James Madison Undergraduate Research Journal

Volume 2 | Issue 1

2014-2015

A Bioinformatics Approach to Revealing the Genetic Basis for Host Range Specificity

Hayley A. Norian
James Madison University

A Bioinformatics Approach to Revealing the Genetic Basis for Host Range Specificity

Hayley A. Norian

Bacteriophages, or phages, are viruses that infect bacteria. Mycobacteriophages are bacteriophages that specifically infect the genus Mycobacterium. *This genus of bacteria includes human pathogens such as* Mycobacterium tuberculosis, Mycobacterium leprae *and* Mycobacterium ulcerans, *which cause tuberculosis, leprosy and Buruli ulcer, respectively. The full genome sequences of 654 mycobacteriophages are currently available. Collectively, these 654 phages encode 69,581 genes. Only 20.25% of these genes have at least one known homologue in NCBI, the National Center for Biotechnology Information, leaving roughly 80% of all known mycobacteriophage genes without even a predicted function. Bacteriophages are highly host-specific and typically only infect a small number of bacterial hosts. The host range of 204 mycobacteriophages, initially isolated on* Mycobacterium smegmatis *strain* mc^2155, *was recently determined on three other bacterial hosts:* M. tuberculosis *and two* M. smegmatis *strains, Jucho and MKD8. The phages that were capable of infecting one or more of the hosts were of particular interest. The host range information was then used in an association study using Phamerator software to examine the relationship between gene products (protein phamilies) and host range of the corresponding phages. With so many uncharacterized genes encoded by these phages, the potential for elucidating key factors involved in the determination of host range is an exciting prospect.*

Introduction

The genus *Mycobacterium* includes numerous human pathogens ranging from the ancient scourge of leprosy to the world's leading infectious cause of death, tuberculosis (Huygen et al., 1996). However, a newly emerging pathogen is taking hold in sub-Saharan Africa, Central America, and Australia. This organism, *Mycobacterium ulcerans*, is the cause of a disfiguring, debilitating disease known as Buruli ulcer. The slow-growing *M. ulcerans* is the cause of the third most common mycobacterial infection in immunocompromised patients worldwide (Weir, 2002). This disease is difficult to diagnose and treat, and research aimed at improving this situation is receiving little funding or recognition. If treatment is delayed, current antibiotic treatment of the topical lesions caused by Buruli ulcer is ineffective; in most cases, the only effective means of removing the disease from the body if antibiotic treatment is not taken early is expensive and invasive surgical removal of the necrotic skin followed by skin grafting (Etuaful et al., 2005). However, an opportunity exists to combat this pathogen with another class of microbe, viruses that infect bacterial cells.

Bacteriophages are viruses that are parasitic on bacteria, and they are the most abundant biological entities on the planet. Recent estimates of their population size suggest that there are 10^{31} particles at any given time (Sabouri & Mohammadi, 2012). Furthermore, these particles are thought to collectively cause 10^{25} infections of bacterial cells per second worldwide (Lima-Mendez et al., 2007). At this time, the complete genomes of 983 bacteriophages have been sequenced and are publicly available through GenBank, the National Institute of Health genetic sequence database. Since these 983 phages represent only a small fraction of the total population, it is potentially misleading to attempt to generalize. However, a comparison of the available genome sequences shows that bacteriophages are a highly genetically diverse population.

Although phages are extremely diverse in nature, most tend to be highly host-specific. An understanding of the infection process of phages, as well as bacterial host defense mechanisms against phage infection, can explain their typically narrow host ranges. As obligate intracellular parasites, phages must penetrate the membrane of the bacterial cell and manipulate its cellular mechanisms in order to replicate and release mature virions (Rakhuba et al., 2010). In order to initiate an infection, phages must bind to specific receptors on the surface of bacteria (Goldberg et al., 1994). When binding is possible, most phages adsorb to the bacterial cell wall, although some are able to adsorb to extracellular components of the bacterium such as flagella or pili. Adsorption in many cases is mediated by tail fibers, or long projections extending from the base plate of tailed bacteriophages (Aksyuk et al., 2009). Tail fibers initiate adsorption by attaching to specific receptors on the membrane of the bacterial cell; the attachment is first reversible and eventually becomes irreversible (Rakhuba et al., 2010). If the phage lacks the unique component required to bind to the particular host cell or is otherwise unable to bind its receptor, it is incapable of infecting the cell and subsequently replicating within the host.

Following adsorption via tail fibers, a conformational change occurs. The sheath portion of the tail contracts, allowing the bacteriophage genome to penetrate the cell membrane of the bacterial host (Kostyuchenko et al., 2005). Once inside the bacterial cell, bacteriophages typically reproduce by one of two methods: the lysogenic cycle or the lytic cycle. Lysogeny involves the integration of the phage genome into the host's genome, where it remains dormant and is replicated as part of the bacterial chromosome (Wittebole et al., 2014). The integrated phage genome is then known as a prophage. Once an event such as host cell damage triggers excision of the integrated prophage, viral replication may then proceed via the lytic cycle. The lytic cycle is characterized by extensive proliferation of the bacteriophage, leading to the eventual lysis, or breakdown, of the bacterial host cell. Lytic phages manipulate the bacterial synthetic machinery to produce viral rather than bacterial nucleic acids and proteins, leading to phage assembly (Labrie et al., 2010).

Adsorption, penetration, and injection of the bacteriophage genetic material into the bacterial host cell do not guarantee productive assembly and release of the virus during a lytic infection. Many host defense mechanisms exist that allow the bacterial cell to recognize foreign genetic material and degrade it or otherwise halt the phage replication cycle, often at the cost of the infected cell. These defense mechanisms further restrict the already narrow host range of bacteriophages. Restriction-modification systems, involving restriction endonucleases, are one well-known mechanism of host defense against bacteriophages. Restriction endonucleases are enzymes that cleave DNA at or near specific recognition sequences as a defense mechanism against viruses (Loenen et al., 2014). When unmethylated phage DNA enters the host cell, a bacterial methylase can add a methyl group to it. The addition of a methyl group allows the DNA to avoid restriction modification and proceed to subsequent steps of the lytic infection life cycle, including DNA replication, transcription, translation, and phage assembly. However, in most cases, the unmethylated viral DNA is recognized by restriction enzymes of the host's restriction-modification system. These enzymes ignore the host DNA but cut the viral DNA at restriction sites, allowing it to be rapidly degraded (Labrie et al., 2010).

The specificity of phages for their target bacteria has proven extremely useful in diagnostics. Methods exploiting this specificity for bacterial identification are typically quick and cost-effective. For example, phage typing is a method used to determine the source of an infection that involves inoculating the bacteria with different phages of known host specificity in order to differentiate the particular strain causing the infection (Schofield et al., 2012). Plaques, or circular clearings in the agar, are visible when the phage successfully infects and lyses the bacterial host cells. More recently, a method

involving the detection of light produced by luciferase or fluorescent reporter phages has been used to observe infections (Rybniker et al., 2006). Specifically, a TM4 mycobacteriophage was engineered to contain a gene encoding enhanced green fluorescent protein (EGFP) in order to detect drug-resistant strains of *Mycobacterium tuberculosis* (Rondon et al., 2011). This method is both a rapid and an economical means of detecting drug-resistant tuberculosis and may be clinically useful.

As public health concerns over antibiotic resistance in pathogenic organisms escalate, a need for novel mechanisms of treating bacterial infection arises. The specificity of phages to their hosts, among other factors, makes them potentially useful as therapeutic agents to combat bacterial infection. Using phages to treat pathogenic bacterial infections is termed phage therapy. While the ultimate lysis of the bacterial cell is common between both phage infections and antibiotics, the mechanism by which it occurs differs. Some antibiotics, including β-lactam antibiotics such as penicillins, cause bacterial cell death by inhibiting cell wall peptidoglycan synthesis (Holten & Onusko, 2000). As the structural integrity of the cell wall is lost, the bacterial cell becomes susceptible to osmotic pressure and eventually lyses. Phages, on the other hand, bind and infect a particular host, utilize the machinery of that host cell to replicate, and eventually lyse bacteria from within the cell with the help of enzymes such as lysozyme, holin, and hydrolase (Young, 1992). Lysis serves to release the replicated phages and is therefore essential to the spread of viral infections.

Exposing bacteria to antibiotics naturally selects for organisms containing antibiotic resistance genes. While the extent of diversity among the bacteriophage population has yet to be determined, it is evident that phages are much more diverse than the very limited number of antibiotics currently in use (Wittebole et al., 2014). Although using phage therapy would inevitably result in some resistance, considering the diversity and number of phages in existence, it is almost a non-issue. Phages are also much more specific for their host than antibiotics, which can act on a much broader spectrum (Kutateladze & Adamia, 2010). Since phages can only infect particular hosts, the likelihood that the normal flora within the body will be affected is greatly reduced. Leaving the human microbiota undisturbed can potentially reduce the risk of opportunistic infection during treatment (Wittebole et al., 2014). The use of phages as therapeutic agents to treat pathogenic bacterial infections in humans has not been approved in the United States. However, phages are currently being used in the United States as a means of controlling growth of bacterial pathogens and spoilage organisms in food and the food-processing environment (Brovko et al., 2012). While phage therapy lost favor in the United States after antibiotics were introduced, phage therapy is still commonly practiced in Georgia and other parts of Eastern Europe (Wittebole et al., 2014).

The diverse population of bacteriophages may be organized by the bacteria that they infect. Mycobacteriophages, viruses that infect the genus *Mycobacterium*, are of particular interest. The complete genome sequences of 654 mycobacteriophages have been determined. Comparative genomic analysis at both the nucleotide and gene content levels shows that while all 654 mycobacteriophages infect *Mycobacterium smegmatis* strain mc^{2}155, the host on which they were isolated, they represent a diverse population as a whole. These phages have been grouped into 62 distinct groups, termed clusters, subclusters, and singletons, based on average nucleotide identity and protein phamily composition. Protein phamilies (phams) are groups of amino acid sequences sharing considerable sequence homology. Clusters are closely related genomes based on these parameters, while subclusters represent genomes within the same cluster that may be further divided based on differences in degrees of nucleotide similarity between members of the same cluster. Singletons are single genomes with little to no sequence similarity with any another sequenced genome at this time. These classifications vary dramatically in size: singletons, the smallest of these distinct groups, are each the only representative of their genome architecture. Meanwhile, there are 72 subcluster A1 genomes representing that architecture. However, it is important to note that there is likely some sampling bias as the mycobacteriophages represented here have been collected over a number of years, from a relatively small number of sites, and mostly in early fall. Therefore, the current distribution probably does not accurately depict the actual diversity of the population at any given time.

Clusters and subclusters are determined by Phamerator, a bioinformatic tool capable of both comparative genome analysis and representation of bacteriophage genomes. Phamerator organizes genomes and proteins into related groups based on nucleotide and amino acid identity. In order to sort proteins into phamilies (phams) of related amino acid sequences, Phamerator performs pairwise amino acid sequence comparisons between predicted protein products of a set of phage genomes (Cresawn et al., 2011). These phams are organized in such a way as to allow the relationships between different phages to be analyzed using genome maps, which, in turn, illustrate the mosaic nature and potentially the evolutionary history of phage genomes.

While all of the sequenced mycobacteriophages infect *M. smegmatis* strain mc^{2}155, a subset also infects other related mycobacterial hosts. These include human pathogens such as *Mycobacterium leprae*, *M. tuberculosis*, and *M. ulcerans*. The host range of 204 mycobacteriophages, initially isolated on *Mycobacterium smegmatis* strain mc^{2}155, was recently determined on *M. tuberculosis* strain mc^{2}7000 and *M. smegmatis* strains Jucho and MKD8 (Jacobs-Sera et al., 2012). The quantification of host range phenotypes was described as an efficiency of plating relative to mc^{2}155. Plating efficiency is a measure of the number of plaques formed on one host relative to another host. Phages with an efficiency of plating of one for a particular host have the same titer, or concentration of virus, on that host as they have on mc^{2}155. Those with an efficiency of plating of zero for a particular host do not detectably infect that

host. Efficiency values between zero and one can indicate either a reduced replication rate or the emergence of viral mutants that can infect a host that is typically non-permissive for that virus. All three types of efficiency were observed, and efficiencies generally correlated with the phage genome clusters.

Two large sets of data existed without any efficient means of correlating the information contained within them: mycobacteriophage genome sequence data and mycobacteriophage host range data. While 654 mycobacteriophages genomes are currently sequenced and the efficiencies of plating of over 204 mycobacteriophages were determined on three hosts, there was no obvious way to draw conclusions or inferences about the relationship between genome composition and host specificity. The majority of bacteriophage genomes consist of genes of unknown function with no known homologues. The few exceptions seem to be comprised of highly conserved structural genes. Of the 69,581 genes encoded by the sequenced mycobacteriophages, only 20.25% have one or more homologues in NCBI, the National Center for Biotechnology Information. This leaves roughly 80% of all sequenced mycobacteriophage genes without even a predicted function. With so many uncharacterized genes encoded by phages, a computational approach was taken by performing an association study to identify individual mycobacteriophage genes or combinations of genes that are linked to host susceptibility or resistance to each phage.

Methods

Phamerator, the bioinformatic software program used for comparative genomic analysis and representation of phage genomes, is written entirely in Python computer programming language. This software organizes related gene products into phamilies based on amino acid sequence similarity using both BLASTP (The Basic Local Alignment Search Tool for proteins) and CLUSTALW (an alignment program that performs multiple protein sequence alignments) to perform pairwise amino acid sequence comparisons. A CLUSTALW threshold of 32.5% identity and a BLASTP e-value cut off of 10^{-50} served as optimal parameters for building phamilies. These values allowed phamilies of homologous proteins to be built of only closely related domains within the proteins themselves without false pham assembly. Phamerator also performs automated searches of GenBank and NCBI to indicate previously identified proteins and conserved domains.

A host range database was established using the current database of bacteriophage genome data within the program Phamerator. Phamerator utilizes MySQL database software to populate phage and gene tables with information in GenBank records. In addition to these tables, another table containing the bacterial hosts used in the host range study was created. A second table comprised of the plating efficiencies of 204 phages on the different bacterial hosts relative to their infection of *M. smegmatis* strain mc^{2}155 was generated. A genome-wide association study (GWAS) concerning host range was performed. Software was written to find conserved protein phamilies existing in any phage that infect a particular host to determine whether or not these conserved proteins correlate with the ability of the phages to infect the host in question. The information from the GWAS was then depicted on the genomic maps generated within the program by selecting the "show host range data" option from the pull-down menu, at which point the program colored the conserved protein phamilies on the genomic maps according to the host(s) that the phage containing the conserved proteins could infect.

Results

After the host range data was added to the MySQL table, a GWAS was performed to determine the number of conserved and non-conserved protein phamilies that existed between mycobacteriophages with an efficiency of plating on *M. tuberculosis* strain mc^{2}7000 and *M. smegmatis* strains Jucho and MKD8 within one order of magnitude of *M. smegmatis* strain mc^{2}155. A Venn diagram depicts the distribution of these phamilies among phages that could infect one or more of the hosts (Figure 1). Somewhat surprisingly, the mycobacteriophages that were capable of infecting *M. smegmatis* strain Jucho were the most distinct; 55% of the protein phamilies found in phages that infect Jucho are not found in phages that infect MKD8 or *M. tuberculosis*. Meanwhile, 21.4% and 0.6% of protein phamilies found in phages infecting MKD8 and *M. tuberculosis* respectively are unique to those groups. This occurrence may be partially explained by the number of mycobacteriophages observed to infect each host: 14 phages infected *M. tuberculosis,* 23 infected MKD8, and 99 infected Jucho. Of these mycobacteriophages capable of successful infection, one was observed to infect both *M. tuberculosis* and MKD8, nine were observed to infect both Jucho and MKD8, 11 were observed to infect both *M. tuberculosis* and Jucho, and one phage was capable of infecting all three hosts. A total of 88 phages were incapable of infecting any of the three hosts other than *M. smegmatis* strain mc^{2}155, the host on which they were originally isolated.

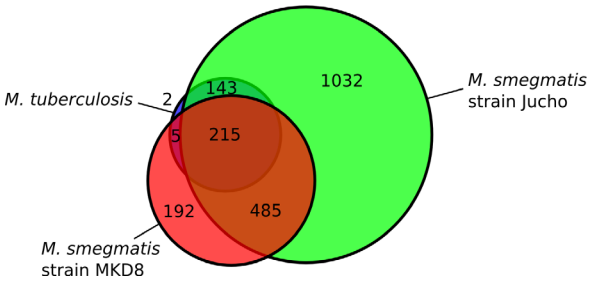

Fig. 1: The number of protein phamilies in phages infecting each of three hosts is shown as a Venn diagram. Two phamilies are found only in genomes that infect the human pathogen M. tuberculosis.

The considerably higher rate at which the tested mycobacteriophages were able to infect Jucho compared to MKD8 and *M. tuberculosis* may be a result of significant diversity within the host Mycobacterium species that were utilized. While a sequenced genome is currently unavailable for *M. smegmatis* strain Jucho, *M. smegmatis* strain MKD8 has been sequenced

and studied. MKD8 appears to be structurally different from *M. smegmatis* mc²155: it lacks a 55.2kb genome duplication present in mc²155 and roughly 1.6% of the genome consists of single-nucleotide polymorphisms, or SNPs (Gray et al., 2013). A total of 649 insertions and deletions greater than 19 basepairs (bp) in length were also observed. Subsets of these insertions and deletions as well as the SNPs present in the genome may be responsible for the phenotypic differences observed within these strains of *M. smegmatis.* Upon analysis of the genome sequences of those 204 phages included in the host range study, pham 982 was identified as a protein phamily of particular interest (Figure 2). Pham 982 is conserved in only five of the 654 mycobacteriophages genomes that have been sequenced to date, and each of the phages containing this protein phamily belong to subcluster A2. Only two of the five phages have had their host range on *M. tuberculosis*, Jucho, and MKD8 tested, but both phages are able to successfully infect Jucho and *M. tuberculosis* (Jacobs-Sera et al., 2012). These two phages, D29 and L5, represent two of the three total A2 phages known to infect *M. tuberculosis*. The third phage, Turbido, lacks pham 982. From this initial analysis, it seems that pham 982 may correlate with the ability of phages L5 and D29 to infect *M. tuberculosis.* HHpred analysis returned a restriction endonuclease as the closest match to the amino acid sequence of pham 982, but with an e-value of 6.6, it remains unclear what the exact function of this protein is. HHpred is a tool used for homology detection and structure prediction.

A recently sequenced and annotated mycobacteriophage, Rover14, shares considerable sequence homology with the Cluster G phage Angel. Angel has been observed to infect both Jucho and *M. tuberculosis* at efficiencies of plating comparable to mc²155 and is the only Cluster G phage known to infect *M. tuberculosis*. Of the three hosts that were tested, the closely related Cluster G phage Halo was only observed to infect Jucho, forming plaques at an efficiency of plating of 1.7 relative to mc²155. When Halo was plated on *M. tuberculosis*, it formed plaques at an efficiency 6.0×10^{-4} lower than on mc²155. However, if plaques were picked from these *M. tuberculosis* plates, harvested and subsequently plated onto mc²155 and *M. tuberculosis*, equivalent titers were observed on both mc²155 and *M. tuberculosis*. The entire genome of the Halo expanded host range mutant was sequenced, and a single non-silent mutation in putative minor tail protein gene product (gp) 22 was identified. This mutation substitutes an alanine residue at position 604 with a glutamic acid residue (Jacobs-Sera et al., 2012).

Angel, which infects *M. tuberculosis* at high efficiency, has an alanine at position 604, suggesting that the glutamic acid residue at this position is not an absolute requirement for infecting *M. tuberculosis* (Figure 3). Interestingly, the homologous gene in Rover14 contains a glutamic acid residue at position 604, just as the mutant Halo phage with enhanced host range does. This observation suggests that like the Halo mutant, Rover14 will infect *M. tuberculosis* at high efficiency.

Utilizing the view by host range data function in Phamerator, comparisons of protein phamily composition between phages belonging to the same cluster or subcluster may be advanced. Upon generation of a genomic map of subcluster A2

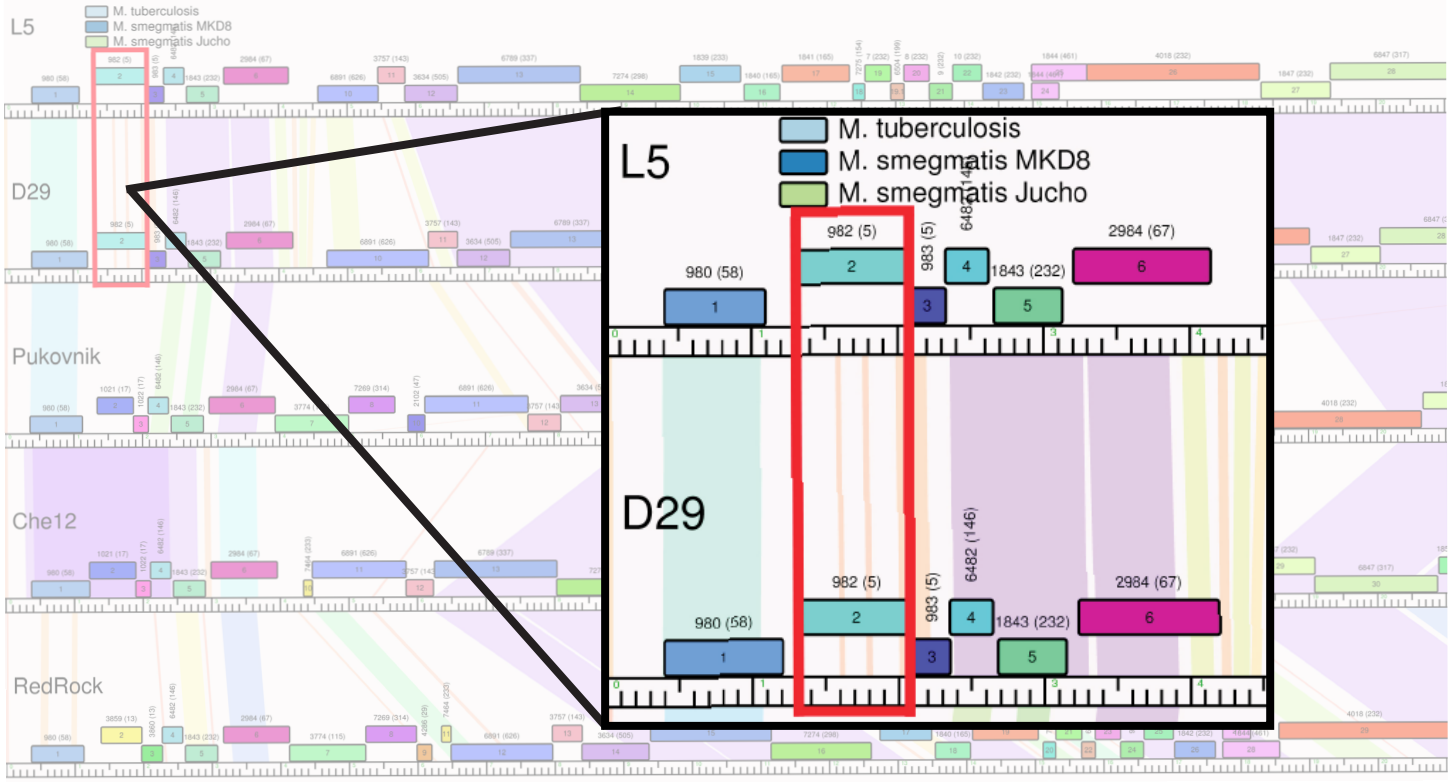

Fig. 2: Genomic map representation of subcluster A2 mycobacteriophages whose efficiencies of plating on M. tuberculosis*, Jucho, and MKD8 were tested in the host range study. Pham 982 (boxed in red) is conserved in only 5 phages, including L5 and D29. Both L5 and D29 infect Jucho and* M. tuberculosis*. Phages Pukovnik, Che12, and RedRock lack Pham 982.*

```
Halo_gp22       PFNAFSITSDSARTFTVSINGTAFDSYTDTAASSSMGANFRNGGWGSSDHNVPGSISQFA  660
Rover14_gp22    PFNEFSITSDSARTFTVSINGTAFDSYTDTAASSSMGANFRNGGWGSSDHNVPGSISQFA  660
Angel_gp22      PFNAFSITSDSARTFTVSINGTAFDSYTDTAASSSMGANFRNGGWGSSDHNVPGSISQFA  660
                ***********************************************************
```

Fig. 3: Clustal Omega amino acid alignment of gene product 22 in cluster G mycobacteriophages Rover14, Angel, and wild-type Halo. While much of the amino acid sequence is conserved between these three gene products, Rover14 contains a glutamic acid residue at position 604, while Angel and wild-type Halo contain an alanine residue (highlighted in red).

mycobacteriophages, an interesting variation in these phages with high nucleotide similarity is observed. While many of the subcluster A2 phages contain the pham 7269, there seem to be two distinct locations at which this protein phamily is found (Figure 4).

Certain mycobacteriophages, such as Trixie, EagleEye, Pukovnik, and RedRock, encode this gene product close to the left end of the genome. Other mycobacteriophages, like Odin, L5, Che12, and D29 encode pham 7269 toward the center of their genomes. Turbido is the only subcluster A2 mycobacteriophage on this map lacking this protein pham. This phage is also the only tested subcluster A2 mycobacteriophage found not to infect the bacterial host Jucho at an efficiency of plating within one order of magnitude of mc^2155.

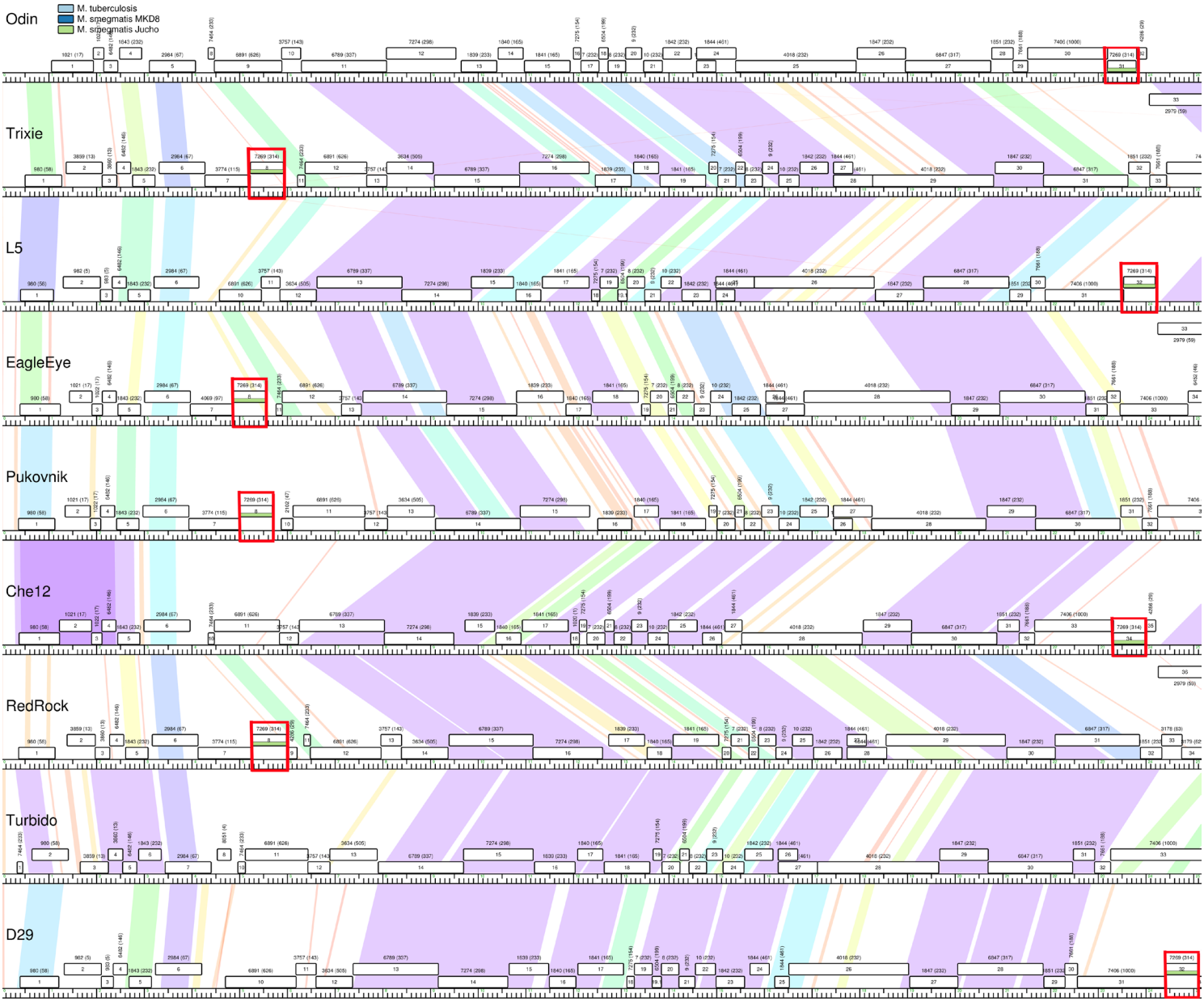

Fig. 4: Genomic map representation of subcluster A2 mycobacteriophages. Pham 7269 (boxed in red) is conserved in eight of the nine phages represented here, with some phages coding for this protein phamily toward the left end of the genome and others toward the middle.

In Figure 4, each instance of pham 7269 is color-coded in green to represent its conservation in phages within the same cluster or subcluster capable of infecting Jucho and lack of conservation in those phages that were not found to infect this host. It is unclear what effect, if any, the positioning of this gene may have on its function, but it is interesting to note that of the four mycobacteriophages with pham 7269 located towards the middle of their genomes, two of these phages are capable of infecting *M. tuberculosis*.

Within this same subcluster, it seems significant that only two of the nine mycobacteriophages representing subcluster A2 here contain pham 3838: Trixie and Turbido. Of the 204 mycobacteriophages tested in the host range study on hosts *M. tuberculosis*, MKD8, and Jucho, Trixie and Turbido were the only two phages belonging to subcluster A2 determined to infect *M. smegmatis* strain MKD8.

The host range data is displayed on the genome maps of these bacteriophages, highlighting the conservation of pham 3838 in subcluster A2 phages capable of infecting MKD8 and non-conservation in those incapable of infection (Figure 5). This phamily is represented in only 18 mycobacteriophage genomes to date.

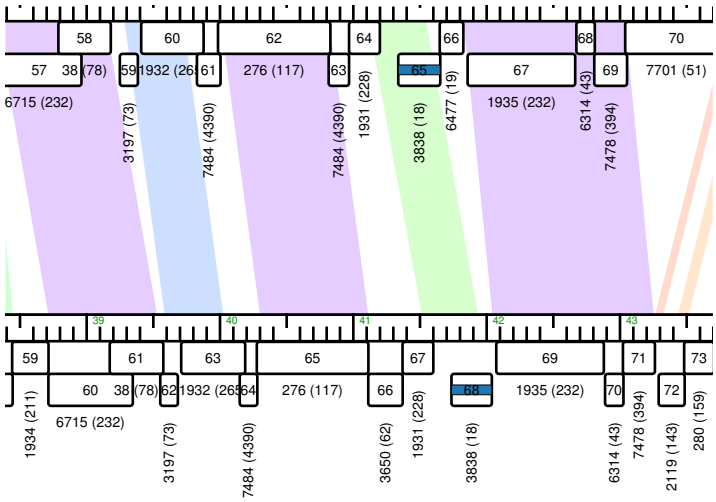

Fig. 5: Subcluster A2 mycobacteriophages Trixie (top) and Turbido (bottom) are the only two tested mycobacteriophages known to infect M. smegmatis *strain MKD8. These two mycobacteriophages both contain pham 3838 while the other A2 phages represented lack it. Pham 3838 is color-coded dark blue.*

Of the three subcluster L1 mycobacteriophages tested in the host range study, only two were capable of infecting *M. smegmatis* strain MKD8 at a plating efficiency comparable to mc^{2}155. When analyzing the genomes of these three mycobacteriophages, JoeDirt, LeBron, and UPIE, pham 3747 was of particular interest (Figure 6). While the phages capable of infecting MKD8, JoeDirt and LeBron, both contain this phamily, this gene product is deleted in the otherwise closely related genome of UPIE, the phage incapable of infecting MKD8.

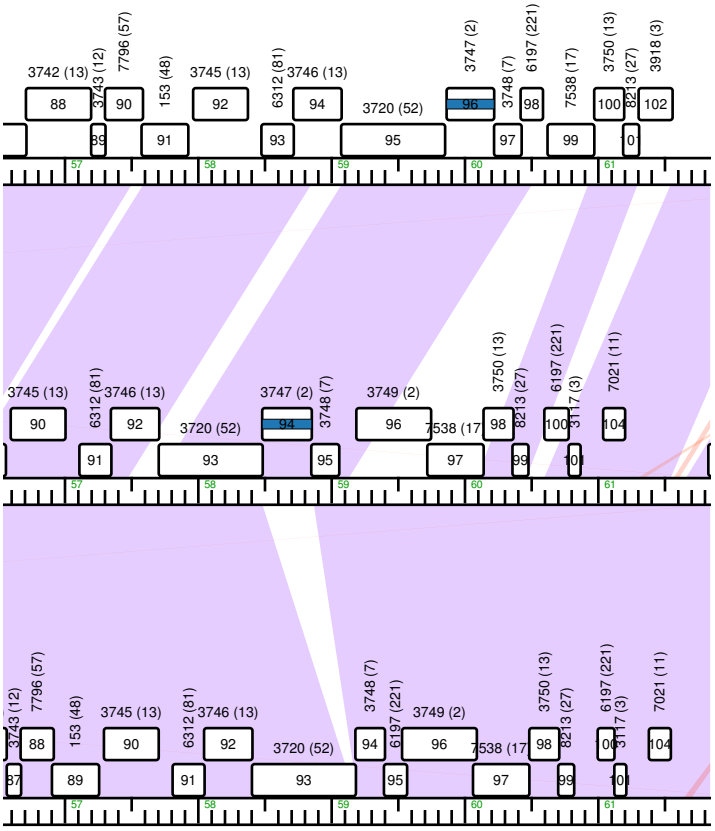

Fig. 6: JoeDirt, LeBron and UPIE represent the three subcluster L1 mycobacteriophages whose host range on M. tuberculosis *and* M. smegmatis *strains MKD8 and Jucho has been tested. Both JoeDirt (top) and LeBron (middle) are capable of infecting MKD8, while UPIE (bottom) lacks this capability. Pham 3747, colored dark blue, is only known to exist in subcluster L1 mycobacteriophages JoeDirt and LeBron.*

JoeDirt and LeBron, the only subcluster L1 mycobacteriophages known to infect MKD8, are also the only two sequenced mycobacteriophages that contain pham 3747. The role of this protein phamily should be investigated further to determine if it plays a role in the specificity of these phages for this host.

Discussion

A number of factors can influence the susceptibility of a given bacterial strain to infection with a particular phage. One well-characterized example is restriction endonucleases, which can degrade the genomic DNA of phages upon its injection into the cell. It is likely that still other systems exist to protect bacterial cells from infection, and that pathways that circumvent these systems remain undiscovered in phage genomes. Thus, it is critical to explore the correlation of protein phamilies in the sequenced mycobacteriophages with the host range of those phages.

All currently sequenced and annotated mycobacteriophages are members of the order Caudovirales. As such, mycobacteriophages share structural similarity in the form of a flexible tail and dsDNA genome contained within an icosahedral

head. Despite their structural similarity, much diversity exists between these phages (Fokine & Rossmann, 2014). The mosaic genomic organization of these phages, as well as the potential evolutionary history, may be observed utilizing Phamerator. Previously, an efficient means of displaying host range data simultaneously with genomic structure did not exist. This software combines host range data compiled in the laboratory with genome arrangement data to search for correlations between the two. Color-coding the conserved protein phamilies in related phages that are capable of infecting a particular host allows for the investigation of the potential role of those proteins in phage infection and host specificity. Considering the relatively high percentage of mycobacteriophage gene products with no known function, potential elucidation of key factors involved in host range determination is an exciting prospect.

A better understanding of this large and diverse population has real world implications. Although bacteriophages may not currently be an integral part of medical care for humans, they have already been implemented in several other fields. Phages are utilized as disinfectants in meat packaging plants and have even been introduced into veterinary medicine (Brovko et al., 2012). Mycobacteriophages in particular provide opportunities for advancement and acceleration of diagnostics for the typically slow-growing genus of bacteria that they infect. Diagnosis of mycobacterial infections and determination of antibiotic-resistant strains of pathogenic bacteria is greatly expedited using mycobacteriophages tagged with EGFP (Rondon et al., 2011). A more thorough understanding of the molecular basis for host range will contribute to the utility of phages as therapeutic and diagnostic tools.

Acknowledgements

I would like to thank my thesis advisor, Dr. Steven G. Cresawn, for his contributions. I would also like to thank Deborah Jacobs-Sera, Dr. Graham F. Hatfull and the Howard Hughes Medical Institute SEA-PHAGES Program. This work was supported by NIH grants 5R24GM093901 and R15AI082527.

References

Aksyuk AA, Leiman PG, Kurochkina LP, Shneider MM, Kostyuchenko VA, Mesyanzhinov VV, et al. 2009. The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. EMBO Journal 28(7): 821-829.

Brovko LY, Anany H, Griffiths, MW. 2012. Bacteriophages for detection and control of bacterial pathogens in food and food-processing environment. Adv. Food Nutr. *Res*. 67: 241-288.

Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics 12: 395.

Etuaful S, Carbonnelle B, Grosset J, Lucas S, Horsfield, C, Phillips R, et al. 2005. Efficacy of the combination rifampin-streptomycin in preventing growth of *Mycobacterium ulcerans* in early lesions of Buruli ulcer in humans. Antimicrob. Agents Chemother. 49(8): 3182-3186.

Fokine A, Rossmann MG. 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4(1): e28281.

Goldberg E, Grinius L, Letellier L. 1994. Recognition, attachment and injection. In: J. D. Karam. Molecular biology of bacteriophage T4. Washington, DC: American Society for Microbiology, p 347–356.

Gray TA, Palumbo MJ, Derbyshire KM. 2013. Draft genome sequence of MKD8, a conjugal recipient of *Mycobacterium smegmatis* strain. Genome Announc. 1(2): 148-13.

Holten KB, Onusko EM. 2000. Appropriate prescribing of oral beta-lactam antibiotics. Am. Fam. Physician 62(3): 611-620.

Huygen K, Content J, Denis O, Montgomery DL, Yawman AM, Deck RR, et al. 1996. Immunogenicity and protective efficacy of a tuberculosis DNA vaccine.Nat. Med. 2(8): 893-898.

Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Bustamante CG, Boyle, MM, et al. 2012. On the nature of mycobacteriophage diversity and host preference. Virology 434: 187-201.

Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov VV, Rossmann MG. 2005. The tail structure of bacteriophage T4 and its mechanism of contraction. Nature Structural & Molecular Biology 12: 810-813.

Kutateladze M, Adamia R. 2010. Bacteriophage as potential new therapeutics to replace or supplement antibiotics. Trends in Biotechnology 28(12): 591-595.

Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nature Reviews: Microbiology 8: 317-327.

Lima-Mendex G, Toussaint A, Leplae R. 2007. Analysis of the phage sequence space: The benefit of structured information. Virology 365: 241-249.

Loenen, WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Research 42(1): 3-19.

Rondon L, Piuri M, Jacobs WR Jr., de Waard J, Hatfull GF, Takiff HE. 2011. Evaluation of fluoromycobacteriophages for detecting drug resistance in *Mycobacterium tuberculosis.* J. Clin. Microbiol. 49(5): 1838-42.

Rybniker J, Kramme S, Small PL. 2006. Host range of 14 mycobacteriophages in *Mycobacterium ulcerans* and seven other mycobacteria including *Mycobacterium tuberculosis* – application for identification and susceptibility testing. J. Medical Microbiology 55: 37-42.

Sabouri GM, Mohammadi A. 2012. Bacteriophage: time to re-evaluate the potential of phage therapy as a promising agent to control multidrug-resistant bacteria. Iran J. Basic Med. Sci. 15(2): 693-701.

Schofield DA, Sharp NJ, Westwater C. 2012. Phage-based platforms for the clinical detection of human bacterial pathogens. Bacteriophage 2(2): 105-283.

Weir E. 2002. Buruli ulcer: the third most common mycobacterial infection. Can. Med. Assoc. J. 166: 1691.

Wittebole X, De Roock, S, Opal SM. 2014. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. Virulence 5(1): 226-235.

Young R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56(3): 430-481.