as the bacterial cell causing the disease is related to the mycobacteriophage’s host, *M. smegmatis*.8 By understanding the genetics of bacteriophages, the scientific community can work towards creating medical applications of phages, such as bacterial killing without affecting non-target bacteria or eukaryotic cells and horizontal gene transfer.1

This paper discusses the characterization of a portion of JewelBug’s genome. Once sequenced, mycobacteriophage genomes are characterized by sorting the species into clusters according to nucleotide sequence similarities.4 JewelBug is a part of Cluster A, Subcluster A6, a subcluster of low genetic diversity.9 The JewelBug genome is 50341 base pairs in length, contains 87 protein-coding genes, and shows similarity with mycobacteriophages VohminGhazi, Isiphiwo, McFly, Kazan, CloudWang3, and Artemis2UCLA. The bacteriophage takes on a siphoviridae morphology, has plaques with turbid edges, and was found at GPS Coordinates 40.42208 N, 86.917246 W in a mildly wet soil environment about 24 centimeters below the surface.3

**METHODS**

**Phage Isolation.** Soil samples were collected and phage were extracted from the soil using an enrichment broth and filter. *Mycobacterium smegmatis* mc2155 cells were infected with the phage and incubated for two days. A plaque created by the lysis of the infected bacterial cells was picked and the phages were mixed with phage buffer. The concentration of phage in the mixture was tested by dropping small samples of serial dilutions of the phage on a bacterial lawn (spot titer) and infecting entire bacterial lawns with the serial dilutions (full plate titer). Lysate from the plates were collected with phage buffer solution and the concentration of the new solution was again tested with spot titers and full plate titers. To ensure that the phage species was completely isolated, streak plates were created by picking a plaque from an infected bacterial lawn and scratching the micropipettor tip across an agar plate, then pouring top agar with bacterial host onto the plate. The plate was inspected after allowing the bacteria to incubate for two days. Webbed plates were created using the titer calculation and new lysate was collected from the webbed plate. The lysate was used for archiving the phage sample, viewing the phage with Transmission Electron Microscopy, and extracting phage DNA. The DNA was sent for genome sequencing and used for restriction enzyme digest characterization. All of the information collected was entered into the Actinobacteriophage Database on the phage’s page.10

**Genome Sequencing.** The DNA is broken up into fragments of 200 to 600 base pairs. Short sequences of DNA are attached to the DNA fragments and the DNA is made single-stranded by incubation with sodium hydroxide. The DNA fragments are washed across a flowcell and the DNA binds to primers on the surface of the flowcell. DNA that does not is washed away. The attached DNA is replicated until the molecules emit a signal that can be detected by a camera. DNA polymerase lengthens the strands of DNA attached to the flowcell between the primers on the flowcell surface. The double-stranded DNA is heated so it breaks down into single strands. Primers and fluorescently labeled terminators are added to the flowcell and attach to the DNA being sequenced. The DNA polymerase binds to the primer and adds the terminators. Lasers pass over the flowcell to activate the fluorescent labels and the colors are detected by a camera and recorded. The terminators have different colors for each base (A, C, G, and T). The terminators are removed and the next base is added to the end of the DNA strand before the terminators are added again. The process continues until the short clusters are sequenced. The sequence is analyzed base-by-base and then aligned to a reference sequence. to create the full sequence.11

**DNAMaster.** This software is used to auto-annotate a genome using a variety of databases as well as user input. The FASTA file, which represents the nucleotide sequence of the genome, is input and compared against two databases, Glimmer and GeneMark, to determine potential genes in the genome. The user then goes through DNAMaster’s auto-annotation and makes edits (moving start sites, adding and deleting features) based on evidence from other pieces of software and databases described below. A user can view the Open Reading Frames of the genome and, when a gene is selected, see the various start site positions a gene might have as well as statistics about the sites, such as the Shine Dalgarno score, start codon, and length of the gene, in order to determine which site is most likely that which is expressed. The user can also see BLAST results of a gene, genes from other species of phage that are closest to that of the target gene, to decide whether there is evidence supporting the existence of a gene at this start site.6

**GeneMark.** This program is a gene prediction software which creates graphic representations of coding potential in the six reading frames of the DNA throughout the genome. A user analyzes the section of DNA in which a gene is believed to exist and determines if there is coding potential in that location and if the gene covers as much coding potential as possible.6

**Glimmer.** This database is a system for finding genes in microbial DNA. It uses Interpolated Markov Models to identify coding regions in the DNA. DNAMaster references this database as it makes its auto-annotation.2

**Phamerator.** This software compares phage genomes, their genes, and gene products and displays the results in a linear fashion, using colors to represent genes within the same phamily and similarities between genomes. It also shows annotated functions on genomes where applicable. Users reference this software to find similarities in other genomes within the same cluster and uses these similarities as evidence to support a gene’s presence.6

**Guiding Principles.** This set of rules for annotating a genome is used in order to determine what is typically feasible for a bacteriophage genome. It references gene overlaps and gaps, frequencies of start sites, and gene orientation switches, among other things. Researchers will reference this when making decisions about genes during the annotation process in order to ensure that the annotation is as accurate as possible.6

**NCBI BLAST.** This database compares the nucleotide or amino acid sequence of a specific gene and finds genes that are most similar. A user determines whether there is evidence supporting the existence of a gene with the current call by analyzing the quality and similarity of the BLAST results. The results can also determine function, if the top result is a close enough match and has other evidence verifying the function.6

**HHpred.** This database is a more sensitive search for gene function. The program returns a list of genes that are highly similar to the target gene and their functions. If the matching gene is highly probable and has a low E value, the user may claim a function for a gene.6

**PECAAN.** This program is used as a quality control check for an annotation. The program shows all of the evidence from the previously mentioned databases and softwares in one place so that the user can check that all evidence is valid and show which parts are being used to verify each decision made about the annotation.7

**GenBank.** GenBank is the database of annotated genomes. A user submits an annotated genome to the database once completed for an expert to review. If accepted by the database, the final annotation will be used as evidence for future annotations.6

**RESULTS**

**Discovery and Isolation**

JewelBug was discovered about 24 centimeters below the surface in the soil of a construction site on Purdue University’s campus between Hansen Hall and Lilly Hall at 2:00 pm on August 24, 2012 by Robert Agee. The soil in which it was discovered was mildly wet and clumpy. The phage creates cloudy plaques which are approximately 0.5 to 1 centimeter in diameter (Fig. 1). Transmission Electron Microscope imaging determined that the phage has the siphoviridae morphotype. Agee named the phage after his mother’s nickname.3



**Fig. 1:** JewelBug plaque image. The diameter of the plaques are approximately 0.5 to 1 centimeter.

**Sequencing**

The genome of JewelBug was sequenced by the Purdue Genomics Facility using Illumina sequencing. The sequencing finished on June 13, 2013 and found that the genome has a length of 50341 base pairs. The GC content is 61.6%. After sequencing, it was determined that JewelBug is part of cluster subcluster A6.3

**Auto-annotation in DNAMaster**

In the final section of the genome, 24 genes were found in the DNAMaster auto-annotation, features 67 - 90. Two of the genes read forward and the other 22 read in reverse on the complementary strand.

**Feature 67 Annotation**

Feature 67 was called by both GeneMark and Glimmer at 41591 base pairs. It has a length of 249 base pairs, starting with the codon ATG and ending with TAA. This call is not the longest open reading frame, though all of the coding potential is covered, according to the GeneMark map. The raw Shine Dalgarno score is -2.611, which is not the highest possible. At this call, there is a 13 base pair overlap between this feature and the next. No change was made to this auto-annotation call due to the agreement of the programs and the fact that this is the longest open reading frame that does not overlap more than 30 base pairs with feature 68. The top BLASTp result is gp69 from Mycobacterium phage Gladiator, which aligns its first base pair with JewelBug’s first base pair for this feature. According to HHpred, the function is a helix-turn-helix domain with a 90.98% probability.

**Feature 68 Annotation**

Feature 68 was called by both GeneMark and Glimmer at 42381 base pairs, though it was decided that the gene could be extended to 42414 base pairs in order to close the gap between this feature and the next. The new call has a length of 837 base pairs, starting with the codon GTG and ending with TAA. This is the longest open reading frame for this feature and all of the coding potential is covered according to the GeneMark map. The raw Shine Dalgarno score is -2.903, which is not the highest possible. At this call, there is a 3 base pair overlap between this feature and the next. The top BLASTp result is gp70 from Mycobacterium phage Gladiator, which aligns its first base pair with JewelBug’s first for this feature. According to HHpred, the function is a phage antirepressor with a 99.80% probability. Phamerator and BLASTp also verified this function.

**Feature 69 Annotation**

Feature 69 was called by Glimmer at 42554 base pairs, but was not called by GeneMark. It was decided that the gene could be extended to 42578 base pairs. This call has a length of 168 base pairs, starting with TTG and ending with TGA. It is the longest open reading frame, and all coding potential is covered, according to the GeneMark map. The raw Shine Dalgarno score is -3.286, which is not the highest possible. At this call, there is a 160 base pair gap following the feature. The top BLASTp result is hypothetical protein PBI\_KAZAN\_75 from Mycobacterium phage Kazan, with its 597th base pair aligning with JewelBug’s first for this feature. There was not enough evidence to declare a function for this feature.

**Feature 70 Annotation**

Feature 70 was called by both GeneMark and Glimmer at 43091 base pairs, though it was decided that the feature should be extended to 43133 base pairs in order to improve the Shine Dalgarno score, increase BLASTp score and alignment, and decrease the gap between this feature and the next. This call has a length of 420 base pairs, which is not the longest possible open reading frame, though all coding potential is covered by this call, according to the GeneMark map. There is an 8 base pair gap following the gene. The raw Shine Dalgarno Score is -3.788, which is the highest possible. The top BLASTp result is gp73 from Mycobacterium phage DaVinci, whose 15th base pair aligns with JewelBug’s first for this feature. According to HHpred, the function of this feature is a resuscitation-promoting factor RfpA, with 89.47% probability.

**Feature 71 Annotation**

Both Glimmer and GeneMark called this feature at 43968 base pairs, and it was decided not to change this call from the auto-annotation as it is the longest open reading frame, has the highest Shine Dalgarno score, and has high quality BLAST results. The call has a length of 828 base pairs, which covers all coding potential, starting with the codon GTG and ending with TAG. The raw Shine Dalgarno score is -2.736. There is a 3 base pair overlap with the next feature. The top BLASTp result is exonuclease/helicase from Mycobacterium phage Isiphiwo, whose first base pair aligns with JewelBug’s first for this feature. According to HHpred, this feature is from the nuclease superfamily, with 99.95% probability, which is validated by Phamerator and BLASTp as well.

**Feature 72 Annotation**

Glimmer called this feature at 44255 base pairs while GeneMark called it at 44351 base pairs. It was decided that this feature should be extended to 44366 base pairs in order to increase the identity and similarity of the BLASTp results, the length of the feature, and the Shine Dalgarno score. This call is 402 base pairs long, the longest possible, starting with the codon TTG and ending with TGA; it covers all coding potential, according to the GeneMark map. The raw Shine Dalgarno score is -2.100, the highest possible. There is a 33 base pair overlap with the next feature; although this overlap is larger than what is usually feasible according to the Guiding Principles, there is enough data to corroborate this decision. The top BLASTp result was hypothetical protein PBI\_MCFLY\_78 from Mycobacterium phage McFly, whose first amino acid aligns with JewelBug’s first for this feature. There is not enough evidence to declare a function for this feature.

**Feature 72.5 Annotation**

Neither Glimmer nor GeneMark called this feature, but it was decided that it should be added in order to close the gap between features 72 and 73. The feature is called at 44488 base pairs and is 156 base pairs long, which covers all of the coding potential, though it is not the longest open reading frame. It starts with the codon GTG and ends with TAG. Its raw Shine Dalgarno score is -4.543 base pairs, which is not the highest. There is a 3 base pair overlap with the next feature. The top BLASTp result is hypothetical protein PBI\_VOHMINGHAZI\_85 from Mycobacterium phage VohminGhazi, whose first amino acid aligns with JewelBug’s first amino acid for this feature. There is not enough evidence to declare a function for this feature.

**Feature 73 Annotation**

Both Glimmer and GeneMark called this feature at 44643 base pairs and no change was made to this annotation. This call is 159 base pairs long, which covers the entirety of the coding potential, though it is not the longest open reading frame possible; the feature starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is -3.178, which is the highest possible. There is a 3 base pair gap following this gene. The top BLASTp result is hypothetical protein PBI\_CLOUDWANG3\_88, from Mycobacterium phage CloudWang3, whose first amino acid aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 74 Annotation**

Both Glimmer and GeneMark called this feature at 44864 base pairs, though it was decided that the feature could be extended to 44888 base pairs in order to raise the SD score, increase alignment, and capture more coding potential. The new call has a length of 243 base pairs, which is not the longest open reading frame, though it covers all of the coding potential according to the GeneMark map. The feature starts with the codon TTG and ends with TAG. The raw Shine Dalgarno score is -2.994, which is not the highest possible. There is a 3 base pair overlap between this and the next feature. The top BLASTp result is gp90 from Mycobacterium phage Blue7, whose 27th base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 75 Annotation**

Both Glimmer and GeneMark called this feature at 45196 base pairs, though it was determined that the feature could be extended to 45535 base pairs in order to increase the length, alignment, identity, and similarity. This extension created a 178 base pair overlap with the next feature, which is typically not feasible according to the Guiding Principles, but had evidence in Phamerator showing similarly long overlaps in related phages supporting this decision. The call is 651 base pairs long, the longest possible call, and covers all coding potential according to the GeneMark map; the feature starts with codon GTG and ends with TGA. The raw Shine Dalgarno score is -2.590, which is not the highest possible. The top BLASTp result is a DNA Methylase from Mycobacterium phage Isiphiwo, whose first base pair aligns with JewelBug’s first for the feature. According to the BLAST results and Phamerator, this feature is a DNA Methylase.

**Feature 76 Annotation**

Both Glimmer and GeneMark called this feature at 45887 base pairs, though it was decided that the feature should be shortened to 45884 base pairs in order to increase the alignment of the feature. The current call has a length of 528 base pairs, which is not the longest possible open reading frame, though it does cover all coding potential, according to the GeneMark map. The feature starts with the codon GTG and ends with TAG. The raw Shine Dalgarno score is -3.372, which is not the highest possible. The feature has a 3 base pair overlap with the next feature. The top BLASTp result is DNA Methylase from Mycobacterium phage VohminGhazi whose 18th base pair aligns with JewelBug’s first for the feature. According to HHpred, this feature is a DNA Methylase, with a 100% probability. BLASTp and Phamerator also corroborate this conclusion.

**Feature 77 Annotation**

Both Glimmer and GeneMark called this feature at 46243 base pairs, and it was decided that this should not be changed as this call is the longest open reading frame and has the highest Shine Dalgarno score. The feature is 363 base pairs long, covers all coding potential according to the GeneMark map, and starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is -2.367. There is a 26 base pair gap following this gene. The top BLASTp result is gp88 from Mycobacterium phage EricB, whose first base pair aligns with JewelBug’s first for the feature. According to HHpred, the feature is a helix-turn-helix DNA binding protein, with 99% probability. Phamerator and BLASTp also verified this conclusion.

**Feature 78 Annotation**

Both Glimmer and GeneMark called this feature at 46649 base pairs, and it was decided that this decision should hold as it had the second highest Shine Dalgarno score of -2.165 and was the longest open reading frame that does not cause an overly-large overlap with the next feature. The feature is 483 base pairs long, covers all coding potential according to the GeneMark map, and starts with the codon GTG and ends with TGA. There is a 3 base pair overlap with the next feature. The top BLASTp result is hypothetical protein PBI\_MCFLY\_91, from Mycobacterium phage McFly, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 79 Annotation**

Both Glimmer and GeneMark call this feature at 46843 base pairs and, as the BLAST results show 100% similarity, identity, and alignment, the call was not changed. The feature is 198 base pairs long, which is not the longest open reading frame, and captures the majority of the coding potential according to the GeneMark map. The feature starts with the codon GTG and ends with TGA. Its raw Shine Dalgarno score is -1.907, which is the highest possible. There is a 3 base pair overlap with the next feature. The top BLASTp result is hypothetical protein PBI\_VOHMINGHAZI\_92 from Mycobacterium phage VohminGhazi, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 80 Annotation**

Both Glimmer and GeneMark call this feature at 46932 base pairs, and this call was not changed as the Shine Dalgarno score is the highest at -3.042 and any longer open reading frame would have caused an overly large overlap with the next feature. The call for the feature is 93 base pairs, which covers the majority of the coding potential, and starts with the codon GTG and ends with TGA. There is a 3 base pair overlap with the next feature. The top BLASTp result is hypothetical protein PBI\_VOHMINGHAZI\_93 from Mycobacterium phage VohminGhazi, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 81 Annotation**

Both Glimmer and GeneMark call this feature at 47111 base pairs. All evidence supports this call, so the start site was not changed. The feature is 183 base pairs long, which is not the longest possible, though it covers all of the coding potential according to the GeneMark map; the feature starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is -2.814, which is the highest possible. There is a 12 base pair overlap with the next feature. The top BLASTp result is hypothetical protein PBI\_VOHMINGHAZI\_94 from Mycobacterium phage VohminGhazi, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 82 Annotation**

Glimmer calls this feature at 47227 base pairs while GeneMark calls the feature at 47233 base pairs. The BLAST results indicated that the Glimmer call is the better call for this feature. The call is 129 base pairs long, which is not the longest possible, though it covers nearly all of the coding potential according to the GeneMark map, and starts with the codon GTG and ends with TGA. The raw Shine Dalgarno score is -1.259, which is the highest possible score. There is a 3 base pair gap following the gene. The top BLASTp result is hypothetical protein PBI\_VOHMINGHAZI\_95 from Mycobacterium phage VohminGhazi, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 83 Annotation**

Both Glimmer and GeneMark call this feature at 47427 base pairs, and due to the alignment data from BLASTp and similarities with other phages seen in Phamerator, it was decided that the call should not be changed. The feature is 198 base pairs long, which is not the longest open reading frame possible, though it does cover all of the coding potential according to the GeneMark map; it starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is -3.880, which is not the highest possible. There is a 9 base pair gap following the feature. The top BLASTp result is hypothetical protein PBI\_MULCIBER\_94 from Mycobacterium phage Mulciber, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for the feature.

**Feature 84 Annotation**

Both Glimmer and GeneMark called this feature at 47582 base pairs. The call was not changed as the coding potential in the GeneMark map supports this gene call, despite there being no BLASTp results for the call. The feature is 147 base pairs long, which is not the longest open reading frame possible and starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is not the highest at -3.880. There is a 9 base pair gap following the feature. There is not enough evidence to declare a function for the feature.

**Feature 85 Annotation**

Both Glimmer and GeneMark call this feature at 48301 base pairs. As this is the longest open reading frame and eliminates a large gap following the gene, the call was not changed. The feature is 711 base pairs long, and starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is not the highest at -4.297. There is a 3 base pair gap following the gene. The top BLASTp result is hypothetical protein PBI\_CLOUDWANG3\_101 from Mycobacterium phage CloudWang3, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 86 Annotation**

Both Glimmer and GeneMark call this feature at 48501 base pairs. As the BLAST data best supports the original call, it was not changed. The feature is 198 base pairs in length, which is not the longest open reading frame possible, but does capture all of the coding potential according to the GeneMark map. It starts with the codon ATG and ends with TAG. The raw Shine Dalgarno score is not the highest at -4.189. There is a 166 base pair gap following the gene. The top BLASTp result is hypothetical protein PBI\_ARTEMIS2UCLA\_101 from Mycobacterium phage Artemis2UCLA, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this gene.

**Feature 87 Annotation**

Only Glimmer called this forward-reading feature at 48545 base pairs; GeneMark did not call a feature in this area. The feature was ultimately deleted as there is no coding potential in the GeneMark map, no BLAST results, and not a large enough gap to justify changing reading direction.

**Feature 88 Annotation**

Both Glimmer and GeneMark call this feature at 48747 base pairs. This call was not changed as there is coding potential in GeneMark and similarity between other phages in the Phamerator Map that supports the gene’s existence. The feature is 81 base pairs long, which is the longest open reading frame possible, and covers all coding potential in the GeneMark map. The feature starts with the codon GTG and ends with TAG. The raw Shine Dalgarno score is the highest possible at -2.814. There is a 33 base pair gap following the feature. The top BLASTp feature is hypothetical protein PBI\_CLOUDWANG3\_105 from Mycobacterium phage CloudWang3, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this feature.

**Feature 89 Annotation**

Only Glimmer called this forward-reading feature at 48768 base pairs; GeneMark did not call a feature in this area. The feature was ultimately deleted as there is no coding potential in the GeneMark map, poor BLAST results, and not a large enough gap to justify changing reading direction.

**Feature 90 Annotation**

Both Glimmer and GeneMark call this feature at 49112 base pairs. The call was agreed with as the alignment is best at this call; all other characteristics of the call stay the same when extended. The feature is 333 base pairs in length, which is not the longest possible open reading frame, though it covers all of the coding potential from the GeneMark map; it starts with the codon ATG and ends with TGA. The raw Shine Dalgarno score is -4.027, which is not the highest possible score. The top BLASTp score is gp100 from Mycobacterium phage EricB, whose first base pair aligns with JewelBug’s first for the feature. There is not enough evidence to declare a function for this gene.

**Overall Genome Section Changes**

By moving start sites within this section of the genome, the identity matches of the BLASTp results improved by 4.7%, the length of the genes was increased by 5.7%, and the average Shine Dalgarno score became 2.5% less negative. All of these improvements make the genome more closely aligned with the guiding principles.

**DISCUSSION**

This semester, the genome of JewelBug was annotated using evidence from a variety of software and databases. From the original auto-annotation made by DNA Master and the Glimmer and GeneMark databases, two gene deletions, one gene addition, and a variety of start site changes were made on the 40900 - 50341 base pair section. The final draft of the JewelBug genome is the most plausible version based on the current information in the databases referenced for evidence.

Next steps for this research include sending the completed annotation to the University of Pittsburgh Howard Hughes Medical Institute for quality control wet lab testing to confirm the decisions made on this annotation, as well as find functions for the various proteins which currently have no known function. Other steps to take with JewelBug’s genome are to determine its host range and compare it to the rest of the A6 subcluster to determine a possible evolutionary path.

**REFERENCES**

**Abedon ST, Kuhl SJ, Blasdel BG, Kutter EM.** Phage treatment of human infections. Bacteriophage. 2011;1(2):66-85. doi:10.4161/bact.1.2.15845.

**About Glimmer.** (n.d.). Retrieved April 29, 2017, from https://ccb.jhu.edu/software/glimmer/

**Agee, R.** (2012). JewelBug. Retrieved April 29, 2017, from http://phagesdb.org/phages/JewelBug/

**Hatfull, G. F.** (2008). Bacteriophage genomics. Current Opinion in Microbiology, 11(5), 447-453. doi:10.1016/j.mib.2008.09.004

**Hatfull, G. F.** (2014). Mycobacteriophages: Windows into Tuberculosis. *PLoS Pathogens,* *10*(3). doi:10.1371/journal.ppat.1003953

**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. Retrieved April 29, 2017, from seaphages.org

**PECAAN.** (n.d.). Retrieved April 29, 2017, from https://discover.kbrinsgd.org/evidence/genes

**Piuri M, Jacobs WR Jr, Hatfull GF** (2009) Fluoromycobacteriophages for Rapid, Specific, and Sensitive Antibiotic Susceptibility Testing of Mycobacterium tuberculosis. PLoS ONE 4(3): e4870. https://doi.org/10.1371/journal.pone.0004870

**Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., . . . Hatfull, G. F.** (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. ELife, 4. doi:10.7554/elife.06416

**Poxleitner, M., Pope, W., Jacobs-Sera, D., Sivanathan, V., & Hatfull, G.** (2016). Phage Discovery Guide. Retrieved April 29, 2017.

**What is the Illumina method of DNA sequencing?** (2015, February 25). Retrieved April 29, 2017, from http://www.yourgenome.org/facts/what-is-the-illumina-method-of-dna-sequencing