Spring 2014

Exploration of genomic determinants for host range in phages of pathogenic and non-pathogenic mycobacteria

Anisa Christine Tracy
James Madison University

Follow this and additional works at: https://commons.lib.jmu.edu/master201019

Part of the Biology Commons

Recommended Citation
Tracy, Anisa Christine, "Exploration of genomic determinants for host range in phages of pathogenic and non-pathogenic mycobacteria" (2014). Masters Theses. 351.
https://commons.libjmu.edu/master201019/351

This Thesis is brought to you for free and open access by the The Graduate School at JMU Scholarly Commons. It has been accepted for inclusion in Masters Theses by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.
Exploration of genomic determinants for host range in phages of pathogenic and non-pathogenic Mycobacteria

Anisa Tracy

A thesis submitted to the Graduate Faculty of JAMES MADISON UNIVERSITY

In Partial Fulfillment of the Requirements for the degree of Master of Science

Biology

May 2014
Acknowledgements

Completing this research and thesis has been not only one of the most challenging things I have done but also one of the most rewarding. I could not be more thankful for my advisor Dr. Cresawn and the countless hours of reflecting, reading, encouraging, and most of all patience throughout the entire process. His wisdom, knowledge, and commitment has inspired and motivated me not only for the last two years but the four years I also spent working with him as an undergraduate.

Additionally, I wish to thank my committee members who were more than generous with their expertise and precious time. Thank you Dr. Tim Bloss, Dr. Alex Bannigan, and Crystal Scott for your contribution and support.

This project would not have been possible without the support of my supportive family, and our phage lab “phamily”. I cannot imagine completing this project without all of their support. You guys will always be phamily.

Lastly, I would like to thank the Biology Department at James Madison University for providing me with the opportunity and financial support to complete this Masters program.
Table of Contents

Acknowledgments...........................................................................................................................................ii

List of Figures...................................................................................................................................................v

Abstract............................................................................................................................................................vi

INTRODUCTION.................................................................................................................................................1

I. Significance....................................................................................................................................................1

Buruli Ulcer......................................................................................................................................................1

Mycobacterium Tuberculosis and Leperae.................................................................2

Mycobacterium ulcerans.....................................................................................................................3

Mycobacteriophages...........................................................................................................................3

Phage Therapy.................................................................................................................................................7

II. Host Restriction in Bacteriophage Infection.........................................................................................8

Current Data...................................................................................................................................................8

Host Restriction at Adsorption..........................................................................................................9

Host Restriction of Infection..............................................................................................................10

III. Specific Aims..........................................................................................................................................12

METHODS..................................................................................................................................................14

Bacterial Strains..........................................................................................................................................14

Bacteriophages...........................................................................................................................................14

Amplification Assay............................................................................................................................15

DNA Sequencing......................................................................................................................................15

Bioinformatics..............................................................................................................................................16

RESULTS.....................................................................................................................................................17

Isolation of novel Bacteriophages........................................................................................................17

Optimization of the Amplification Assay..........................................................................................18

Sequence of novel bacteriophages....................................................................................................22

Bioinformatic Analysis..........................................................................................................................28
DISCUSSION..................................................................................................................30

Isolation of novel Bacteriophages.............................................................................30

Amplification assay....................................................................................................30

Sequencing................................................................................................................31

Bioinformatics.........................................................................................................32

Summary...................................................................................................................33

REFERENCES...........................................................................................................37
List of Figures

Figure 1. Phages within a cluster are closely related, while phages in two different clusters have little of no recognizable similarity………………………………………….6

Figure 2. Map of Uganda where soil samples were collected…………………………………….17

Figure 3. Plaque assay of Espe……………………………………………………………………….18

Figure 4. M. ulcerans liquid culture……………………………………………………………….19

Figure 5. Resuspension of M. ulcerans liquid culture with optical density of 0.6 ........19

Figure 6. Elimination of exogenous phage particles with 2 mM FAS and MOI of 1……20

Figure 7. Amplification assay with M. ulcerans……………………………………………………21

Figure 8. Amplification assay with M. smegmatis………………………………………………22

Figure 9. Loading metrics from the Ion Torrent PGM for the first run…………………24

Figure 10. Sample specific metrics for run one on the Ion Torrent PGM…………………25

Figure 11. Loading metrics from the Ion Torrent PGM for the second run………………26

Figure 12. Sample specific metrics for the second run on the Ion Torrent PGM.......27

Figure 13. Phamerator maps for Espe and OrionPax………………………………………29

Figure 14. Conservation in the left arm of Bxb1, Espe and U2………………….35
Abstract

The genus *Mycobacterium* is diverse, ranging from slow-growing human pathogens such as *M. tuberculosis* to fast-growing saprophytes such as *M. smegmatis*. *M. ulcerans* is the causative agent of the neglected disease known as Buruli ulcer, an emerging pathogen first described in Buruli County, Uganda. Antibiotic resistance among Mycobacterial pathogens is growing at an alarming rate and is creating an urgent need for novel diagnostic and therapeutic approaches. Elucidating the details of virus/host interactions among the mycobacteria and their phages thus has the potential to affect how we think about the role of these bacteria in both clinical and environmental settings.

Towards this end, we are analyzing the complete genome sequence of 654 mycobacteriophages isolated on *Mycobacterium smegmatis mc²155*. Here we report on six novel bacteriophages from Ugandan soil, in addition to three bacteriophages isolated from American soil, and using pairwise nucleotide identity and phamily composition of the genomes work to elucidate the genomic determinants for their ability or inability to infect clinically relevant Mycobacteria, including *M. ulcerans* and *M. tuberculosis*. 
INTRODUCTION

I. Significance

Buruli Ulcer

Buruli ulcer, a disease caused by *Mycobacterium ulcerans*, is now considered a neglected tropical disease by the World Health Organization (WHO) and is resulting in significant economic and social impact (Hamzat & Boakye-Afram, 2011; World Health Organization, 2012a). This disease leads to large skin lesions that are generally found on the arms and legs of patients. Before the ulcerative stage of the infection a painless nodule appears. The ulcerative stage leads to still painless lesions that may reach the bone. The recommended treatment is a long-term combination of antibiotics including floroquinolones and rifampicin as well as surgical debridement and grafting (O'Brien et al., 2012). While these treatments have been reasonably successful in the treatment of Buruli ulcer they are not always feasible, due primarily to the fact that Buruli ulcer is most prevalent in poor and rural communities in Africa and Latin America. Early diagnosis is important to successful treatment with antibiotics, but because of the rural location of most patients antibiotics are often not accessible early enough, if at all. In areas most affected by Buruli ulcer, etiology is also often attributed to mystical causes, leading individuals first to seek traditional healers, and then medical treatment when it is too late, leaving disfiguring surgery as the only option (Peeters Grietens et al., 2012). Treatment is also often avoided due to the reputation of the surgery causing greater complications than the disease itself.
M. ulcerans pathogenesis is still unclear, making diagnosis of Buruli ulcer difficult (World Health Organization, 2012a). In addition to difficulty in treatment Buruli ulcer remains difficult to diagnose early. In addition the painless lesions, and disfigurement associated with treatment result in Due to M. ulcerans slow doubling time, culturing the bacteria takes six to eight weeks. While PCR remains the gold standard for diagnosis, it remains difficult due to the high cost, reliance on trained professionals, and availability of laboratory equipment.

Mycobacterium Tuberculosis and Leperae

In addition to M. ulcerans, the genus Mycobacterium also includes Mycobacterium tuberculosis and Mycobacterium leprae. According to the WHO, in 2012 there were an estimated 8.7 million new cases of tuberculosis, leading to 1.4 million deaths (World Health Organization, 2012b). M. leprae has been characterized as having an intracellular lifestyle, infecting macrophages, and Schwann cells of neurons (Bhat & Prakash, 2012). M. tuberculosis also resides within macrophages of the alveoli of the lung; within these cells fusion of the phagosome with the lysosome is inhibited, inhibiting lysis and allowing the pathogen to evade the immune system and persist within the cell (Wong, Chao, & Av-Gay, 2012). One way in which M. tuberculosis and M. leprae become intracellular is through the use of complement receptors and mannose receptors to mediate their entry into macrophages (Van der Werf, Van der Graaf, Tappero, & Asiedu, 1999). This intracellular lifestyle makes access to the bacteria difficult in treatment.
Mycobacterium ulcerans

While *M. tuberculosis* and *M. leprae* are intracellular pathogens, *M. ulcerans* is an extracellular pathogen (Yeboah-Manu et al., 2006). *M. ulcerans* has been shown to live an extracellular life due to the loss of the ESX-1 protein secretion system, and antiphagocytic properties of mycolactone (the toxin secreted by *M. ulcerans*) (Stinear et al., 2007a). This means that the bacteria can be more easily accessed for treatment.

The genomes of clinical isolates have been sequenced revealing that *M. ulcerans* consists of a 5632 kb circular chromosome and a plasmid pMUM001 that is 174 kb. The plasmid has a G+C content of 62.5% and contains 81 protein coding sequences (Stinear et al., 2007b). Comparative genomics has shown that *M. ulcerans* likely evolved over a million years ago from the fish pathogen progenitor *M. marinum*, having 98% nucleotide similarity (Stinear, Jenkin, Johnson, & Davies, 2000). Further analysis has shown that divergence likely occurred with the acquisition of the pUMUM001 plasmid encoding for a mycolactone that is responsible for the phenotypic differences from *M. marinum* (Doig et al., 2012). Mycolactone is the immunosuppressive polyketide toxin that is responsible for the pathogenesis associated with infection with *M. ulcerans*, and is believed to have given *M. ulcerans* the ability to cause this global neglected disease, Buruli ulcer (George, Barker, Welty, & Small, 1998).

Mycobacteriophages

Bacteriophages are viruses that infect bacteria, and are estimated to be the most abundant biological entities on planet earth with an estimated $10^{31}$ particles globally (Wommack & Colwell, 2000). After infection of a host cell, the bacteriophage may
undergo one of two lifestyles. In the lytic cycle replication begins immediately upon infection of the cell, leading to rapid lysis of the bacterial host. Alternatively, in the lysogenic cycle bacteriophages may integrate into the host genome and persist for generations before beginning active replication and lysis of the host cell, this is believed to be primarily due to the state of the bacterial cell but the clear mechanism is not yet defined.

Mycobacteriophages specifically are viruses that infect mycobacterial hosts. Currently there are 654 completely sequenced Mycobacteriophage genomes, each of which was isolated using the non-pathogenic *Mycobacterium smegmatis* as a host (Henry et al., 2010). This work has been largely due to the SEA-PHAGES program, a course sponsored by the Howard Hughes Medical Institute (HHMI) to help encourage undergraduate involvement in research (Hatful et al., 2006). Sequenced and annotated genomes have been compiled into a single database allowing for bioinformatic analysis. Effective bioinformatic tools are becoming more important due to the improvements in technology making sequencing more efficient and cost effective. Even as the number of sequenced mycobacteriophage genomes increases, the data continue to reveal a higher than estimated diversity, as well as “mosaic” genome structure (Pedulla et al., 2003).

Phamerator, a bioinformatic software application, uses pairwise comparisons to sort protein-coding genes into “phamilies” of related sequence (Cresawn et al., 2011). Creation of genome maps allows for visualization of relatedness between phages, as well as analysis of different evolutionary histories of individual genes of these mosaic genomes. Using Phamerator, the mycobacteriophage genomes within the current database can be divided into clusters. A cluster is defined based on the average pairwise nucleotide
identity and protein family composition. Currently Mycobacteriophages are divided into 18 clusters, given alphabetic designation, and further divided into 28 subclusters, noted by a numeric designation. Phages within a specified cluster have highly conserved genomes, whereas when comparing phages between clusters there may be little similarity (Figure 1). Bioinformatic analysis can be utilized to gain further understanding of the mechanisms behind host restriction in *M. ulcerans.* Even among phages from the same subcluster, gene content and host range phenotypes vary considerably. While the genetic basis of host range remains unclear, a correlative study of gene content and host range is a first step toward identifying causal relationships.
Figure 1. Phages within a cluster are closely related, while phages in two different clusters have little or no recognizable similarity. Phages D29 and L5 represent phages from subcluster A2. Containing many similarities as denoted by the colors between the genome maps. Phages BPs and Halo are both cluster G phages and as noted by the purple showing the highest level of conservation. While both cluster A and both cluster G phage are closely related within their respective clusters, conservation is not observed between cluster A and G. (Colors between maps represent level of similarity with the no similarity is shown by white while lowest is red, followed by orange, yellow, green, blue, indigo, and violet being the greatest following the colors of the rainbow)
Phage Therapy

Phage therapy, or the use of bacteriophages in the treatment of bacterial infections is not a new idea, and may hold immense potential with the threat of a post-antibiotic era. In the early 1900s, Felix d’Herelle and Fredrick Twort independently discovered bacteriophages. Felix d’Herelle immediately recognized their potential, and soon after their discovery bacteriophages were used to treat dysentery, cholera, bubonic plague and staphylococcus skin infections (Summers, 1999). Many of these early trials have received much criticism from the west due to their lack of documentation or credibility. In addition, by the mid 1900’s and the emergence of antibiotics, the idea was abandoned by the Western medicine.

While the Western world has turned away from phage therapy, its use has continued in Eastern European and former Soviet Union countries. The Eliava institute of Tbilisi, Georgia is known as one of the leaders in phage therapy having continually used phage therapy in the clinic since the 1900s (http://www.eliava-institute.org). Although phage therapy is not being used in the most traditional sense in the Western world, several companies are using bacteriophage preparations in other novel ways including preparations that have been approved by the FDA for the treatment of surfaces in food processing facilities (http://www.intralytix.com). Additionally, bacteriophages are being explored in diagnostics, especially in very-slow growing or difficult to culture bacterial strains.

Several aspects of bacteriophages make them attractive therapeutic use, including the fact that they are highly host-specific (minimizing damage to the normal flora of the
host), they are self-limiting in that they require their specific bacterial host to replicate, and they are able to replicate at the site of infection due to the presence of the specific host (Clark & March, 2006). *M. ulcerans* has demonstrated resistance to antibiotics including rifampicin, and while in the future resistance to bacteriophages could also be observed, there are an estimated $10^{31}$ bacteriophage particles globally (Wommack & Colwell, 2000). Mixed populations of mycobacteriophages that infect *M. ulcerans* could be used, offering a potentially large number of combinations for therapy. The need for an affordable and accessible novel treatment of *M. ulcerans* is great and due to its localized infection, extracellular life cycle, and poor response to current treatments, bacteriophage therapy could be feasible.

II Host Restriction in Bacteriophage Infection

Current Data

While genomic data for mycobacteriophages are increasing dramatically, host range information is still limited. The only research currently published pertaining to *M. ulcerans* host range evaluated this host’s capacity to support replication of fourteen Mycobacteriophages. Four of which infected *M. ulcerans* (Rybniker J, Kramme S., Small P.L., 2005). Of these four bacteriophages, three were from cluster A, with bacteriophages D29 and L5 belonging to cluster A2 and bacteriophage Bxz2 belonging to cluster A3. The fourth phage, TM4, belongs to cluster K2. All four bacteriophages also demonstrated the ability to infect two different strains of *M. tuberculosis*. While phages from cluster A appear to have an advantage in infecting *M. ulcerans*, it is important that additional phages from cluster A, as well as other clusters are further explored, as it is clear that host
range can differ significantly at even the sub cluster designation. This will help gain a greater insight to weather some specific clusters actually infect *M. ulcerans* more readily than others. In a second host range paper by Deborah Jacobs-Sera *et al.*, phages that are able to infect *M. tuberculosis* were investigated, finding that cluster K and subcluster A2 and A3 appear to most readily infect *M. tuberculosis* (Jacobs-Sera *et al.*, 2012). Phages utilized in this study are included in the Mycobacteriophage database accessed by Phamerator. Results also indicate that host restriction is likely through single residue substitutions in tail proteins (Jacobs-Sera *et al.*, 2012). Phages from cluster A2, A3, K and G all show an ability to infect the genus Mycobacterium. While the viral particles viewed with a transmission electron microscope are similar, the gene content is highly varied. Both of these papers indicate a potential correlation between cluster and host range restriction in the genus Mycobacterium. Furthermore, We propose a study of a broad phage collection like Jacobs-Sera using the host *M. ulcerans*.

**Host restriction at Adsorption**

The first step of infection by a bacteriophage is attachment to the host cell, also known as adsorption. This attachment of the phage receptor binding protein is imperative for subsequent injection of the viral DNA into the host therefore, attachment is believed to be one of the largest determinants in bacteriophage infection (Henry & Debarbieux, 2012; Moldovan, Chapman-McQuiston, & Wu, 2007). In one *Escherichia coli* bacteriophage, adaption to new phage binding receptors has been observed, demonstrating that when a primary receptor is not available, the bacteriophage is able to target a secondary receptor on the cell (Meyer *et al.*, 2012). We therefore propose that host restriction in *M. ulcerans* may be primarily due to differences between the tail fibers
of mycobacteriophages. The Mycobacterial cell membrane of gram-negative bacteria contains a high level of various lipids. Cell membrane content not only differs between genera, but also by species. One example of this is arabinogalactan (AG), a lipid specific to Mycobacteria. Slow growing mycobacteria also contain the phthiocerol lipid family, which is not observed in fast-growing Mycobacteria (Guenin-Mace, Simeone, & Demangle, 2009). One study demonstrated the importance of glycolipid molecules for attachment of phage in *M. smegmatis*, predicting that phage binding capacity could be to strain-specific glycolipid antigens (Khoo et al., 1996). In addition, the unique mycolactone expressed in *M. ulcerans* could potentially play a similar role in host restriction. The host range could be limited, not only by the specificity of these receptors, but also by mutation in the tail fibers, which could prevent binding, and therefore subsequent productive infection.

**Host Restriction of Infection**

Host restriction is not only observed in the initial adsorption stage, but also during the stages of infection through various bacterial defense methods including Clustered, Regularly Interspaced Short Palindromic Repeats (CRISPRs), toxin anti-toxin systems and restriction endonucleases. Host restriction has been observed where Clustered Regularly Interspaced Short Palindromic Repeats CRISPRs are present. These systems offer a horizontally-acquired, but heritable “adaptive” immunity for bacteria and are composed of CRISPR loci and CRISPR associated or *cas* genes (Richter, Chang, & Fineran, 2012). In this system the CRISPR loci contain short non-repetitive DNA stretches that were likely derived from the invading bacteriophage genome. This allows for the recognition of the invading bacteriophages and subsequent sequence-specific
cleavage of the bacteriophage DNA (Deveau et al., 2008). While CRISPR systems have been shown to play a powerful role as the bacterial defense, this is not likely to contribute to host restriction in *M. ulcerans* due to the fact that no defined CRISPR or CRISPR associated (Cas) Systems have been identified in this host.

Toxin-antitoxin systems could play a potential role in host restriction. In these systems, toxins are produced along with antitoxins that are capable of neutralizing the toxin in normal situations. When the cell undergoes stress, such as infection by a phage, the antitoxin can readily degrade. Upon degradation the toxin accumulates and acts to induce cell death (Yamaguchi, Park, & Inouye, 2011). These systems have been identified in systems in which a plasmid relies on the constant production of an anti-toxin, if production of the anti-toxin stops the cell may then die due to accumulation of the toxin without the complementary anti-toxin therefore stabilizing the plasmid in the cell (Frampton, Aggio, Villas-Boas, Arcus, & Cook, 2012). While these systems could effectively restrict host range, it is at the population level rather than conferring a benefit to an individual cell. This leads us to ask the question: is the cost of phage infection higher than death of an individual cell? While more than 60 toxin-antitoxin systems have been identified in *M. tuberculosis*, there are not any currently identified in *M. ulcerans*.

Other possible means of host restriction include restriction endonucleases. Restriction endonuclease systems are commonly observed in bacteria and are able to restrict the host range by targeting and cleaving specific sequences with the invading bacteriophage genome, terminating the infection of the host (Bickle & Kruger, 1993). Bacteriophages are not likely contain restriction sites specific to their bacterial host, making them less prone to cleavage by bacterial endonucleases and giving them an
advantage in infecting that specific bacterium (Kruger & Bickle, 1983). Several potential restriction endonuclease-like proteins have been identified within the *M. ulcerans* genome (Stinear et al., 2007a). These restriction enzymes could potentially target specific sequences in phage DNA for degradation before they are able to replicate within the cell.

While this system of restriction may be observed in *M. ulcerans* enzymes involved are not well characterized and are not believed to play a large role in host restriction.

**III Specific Aims**

This research aims to add to the current host range data available for mycobacteriophages, specifically for the host *M. ulcerans*. This was achieved by first identifying novel phages that have the ability to infect *M. ulcerans*. These novel phages were isolated from soil using a mycobacterium lab strain *Mycobacterium smegmatis* (mc²155) and then tested using the *M. ulcerans* host. A second aim was to sequence and annotate a portion of novel phages using next generation sequencing techniques and annotation tools. The final aim was to use bioinformatics to further define host range in *M. ulcerans*. In addition to the novel environmental isolates, other mycobacteriophages from various geographical regions as well as other mycobacteriophages that have already been sequenced were also obtained and used for testing on the *M. ulcerans* host.

We propose that bioinformatic comparisons can be used to predict host range phenotypes of mycobacteriophages. By comparison of the two subsets of gene content data for bacteriophages with a phenotype able to infect vs. unable to infect the host within a single sub cluster, individual tail protein families are predicted to cluster according to
host range. Identification of point mutations within these tail fiber proteins can be further investigated in the lab by restoring mutations, thereby restoring the ability of that phage to infect *M. ulcerans*. A genome wide-association approach was used to investigate further conservation within phages that are able to infect *M. ulcerans*. Using bioinformatics to guide further laboratory investigation, we aim to develop an increased understanding of *M. ulcerans* host range, allowing for the incorporation of host range predictions into bioinformatic tools like Phamerator.
METHODS

Bacterial Strains

Two *M. ulcerans* strains were obtained as a gracious gift from Pamela Small (University of Tennessee). *M. ulcerans* 1615 is a pathogenic clinical isolate from a Malaysian patient (American Type Culture Collection 35840), and *M. ulcerans* 118 is a derivative that is lacking in the pMUM001 virulence plasmid (Stinear et al., 2004). Both *M. ulcerans* strains were grown at 30°C for 1-3 months on 10% OADC enriched Middlebrook 7H10 agar. Liquid cultures were made by passaging cells from plates in 10% OADC enriched Middlebrook 7H9 media and incubated at 30°C while shaking at 50 RPM for 1-2 weeks. *Mycobacterium smegmatis* mc²155 (Snapper, Melton, Mustafa, Kieser, & Jacobs, 1990) was used for isolation of novel phages. *M. smegmatis* was grown on Middlebrook 7H10 agar 10% ADC, at 37°C shaking at 250 rpm for 24-48 hours. Carbenicillin (50 μg/ml) and cycloheximide (10 μg/ml) were added to all media.

Bacteriophages

All novel bacteriophages were isolated by enrichment of soil samples with *M. smegmatis* for 24 hours before plating on top agar overlays as described below. Putative plaques were then selected and subjected to repeated plaque purification through a series of plating until plaque morphology indicates a single phage population. All mycobacteriophages were propagated using *M. smegmatis* to generate high titer lysates. This was achieved by seeding top agar overlays with 500 μl *M. smegmatis* and plating bacteriophages diluted in phage buffer (10 mM Tris-HCl, pH 7.5; 10 mM MgSO₄; 68.5
mM NaCl; 1 mM CaCl$_2$). Lysates were collected by flooding, collecting, and filtering phage buffer from high titer plates.

**Amplification Assay**

To rapidly identify phages that have the ability to infect the *M. ulcerans* host, a phage amplification assay was performed as outlined by Small (Rybniker J, Kramme S., Small P.L., 2005). Phages tested included unsequenced novel phages as well as a collection of previously sequenced phages obtained as a generous gift from Graham Hatfull (University of Pittsburgh). D29 was used as a positive control bacteriophage capable of infecting *M. ulcerans* (Rybniker J, Kramme S., Small P.L., 2005). The amplification assay was performed by adding 50 µl of each phage tested to 50 µl of bacterial culture at an OD of 0.6. Each phage was diluted such that the final concentration in the infection was $10^7$ pfu/ml, establishing a multiplicity of infection (MOI) of 1. Infection was allowed to take place at 30°C for 90 min (Giri, Bhowmik, Bhattacharya, Mitra, & Das Gupta, 2009). After infection exogenous phage particles were deactivated using 2 mM ferrous ammonium sulfate (FAS) as previously described (McNerney et al., 1998). After vigorous vortexing and incubation at room temperature for 5 min, 20 µl of each amplification product was plated on top agar overlays seeded with 500 µl *M. smegmatis* and incubated for 24 hrs at 37°C.

**DNA Sequencing**

Genomic DNA was prepared for next generation sequencing by first obtaining 10 ml of high titer lysate prepared by infecting lawns of *M. smegmatis* and incubating at 37°C for 24 hours. The plates were flooded with phage buffer and allowed to further
incubate overnight at 4°C. Bacteriophage particles were then filtered through a 0.22 μm filter to remove any bacterial cells. DNase was then added to degrade any bacterial DNA. After incubation with DNase at 37°C phage particles were separated from the mixture using polyethylene glycol precipitation and centrifugation. Bacteriophages were then resuspended in buffer and the DNA was extracted using guanidinium thiocyanate and syringe-mounted columns (Promega Wizard® Genomic DNA Purification Kit.). Genomic DNA was quantified using a standard fluorometer assay.

A bar-coded 200 bp DNA library was then prepared using enzymatic shearing (IonXpress™ Plus gDNA Fragment Library Preparation kit). Appropriate sized fragments were size selected using the E-Gel® SizeSelect™ Agarose Gel. The sample libraries were equalized using the Ion Library Equalizer™ Kit before templating and sequencing. This was achieved using the Ion torrent Personal Genome Machine (PGM) with an Ion 314™ chip, and the Ion PGM™ Template OT2 200 templating and sequencing kits.

**Bioinformatics**

Genomic sequence was assembled using the GS de Novo assembler. After the genomes were assembled they were annotated using gene prediction and annotation software including DNA Master, GeneMark, and Gene Locator and Interpolated Markov ModelER (GLIMMER). These genomes were then added to the existing Mycobacterium database in Phamerator. This was achieved by assigning protein families to each gene using nucleotide blast alignments with an evalue threshold of 1 x 10⁻⁵. Proteins with no database matches are assigned a new protein family and are considered “orphans”.
RESULTS

Isolation of novel bacteriophages

Six novel mycobacteriophages were isolated from soil samples obtained from Ugandan soil taken from three locations in Gulu, Uganda (GPS coordinates 2.788194, 32.305600, 2.788194, 32.305637 and 2.775200, 32.290647) (figure 2). The names given to these bacteriophages are Espe, Naomi, Gloria, Mercy, Fatuma and Aloyo. All phages produced clear plaques on *M. smegmatis* (Figure 3).

Figure 2. Map of Uganda showing were soil samples were collected (Red pin), in relation to Buruli county Uganda (Blue pin) where Buruli ulcer was first identified.
In addition four other phages were isolated from soil from Harrisonburg, Virginia through collaboration with undergraduate students. These phages also isolated on *M. smegmatis mc²*155 were named Lauren 3(1), Burke, OrionPax and Rover14. OrionPax produced halo-like plaques, while Lauren 3(1), Burke and Rover14 all produced clear plaques similar to those of the phages isolated from the Ugandan soil.

**Optimization of the Amplification Assay:**

Pathogenic and non-pathogenic *M. ulcerans* strains were first plated on 7H9 with 10% OADC before suspensions were made for use in the assay. Inoculation of liquid media was not effective in growth of liquid cultures (Figure 4). In order to obtain liquid cultures with an optical density of 0.6, suspensions of *M. ulcerans* 118 cells were made by passaging of cells through a 25 gauge needle (figure5).
Figure 4. *M. ulcerans* liquid culture grown for one year at 30° C. Shaking at 50 RPM. Cell growth produces colonies that adhere to the flask rather than producing a turbid culture.

Figure 5. Resuspension of *M. ulcerans* cells to produce a culture with an optical density of 0.6 at 600 nm. After passage of colonies grown in liquid culture through a 25 gauge needle, the culture is turbid as compared to the culture shown above.
Elimination of exogenous phages with ferrous ammonium sulfate (FAS) at a molarity of 2 mM was shown effective using *M. smegmatis* mc²155 and D29. Infections with a ratio of one bacterial cell to one bacteriophage, or an MOI of one were used, and infection proceeded for 100 min. When no bacterial cells were are added, all phages were successfully eliminated, while when bacterial cells were added successful infection occurred on the indicator plate. When no FAS was added initial added, as well as those produced by infection are plated on the indicator plate (figure 6).

Figure 6. Elimination of exogenous phage particles with 2mM FAS and a MOI: 1 with D29. Top row: *M. smegmatis* mc²155- D29+ FAS+. Middle row: *M. smegmatis* mc²155+ D29+ FAS+ (6-fold dilutions from left to right) Bottom row: *M. smegmatis* mc²155 + D29+ FAS- (6-fold dilutions as shown above).
Control amplification assays were preformed with D29 and Rosebush, both isolated and able to infect *M. smegmatis mc²155*. D29 was also used as a positive control in the *M. ulcerans* 118 strain while Rosebush was used as a negative control for *M. ulcerans* 118 as described by Pam Small (Figure 7 & 8). It should be noted that in the *M. smegmatis mc²155* plate the amplification product ran, resulting in the smeared appearance, as compared to the spots typically seen. Infections were at a MOI of one and allowed to precede for 100 min before deactivation with 2 mM FAS.

Figure 7. Amplification assay with *M. ulcerans*118, 2mM FAS, MOI: 1 and 100 min infection time. Top row: Negative control Rosebush, Left: *M. ulcerans* 118 and Rosebush + no FAS, middle: *M. ulcerans* 118 – Rosebush + FAS +, Right *M. ulcerans* 118 and Rosebush + FAS +. Bottom row: same as above with positive control bacteriophage D29.
Figure 8. Amplification assay with *M. smegmatis* mc²155, 2mM FAS, MOI:1 and 100 min infection time. Top row: Negative control Rosebush, Left: *M. smegmatis* mc²155 and Rosebush + no FAS, middle: *M. smegmatis* mc²155 – Rosebush + FAS +, Right *M. smegmatis* mc²155 and Rosebush + FAS +. Bottom row: same as above with positive control bacteriophage D29.

**Sequence of novel bacteriophages**

Genomic DNA was isolated from Espe, Naomi, OrionPax, Rover14 and Burke. The concentration of the genomic DNA was then calculated using a standard fluorometer. Espe had a concentration of 114.97 ng/μl, Naomi 14.70 ng/μl, OrionPax 276.50 ng/μl, Rover14 302.83 ng/μl and Burke 139.45 ng/μl.

After creation of the 200bp library by enzymatic shearing the individual libraries were pooled together for templating and enrichment. The final template
library concentration as calculated using the qubit ion sphere quality control kit was 3.31%. The library was then run on two 314 ion chips.

The first run produced 183,217 total sequencing reads with a mean length of 216 base pairs. Ion sphere particles (ISP) were loaded into 39% of the wells on the 314 ion chip with 80% enrichment and 86% clonal spheres (Figure 9). Of the total reads, 26,273 corresponded to the Espe genome, 9,296 to Rover14, 22,388 to Burke and 123,994 to OrionPax (figure 10). In the second run 114,859 total reads were sequenced with a mean length of 188 base pairs. 25% of the wells on the second 314 ion chip were loaded with ISP’s with a 79% enrichment and 86% clonal spheres. In this run 16,211 sequencing reads corresponded to the Espe genome, 6,105 to Rover14, 13,853 to Burke and 77,818 to OrionPax.
Figure 9. Loading metrics from the Ion Torrent PGM for the first run of Espe, Rover14, Burke and OrionPax. Ion Sphere Particle (ISP) loading density of the 314 chip is shown on the left where blue indicates a low ISP density and red indicating the highest ISP density. Center panel shows information regarding the ISPs and whether they were usable in the final library. Right panel shows average read length.
Figure 10. Sample-specific metrics retrieved by barcode for run one on the Ion Torrent PGM of Espe, Rover14, Burke and OrionPax. For each sample the number of bases sequenced, quality bases sequenced, reads sequenced and the mean length of those reads is shown.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Bases</th>
<th>&gt;=Q20 Bases</th>
<th>Reads</th>
<th>Mean Read Length</th>
<th>Read Length Histogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>238,643</td>
<td>207,198</td>
<td>1,265</td>
<td>186 bp</td>
<td></td>
</tr>
<tr>
<td>Espe</td>
<td>5,095,011</td>
<td>4,497,261</td>
<td>26,273</td>
<td>193 bp</td>
<td></td>
</tr>
<tr>
<td>Rover</td>
<td>1,653,518</td>
<td>1,480,261</td>
<td>9,296</td>
<td>177 bp</td>
<td></td>
</tr>
<tr>
<td>Burke</td>
<td>4,314,067</td>
<td>3,818,650</td>
<td>22,388</td>
<td>192 bp</td>
<td></td>
</tr>
<tr>
<td>OrionPax</td>
<td>24,608,555</td>
<td>21,744,210</td>
<td>123,994</td>
<td>198 bp</td>
<td></td>
</tr>
</tbody>
</table>
Figure 11. Loading metrics from the Ion Torrent PGM for the second run of Espe, Rover14, Burke and OrionPax. Ion Sphere Particle (ISP) loading density of the 314 chip is shown on the left where blue indicates a low ISP density and red indicating the highest ISP density. Center panel shows information regarding the ISPs and whether they were usable in the final library. Right panel shows average read length.
Figure 12. Sample specific metrics for run two on the Ion Torrent PGM of Espe, Rover14, Burke and OrionPax. For each sample the number of bases sequenced, quality bases sequenced, reads sequenced and the mean length of those reads is shown.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Bases</th>
<th>&gt;=Q20 Bases</th>
<th>Reads</th>
<th>Mean Read Length</th>
<th>Read Length Histogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>152,072</td>
<td>127,655</td>
<td>872</td>
<td>174 bp</td>
<td></td>
</tr>
<tr>
<td>Espe</td>
<td>3,008,292</td>
<td>2,611,110</td>
<td>16,211</td>
<td>185 bp</td>
<td></td>
</tr>
<tr>
<td>Rover14</td>
<td>1,051,305</td>
<td>927,724</td>
<td>6,105</td>
<td>172 bp</td>
<td></td>
</tr>
<tr>
<td>Burke</td>
<td>2,557,714</td>
<td>2,221,464</td>
<td>13,853</td>
<td>184 bp</td>
<td></td>
</tr>
<tr>
<td>OrionPax</td>
<td>14,864,536</td>
<td>12,909,484</td>
<td>77,818</td>
<td>191 bp</td>
<td></td>
</tr>
</tbody>
</table>
Each genome was assembled using the GS de Novo assembler. Espe, OrionPax and Rover14 each assembled into a single contig of 51,099 base pairs, 75,059 base pairs and 41,440 base pairs respectively. Burke assembled into four contigs of 134,463 base pairs, 13,599 base pairs, 4,355 base pairs and 2,680 base pairs.

The average depth of coverage for each of the genomes was then calculated. OrionPax had the greatest depth of coverage with 526, while Espe had an average depth of coverage of 158 and Rover14 had an average depth of coverage of 65.

**Bioinformatic Analysis**

Annotations were completed for Espe and OrionPax. The Espe genome was determined to have GC content of 63.85% and 100 genes. Twenty of these genes were assigned putative functions based on BLAST, HHpred, and RPS-BLAST searches. This genome belongs to the mycobacteriophage cluster A1.

OrionPax has a GC content of 63% and 143 genes, with 30 where function could be assigned. OrionPax belongs to the mycobacteriophage cluster E.
Figure 13. Phamerator genome maps for Espe and OrionPax. The left end of each genome is shown. Genes are displayed as colored boxes above and below the ruler, forward transcribed above and reverse below. Color indicates conserved protein families.
DISCUSSION

Isolation of novel Bacteriophages

While not all of the novel phages isolated have been sequenced, the five unsequenced phages have been archived for use in subsequent experiments and will be tested on the M. ulcerans host. These phages are particularly valuable because they were isolated from Gulu Uganda, which is near where Buruli ulcer was first identified and is still endemic.

Amplification assay

Due to several factors, the amplification assay method to test bacteriophages on M. ulcerans has proven to be more difficult than expected, and has not given reproducible results for most tested phages. One limitation of this method is the host itself. M. ulcerans is notoriously difficult to culture, and our experiences are consistent with this. With a doubling time of 30-80 hours it is a very slow growing organism. In addition to the slow growth, cells also tend to clump significantly in liquid cultures (figure 4) making it difficult to obtain a turbid culture. In order to circumvent some of these difficulties we were able to create liquid cultures by suspending cells grown on a plate. We found that passing cells through a 25-gauge needle repeatedly, in new media, significantly reduced clumping and allowed for a liquid culture with an optical density of 0.6 at 600 nm to be achieved (figure 5).

Elimination of exogenous phage particles is important to ensure that the phages seen on the indicator plate are from active infection of the pathogenic bacteria and not from the initial infection. In order to eliminate exogenous bacteriophages, FAS was used
at a molarity of 2 mM as previously described. We have found that elimination of bacteriophages with FAS can be inconsistent and sometimes insufficient at 2 mM when infecting at a multiplicity of infection of 1. We have also noted that elimination with FAS may vary depending on the bacteriophage itself. As the mechanism behind inactivation with FAS is unknown it is difficult to determine why effectiveness could vary.

A third variable is the time that infection is allowed to proceed. It has been previously reported that temperature has the ability to effect the time taken from attachment to lysis of the bacterial cell. Time points of 40 min and 100 min were routinely used in our assay. Inconsistencies in results have made it difficult to determine if the bacteriophage particles plated were due to amplification (lysis) of the bacterial host or plating of bacterial cells that have been infected. This hinders the ability of this test to be used in a quantifiable manner.

Phage amplification still holds potential for testing many bacteriophages on the M. ulcerans host. Although results have been inconsistent, the positive control (mycobacteriophage D29) and negative control (mycobacteriophage Rosebush) have been successful multiple times (figure 7 & 8).

**Sequencing**

Here we report on the first sequence produced on the ion torrent PGM at James Madison University. Of the four phages sequenced in the first two runs, Espe and IronPax were completely finished and annotated using the PGM sequence alone. Rover14 assembled into a single contig, but due to the fact that the average depth of coverage was only 65, some areas contain low quality consensus sequence. In order to complete this
genome, primers have been designed for Sanger sequencing. The fourth genome Burke, assembled in to four contigs. After evaluation of each of the four contigs it appears that Burke belongs to the mycobacteriophage cluster C. Cluster C phages are relatively large compared to other phages with an average genome size of 155,504 base pairs. Due to the ISP loading densities being low for both runs (figure 9 & 11), and low final template library concentration with only 3.31% templated we were not able to generate enough depth of coverage to obtain a quality consensus sequence for the entire genome with an average depth of coverage of 44 reads.

When finishing the genomes it was also noted that strand specific insertions and deletions were common in all genomes. This strand bias has been noted in many ion torrent sequencing applications, but it has yet to be explained (Quail et al, 2012).

**Bioinformatics**

Mycobacteriophages are known for their mosaic and diverse genomes. Here we report on the complete genome sequencing and annotation of two novel mycobacteriophages: Espe and OrionPax. Using pairwise nucleotide alignment and gene content analysis, Espe was determined to belong to the A1 subcluster of mycobacteriophages. This genome contains one novel gene, which is the basis for a new “orpham.” When comparing Espe to phages of cluster A, a number of sequence repeats were observed. These are consistent in terms of their genomic locations (near 21,000, 41,000, and 47,000 base pairs) with “stoperators,” sequences that act in cis to regulate transcriptional elongation. OrionPax belongs to cluster E. There were no orphans
identified in this genome. This is typical for cluster E phages as they are highly conserved at the nucleotide and protein level.

**Summary**

Some cluster A, K, and G phages isolated on *M. smegmatis* have since been shown to infect *M. tuberculosis* or *M. ulcerans* (Jacobs-Sera et al., 2012; Rybniker J, Kramme S., Small P.L., 2005). There is a strong correlation between phage sub cluster and the ability to infect these alternative hosts, however this correlation is not absolute. Cluster A is large and diverse, with eleven sub clusters in the current database. Genomes within the same sub cluster are more closely related to one another than to genomes with different sub clusters of cluster A. However, when comparing protein conservation within cluster A phages a much greater level of conservation is seen at both cluster and sub cluster designations.

Previous data suggest that the set of phages that can replicate on *M. ulcerans* is very similar to that for *M. tuberculosis*. However, we have found this not to be the case. Within cluster A, several bacteriophages from sub clusters A2 and A3 have the ability to infect *M. tuberculosis* and *M. ulcerans*. Two additional phages from the A1 cluster, Bxb1 and U2, have also shown a limited ability to infect *M. tuberculosis*. Espe belongs to the A1 sub cluster. When comparing genomic data and preliminary amplification assay data for Espe, we propose that this bacteriophage may also be able to infect *M. ulcerans*. When using Phamerator maps for comparisons of Espe, Bxb1 and U2 a high level of conservation is observed at the protein level, especially between Espe and Bxb1. In addition, the highest level of conservation at both the nucleotide and protein level is
observed in the region where tail fiber proteins are predicted (Figure 14). KSSJEB, belonging to the A1 sub cluster has not previously been shown to infect mycobacterial hosts. This is also consistent with the preliminary data that we have obtained using the amplification assay.
Figure 14. Conservation in the left arm of Bxb1, Espe and U2, three cluster A1 phages. Genes shown by boxes above the ruler are translated in the forward direction. Shading between genomes shows nucleotide conservation (purple being the highest degree of conservation followed by indigo, blue, green, yellow, orange and red being the lowest). Protein conservation is shown by the color of the boxes. Each protein family is assigned a unique identifying color. The box shows area of predicted tail fiber function.
Although final annotations have not been completed for Rover14, we predict that this bacteriophage will have the ability to infect *M. ulcerans* due to preliminary data from the amplification assay and that all previously tested cluster G phages have been able to infect pathogenic mycobacterial hosts. While the OrionPax genome was completely finished, host range data is limited for cluster E phages.

Mycobacterial infections pose a serious health threat globally, and have a serious impact in many middle to low income countries. Here we propose that bacteriophages could offer an affordable and effective tool in combating these serious diseases. Not only do bacteriophages provide potential as a traditional therapeutic, but also could play an important role in the diagnosis of these slow-growing and difficult-to-diagnose diseases. In the case of Buruli ulcer the potential of a topical treatment could provide a unique opportunity to reintroduce phage therapy. In order to fulfill the potential of this large, diverse and dynamic population, investigation of host restriction and functionality of proteins in bacteriophage genomes must be completed. We hope that the amplification assay coupled with genomic analysis will help to elucidate these complex virus/host interactions.
References


doi:[http://dx.doi.org/10.1016/j.virol.2012.09.017](http://dx.doi.org/10.1016/j.virol.2012.09.017)

Henry, M., O'Sullivan, O., Sleator, R. D., Coffey, A., Ross, R. P., McAuliffe, O., & O'Mahony, J. M. (2010). In silico analysis of ardmore, a novel mycobacteriophage isolated

10.1016/j.gene.2009.12.007


doi:[http://dx.doi.org/10.1016/j.virol.2012.09.026](http://dx.doi.org/10.1016/j.virol.2012.09.026)


doi: 10.1021/bi961055+


doi:10.1126/science.1214449; 10.1126/science.1214449


mycobacterium ulcersans, the causative agent of buruli ulcer. *Genome Research, 17*(2), 192-200. doi:10.1101/gr.5942807


lesions caused by the extracellular pathogen mycobacterium ulcerans. Journal of Leukocyte Biology, 79(6), 1150-1156. doi:10.1189/jlb.1005581