Submitted: 01/24/2022
Accepted: 03/27/2022

Research Article

[1]Department of Microbiology and Immunology McGill University, Montreal, QC, Canada.

Keywords

Bacteriophage (phage), phage cluster, genome annotation, SEA-PHAGES, *Mycobacterium smegmatis mc² 155*

Email Correspondence

morgane.brouillard-galipeau@mail.mcgill.ca

Morgane Brouillard-Galipeau[1], Bao-An Chau[1], Jamie Cyr[1], Rafael Intrevado[1], Sunu (Isabel) Kim[1], Cal Koger-Pease[1], Elizabeth Lapshina[1], Alexandra Mircescu[1], Daniella Serrador[1], Michael Slattery[1], Benjamin Vonniessen[1], Michael Shamash[1], Dr. Jasmin Chahal[1]

Genome Annotation of Novel K1 Subcluster Mycobacteriophage Blizzard

Abstract

The evolution of antimicrobial resistant pathogens constitutes a significant global public health threat. Combined with the lack of incentive for pharmaceutical companies to invest in developing new antibiotics, it is clear alternative treatments are needed. Bacteriophages present one possible avenue as they harness the diversity and specificity of a microorganism that has coevolved with bacteria. However, little is known about these bacterial viruses. The SEA-PHAGES program was designed to identify and characterize novel bacteriophages and their associated gene functions. Herein, we report the genome annotation of one such novel phage: Mycobacteriophage Blizzard (GenBank accession number MW712733). Blizzard's gene content was functionally annotated using bioinformatic tools including DNA Master, Phamerator, and NCBI BLAST, to call start sites as well as predict gene function. Overall, 96 genes were identified, including a tRNA and a translational frameshift, using highly similar reference phages BEEST, Belladonna, and CREW. From the 96 genes identified, 46 were functionally annotated. The remaining 50 genes have unknown functions due to the lack of significant matches in the databases. Our results demonstrate a novel annotated phage, whose genome serves to expand the understanding of phage biology and potential implications as alternative treatment to antibiotics.

Introduction

Bacteriophages (phages) are the most abundant, ubiquitous, and diverse microorganisms on Earth[2]. Phages are viruses that infect bacteria and have been isolated from every biome where their bacterial hosts are found[1]. Their host range can span from a considerable breadth of numerous strains across bacterial species or genera to a narrow specificity of a single strain within a bacterial species. Phages that bind to a unique receptor are prone to show a narrow host range, while those that bind to multiple receptors tend to have a larger range[3].

Phages are differentially classified according to their physical structure. The largest of these classifications, the *Caudovirales* order, represents over 96% of the phages known to date. *Caudovirales* are characterized as non-enveloped, tailed phages with a double-stranded DNA (dsDNA) genome contained in an icosahedral protein capsid[4, 5]. A phage that is shown to infect pathogenic bacteria but does not kill commensal organisms can be employed to develop phage therapies, which use phages to treat bacterial infections[3]. This therapy presents an alternative option relevant to the antibiotic resistance crisis.

With the rise in antimicrobial-resistant infections and the pipeline for new antibiotics growing dry, phage therapy has become a more relevant solution. It has been shown to successfully clear multidrug-resistant mycobacteria both *in vivo* and *in vitro*[6, 7, 8]. Interestingly, the selected phages used to target these bacteria were isolated through the SEA-PHAGES program. SEA-PHAGES (Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science) is a program dedicated to cataloguing novel bacteriophages in the public Actinobacteriophage Database (PhagesDB)[9]. To isolate phages, the SEA-PHAGES program uses Actinobacteria, such as mc² 155, as bacterial hosts. To date, 17,000 actinobacteriophages have been isolated and 3,000 have been sequenced by SEA-PHAGES[10]. This expansion of phage gene sequencing has necessitated the grouping of mycobacteriophages into clusters and subclusters according to their nucleotide similarity[11]. Following the sequencing of novel phages, the genome is annotated. Annotated genomes improve efficiency when developing phage therapies, which is essential when treating patients with critical bacterial infections[12]. The more phages are characterized, the more options there are for researchers attempting phage therapy, and the more rapidly we are able to identify and gather phages to target specific bacteria[12].

Figure 1. Phylogenetic tree comparing the full genome of Blizzard (Query_44707 - red box) to the full genome of similar mycobacteriophages. BEEST (green box) appeared as the closest relative to Blizzard, sharing the closest ancestor. CREW (blue box) and Belladonna (purple box) appeared further along the branches. Obtained from BLASTn[14].

Genome annotation utilizes bioinformatics, a multidisciplinary field of science that focuses on computational analysis of biological data[13]. To annotate a phage genome, various tools including DNA Master[14], NCBI BLAST[15], Phamerator[16], and Aragorn[17] are used to align DNA and predicted protein sequences of interest to reference phages. This comparison is used to subsequently infer gene start coordinates, gene functions, as well as genetic elements such as frameshifts and tRNAs. The annotated genome must then be reviewed manually in accordance with guidelines set by GenBank to standardize annotation and generate admissible data.

Our chosen novel phage, Blizzard, was discovered in 2013 by Jean Klonchko Bull at Hope College. It was isolated from an enriched soil sample using the host *Mycobacterium smegmatis* mc^2 155, a nonvirulent relative of *M. tuberculosis*, and sequenced via Illumina sequencing[18]. We aim to annotate the Blizzard genome by finding putative protein coding and tRNA genes as well as predicting their start sites and functions to contribute to the PhagesDB database. Blizzard has over 97% homology with multiple members of its K1 subcluster[15]. We hypothesize that upon analysis, according to the annotation guidelines set by GenBank, Blizzard's genome will reveal gene functions, start sites, and tRNAs similar to many subcluster K1 phages, some of which are promising therapeutic agents.

Methodology

Retrieving the Complete Blizzard Sequence

The discovery and sequencing output data of Blizzard was obtained from PhagesDB[18]. From here, we extracted the fasta file of the complete Blizzard sequence as well as other characterization information.

DNA Master version 5.23.5 was used as the primary program for the genome annotation[14]. DNA Master setting preferences were updated to reflect the SEA-PHAGES Bioinformatics Guide recommendations[19]. The auto-annotate function of DNA Master was used to add gene predictions to the draft annotation of the phage genome[14]. We then performed a BLAST search on all the genes against the NCBI public database[15].

Gene Calling

To determine the validity of the gene calls made by DNA Master, we assessed which auto-annotation programs called the gene, the evidence of coding potential, and the presence of the same gene in other phages from the same cluster. The two auto-annotation programs used were Glimmer and GeneMark[20, 21]. Both systems specialize in determining the gene start sites of microbial genomes and can distinguish between coding and non-coding sequences[20, 21]. Coding potential was also determined using GeneMarkS graphs. The bioinformatic tool Phamerator was used to compare the Blizzard genome to BEEST (Accession number: MH509444.1, 9), CREW (Accession number: KY380102.1, 9), and Belladonna (Accession number: MH697578.1, GenBank direct submission), three annotated phages from the same cluster as Blizzard[16].

These phages share 99.95%, 96.36%, and 95.34% in nucleotide identity, respectively, with Blizzard[15]. Percent identity and e-value, indicators of quality and similarity to the query sequence, were also recorded for each gene. Additionally, every gene was assessed on their adherence to the major guiding principles as defined by the SEA-PHAGES bioinformatics guide. An overlap of 4 base pairs (bp) between genes is optimal, while over 30 bp may be unfavorable. The length of a gene is usually above 200 bp, but they can still be considered if they are over 120 bp[19].

Gene Start Site

To determine the start site of the genes, the suggested calls of Glimmer and GeneMark were noted[20, 21]. Next, using the Ribosome Binding Site (RBS) or Shine-Dalgarno Sequence Finder within DNA Master, the start sites with the most favorable scores were selected[14]. The RBS Sequence Finder is an algorithm assessing the upstream sequence of start codons, evaluating multiple variables such as the distance between this sequence and the start codon. The Z-Score and final score were used to determine the caliber of each start call[14]. The program Starterator was then used to examine all the genes within the same cluster as Blizzard. Starterator examines the alignment of the longest open reading frame (ORF) of each gene in related phages and assists in determining which start is present in most annotated genomes[19]. Then, the local BLASTp results were assessed to determine if the start site was conserved in other phage genomes[15]. The consensus of the percentage alignment, percentage similarity, e-value, and start position alignment were used to call the start sites.

Gene Function

To determine the putative function of genes, the amino acid sequences were compared by sequence alignment via PhagesDB, GenBank, and NCBI databases[15, 18, 22]. Only proteins with an e-value of 10^{-4} or less, with an appropriate query coverage, were selected. HHpred was then used to analyze the best match to the selected database sources (PDB, SCOPe7-A, Pfam-A, and NCBI CD) with a high probability score (>90%)[23]. Synteny, the use of the location of a gene to ascribe its function, was evaluated using Phamerator with reference phages BEEST, Belladonna, and CREW[16]. The general consensus of these tools determined the function of the genes.

Identifying tRNA in the genome

Aragorn and tRNAscan-SE were used to assess the presence of tRNAs through the prediction programs[17, 24]. Both programs were run during manual annotation to call for the stop site of the tRNA, if present, and scan the phage genome for a conserved region in the tRNA. Results from Aragorn embedded in DNA Master were compared to those from tRNA-scan-SE and the updated online version of Aragorn.

Evaluating the Presence of Frameshifts

Phamerator and BLAST were used to determine the location of the shift in the protein sequences[15, 16]. Then, the direction of the frameshift and the coordinates of the slippery sequence were identified[19]. To edit frameshifts for Blizzard, we used the Six Frame Translation window in DNA Master[14]. The gene slippage was identified by observing a 5'-GGAAAA-3' sequence common to all K1 phages[25].

Adding Genes

After manual annotation of genes called by DNA Master, large gaps (>120 bp) between genes were evaluated using the same protocol employed for the manual verifications, to see whether any genes were missed during auto-annotation.

Results

Overall Genome Characteristics

Blizzard is part of the K1 subcluster. Its genome is 59,905 bp and has a G/C content of 66.6%[18]. The auto-annotation resulted in a list of 97 protein-coding genes and one tRNA sequence. After the manual annotation, a total of 95 protein-coding genes and one $tRNA^{Trp}$ were confirmed. Functions were assigned to 46 genes, including all the genes that qualified for functional assignment via synteny. Additionally, functions were called for all required genes as outlined by the SEA-PHAGES Bioinformatics Guide.

tRNA

The tRNA gene identified in the original auto-annotation was located at 1052-1128 bp, and was identified as a $tRNA^{Trp}$, carrying a CCA anticodon (Figure 2). This gene was 76 bp long, within the normal range for a tRNA. However, its 3' end did not include the necessary terminal sequence. The Aragorn software output resulted in a correctly trimmed $tRNA^{Trp}$ sequence at 1053-1126 bp, containing one C base of the conserved 3' terminal sequence.

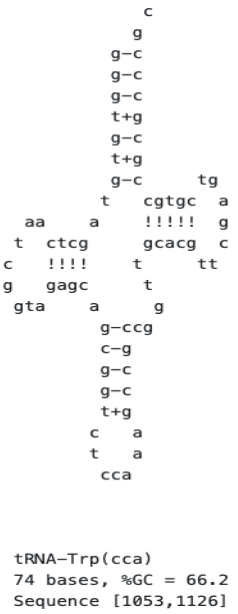

Figure 2. The predicted structure of the trp-tRNA. Structure predicted by Aragorn 1.2.38[22].

Modifications to the Auto-Annotation

Deleted Genes

Predicted genes 5, 18, and 42 were determined to not be true protein-encoding genes as they all had only one or zero BLASTp matches, were not called by GeneMark, and had no similar genes in Belladonna or CREW (Figure 3). The genes were all shorter than 200 bp and genes 18 and 42 faced the reverse direction of those flanking them, which is uncommon for putative protein-encoding genes[16].

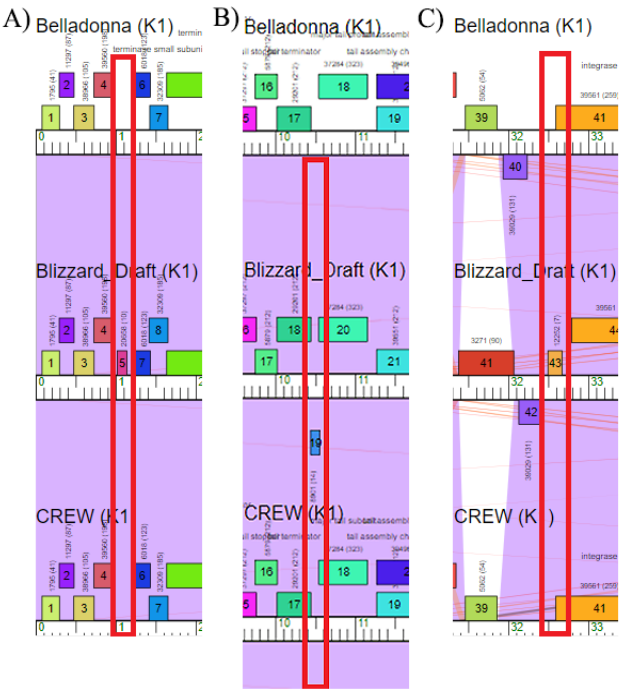

Figure 3. Phamerator maps comparing (A) Blizzard's auto-annotated gene 5 (labelled 5) to gaps in Belladonna and CREW, (B) Blizzard's auto-annotated gene 18 (labelled 19) to gaps in Belladonna and CREW, and (C) Blizzard's auto-annotated gene 42 (labelled 43) to gaps in Belladonna and CREW. Each box represents a gene, and the scale indicates kbp. The colour of the genes indicates their gene family, and background purple indicates sequence similarity, while background white indicates dissimilarity. Map generated by Phamerator[19].

Frameshift

Genes 20 and 21 encode the tail assembly chaperone genes. These genes contain a programmed -1 translational frameshift across a slippery sequence (5'-GGAAAA-3'). Such a frameshift is conserved across K1 phages, including BEEST, Belladonna, and CREW (Figure 4)[16, 18, 25]. The slippery sequence was identified with the first "A" nucleotide located at 11,670 bp (Figure 5). The annotation for gene 21 was modified to include two regions. The first region began at the start site for gene 20 (11,629 bp) and ended at the first "A" (11,670 bp) while the second region began at 11,629 bp and ended at the gene 21 stop site, 12,116 bp.

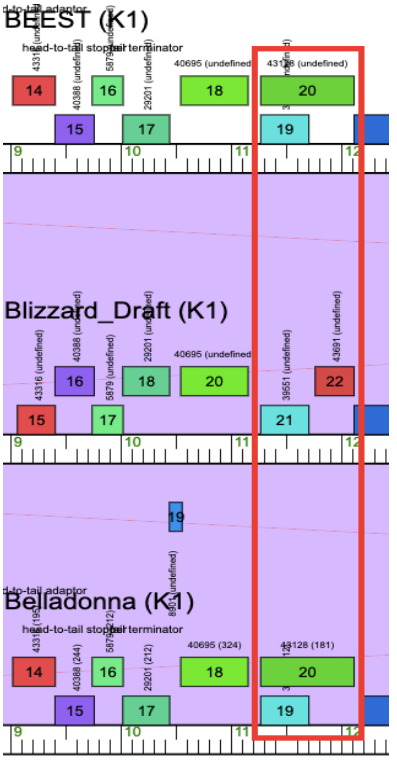

Figure 4. Phamerator genome map showing tail assembly genes in Blizzard (auto-annotated), Belladonna, and BEEST. The programmed translational frameshift is visible in Belladonna and BEEST genes 19 and 20 but is not yet edited in Blizzard_Draft genes 21 and 22 (within the red box). Each box represents a gene, and the scale indicates kbp. The colour of each gene indicates their gene family and background purple indicates sequence similarity between the two genomes. Map generated by Phamerator[19].

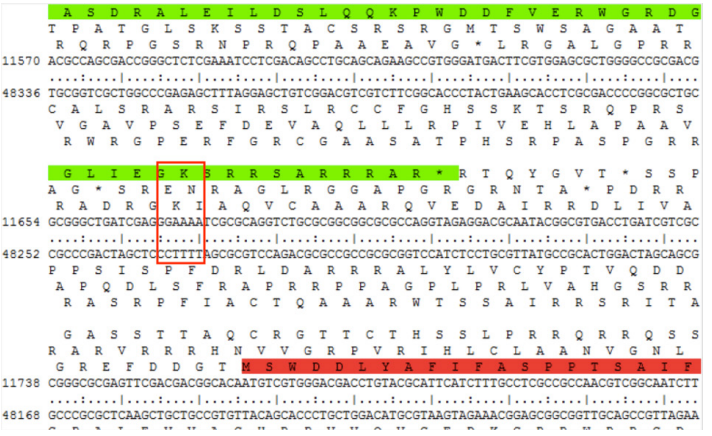

Figure 5. DNA Master Six frame view showing gene 20 in green and gene 21 in red, with the slippery sequence in the red box. Figure from DNA Master, version 5.23.5[15].

Added Gene

Large gaps are uncommon in phage genomes, such that stretches of unassigned DNA greater than 120 bp were examined for putative genes[19]. An additional gene not called by the auto-annotation was identified between

auto-called genes 71 and 72. This region of 409 bp did not show coding potential on the GeneMark maps (Figure 6), but it had well-aligned BLASTp matches to several putative proteins. There are also homologous genes in Belladonna and BEEST, but not in CREW (Figure 7). According to the Starterator report, this protein was annotated in 46 of subcluster K1's 83 non-draft members, excluding Blizzard, in 1 of subcluster K3's members, and in 10 of subcluster K6's members[26]. However, few BLASTp matches had a percent identity greater than 90% with this protein's sequence[15], indicating that this protein sequence is weakly conserved.

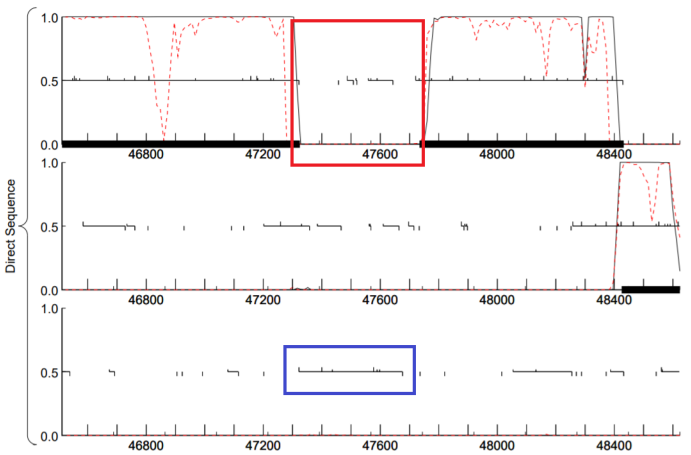

Figure 6. GeneMarkS graph of the gap between genes 71 and 72 (red box) and the ORF where gene 71.5 was inserted (green box). Horizontal lines indicate ORFs, with upward ticks indicating potential start sites and downward ticks indicating stop sites. Black curves indicate typical coding potential, white red dotted curves indicate atypical coding potential. Scale is measured in base pairs. Figure from GeneMarkS version 2.5p[18].

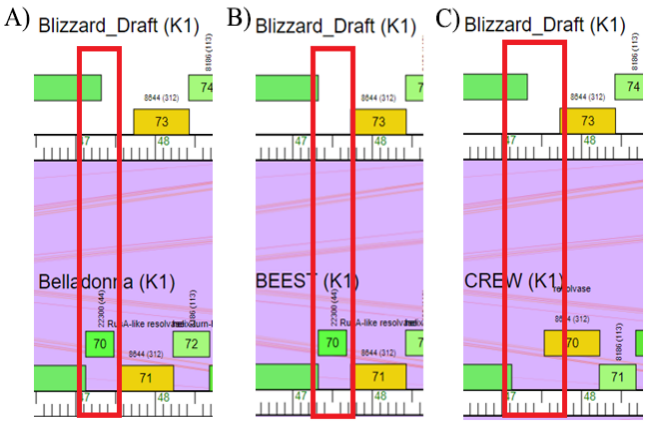

Figure 7. Phamerator maps comparing the gap in Blizzard auto-annotation where gene 71.5 was inserted with (A) gene 70 in Belladonna, (B) gene 70 in BEEST, and (C) a gap in CREW. Each box represents a gene, and the scale indicates kbp. The colour of the genes indicates their gene family, and background purple indicates sequence similarity. Map generated by Phamerator[19].

Ambiguous Gene Calls

Several genes had conflicting evidence for different potential start sites, such as gene 46. The gene was not called by GeneMark's auto-annotation. However, it was called by Glimmer, had atypical coding potential in the GeneMarkS graph (Figure 8), and had strong BLASTp similarity to the well characterized Cro protein, and therefore it was determined to be a gene. Glimmer called the start site at 35,432 bp, but the start site with the best RBS score was 35,456 bp. Starterator called a third start site at 35,474 bp as the consensus, present in 20.3% of annotated homologues. The 35,432 bp start resulted in the longest reasonable ORF. In the K1 phage BEEST, a similar gene had a start site at 35,474 bp (Figure 9). Gene sequences with the different start sites were analyzed by BLASTp, and the sequence starting at 35,432 bp had the strongest results. Ultimately, the Glimmer call, BLASTp results, and the ORF length were determined to be the stronger pieces of evidence, and the start site was called at 35,432 bp.

Another ambiguous start site call was gene 62. Glimmer called the start at 41,471 bp, while GeneMark called it at 41,453 bp. The Starterator report called the consensus start at 41,447 bp. 41,447 bp also had the highest RBS score and longest possible ORF. Additionally, it resulted in the ideal 4 bp overlap with the previous gene (bioinformatics guide). For these reasons, 41,447 bp was called as the start site for gene 62.

Some gene functions were difficult to call due to conflicting evidence. Gene 71's strongest BLASTp match was DNA primase from Adephagia, however, further analysis suggested the presence of additional domains: there were BLASTp matches to DNA primase/helicase and DNA primase/polymerase. The HHPred data support the presence of additional domains, as the 5' end sequence of gene 71 best matched a DNA primase, and the 3' end matched a helicase. We sought to confirm a helicase domain running only the suspected domain's sequence in BLASTp but obtained no different results than for the entire gene 71. Thus, there was insufficient evidence to conclude that gene 71 had a helicase domain, and it was called a DNA primase.

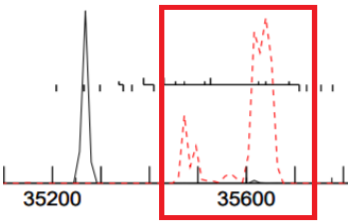

Figure 8. GeneMarkS graph of gene 46 (35432-35713 bp). Horizontal lines indicate ORFs, with upward ticks indicating potential start sites and downward ticks indicating stop sites. Black curves indicate typical coding potential, white red dotted curves indicate atypical coding potential. Scale is measured in base pairs. Figure from GeneMarkS version 2.5p[18].

Gene 92's function was predicted to be "antitoxin in a toxin/antitoxin system, HicB-like". BLASTp results for this protein showed primarily hypothetical protein matches, with some helix-turn-helix DNA binding domain and HicB-like antitoxin matches. HHPred had multiple matches to bacterial HicB antitoxin. Although non-phage matches are generally less convincing, here the gene function in question was HicB-like, indicating similarity to HicB. Given that the data for gene 93 was confidently annotated as HicA-like toxin, synteny suggests that its obligatory counterpart HicB-like antitoxin should be the next gene, despite poor BLASTp data.

Discussion

In our annotation of Blizzard, we identified 96 putative genes, including one tRNA. We were able to assign known functions to 46 of the genes. In accordance with our hypothesis, the called genes and their locations were like other K1 phages. Our manual annotation was revised by the SEA-PHAGES review board, who completed the annotation of gene 1 and added gene 94, bringing the total gene count to 97. Blizzard's 5' end showed a conserved sequence of structural genes and many hypothetical proteins directly downstream of that region. Closer to the 3' end of the genome, non-structural proteins were identified. These are required for the phage's lysogenic life cycle and DNA replication.

Blizzard has one more gene than its closest relative BEEST but otherwise contains the same gene families in the same locations[16]. Adephagia and Blizzard are also nearly identical; Blizzard has two additional genes, genes 70 and 94, as well as a non-homologous gene number 64[16]. The prominent presence of gene 70 in the K1 subcluster compared to its sparse occurrence in the other K subclusters despite its low BLAST matches could indicate that the gene is likely not involved in a critical phage function but that its function may still be somewhat advantageous to K1 phages specifically. However, it should be noted that there is a possibility for this gene to be an artifact; *in vitro* experiments would be needed to confirm the annotation's validity.

Notably, we identified a translational frameshift in the 5' region, in the tail assembly chaperone genes. This corresponds with the highly conserved translational frameshift found in dsDNA bacteriophage tail assembly

genes[25]. Although other phage genes can undergo translational frameshifts, we only considered the one in the tail assembly chaperone genes, as it is the only programmed translational frameshift in phage genomes supported by *in vitro* experiments[25].

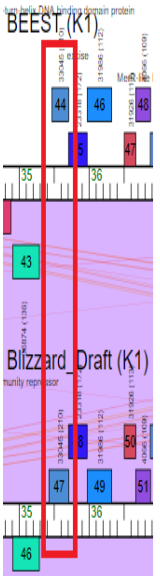

Figure 9. Phamerator map comparing Blizzard's auto-annotated gene 46 (labelled 47), starting at 35432 bp, with BEEST's gene 44 (labelled 44), which starts at 35474 bp. Each box represents a gene, and the scale indicates kbp. The colour of the genes indicates their gene family, and background purple indicates sequence similarity. Map generated by Phamerator[19].

Blizzard's genome contains a single $tRNA^{Trp}$ gene, as do most phages in sub-cluster K1[16, 18]. Phages use host translation machinery, so they do not typically encode translational proteins, although some phages encode tRNA molecules[27]. Phages may encode tRNAs for amino acids used commonly in phage proteins but not host proteins in order to make translation more efficient[27]. This would suggest that Blizzard's proteins contain more tryptophan codons than its host(s). This result aligns with phage biology as Blizzard is a temperate phage and thus integrates its genome within that of the host. Temperate phages encode fewer tRNAs than lytic phages, which replicate in and lyse their host cell, as temperate phages have lower codon biases and few compositional differences from their host genome, facilitating genome integration[27].

We called the function of gene 71 as DNA primase, but HHPred data suggests that this gene contains a helicase domain as well. Gene 71 could contain a primase/helicase polyprotein that is separated post-translationally via proteolytic cleavage. If gene 71 codes for a polyprotein, there is likely a protease gene present elsewhere in Blizzard's genome. Alternatively, this gene could encode a single bifunctional primase/helicase enzyme, as observed in other *Caudovirales* phages such as T7[28].

Blizzard's gene 40 did not have a homologue in many K1 phages, however there are homologues present in some K3 and K6 phages. This implies that the gene may have been present in the last common ancestor of K cluster phages and was lost in most K1 phages. Alternatively, Blizzard and other cluster K phages could have acquired it horizontally[29]. *In vitro* experimentation involving gene 40 knockout strain of Blizzard could determine the function and necessity of this gene.

While manual annotation allows for integration of results from various bioinformatics and auto-annotation software, the methods can still be perfected as the estimates of gene functions are mainly based on synteny and comparison with closely related phage genomes. Since many of the tools used also rely on similar algorithms, it is also possible that if one tool provides erroneous data, other tools may also have made the same error. This limitation illustrates the importance of using a variety of types of tools with differing algorithms to come to a well-supported decision when making a conclusion. Phage-specific auto-annotation software, PHANOTATE and Prodigal, can be used to confirm our results[30]. Another limitation of this genome analysis is that it is not possible to determine the functions of many of the proteins present in Blizzard. To further confirm our predictions and function assignments and determine the function of hypothetical proteins, in vitro proteomic studies should be performed.

Figure 10. The top HHPred search results for gene 71's product. These results are visualized as thick, coloured bars. Red-coloured bars indicate a strong match to the query sequence. The length of the bar corresponds to the section of the query (the thin green bar at the top) that the result matches with. Figure from HHPred[21].

Significance of Genome Annotation

As phages play a crucial role in ecological processes, the characterization of isolates contributes to the understanding of the diversity of phages in the environment. By studying genome organization and gene conservation, we can better understand phage biology. Furthermore, the screening and identification of phages adds to the library of characterized genomes required for phage therapy[31].

In fact, some phages isolated and annotated by SEA-PHAGES have been used for the treatment of drug-resistant mycobacterial strains[7]. Recently, Guerrero-Bustamante et al. created a five-phage cocktail that efficiently kills several strains of *M. tuberculosis* while minimizing phage resistance[8]. One of the five phages was an engineered strain of Adephagia, a phage discovered through the SEA-PHAGES program. The tyrosine integrase and immunity repressor genes necessary for lysogeny were deleted in the strain AdephagiaΔ41Δ43, creating a lytic derivative[8]. Adephagia and Blizzard are both K1 phages with similar gene functions, as previously discussed. This makes Blizzard an ideal candidate to test against *M. tuberculosis*. Our annotation contains the putative locations and sequences of the immunity repressor and integrase genes and could be used to target these genes to engineer a lytic mutant. Furthermore, Blizzard has first been isolated using the *Mycobacterium smegmatis* mc^2 155 host, a laboratory strain of a non-pathogenic relative of *M. tuberculosis*[18, 32, 33]. This host has been used to isolate 11,752 phages thus far, across at least 31 different clusters, yet despite this wide diversity and large number, only phages of cluster K and subclusters A2 and A3 are known to also infect *M. tuberculosis* efficiently[18, 34]. Little is currently known about receptors phages utilize to invade *Mycobacterium* hosts[34], but it can be hypothesized that *M. tuberculosis* and *M. smegmatis* share a surface protein that is used as a receptor by K-cluster phages. Because Blizzard is contained in subcluster K1, it is possible that Blizzard can infect *M. tuberculosis* in a similar way it can penetrate the *M. smegmatis* host. Thus, creating a lytic Blizzard derivative and testing it for activity against *M. tuberculosis* through plaque assays is a promising future direction, as this could provide another potential phage for use in a cocktail against tuberculosis.

Finally, when choosing bacteriophages to use in a cocktail, it is important to determine whether the phages in question contain any genes that would promote the virulence of the bacteria, such as antibiotic resistance genes (ARGs). Current data suggests that phages, especially phages with lytic activity, rarely encode ARGs[35, 36]: a recent study showed that of 5295 viral sequences reconstructed from river samples, only 25 encoded ARGs[35]. This study also found that in viromes of various habitats such as soil and ocean water, only 0.001 to 0.440% of viral reads were predicted to be ARGs[35]. Because so few ARGs were annotated from phage genomes, and because

our phage annotation depends on previously annotated phages, it is not possible to identify ARGs in Blizzard based on bioinformatic annotation alone; wet lab experiments would need to be performed. However, even if it is found that Blizzard contains an ARG, it would be possible to modify its genome to remove the gene for use in therapeutics.

Conclusion

Through this project, we annotated the genome of Blizzard, a temperate K1 phage. Blizzard has 96 genes; protein functions were identified for 46 of these genes, including several genes characteristic of temperate phages, as well as a $tRNA^{Trp}$. Though the annotation relies on a limited number of tools and techniques, further *in silico* or *in vitro* experiments could be performed to confirm the gene positions and functions. The annotated genome of Blizzard furthers our understanding of phage biology and allows better characterization of the phage for use in therapy against antibiotic-resistant bacterial infections. Additionally, the annotation facilitates the creation of targeted knockouts of Blizzard, which allows for creation of a lytic derivative, as was done with Adephagia[7]. Blizzard or its lytic derivative could be tested against drug-resistant pathogens to examine their possible use in phage therapy. Phages such as Blizzard provide an alternative treatment for bacterial infections that could prove invaluable in the fight against the antibiotic-resistance crisis.

Acknowledgments

We would like to thank our supervisors Dr. Jasmin Chahal and Michael Shamash for their guidance. We would also like to thank the SEA-PHAGES program, Dr. Samantha Gruenheid, and the Microbiology and Immunology Department at McGill University for this research opportunity. Additionally, thank you to Drs. Jesse Shapiro and Ken Dewar for their contributions to the course. Finally, we would like to extend our thanks to Hope College for isolating Blizzard and to Pittsburgh Bacteriophage Institute for sequencing the phage genome.

References

1. Salmond, G.P.C. & Fineran, P.C. A century of the phage: past, present and future. Nat. Rev. Microbiol. 13, 777-786 (2015).

2. Keen, E.C. A century of phage research: bacteriophages and the shaping of modern biology. Bioessays 37, 6-9 (2015).

3. Ross, A., Ward, S. & Hyman, P. More Is Better: Selecting for Broad Host Range Bacteriophages. Front. Microbiol. 7 (2016).

4. de Jonge, P.A., Nobrega, F.L., Brouns, S.J.J. & Dutilh, B.E. Molecular and Evolutionary Determinants of Bacteriophage Host Range. Trends Microbiol. 27, 51-63 (2019).

5. Ackermann, H.W. Bacteriophage taxonomy. Microbiol. Aust. 32, 90-94 (2011).

6. World Health Organization. Antimicrobial resistance (2021); https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

7. Dedrick, R.M. et al. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. Nat. Med. 25, 730-733 (2019).

8. Guerrero-Bustamante, C.A. et al. Toward a Phage Cocktail for Tuberculosis: Susceptibility and Tuberculocidal Action of Mycobacteriophages against Diverse Mycobacterium tuberculosis Strains. mBio 12, e00973-00921 (2021).

9. Jordan, T.C. et al. A Broadly Implementable Research Course in Phage Discovery and Genomics for First-Year Undergraduate Students. mBio 5, e01051-01013 (2014).

10. Hatfull, G.F. Actinobacteriophages: Genomics, Dynamics, and Applications. Annu. Rev. Virol. 7, 37-61 (2020).

11. Hatfull, G.F. et al. Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J. Mol. Biol. 397, 119-143 (2010).

12. Bárdy, P., Pantůček, R., Benešík, M. & Doškař, J. Genetically modified bacteriophages in applied microbiology. J. Appl. Microbiol. 121, 618-633 (2016).

13. Third international conference "Bioinformatics: from Algorithms to Applications" (BiATA 2019). BMC Bioinformatics 20, 516 (2019).

14. Lawrence, J. DNA Master. 5.23.5 ed. University of Pittsburgh, Pittsburgh, PA; 2020.

15. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 215, 403-410 (1990).

16. Cresawn, S.G. et al. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12, 395 (2011).

17. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11-16 (2004).

18. Russell, D.A. & Hatfull, G.F. PhagesDB: the actinobacteriophage database. Bioinformatics 33, 784-786 (2020).

19. Pope, W.H., Jacobs-Sera, D., Russel, D.A., Cresawn, S.G. & Hatfull, G.F. SEA-PHAGES Bioinformatics Guide. University of Pittsburgh; 2017.

20. Delcher, A.L., Bratke, K.A., Powers, E.C. & Salzberg, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23, 673-679 (2007).

21. Besemer, J. & Borodovsky, M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33, W451-W454 (2005).

22. Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. GenBank. Nucleic Acids Res. 44, D67-D72 (2016).

23. Söding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244-W248 (2005).

24. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. 25, 955-964 (1997).

25. Xu, J., Hendrix, R.W. & Duda, R.L. Conserved Translational Frameshift in dsDNA Bacteriophage Tail Assembly Genes. Mol. Cell 16, 11-21 (2004).

26. Starterator. http://phages.wustl.edu/starterator/.

27. Bailly-Bechet, M., Vergassola, M. & Rocha, E. Causes for the intriguing presence of tRNAs in phages. Genome Res. 17, 1486-1495 (2007).

28. Toth, E.A., Li, Y., Sawaya, M.R., Cheng, Y. & Ellenberger, T. The crystal structure of the bifunctional primase-helicase of bacteriophage T7. Mol. Cell 12, 1113-1123 (2003).

29. Hendrix, R.W., Hatfull, G.F. & Smith, M.C.M. Bacteriophages with tails: chasing their origins and evolution. Res. Microbiol. 154, 253-257 (2003).

30. Salisbury, A. & Tsourkas, P.K. A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation. Int. J. Mol. Sci. 20, 3391 (2019).

31. Gibson, S.B. et al. Constructing and Characterizing Bacteriophage

Libraries for Phage Therapy of Human Infections. Front. Microbiol. 10 (2019).

32. Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T. & Jacobs Jr, W.R. Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. Mol. Microbiol. 4, 1911-1919 (1990).

33. Li, Q., Ge, F., Tan, Y., Zhang, G. & Li, W. Genome-Wide Transcriptome Profiling of Mycobacterium smegmatis MC^2 155 Cultivated in Minimal Media Supplemented with Cholesterol, Androstenedione or Glycerol. Int. J. Mol. Sci. 17, 689 (2016).

34. Allué-Guardia, A., Saranathan, R., Chan, J. & Torrelles, J.B. Mycobacteriophages as Potential Therapeutic Agents against Drug-Resistant Tuberculosis. Int. J. Mol. Sci. 22, 735 (2021).

35. Moon, K. et al. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. Microbiome 8, 75 (2020).

36. Enault, F. et al. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. 11, 237-247 (2017).