GLOBAL PROTEOME INVESTIGATION OF MYCOBACTERIOPHAGE OCHI17-MYCOBACTERIUM SMEGMATIS INTERACTIONS

by

Ikenna O. Okekeogbu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Agricultural and Biological Engineering

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Kari Clase, Chair

School of Agricultural and Biological Engineering

Dr. Kevin Solomon

School of Agricultural and Biological Engineering

Dr. Stephen Byrn

Department of Industrial and Physical Pharmacy

Dr. Uma Aryal

Department of Comparative Pathobiology

Approved by:

Dr. Nathan S. Mosier

Dedicated to my parents, Mr. Isaiah and Mrs. Angela Okekeogbu, and my siblings- Nnaama Okekeogbu, Onuchukwu Azozie, Chibuike Okekeogbu, and Arize Okekeogbu.

ACKNOWLEDGMENTS

First of all, I will like to thank the Almighty God for His faithfulness and mercies over me throughout the course of my Ph.D. I wouldn't have made it this far without Him.

I will like to express my gratitude to my parents for their love, prayers, care and support. Their passion for education and excellence has continued to inspire me to be the best I can.

My sincere gratitude goes to my research advisor, Dr. Kari Clase for her support and guidance in my research and my life goals since I joined her lab. Thank you for the opportunity you gave me to direct the course of my research and mentor undergraduate students in the lab as well as the opportunity to be a Teaching Assistant.

I also want to thank my committee members, Dr. Kevin Solomon, Dr. Stephen Byrn and Dr. Uma Aryal. Thank you for the support, feedbacks, and help to making my research better.

I was blessed to be a part of a family here at West Lafayette called the African Christian Fellowship. I couldn't have asked for a better group of people to be associated with and to be called my family. Their support and prayers were invaluable for my success- Dr. Kola and Mrs. Mary Ajuwon, Mr. Charles and Mrs. Elizabeth Nderitu, Dr. Betty Kanani, Dr. Adedayo Adebanjo, Ayodeji and Susan Adebirigbe, Fola and Bunmi Bolaji, Mercel and Mercy Okezue, Femi and Motun Babatunde, Abidemi Adekoya, Akhere Olenloa, Edeoba Edobor, Paul Oladele, Opeadura Osunbami, Emmanuel Alagbe, Kehinde and Folake Ayano, Rita Appiah and Damilola Attiko. I also want to thank my friends for their encouragement and moral support- Ikechukwu Ozobia, Kenneth Onyedibe, Saheed Osho, Temi Wright, Carine Ajua, Isaiah Mensah, Nuela Enebechi, Alycia Berman, Jumbam Blaise, Luke Brehm, Naa Alabi, Ruth Anyaeche, Adegoke Adetunji, and David Scantlin.

I also want to appreciate the pastors and elders of the Upper Room Christian Fellowship (URCF): Pastor Andy and Mrs. Shalyn Byrn, Pastor Del and Mrs. Marilyn Broersma, Dr. Don Blake, Dr. Nelson and Mrs. Janet Moore, Mr. Kevin and Mrs. Ramona Kuhn, and Mr. John and Mrs. Nancy Mandeville who were all helpful and prayed for my success. Special thanks to Dr. Don Blake who was always praying with me and mentoring me. I also thank all the men of URCF for their love and support.

I thank Victoria Hedrick, Rodrigo Mohallem and Jackeline Franco Marmolejo from the Purdue Proteomics Facility, for their technical assistance in LC-MS/MS analysis, and Saravana

Kumar for help with RT-qPCR data analysis. Special thanks to the Biotechnology Innovation and Regulatory Science (BIRS) Center for providing funds for instrument time, and reagents and supplies. I also thank the department of Agricultural and Biological Engineering for providing me with TA funding.

Phage Ochi17 was isolated in 2017 as part of the Howard Hughes Medical Institute Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science program at Purdue University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Bacteriophages (phages) have broad applications in diverse areas including phage therapy, agriculture, food safety, and environmental protection. In order to fully realize the potential for phage applications, it is critical to understand phage-bacteria interactions and characterize bacterial responses/targets to phage infection. Previous studies have largely focused on other classes of phages other than mycobacteriophages. This research provides the first global proteome investigation of the dynamic relationship between a mycobacteriophage and a mycobacterial host. Mycobacteriophages are viruses that infect mycobacteria. They have been reported to have vital potential uses in various fields, especially as an alternative in the prevention and treatment of mycobacterial diseases such as tuberculosis. Despite their potential, not much is known about the molecular interaction with mycobacteria during a mycobacteriophage infection, especially at the translational level. To better understand this, a novel mycobacteriophage, Ochi17 was first isolated and characterized based on the genome and structure. I then applied label-free quantitative proteomics using the model host, *Mycobacteria smegmatis*, which was infected with Ochi17 at different infection time points. Phage Ochi17 was found to be a temperate phage and classified as a Siphoviridae. The proteome changes occurring at the mid-lytic stage of Phage Ochi17 infection was first examined followed by a temporal study of the global changes. More than 2,000 *M. smegmatis* proteins and at least 50 Ochi17 proteins were identified across all time points. Homologous recombination and host macromolecular synthetic processes were significantly upregulated, while lipid metabolism was significantly downregulated. The results suggested that Ochi17 suppressed the growth of Mycobacterium smegmatis not just by utilizing the macromolecular synthesis of the host, but also by suppressing host transcription, and fatty acid biosynthesis, in addition to the degradation of fatty acids irrespective of infection time. The two-component system was a target at only 24 h post infection. I also showed that phage Ochi17 proteome expression is time-dependent and the proteins typically cluster based on functional relatedness. The results presented here may contribute in the development of mycobacteriophages as antimicrobial therapies that can overcome various defense strategies employed by host mycobacteria.

1. INTRODUCTION

1.1 The challenge of antimicrobial resistance

Over time, bacteria have developed antibiotic resistance genes to common antibiotics such as tetracycline, chloramphenicol, aminoglycosides, and β-lactams, largely as a result of indiscriminate use of antibiotics[1,2]. Acquisition of these resistance genes constitutes a significant menace to the treatment of diseases and has become a serious public health concern[1]. The Center for Disease Control and World health Organization have declared antibiotic resistance as a threat to global health[3,4] with as much as two million illnesses and at least 23, 000 deaths occurring yearly, thus costing the United States $55 billion annually[5]. About 700, 000 annual deaths are recorded globally as a result of bacteria resistance infections and has been projected to cost $100 trillion and account for 10 million deaths by 2050[5]. The problem has been compounded due to the refusal of pharmaceutical companies to invest in research and development of novel compounds as a result of the rate at which bacteria evolve resistance to antibiotics[6]. In 2016, the United Nations General Assembly called the problem of antibiotic resistance, 'the greatest and most urgent global risk'[7]. For instance, mycobacterium species such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*, known to cause serious diseases such as human tuberculosis and leprosy respectively have been reported to be resistant to antibiotics and will require surgery to treat[8,9]. *Mycobacterium tuberculosis*, the world's deadliest human pathogen is the causative agent of Tuberculosis (TB). TB is one of the deadliest infectious diseases in the world, causing 10 million people to fall ill and killing about 1.6 million people yearly[10]. The use of drugs in the treatment of TB has become challenging and ineffective as a result of the emergence of multidrug-resistant and extensively drug-resistant TB[11].

1.1.1 Potential solution

In the search for alternative strategies to control and treat bacterial infections and diseases, there have been renewed interests in the application of phages to treat these diseases. Host specificity, self-amplification, biofilm degradation, and little or no adverse effects on humans are some of the major advantages of phage therapy[12,13]. With the recent developments of analytic and molecular tools capable of studying small biological entities, such as polymerase chain reaction,

next generation sequencing, electron microscopy, microarray, yeast two hybrid, and mass spectrometry, the study of phages have been advancing astronomically with researchers looking into the molecular aspect of phage biology to fully understand its mechanism of actions.

1.2 Bacteriophages

Bacteriophages (or phages) are viruses that infect bacteria and are regarded as the oldest viruses on earth[14-15]. They are simple non-living biological entities consisting of DNA or RNA enclosed within a protein capsid and rely on a bacterial host for survival[6]. The phage capsid head is attached a tail containing fibers which it uses to attach to receptors on host bacterial surfaces[16]. They are ubiquitous and can be found in places such as sewage, oceans, agricultural and forest soil samples[17].

Phages were initially discovered in 1915 by Frederick Twort, and then in 1917, Felix d'Herelle gave a clear description of the phage phenomenon[18,19]. D'Herelle further extended his work with phages with the introduction of phages as therapeutic agents by testing them with avian typhosis (*Salmonella gallinarum*) and with *Shigella dysenteriae* infection of rabbits[20]. They were later applied as a therapy to wound recovery due to the accessibility of the infection[18]. However, with the emergence of antibiotics in the mid-20th century and as a result of World War II, antibiotics became widely preferred over phages due to their ease of production, the relatively broad spectrum of action, and the stability of their preparations. As a result, phage research as a therapeutic agent was largely abandoned[18]. Instead, they were used as molecular tools to understand the basic principle of molecular biology and several biotechnological applications such as phage display, gene therapy, production of recombinant antibodies, and drug delivery[2, 21]. Apart from their use as therapeutic agents, phages also have potential uses in disease prevention (as phage vaccine), biocontrol of plant pathogens and bioremediation[22].

1.2.1 Structural classification of phages

Phages have generally been classified into 10 families based on the type of nucleic acid and virion morphology by the International Committee on Taxonomy of Viruses (ICTV). They include Myoviridae, Siphoviridae, Podoviridae, Tectiviridae, Corticoviridae, Plasmaviridae, Microviridae, Inoviridae, Leviviridae, Cystoviridae[23]. Most phages belong to the Siphoviridae

family, having icosahedral capsid and noncontractile tails. Others belong to Myoviridae, which consists of myoviruses with isometric head and contractile tail or Podoviridae, which are viruses with short noncontractile tails[24] (Fig. 1.1).

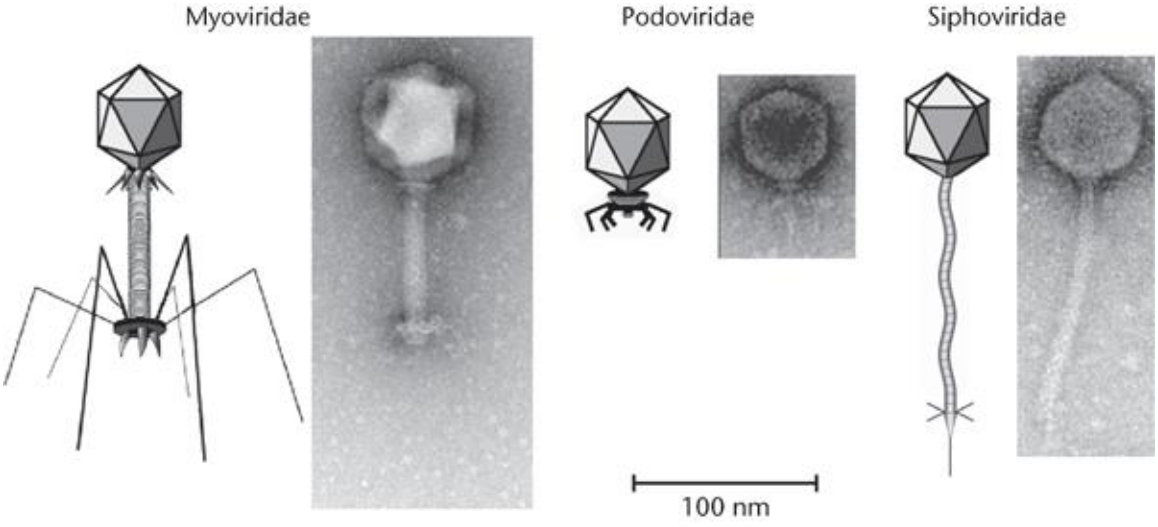


Figure 1.1. The three commonest phage families- myoviridae, podoviridae, and siphoviridae. Adapted from Harper et al. (2014)[25].

1.2.2 Phage life cycle

Phages can only survive based on their potential to infect bacterial hosts upon entry into the cells[26]. Phage interaction with its bacterial host begins with binding to specific receptors on the bacterial host cell surface. These cell surface receptors, such as proteins, polysaccharide, or lipopolysaccharide are specific in nature and must be accessible and spatially distributed[27,28]. This initial interaction with the host cell surface receptors is a fundamental factor in determining the successful infection of the host bacteria[27,29]. Following phage attachment to these surface receptors, penetration of the bacterial envelope is initiated, and then hydrolytic proteins are released to locally digest the cell wall murein/peptidoglycan layer of the bacteria[30-32]. The genetic materials are injected into the host cell, followed by either integration of the genetic material into the host genome and reproduction vertically from mother to daughter cell (lysogenic life cycle), or hijack the bacterial replication machinery to give rise to the next generation of phage progeny and lyse the cell (lytic life cycle) (Fig. 1.2). Once a critical mass of phage progeny is reached, the lytic proteins become active and hydrolyze the peptidoglycan cell wall, releasing novel phage to begin the lytic cycle again[33,34]. Phages that undergo the lysogenic life cycle are called temperate phages, while phages that undergo the lytic life cycle are called lytic phages[27]. Other less characterized life

cycles are the pseudolysogeny and the carrier state[35]. The pseudolysogeny state has been majorly defined as a state where the phage neither integrates into the host genome nor enters the lytic cycle. This could be as a result of conditions of cell starvation, and the phage could end up entering the lytic or lysogenic life cycles[36]. The carrier state occurs when a phage establishes a chronic and persistent infection of bacteria and neither integrates into the host genome nor induces lysis. The resulting progenies are instead budded off the cell or passed down to daughter cells assymetrically after division[37]. This state is induced in rich nutrient conditions and can persist during exponential growth[35].

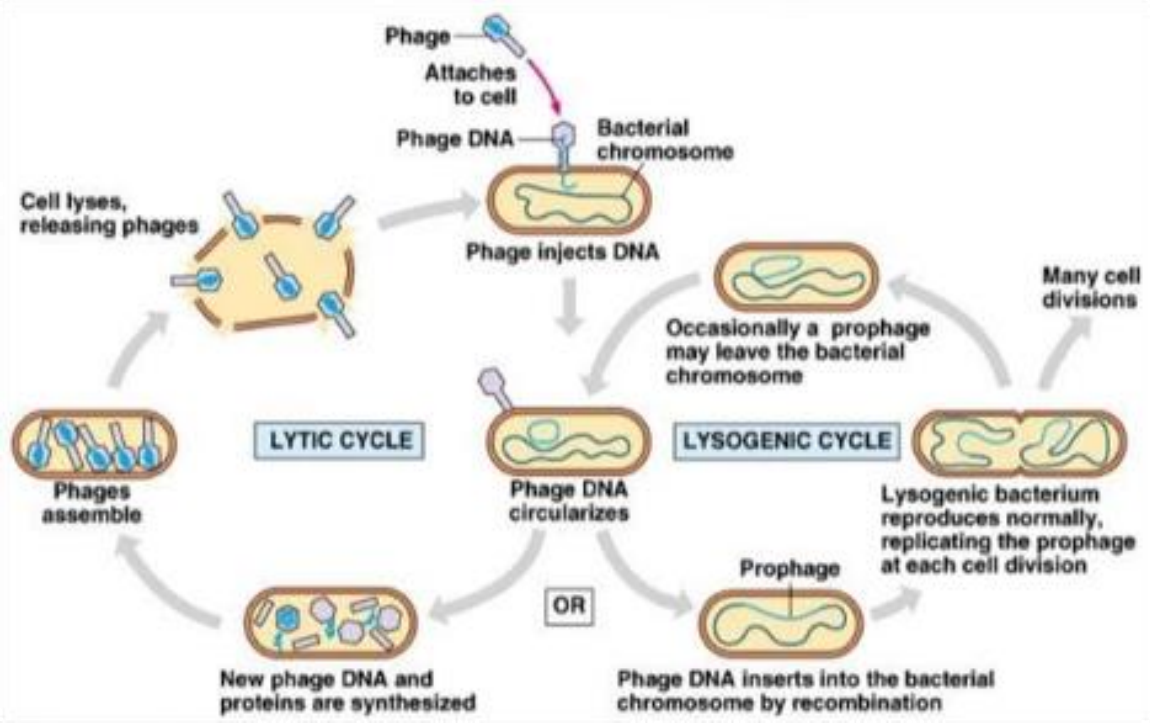


Figure 1.2. A typical phage life cycle in a host bacterium (Adapted from https://www.slideshare.net/suganyakunju/bacteriophages-71259201)

1.3 Bacteriophage-bacteria interactions

Phages coexist with their bacterial hosts towards a limitless co-evolutionary equilibrium in natural and man-made environments[27]. They have been reported to be ten times more abundant than bacteria in any particular environment[38]. This inflicts a tremendous selection pressure on bacteria, which eventually results in the lysis of the bacterial cells and phage multiplication[39]. There have been renewed interests in understanding how phages and bacteria coexist due to the potential of phages to serve as therapeutic agents in medicine and veterinary practices, and also in

applications ranging from food bio-preservation, wastewater treatment, and disease diagnosis[40]. Understanding how phages attack and how bacteria defend themselves against these phage attacks will help improve phage applications in these different fields.

1.3.1 Bacteria defense strategies

In reaction to invading phages, bacteria have evolved various antiphage mechanisms for survival[27]. These mechanisms include the inhibition of phage attachment to cell receptors, blockage of phage DNA entry, cutting of the invading phage genome, and abortive infection systems[27,41,42] (Fig. 1.3). These bacteria defense mechanisms constitute a major disadvantage to phage therapy as they reduce the ability of phages to attack their hosts successfully[41]. In limiting or blocking phage attachment to receptors, bacteria generally adjust the structure or availability of their surface receptors by mutating or masking the receptors, synthesis of extracellular matrix for masking receptors and taking advantage of competitive receptor inhibitors[41]. If phage attachment to receptors is successful, bacteria then use phage-encoded superinfection exclusion (Sie) systems to block injection of the phage DNA. Following a successful entry of phage DNA into a bacterium, the bacterium typically protects itself against invading DNA using various innate nucleic acid degrading systems such as restriction-modification (R-M) systems, and Clustered, regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems. These systems are used to destroy the phage DNA and prevent its replication and release[28,41,42]. R-M systems consist of a methyltransferase (MTase), which methylates the bacteria DNA at specific recognition sites, and a restriction endonuclease (REase), which recognizes the unmethylated foreign DNA and cleaves it[43]. The CRISPR-Cas systems are innate immunity systems that targets foreign nucleic acids such as phages and other mobile genetic elements. The CRISPR arrays integrate short sequences from foreign nucleic acids to remember past infections[28]. The CRISPR array transcripts, also known as precursor crRNAs (pre-crRNAs), produce small CRISPR RNAs (crRNAs) which bind to complementary sequences (or protospacers) in the foreign genome. This eventually results in the degradation of the invading genome[28]. Bacteria can also use self-destructive systems to prevent phage infection and multiplication[41]. These systems also lead to the death of the host bacterium. One such system is the abortive infection (Abi) system that are encoded by mobile genetic elements, and targets the process of phage replication, transcription or translation[27,41]. The Toxin-Antitoxin (TA) system is another self-abortive system that encodes a

toxin gene and a preceding antitoxin gene both transcribed from a single promoter[44,45]. Like Abi systems, they target cellular processes such as replication and translation, and also cytoskeletal or cell wall formation[28].

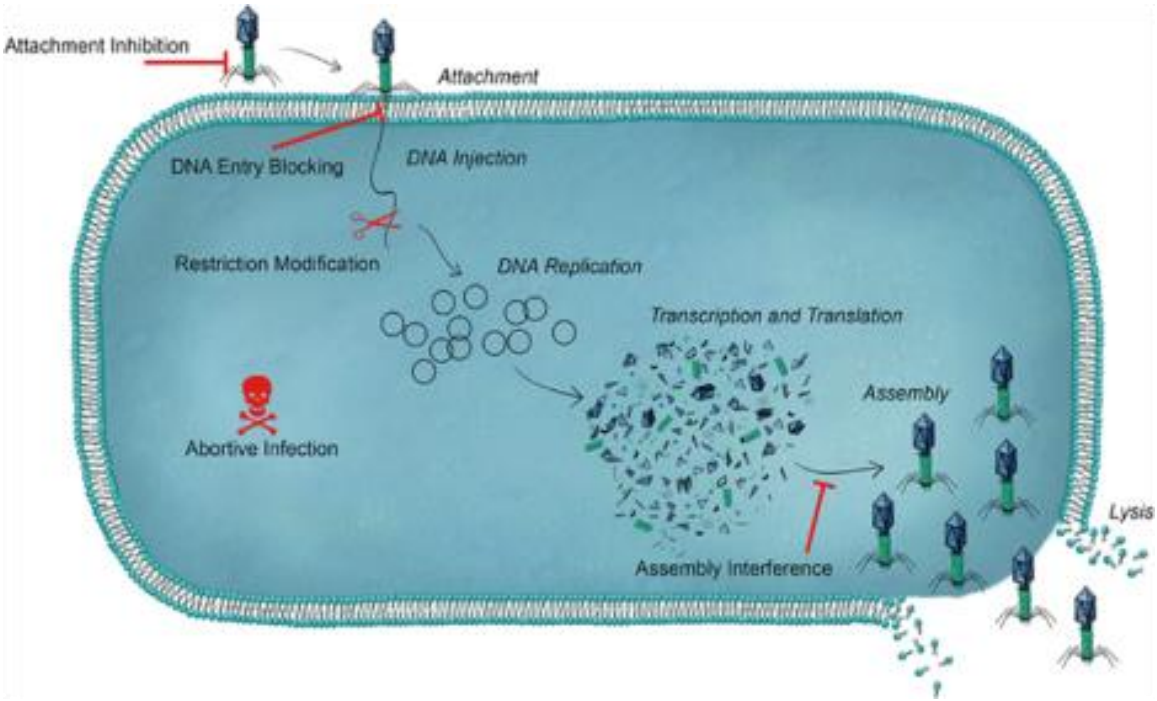


Figure 1.3. Bacterial defense systems against phage attacks starting from the interaction at the cell wall. Red arrows denote inhibition[44]

1.3.2 Phage counter-attack strategies

Through co-evolution, phages on their part, have developed strategies to evade these bacterial defense mechanisms and infect them[27]. Phage counter adaptive mechanisms include point mutations in specific genes, genome rearrangements, and acquisition of novel traits via genomic exchange with other viral or microbial genomes[27]. In response to bacteria resistance at the cell surface level, phages have been reported to modify their receptor binding proteins (RBPs) by acquiring mutations in the genes encoding the RBP or tail fibers in order to recognize a mutated bacteria surface receptor[27,46,47]. Phages can also hydrolyze any barrier like capsule or other exopolysaccharide (EPS) that the bacterial host might want to use to mask the surface receptors. They do this by expressing a depolymerase or acquiring an EPS-degrading enzyme by genetic transfer. Additionally, phages can mutate their RBPs to allow interaction with a cell surface receptor that is expressed stochastically through phase variation or physiological regulation[27].

Phages can also escape bacterial hosts R-M and CRISPR-Cas systems by using various antirestriction and mutation strategies. For example, the host REase is unable to target some phages because their restriction sites are few, too far apart, in unfavorable orientation, or are masked by some phage proteins[48-51]. Also, the phage genome can be modified by MTase expressed by the host or produced by the phage, thereby protecting it against the specific host REase during infection. Phages could also express proteins that mimic the target DNA and shut off the restriction enzyme or that activate the activity of the MTase to promote protection of the phage DNA[52-54]. The bacteria CRISPR-Cas system can be evaded by phages via a single-nucleotide substitution in the protospacer region or in the conserved protospacer-adjacent motif. Phage evasion can also take place when the protospacers and/or PAM sequences are deleted in the phage genome[55]. Phages that attack *Pseudomonas aeruginosa* have been found to possess anti-CRISPR proteins that prevents the interference of CRISPR-Cas system by blocking the formation or action of the complex[56]. *Vibro cholerae* phages encode functional innate CRISPR-Cas systems that are expressed once the phage genome gets into the *V. cholerae* cell, and target the antiphage system[57,58]. One of the last resorts in bacteria defense against phages is abortive infection via toxin-antitoxin systems. Phages have also evolved mechanisms to overcome this bacteria strategy. For instance, coliphage T4 encodes its own antitoxin protein Dmd that functionally replaces the host antitoxin to counter the toxin activity and prevent cell abortion[59]. Also, *Pectobacterium atrosepticum* phage ϕTE expresses a pseudo-antitoxin RNA or hijacks the native antitoxin ToxI to neutralize the toxin ToxN during infection of the host[60].

1.4 Global multi-omics study of the phage-bacteria molecular arms race

Several studies like the ones above have looked at phage and bacteria interaction mechanisms at the individual protein or functional level under controlled laboratory conditions with meaningful insights into phage and bacteria response mechanisms[27,41]. However, in order to have a full understanding of the phage-bacteria dynamics, it is pertinent that a global experimental approach be taken to study the comprehensive interaction dynamics between these two organisms. This in turn can be of immense value not just in the re-emerging field of phage therapy but also in the food and biotechnology industries, which depend on phage-resistant bacteria to engineer peculiar products and also in applying phages as biocontrol agents[27].

Phage infection of host bacteria first targets the host's DNA transcription, thereby causing modifications in mRNA and protein expression[61,62]. The different transcriptomics studies showed that phage attack disrupted less than 10% of the host bacteria genes. For instance, about 2% *B. subtilis* genes were significantly regulated after infection with phage ϕ29[63]. Less than 3% of *L. lactis* IL1403 genes were significantly affected upon infection with phage c2[64]. About 7.1% of the host *P. aeruginosa* genes significantly regulated when infected by the lytic phage, PaP1[65]. In another study using a global proteomics approach, 16% of the host *L. lactis* MG1363 genes was identified[66].

Previous studies on the global impact of the molecular interactions of phage-bacteria interactions on the host bacteria have largely focused on the host transcriptome regulations[67-71]. In a global transcriptome study of *Bacillus subtilis* infected by lytic phage ϕ29, genes associated with nucleic acid metabolism, carbohydrate metabolism, and transport were primarily regulated during early infection stages (8 and 16 min post infection) at a significant level[63]. More proteins involved in carbohydrate metabolism and transport were upregulated than downregulated. Proteins involved in lipid and protein metabolism were exclusively upregulated. Whole-genome microarrays to understand the response of *Lactococcus lactis* IL1403 during early stage infection with lytic phage c2 showed that *L. lactis* IL1403 responded significantly mostly in functions related to cell envelope processes, carbohydrate metabolism, and regulatory functions[64]. Others include starch and sucrose metabolism, lipid metabolism, protein and amino acid metabolism, purine, pyrimidine and nucleotide metabolism, DNA replication and repair, transcription, energy metabolism, and metabolism of cofactors and vitamins.

A recent global transcriptomics study of phage φAbp1-infected *Acinetobacter baumannii* showed about 15% of the host's genes affected by the infection at 5, 10 and 20 min after infection[71]. They reported more upregulated differentially expressed genes than downregulated. Upregulated genes represented pathways and functional groups such as oxidation-reduction process and proteolysis involved in stress responses, translation, ribosome and amino acid pathways involved in protein synthesis, and metabolic processes. Downregulated groups and functions include nucleic acid metabolic processes and components such as purine and pyrimidine metabolism, material transport, and host biosynthetic processes[71].

Another transcriptomics study of lytic phage NCTC 12673 infection of a foodborne human pathogen, *Campylobacter jejuni* at 30, 60 and 120 min showed more genes upregulated than

downregulated. Genes involved in DNA synthesis, replication and repair, transcription, and protein synthesis (translation and ribosome pathway), and amino acid metabolism were upregulated. Additionally, genes involved in oxidative stress and iron metabolism, and multi-drug efflux pump CmeABC were reported to be significantly upregulated. Whereas genes associated with energy metabolism (pyruvate, propanoate and butanoate metabolism pathways, and the TCA cycle) were downregulated[72].

Microarray analyses was used to observe *Pseudomonas* responses to phage infection over several time points and reported that there was no major reprogramming of the host in early infection by a double-stranded DNA phage and a single-stranded RNA phage[67,68]. Also observed was that most phage-induced changes occurred after the synthesis of virion components[73]. However, the early genes of a temperate phage, PaP3 had the strongest effect on the host gene expression. Genes involved in amino acid metabolism seemed to be the most affected[70].

Studies on phage-bacteria interactions have largely focused on the host transcriptome regulation[62,73-76], and are unable to capture the direct changes to a cell's functional activities[77]. Consequently, the use of transcript abundance in analyzing phage-bacteria interaction does not give a complete picture of the phage-bacteria interaction since mRNA transcript abundance often partially correlates with protein abundance[78]. Many proteins are post-translationally regulated and such information cannot be gleaned from transcript level analysis. Therefore, there is a need for the application of quantitative proteomics to study phage-bacteria interaction in order to get a comprehensive assessment of the interaction.

There have been some studies that have looked at proteome level research in phage bacteria interactions. However, these studies only focused on elucidating the interactions among host proteins and various phage proteins using tools such as yeast-two hybrid screening[79,80]. Most recently, two global label-free quantitative proteomics studies have been carried out on *Salmonella* Typhimurium and *Lactococcus lactis* MG1363 infected with a giant phage SPN3US and phage p2 respectively[66,81]. *Salmonella* infection with giant phage SPN3US resulted in the identification of proteins involved in host macromolecular synthesis, such as DNA replication, transcription and translation proteins. They include 30S ribosomal proteins S1, S2, S3 and S4, 50S ribosomal proteins L1, L3, L5 and L9, DNA polymerases I and III, RNA polymerases B (*rpoB*) and C (*rpoC*), and major outer membrane proteins A (*OmpA*), C (*OmpC*), D (*OmpD*) and F (*OmpF*). Overtime, with the exception of 30S ribosomal protein S3, the ribosomal and chaperone proteins required for

translation increased in abundance, especially towards the late stage of infection (60 min). This is justified since phages depend on the host translation machinery for propagation, and as infection time increases, so does the amount of phages[81]. However, DNA replication and transcription proteins decreased as time of infection increased. Phages such as SPN3US encode their own DNA and RNA polymerases, and so do not rely on the host polymerases for propagation. Thus, accounting for the decrease in abundance over time[81].

Host *Lactococcus lactis* MG1363 proteome response during a time-course infection with phage p2 was characterized[66]. The authors reported exclusive representation of a little over 30% of identified proteins involved in transport and binding, regulatory and cell envelope functions in the infected host. It was reported that after 10 min of infection, cell envelope and signal transduction proteins, as well as proteins involved in pyrimidine ribonucleotide biosynthesis had increased protein concentrations with respect to the uninfected control. After 20 min, protein concentrations in restriction-modification and DNA degradation pathways decreased. However, a number of proteins involved in the synthesis and modification of ribosomal proteins increased within the same period. Interestingly, after infection at 40 min, there was a reduction in protein synthesis as proteins involved were less abundant. Phage p2 suppression of host genes are not functionally related based on proteins identified in their study[66]. Previous studies on the global impact of phage-bacteria molecular interactions on the host bacteria have largely focused on the host transcriptome and metabolome regulations[82-84].

1.5 Problem statement and specific aims

The ability of mycobacteriophages to attack mycobacteria has made them of vital interest to the scientific community, especially as an investigative tool in studying mycobacteria species[85]. There has been some progress in the elucidation of global phage-bacteria interactions from the literatures reviewed above. These studies have focused on phage interactions with host pathogens responsible for food-borne and other forms of human diseases like gastroenteritis, pneumonia, and diarrhea[62,63,69,72-76]. However, none has been reported for the causative agent of the top infectious killer in the world, *Mycobacterium tuberculosis* or any other mycobacterial host. Phages are known to be specific in nature, attacking only a particular bacterium or bacteria from the same class. Therefore, a study to examine the effects of a mycobacteriophage on a mycobacterial host will be necessary to understand and evaluate the effectiveness of proposing mycobacteriophage as a

molecular tool and as a potential alternative to antibiotics in treating tuberculosis. A model organism to study mycobacterial physiology and other relevant processes applicable to the pathogenic mycobacteria species is *Mycobacterium smegmatis* (*M. smegmatis*). It is a gram-positive, acid-fast and fast-growing mycobacterium originally isolated from humans that is mostly non-pathogenic[85,86].

In studying the interactions between *Mycobacterium smegmatis* and a mycobacteriophage, the research was focused on two major objectives:

In objective I, the research was geared towards understanding how *Mycobacterium smegmatis* responds to a mycobacteriophage attack at the molecular level. In order to achieve this objective, a novel mycobacteriophage was first isolated, characterized, and functionally annotated (chapter 2). Afterwards, this isolated phage was then used to infect *Mycobacterium smegmatis* and the proteomic responses of the host was characterized, as well as the phage proteome expression during the infection. Transcript expression of few of the identified proteins were also characterized (chapter 3). This was the first global proteome study of a mycobacterial host during phage infection.

In objective II, following the observations from objective I, the next question to answer became, does the interaction between the isolated mycobacteriophage and *M. smegmatis* change over time? To answer this question, a temporal (time-series) global proteomic study was carried out over 24 h to understand the dynamic proteome expression during their interaction (chapter 4).

In summary, the studies presented here suggested that the mycobacteriophage, Ochi17 predominantly targets fatty acid and lipid metabolism of the host. Ochi17 infection causes little or no change in carrying capacity of the host until 6 h post infection. After 6 h post infection, the carrying capacity of the host is significantly reduced by more than 50% (chapters 3 and 4). While carrying capacity was being reduced, the macromolecular synthesis pathways were being upregulated at the proteome level. This research has revealed important mycobacterial targets for phage Ochi17 and further research will be required to study the direct impact of phage Ochi17 on these targets.

1.6 References

1. Zhang, X.X., Zhang, T. & Fang, H.H. Antibiotic resistance genes in water environment. *Appl Microbiol Biotechnol* **82**, 397-414 (2009).
2. Domingo-Calap, P., Georgel, P. & Bahram, S. Back to the future: bacteriophages as promising therapeutic tools. *HLA* **87**, 133-140 (2016).
3. Centers for Disease Control. Antibiotic resistance threats in the United States. CDC, Atlanta, GA (2015).
4. World Health Organization. Antimicrobial resistance: global report on surveillance. WHO, Geneva, Switzerland (2015).
5. Laxminarayan, R. *et al*. Antibiotic resistance- the need for global solutions. *The Lancet Infectious Disease Commission* **13**, 1057-1098 (2013).
6. Lin, D.M., Koskella, B. & Lin, H.C. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World J Gastrointest Pharmacol Ther* **8**, 162-173 (2017).
7. United Nations. United Nations high-level meeting on antimicrobial resistance. UNGA, New York, USA (2016).
8. Hatfull, G. The secret lives of mycobacteriophages. *Adv Virus Res* **82**, 179-288 (2012).
9. Sassi, M., Bebeacua, C., Drancourt, M. & Cambillau, C. The first structure of a mycobacteriophage, the *Mycobacterium abscessus* subsp. *Bolletii* phage Araucaria. *J Virol* **87**, 8099-8109 (2013).
10. World Health Organization. Global tuberculosis report. *WHO* **15**, 1-261 (2019).
11. Broxmeyer, L. *et al*. Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a mycobacteriophage delivered by a nonvirulent mycobacterium: a model for phage therapy of intracellular bacterial pathogens. *The J Infectious Disease* **186**, 1155-1160 (2002).
12. Bourdin, G. *et al*. Coverage of diarrhoea-associated *Escherichia coli* isolates from different origins with two types of phage cocktails. *Microbiol Biotechnol* **7**, 165-176 (2014).
13. Donlan, R.M. Preventing biofilms of clinically relevant organisms using bacteriophage. *Trends Microbiol* **17**, 66-72 (2009).
14. Clokie, M.R.J., Millard, A.D., Letarov, A.V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31-45 (2011).
15. Summers, W.C. Bateriophage therapy. *Annu Rev Microbiol* **55**, 437-451 (2001).

16. Ackerman, H.W. Tailed bacteriophages: the order caudovirales. *Adv Virus Res* **51**, 135-201 (1998).
17. Gardener, G.M. & Weiser, R.S. A bacteriophage for Mycobacterium smegmatis. *Exp Biol Med* **66**, 205-206 (1947).
18. Summers, W.C. Bacteriophage therapy. *Annu Rev Microbiol* **55**, 437-451 (2001).
19. Abedon, S.T., Kuhl, S.J., Blasdel, B.G. & Kutter, E.M. Phage treatment of human infections. *Bacteriophage* **1**, 66–85 (2011).
20. d'Herelle, F. The bacteriophage and its behavior. *Nature* **118**, 183-185 (1926).
21. Weisberg, R.A. A genetic switch: phage Lambda revisited. Third Edition. By Ptashne, M. *The Quarterly Rev Biol* **79**, 427-428 (2004).
22. Haq, I.U., Chaudhry, W.N., Akhtar, M.N., Andleeb, S. & Qadri, I. Bacteriophages and their implications on future biotechnology: a review. *Virol* **9**, 9 (2012).
23. Fauquet, M., Mayo, M.A., Maniloff, J.D., Desselberger, U. & Ball, L.A. Virus taxonomy: VIIIth report on the international committee on taxonomy of viruses. *Elsevier Academic Press* San Diego, CA (2005).
24. Kozloff, L.M., Sundar, R.C.V., Nagaraja, R.R., Chapman, V.A. & DeLong, S. Structure of a transducing mycobacteriophage. *J Virol* **9**, 390 (1972).
25. Harper, D. *et al*. Bacteriophages and biofilms. *Antibiotics* **3**, 270-284 (2014)
26. Chaturongakul, S. & Ounjal, P. Phage-host interplay: examples from tailed phages and gram-negative bacterial pathogens. Front Microbiol **5**, 442, doi: 10.3389/fmicb.2014.00442 (2014).
27. Samson, J.E., Magadan, A.H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defenses. *Nat Rev Microbiol* **10**, 675-687 (2013).
28. Dy, R.L., Ritcher, C., Salmond, G.P.C. & Fineran, P.C. Remarkable mechanisms in microbes to resist phage infections. *Annual Rev Virol* **1**, 307-331 (2014).
29. Azam, A.H. & Tanji, Y. Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. *Appl Microbiol Biotechnol* **103**, 2121-2131 (2019).
30. Kanamaru, S., *et al*. Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553-557 (2002).

31. Mosig, G., Lin, W., Franklin, J. & Fan, W.H. Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene *e*) and a baseplate-associated (gene *5*) protein. *New Biol* **1**, 171-179 (1989).
32. Moak, M. & Molineux, I.J. Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol Microbiol* **51**, 1169-1183 (2004).
33. Delbrück, M. The growth of bacteriophage and lysis of the host. *J Gen Physiol* 23, 643-660 (1940).
34. Weinbauer, M.G. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**, 127-181 (2004).
35. Diaz-Munoz, S.L. & Koskella, B. Bacteria-phage interactions in natural environments. *Adv Appl Microbiol* **89**, 135-183 (2014).
36. Los, M. & Wegrzyn, G. Pseudolysogeny. *Adv Virus Res* **82**, 339-349 (2012).
37. Cenens, W. *et al*. Expression of a novel P22 ORFan gene reveals the phage carrier state in *Salmonella typhimurium*. *PLoS Genet* **9**, e1003269 (2013).
38. Brussow, H. & Hendrix, R.W. Phage genomics: small is beautiful. *Cell* **108**, 13-16 (2002).
39. Fuhrman, J.A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541-548 (1999).
40. Dedrick, R.M. *et al.* Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus. Nature Medicine* 25, 730-733 (2019).
41. Labrie, S.J., Samson, J.E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Rev Microbiol* **8**, 317-327 (2010).
42. Seed, K. Battling phages: how bacteria defend against viral attack. *PLoS Pathogens* 11, e1004847, doi:10.1371/journal.ppat.1004847 (2015).
43. Tock, M.R. & Dryden, D.T.F. The biology of restriction and anti-restriction. *Curr Opin Microbiol* **8**, 466-472 (2005).
44. Gerdes, K., Christensen, S.K. & Løbner-Olesen, A. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* **3**, 371-382 (2005).
45. Schuster, C.F. & Bertram, R. Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. *FEMS Microbiol Lett* **340**, 73-85 (2013).
46. Michel, A., Clermont, O., Denamur, E. & Tenaillon, O. Bacteriophage PhiX174's ecological niche and the flexibility of its *Escherichia coli* lipopolysaccharide receptor. *Appl Environ Microbiol* **76**, 7310-7313 (2010).

47. Meyer, J.R. *et al.* Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428-432 (2012).
48. Kruger, D.H. & Bickle, T.A. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* 47, 345-360 (1983).
49. Bickle, T.A. & Krüger, D.H. Biology of DNA restriction. *Microbiol Rev* **57**, 434-450 (1993).
50. Krüger, D.H. *et al*. Use of bacterial virus T7 as a tool for the study of DNA methylation. *Gene* **74**, 85-87 (1988).
51. Meisel, A., Bickle, T.A., Kriiger, D.H. & Schroeder, C. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**, 467-469 (1992).
52. Loenen, W.A. & Murray, N.E. Modification enhancement by the restriction alleviation protein (RaI) of bacteriophage λ. *Mol Biol* **190**, 11-22 (1986).
53. Iida, S., Streiff, M.B., Bickle, T.A. & Arber, W. Two DNA antirestriction systems of bacteriophage P1, *dar*A, and *dar*B: characterization of *dar*A- phages. *Virol* **157**, 156-166 (1987).
54. Atanasiu, C., Su, T.J., Sturrock, S.S. & Dryden, D.T.F. Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res* **30**, 3936-3944 (2002).
55. Deveau, H. *et al*. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus. J Bacteriol* **190**, 1390-1400 (2008).
56. Bondy-Denomy, J., Pawluk, A., Maxwell, K.L. & Davidson, A.R. Bacteriophage genes that inactivate the CRISPR?Cas bacterial immune system. *Nature* **493**, 429-432 (2013).
57. Seed, K.D., Lazinski, D.W., Calderwood, S.B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489-491 (2013).
58. Villion, M. & Moineau, S. Virology: phages hijack a host's defence. *Nature* **494**, 433-434 (2013).
59. Otsuka, Y. & Yonesaki, T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Mol Microbiol* **83**, 669-681 (2012).
60. Blower, T.R., Evans, T.J., Przybilski, R., Fineran, P.C. & Salmond, G.P.C. Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. *PLoS Genet* **8**, e1003023, doi:10.1371/journal.pgen.1003023 (2012).

61. Nechaev, S. & Severinov, K. Bacteriophage-induced modifications of host RNA polymerase. *Annu Rev Microbiol* **57**, 301-322 (2003).
62. Zhao, X. *et al.* Global transcriptomic analysis of interactions between *Pseudomonas aeruginosa* and bacteriophage PaP3. *Sci Reports* **6**, 1-12 (2016).
63. Mojardin, L., & Salas, M. Global transcriptional analysis of virus-host interactions between phage ϕ29 and *Bacillis subtilis*. *J Virol* **90**, 9293-9304 (2016).
64. Fallico, V., Ross, R.P., Fitzgerald, G.F. & McAuliffe, O. Genetic response to bacteriophage infection in *Lactococcus lactis* reveals a four-strand approach involving induction of membrane stress proteins, d-alanylation of the cell wall, maintenance of proton motive force, and energy conservation. *J Virol* **85**, 12032-12042 (2011).
65. Zhao, X. *et al*. Transcriptomics and metabolomics profiling of phage-host interactions between phage PaP1 and *Pseudomonas aeruginosa*. *Front Microbiol* **8**, 548, doi: 10.3389/fmicb.2017.00548 (2017).
66. Lemay, M-L. *et al*. Investigating Lactococcus lactis MG1363 response to phage p2 infection at the proteome level. *Mol Cel Prot* **18**, 704-714 (2019).
67. Poranen, M.M. *et al*. Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* **80**, 8081–8088 (2007).
68. Ravantti, J.J., Ruokoranta, T.M., Alapuranen, A.M. & Bamford D.H. Global transcriptional responses of *Pseudomonas aeruginosa* to phage PRR1 infection. *J Virol* **82**, 2324–2329 (2008).
69. Leskinen, K, Blasdel, B.G., Lavigne, R. & Skurnik, M. RNA-sequencing reveals the progression of phage-host interactions between ωR1-37 and Yersinia enterocolitica. *Viruses* **8**, 111, doi:10.3390/v8040111 (2016).
70. Zhao, X. *et al*. Global transcriptomic analysis of interactions between *Pseudomonas aeruginosa* and bacteriophage PaP3. *Sci Reports* **6**, 19237, doi:10.1038/srep19237 (2016).
71. Yang, Z. *et al*. Global transcriptomic analysis of the interactions between phage φAbp1 and extensively drug-resistant *Acinetobacter baumannii*. *mSystems* **4**, e00068-19 (2019).
72. Sacher, J.C. *et al.* Transcriptomic analysis of the *Campylobacter jejuni* response to T4-like phage NCTC 12673 infection. *Viruses* **10**, 332, doi:10.3390/v10060332 (2018).
73. Poranen, M.M. *et al*. Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* **80**, 8081-8088 (2006).

74. Ravantti, J.J., Ruokoranta, T.M., Alapuranen, A.M. & Bamford, D.H. Global transcriptional responses of Pseudomonas aeruginosa to phage PRR1 infection. *J Virol* **82**, 2324–2329 (2008).
75. Leskinen, K., Blasdel, B.G., Lavigne, R. & Skurnik, M. RNA-sequencing reveals the progression of phage-host interactions between ωR1-37 and Yersinia enterocolitica. *Viruses* **8**, 111, doi:/10.3390/v8040111 (2016).
76. Yang, Z. *et al.* Global transcriptomic analysis of the interactions between phage φAbp1 and extensively drug-resistant *Acinetobacter baumannii*. *mSystems* **4**, e00068-19 (2019).
77. Schwanhäusser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
78. Vogel, C. & Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**, 227-232 (2012).
79. Mariano, R., Wuchty, S., Vizoso-Pinto, M.G., Hauser, R. & Uetz, P. The interactome of *Streptococcus pneumoniae* and its bacteriophages show highly specific patterns of interactions among bacteria and their phages. *Sci Rep* **6**, 24597, doi: 10.1038/srep24597 (2016).
80. Mehla, J. *et al*. Virus-host protein-protein interactions of mycobacteriophage Giles. *Sci Rep* **7**, 16514, doi: 10.1038/s41598-017-16303-7 (2017).
81. Weintraub, S.T. *et al*. Global proteomic profiling of *Salmonella* infection by a giant phage. *J Virol* **93**, e01833-18, doi: 10.1128/JVI.01833-18 (2019).
82. Ankrah, N.Y.D. *et al*. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *The ISME J* **8**, 1089-1100 (2014).
83. Chevallereau, A. *et al.* Next-Generation "-omics" approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa. PLoS Genet* **12**, e1006134, doi: 10.1371/journal.pgen.1006134 (2016).
84. De Smet, J. *et al*. High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *ISME J* **10**, 1823-1835 (2016).
85. Fu, X., Ding, M., Zhang, N. & Li, J. Mycobacteriophages: an important tool for the diagnosis of *Mycobacterium tuberculosis* (review). *Mol Med Rep* **12**, 13-19 (2015).
86. Hatfull, G. The secret lives of mycobacteriophages. *Adv Virus Res* **82**, 179-288 (2012).

2. ISOLATION AND CHARACTERIZATION OF THE NOVEL MYCOBATERIOPHAGE, OCHI17

2.1 Abstract

Mycobacteriophages are ubiquitous viruses that infect mycobacteria. They have been reported to have vital potential uses in the field of biotechnology, molecular biology and medical science, especially as an alternative in the prevention and treatment of tuberculosis. Meanwhile, only a meager number of mycobacteriophages have been identified and characterized out of the multitudes present in the biosphere. In this study, as part of the Howard Hughes Medical Institute Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (HHMI SEA- PHAGES), a novel mycobacteriophage that infects *Mycobacterium smegmatis* was isolated directly from the soil, purified and then amplified using the plaque assay to get a single clonal population. The structure and genome of the isolated phage was then characterized using Transmission Electron Microscopy (TEM) at Purdue University and Illumina sequencing at the Pittsburgh Bacteriophage Institute. TEM revealed that the mycobacteriophage, named Ochi17 has a capsid head (56.67 nm) and a non-contractile tail that is 200 nm in length. Based on this structure, it was classified into the Siphoviridae morphotype. The genome sequencing showed that it has a genome size of 58kbp and a high GC content of 61%. The identification and characterization of mycobacteriophage, Ochi17 is a major step towards advancing mycobacteriophages as potential solution in combating antibiotic resistance in mycobacteria.

2.2 Introduction

Bacteriophages (or phages) are viruses that infect bacteria[1]. They can be found everywhere and are the most abundant organisms in the biosphere[2]. William Twort initially discovered them in 1915 and by 1917 Felix d'Herelle discovered their ability to kill bacteria[1]. Mycobacteriophage, the phage that infects mycobacteria was initially isolated in 1946 from soil and leaf mold samples[3]. Most mycobacteriophages have capsids with double stranded DNA and noncontractile tails belonging to the Siphoviridae family. Others belong to Myoviridae, which consists of myoviruses with isometric head and contractile tail or Podoviridae that comprises viruses with short noncontractile tails[4].

Mycobacteria such as *Mycobacterium tuberculosis* and *Mycobacterium leprae* are known to cause a wide range of infections such as respiratory tract diseases and some have been reported to be resistant to antibiotics and will only require surgery to cure[5,6]. The ability of mycobacteriophages to attack mycobacterial hosts and the potential uses in the field of biotechnology and medical science with applications ranging from disease diagnosis, through phage typing, phage vaccine and phage therapy has made them of vital interest to the scientific community[7]. The abundance of mycobacteriophages in the biosphere makes it possible to explore the wide-ranging usefulness they can be to human beings[7]. However, only a meager number of mycobacteriophage have been identified and characterized out of the multitudes present[8].

In this study, the isolation and characterization of the genome and structure of a novel mycobacteriophage that infects *Mycobacterium smegmatis* was described. *M. smegmatis* was used as a mycobacteria model because it is non-pathogenic and relatively fast-growing, compared to the pathogenic *M. tuberculosis*, which takes 24 h to double, and *M. leprae* that cannot be readily cultured in controlled lab conditions[5,9]. The characterized mycobacteriophage has capsid size of 56.67 nm and a non-contractile tail length of 200 nm. It was classified into the Siphoviridae morphotype based on the structure. Therefore, this study is a step closer to increasing the number of identified and characterized mycobacteriophages known for their specific host range.

2.3 Results

2.3.1 Phage isolation and purification.

After more than 24 h incubation at 37$_{o}$C, plaques were observed on the agar plates (Fig. 2.1). The presence of plaques on the agar plate is an evidence of a successful phage isolation and three rounds of phage purification. The plaques are circular and clear and having no regular margin. They measure about 2-3 mm in diameter. Due to the plaques having cloudy area around the boundary, Ochi17 was classified as a temperate phage (Fig. 2.1). This information can be found on the actinobacteriophage database (http://phagesdb.org/phages/Ochi17/). The phage was named ‘Ochi17’- which is a combination of the first part of my middle name (Ochiagha) and the year it was isolated (2017).

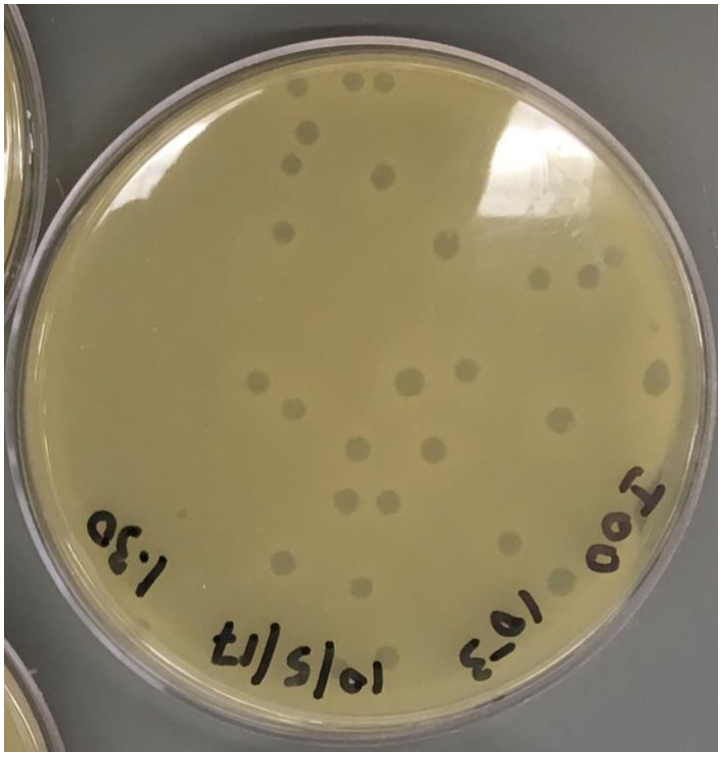


Figure 2.1. Ochi17 plaque assay on LB agar plate shown upside down after isolation and three rounds of purification.

2.3.2 Structural characterization of phage Ochi17.

Transmission electron microscopy (TEM) observation of Ochi17 revealed a long noncontractile flexible tail of length 200 nm and an icosahedral head (capsid) of diameter 56.67 nm (Fig. 2.2). This is a typical feature of the morphotype Siphoviridae. The collar was not visible neither was the tail striation. The phage belongs to the order Caudovirales.

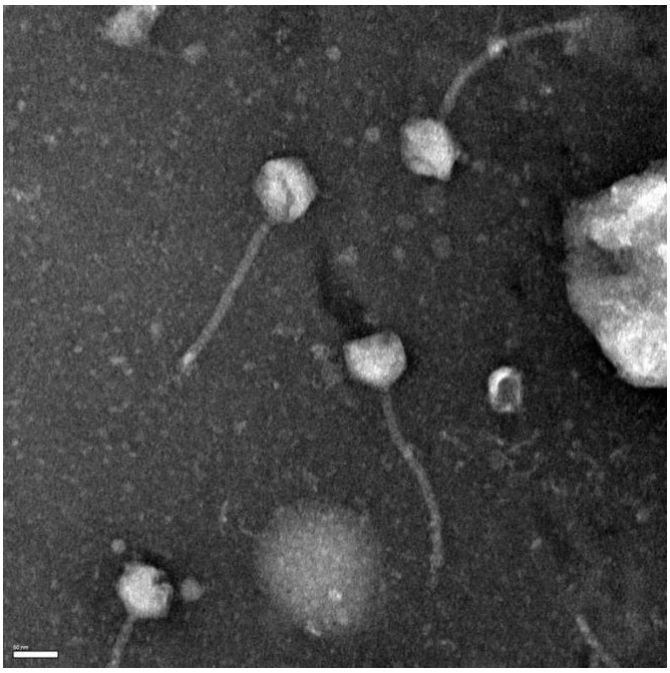


Figure 2.2. TEM Structure of the isolated phage as seen with a FEI Tecnai G2 20 TEM equipped with LaB6 filament.

2.3.3 Phage Ochi17 genome sequencing and characterization.

The extracted genomic DNA had a high concentration of 289.2 ng/ul. The DNA was digested with six restriction enzymes, analyzed by agarose gel electrophoresis and then sequenced using Illumina sequencing. Genome sequencing showed the genome to be 58 kbp long and has GC content of 61% (Fig. 2.3). Functional annotation of the genome using DNAMaster and PECCAN[10] showed the genome is made up of 110 genes (Fig. 2.3). Of these, 67 of them were without any annotated functions (Fig. 2.3). The first 23 genes towards the 5' region are all structural proteins, while the rest of the genome (especially towards the 3' end) is made up of either host interaction genes or genes involved in DNA metabolism and replication. The host lysis genes (genes 28-30) are located just after the structural genes (Fig. 2.3). Gene 43, annotated as an integrase- a unique gene in temperate phages, confirmed Ochi17 as a temperate phage (Fig. 2.3). As a result of having about 50% sequence similarity with other mycobacteriophages in cluster F, Ochi17 was also grouped as a cluster F phage[5].

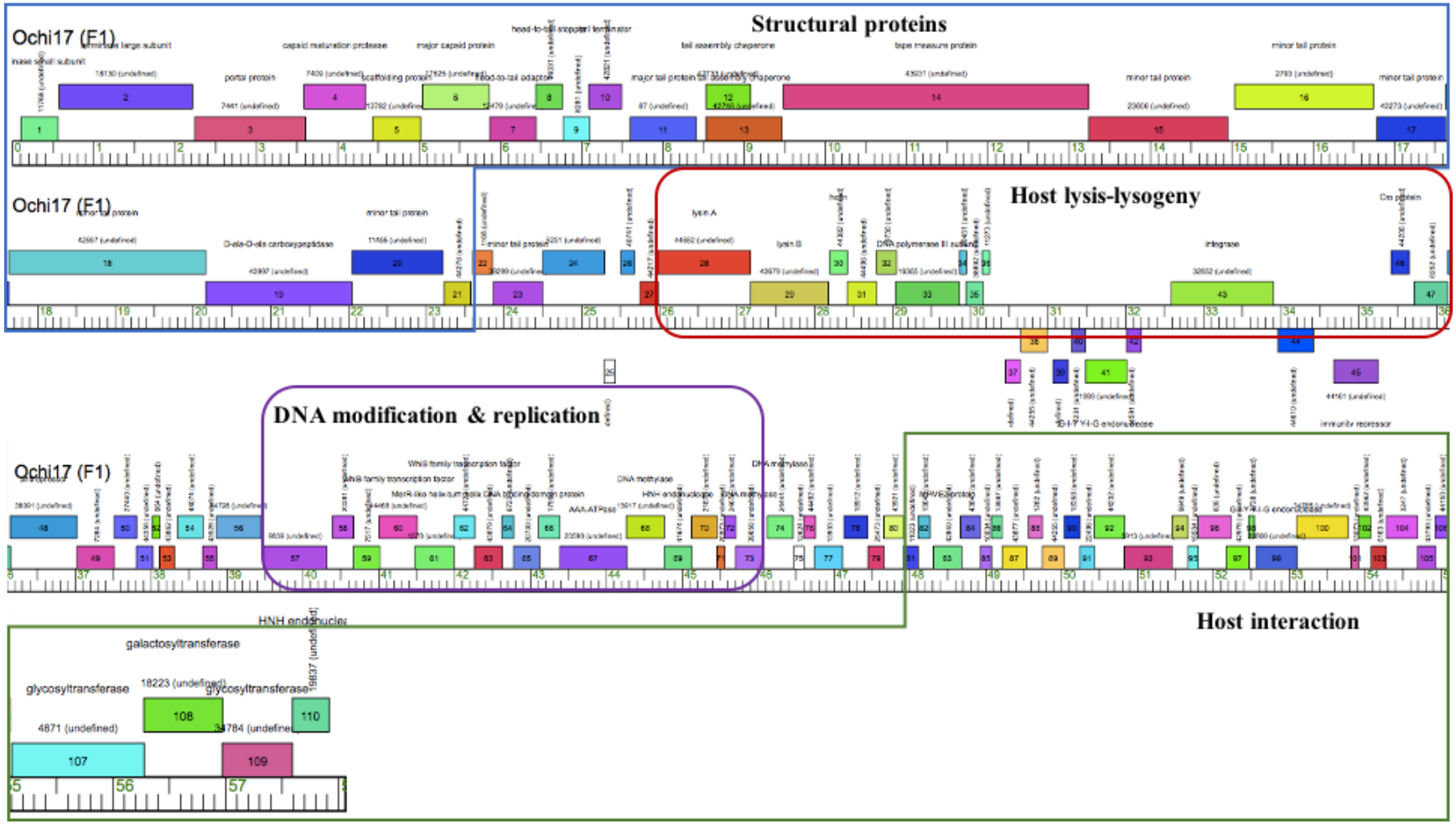


Figure 2.3. Phage Ochi17 genome organization showing functionally annotated genes. The genome map was constructed using Phamerator[11].

2.4 Discussion

The isolation, identification and characterization of the novel mycobacteriophage Ochi17 has further increased the genetic diversity of available mycobacteriophages in the Actinobacteriophage database. The database showed that there are about 182 phages in Cluster F, belonging to five sub-clusters (F1 to F5). Ochi17 belongs to sub-cluster F1, including 169 other mycobacteriophages[12]. Nucleotide BLAST search of closely related phages showed the most aligned F1 phages as Modragons, Seabastian, and Llama[13]. Phamerator comparison of their genome organizations showed a very conserved sub-cluster F1 phages in terms of gene composition and genome length[11]. They all encode similar structural proteins, host lysis-lysogeny proteins, DNA modification and replication proteins, and host interaction proteins as Ochi17, as well as having similar genome arrangements (Fig. 2.3).

Ochi17 was classified as a temperate phage based on the appearance of the formed plaques, and also the presence of an integrase gene in its genome (Fig. 2.1, 2.3). Cluster F phages form turbid plaques and can undergo lysogeny[5]. Temperate phages typically undergo the lysogenic life cycle and only lyse the cell when conditions are favorable. They are preferred over lytic phages for phage therapy because the latter could lead to the release of bacterial endotoxins[14]. Thus, Ochi17 could be a potential candidate for phage therapy.

Genes of similar functional class were clustered together. Structural proteins were clustered towards the 5' end. In the middle were host lysis-lysogeny related proteins. Towards the 3' end are proteins involved in DNA modification/replication and host interaction (Fig. 3). This function-based organization of the genes could translate to functional-based expression of the genes at specific periods during infection of the host. For instance, the immunity repressor gene, which maintains lysogeny by downregulating lytic promoters and conferring superinfection immunity[15], and the Cro repressor gene that promotes transition from lysogenic to lytic phase by promoting the de-repression of the lytic promoters from c1[16] are located side-by-side. They could both be expressed at different infection periods due to their opposing functions. The antirepressor gene which plays similar role as Cro repressor gene but inhibits the action of gene c2[17] might be expressed together with Cro repressor. Further studies looking at the temporal expression of these Ochi17 proteins would be necessary to answer this question (see chapter 4).

The addition of the novel mycobacteriophage, Ochi17 to the actinobacteriophages database is a step closer to identifying as many phages as possible so as to explore their wide

host range and genetic diversity[5,6]. Beside the fact that this is a novel phage, this study has shown the host specificity of another known phage. Future studies will look at the patterns in which this phage genes express themselves during infection of *M. smegmatis* as well as the response of the host during the infection period.

2.5 Materials and Methods

2.5.1 Environmental sample collection.

Following protocol described[18], soil sample was collected in front of the Purdue Horticultural garden (40°25'17"N, 86°54'50"W) at 10 cm depth and put into 50 ml Corning tube until half full, capped and then labeled.

2.5.2 Isolation of mycobacteriophage Ochi17.

Direct isolation was carried out as described[18]. Briefly, a 15 ml conical tube was filled to half volume with collected soil sample using a spatula. Enrichment broth was added aseptically until the samples were submerged in the broth. The tube was then mixed thoroughly by inverting and vortexing. The sample in phage buffer was allowed to sit for about 45 min with intermittent shaking every 15 min until soil particles are mostly settled. Then a 5 ml syringe was used to aseptically withdraw 1 ml supernatant from the top of the sample tube. The syringe was then attached to the top of a 0.22 μm filter and then 0.75 ml of the filtrate was aseptically dispensed into a sterile microcentrifuge tube.

In order to confirm the presence of phage in the isolated filtrate, a method known as plaque assay was used because phages form plaques on agar plates when infected with bacteria[15]. This assay was done by aseptically mixing 500 μl of the isolated phage into 250-μl host bacteria culture (*M. smegmatis*) and then allowed to sit for 10 min to allow for attachment. With a 5 ml pipette, 3 ml of warm top agar was aseptically transferred into the tube containing the phage-infected bacterium and then poured (aseptically) into agar plates. The plates were quickly tilted in multiple directions to spread the top agar mixture evenly. Plates were then allowed to sit for some time and then incubated at 37°C for at least 24 h.

2.5.3 Purification of isolated phage Ochi17.

Due to the possibility of having more than one kind of phage in my sample, the phage sample was purified to get a single clonal phage population. This was done following protocols described[18]. Briefly, a plaque of interest was picked from the incubated plate with a sterile micropipette tip and dislodged in a microcentrifuge tube containing 100 µl phage buffer. Three rounds of 10-fold serial dilutions ranging from 100 (undiluted phage stock) to 10^{-6} dilution were carried out by diluting each concentration using a 1:10 lysate to phage buffer ratio to a total volume of 100 µl. Using 10 µl of each phage lysate dilution, plaque assay was carried out to check for uniform plaque morphology on agar plates as described above.

After 24-48 h incubation, the plate with the highest concentration of plaques (webbed plate) was flooded by aseptically adding 8 ml of sterile phage buffer to the plate and allowed to sit for 2.5 h at room temperature. After incubation, the plate was tilted by placing one side on the lid of the plate. Using a 5 ml syringe, the lysate pool on one side of the plate was aspirated and then the syringe was attached to a 0.22 µm filter. Filtrate was then collected in a 15 ml sterile conical tube, volume recorded and then stored in 4^{o}C.

In order to determine the concentration of phage particles in the collected lysate (full plate titer), serial dilutions of the collected phage lysate were performed followed by assaying of plaques as described in the isolation of phage section above according to protocols[18].

After incubating for at least 24 h, the number of plaques on the 10^{-6} plate were counted and then divided by the volume of sample that was plated (10 µl). This was then multiplied by the reciprocal of the dilutions used to make the plate. This value was then converted from µl to ml by multiplying by 1000.

2.5.4 Amplification of phage Ochi17.

The number of plaques I needed to obtain a webbed plate was estimated by dividing the size of the plate by the size of the plaques. Then to find the volume of lysate needed to create a webbed plate, the estimated number of plaques was divided by the lysate titer. 10-fold serial dilutions were set up to 10^{-5} by first adding 90 µl to each tube. 10 µl of the undiluted phage sample was added to the 10^{-1} tube and vortexed, and then 10 µl of the 10^{-1} sample was added into the 10^{-2} tube and so on until I reached the 10^{-5} tube. Bracket dilutions of 0.5, 2, 5, and 10 -fold were also made from the

10^{-4} and 10^{-5} tubes. The 10^{-5} dilution and the bracket dilutions were used to infect the bacteria host as follows: 12.4, 6.2, and 24.8 µl of the 10^{-5} phage dilution was aseptically dispensed into the 250 µl host bacterium to generate the 0-fold, 0.5-fold, and 2-fold bracket dilution, while 6.2 and 12.4 µl of the 10^{-4} phage dilution was aseptically dispensed into the 250 µl host bacterium to generate the 5-fold, and 10-fold bracket dilution. The culture tubes were mixed and then allowed to sit for 10 min to allow for attachment. Then 3 ml of warm top agar (directly from water bath) was aseptically transferred into the tube containing the phage-infected bacterium with a 5 ml pipette. The tube was tapped gently and then poured (aseptically) into agar plates. The plates were quickly tilted in multiple directions to spread the top agar mixture evenly. Plates were then allowed to sit for some time and then incubated without inverting them for at least 24 h.

Following previously described protocol[18], the webbed plate was flooded by aseptically putting 8 ml of sterile phage buffer to the webbed plate, swirling the plate content gently and then allowed plate to sit for 2.5 h at room temperature. After incubation, the plate was tilted by placing one side on the lid of the plate. Using a 5 ml syringe, the lysate pool on one side of the plate was aspirated and then the syringe was attached to a 0.22 µm filter. Filtrate was then collected in a 15 ml sterile conical tube, volume recorded and then stored in 4^{o}C. Full plate titer was calculated as described above in the purification of phage section.

2.5.5 Transmission electron microscopy (TEM) of phage Ochi17.

The isolated phage was structurally characterized using TEM[18]. Basically, 100 µl of the high-titer phage lysate was transferred into a sterile microcentrifuge tube and centrifuge at top speed to pellet the phage particle. The supernatant was then removed using a micropipettor and then the pellet was resuspended in 100 µl of phage buffer and mixed gently using the pipette tip. Then the EM forceps was used to remove a fresh grid from a box/plate of unused grids. The forceps were placed on the paper with the dark and shiny side of the grid facing up. 5 µl of the lysate was placed onto the grid without touching the tip of the pipette to the grid. The lysate was allowed to sit on the grid for at least 2 min and then the grid was rinsed by pipetting 60 µl of sterile water across the dark-and-shiny face of the grid held at 45^{o} angle. Excess water was wicked away by placing a fresh wedge of filter paper against one edge of the grid. 5 µl of 1% uranyl acetate was added to the grid and then excess stain was wicked off using a wedge of filter paper. The grid was then imaged at

the Purdue EM facility using a FEI Tecnai G2 20 electron microscope equipped with electron source with LaB6 filament emitter and high tension between 20 and 200 kV.

2.5.6 DNA extraction of phage Ochi17.

Following protocol described[18], workbench was prepared for aseptic work by cleaning with 45% Cidecon and 70% ethanol. With gloves on, the bacterial DNA/RNA was degraded by aliquoting 1 ml phage lysate into a microcentrifuge tube and 5 µl nuclease mix added to the tube and then tube was inverted thoroughly and incubated at room temperature for 30 min. The nuclease-treated phage lysate was transferred into a 15 ml conical tube containing 2 ml of DNA clean-up resin and the solution mixed gently by inverting repeatedly for 2 min.

Syringe barrel attached to a column was set on a new microcentrifuge tube and 1.5 ml of the phage DNA/resin solution was transferred from the 15 ml tube to the column using a pipette. A plunger was inserted into the syringe and was carefully pushed down to remove all the liquid. The column was unscrewed from the syringe barrel, the plunger was removed from the syringe barrel, and then the syringe barrel was re-attached to the column. The salts from the DNA was washed by adding 2 ml of 80% isopropanol to each syringe barrel/column and the liquid was pushed through the column using a plunger. The column was unscrewed from the syringe barrel, the plunger was removed from the syringe barrel, and then the syringe barrel was re-attached to the column. This salt removal step was done for a total of three times. In order to remove any residual isopropanol, column was put in a fresh 1.5 ml microcentrifuge tube and spun at 10,000 xg for 5 min. Additional 1 min centrifugation was done at the same speed. Placing the column directly in a 90^{o}C heating block for 60 s evaporated every last trace of isopropanol.

Finally, the DNA was eluted from the columns by first placing each column in a clean microcentrifuge tube and then adding 50 µl of sterile ddH_2O directly to each column, followed by incubation at room temperature for 1 min, and then spun at 10,000 xg for 1 min. The phage DNA concentration was determined using a Nanodrop to quantify the phage DNA. This DNA stock was then stored in 4^{o}C for phage genome characterization.

2.5.7 Characterization of phage DNA.

The DNA sample was gently mixed, incubated at 65^{o}C for 10 min, and then quickly placed on ice. Restriction digest reactions were set up in microcentrifuge tubes for the six restriction enzymes and an additional 'uncut' control by adding the following solution sequentially: 20.27 µl of sterile diH_2O, 0.5 ul restriction enzyme, 2.5 µl 10X reaction buffer, and then finally 1.73 µl (0.5 µg) of DNA sample. For the uncut control tube, 3.0 µl of additional diH_2O was added instead. Contents were mixed gently and spun quickly for less than a minute to move all liquid to the bottom. Then they were incubated for up to 15 - 60 min depending on the enzyme[18].

For gel electrophoresis of restriction enzyme digests, gel apparatus was set up following the manufacturer's directions. 100 ml of 0.8 % agarose gel was prepared by measuring 0.8 g agarose and transferred to an Erlenmeyer flask. 100 ml TBE buffer was added to the agarose and then the contents were swirled gently to mix. The mixture was heated in the microwave until it begins to boil or until clumps are no longer visible in the mixture. The solution was allowed to cool for about 10 - 15 min and with gloves on, 5 µl (0.5 µg/ml) ethidium bromide (EtBr) was added to the solution and mixed. Then the agarose/EtBr mixture was poured into the prepared gel apparatus and the comb was inserted to cast the wells. This was left to solidify for about 30 min and then the comb was removed, and gel platform removed from the casting tray and put in the gel box with the wells at the cathode end of the box. The 1X TBE buffer was poured into the box until the gel was submerged in the buffer. Before the samples were loaded, 5 µl of concentrated 6x loading dye was added to each restriction enzyme sample and they were placed in a water bath at 65^{o}C for 5 min after which they were placed on ice to cool and then spun for 15 s at 10, 000 rpm. Then 20 µl of the samples were loaded into the wells in the following order: DNA ladder - Uncut DNA - BamHI - ClaI - EcoRI - HaeIII - HindIII - SaII. The electrodes were plugged, and the gel ran at 100 V for 1 h. After the run, the gel was removed from the chamber and photographed.

2.6 References

1. Clokie, M.R.J., Millard, A.D., Letarov, A.V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31-45 (2011).
2. Pedulla, M.L. *et al.* Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**, 171-182 (2003).

3. Gardner, G.M. & Weiser, R.S. A bacteriophage for *Mycobacterium smegmatis*. *Proc Soc Exp Biol Med* **66**, 205-206 (1947).
4. Kozloff, L.M., Sundar Raj, C.V., Nagaraja Rao, R., Chapman, V.A. & DeLong, S. Structure of a transducing mycobacteriophage. *J Virol* **9**, 390-393 (1972).
5. Hatfull, G. The secret lives of mycobacteriophages. *Adv Virus Res* **82**, 179-288 (2012).
6. Sassi, M., Bebeacua, C., Drancourt, M. & Cambillau, C. The first structure of a mycobacteriophage, the *Mycobacterium abscessus* subsp. bolletii phage Araucaria. *J Virol* **87**, 8099-8109 (2013).
7. Haq, I.U., Chaudhry, W.N., Akhtar, M.N., Andleeb, S. & Qadri, I. Bacteriophages and their implications on future biotechnology: a review. *J Virol* **9**, 9, doi:10.1186/1743-422X-9-9 (2012).
8. McNerney, R. & Traoré, H. Mycobacteriophage and their application to disease control. *J Appl Microbiol* **99**, 223-233 (2005).
9. Scollard, D.M. *et al.* The continuing challenges of leprosy. *Clin Microbiol Rev* **19**, 338-381 (2006).
10. Pope, W.H. & Jacobs-Sera, D. Annotation of bacteriophage genome sequences using DNA Master: an overview. *Methods Mol Biol* **1681**, 217-229 (2018).
11. Cresawn, S.G. *et al.* Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* **12**, 395, doi:10.1186/1471-2105-12-395 (2011).
12. Rusell, D.A. & Hatfull, G.F. PhagesDB: the actinobacteriophage database. *Bioinformatics* **33**, 784-786 (2017).
13. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
14. Cisek, A.A., Dąbrowska, I., Gregorczyk, K.P. & Wyżewski, Z. Phage therapy in bacterial infections treatment: one hundred years after the discovery of bacteriophages. *Curr Microbiol* **74**, 277-283 (2017).
15. Villanueva, V.M., Oldfield, L.M. & Hatfull, G.F. An unusual phage repressor encoded by mycobacteriophage BPs. *PLoS ONE* **10**, e0137187, doi:10.1371/journal.pone.0137187 (2015).
16. Lee, S., Lewis, D.E.A. & Adhya, S. The developmental switch in bacteriophage lambda: a critical role of the Cro protein. *J Mol Biol* **430**, 58-68 (2018).

17. Kim, M. & Ryu, S. Antirepression system associated with the life cycle switch in the temperate *Podoviridae* phage SPC32H. *J Virol* **87**, 11775-11786 (2013).
18. Poxleitner, M., Pope, W., Jacobs-Sera, D., Sivanathan, V. & Hatfull, G. Phage discovery guide for use with Mycobacterium smegmatis. *HHMI* 1-118 (2016).

3. GLOBAL PROTEOME INVESTIGATION OF THE RESPONSE OF *MYCOBACTERIUM SMEGMATIS* DURING INFECTION BY MYCOBACTERIOPHAGE OCHI17

3.1 Abstract

The potential for mycobacteriophages to serve as a therapeutic agent in the treatment of the deadly pathogen, *Mycobacterium tuberculosis* has led to renewed interest in the study of mycobacteriophage-mycobacteria interactions. Mycobacteriophages are phages that attack mycobacteria. Not much is known, however, about the molecular interactions of both species, especially at the translational level. To better understand this interaction, we applied label-free quantitative proteomics using the model host, *Mycobacteria smegmatis*, infected with the phage, Ochi17 at ~3.5 h. A total of 2,188 proteins were identified, with 299 displaying significant upregulation and 135 significant downregulation. Homologous recombination and processes of macromolecular synthesis such as amino acid metabolism, DNA replication, transcription, translation, and vitamin metabolism were significantly upregulated. Fatty acid and lipid metabolism were also primary targets of phage Ochi17. This study provides the first global proteome investigation of any mycobacteriophage-infected mycobacterium with respect to the host mycobacterium response and may contribute in the development of mycobacteriophages as antimicrobial therapies that can overcome various defense strategies employed by host mycobacteria.

3.2 Introduction

Tuberculosis (TB) is currently the most important public health crisis causing millions of death each year[1,2]. The causative organism of TB in humans is *Mycobacterium tuberculosis* (*M. tuberculosis*), the world's deadliest human pathogen[3]. It is a slow growing (doubling every 24 h) pathogenic bacterium belonging to the family Mycobacteriaceae[3,4].

The use of drugs in the treatment of TB has become challenging and ineffective due to the emergence of multidrug-resistant and extensively drug-resistant TB[5,6]. Consequently, outbreaks of multidrug resistant *M. tuberculosis* infections has led to renewed calls by scientists to come up with a better alternative treatment strategy for the treatment of TB[7,8]. The application of

(myco)bacteriophages is an alternative that has gained traction not just as therapeutic agents, but also in broad applications ranging from food bio-preservation, wastewater treatment, and disease diagnosis[1,9]. This is due to several advantages over antibiotics, such as host specificity, high genetic diversity, and safety[9].

Bacteriophages (phages) are ubiquitous viruses that hijack bacterial host metabolism for replication and are reported to be ten times more abundant than bacteria in any particular environment[10]. Specifically, mycobacteriophages are viruses that infect mycobacterial species[4]. Phage infection of host bacteria first targets the host's DNA transcription, thereby altering host gene and protein expression[11,12]. This attack from the phage also results in various responses from the host bacteria, thereby leading to complex interactions happening between them[13,14]. As a result of this, it is vital to undertake a global omics study to understand the molecular changes in phage-infected hosts, and identify actively expressed phage proteins.

The majority of the studies on phage-bacteria interactions have focused largely on the transcriptomic regulation in pathogens such as *Pseudomonas aeruginosa*, *Escherichia coli*, *Campylobacter jejuni*, and others which cause food-borne illnesses and other human diseases like gastroenteritis, pneumonia, and diarrhea[12, 15-20]. Very few studies have reported global proteomics profiling of bacterial hosts after infection by lytic phages at various time points[21-23]. However, no omics study has been reported on the response of any mycobacteria species to a mycobacteriophage attack. Although, expression of phage Patience proteins during early and late stage infection of *M. smegmatis* has been reported, the study did not include the host proteome response[24]. Additionally, the focus of these studies has primarily been on lytic phages. Host responses vary by the type of phage used- whether lytic or temperate[12,25]. Temperate phages are preferred over lytic phages for phage therapy because the latter could lead to the release of bacterial endotoxins[26].

As part of a diverse research community effort to examine phage genetic diversity with potential impact on human health and future therapeutic treatments, mycobacteriophages, such as Ochi17 are isolated, sequenced, annotated and studied extensively by the Howard Hughes Medical Institute Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (HHMI SEA-PHAGES) diverse research community[9]. Ochi17 is a cluster F *Siphoviridae* temperate mycobacteriophage isolated at Purdue University in 2017 (https://phagesdb.org/phages/Ochi17/) that infects *Mycobacterium smegmatis (M. smegmatis)*, a

model organism to study mycobacterial physiology and other relevant processes applicable to the pathogenic species of mycobacteria. It is a gram-positive, acid-fast and fast-growing mycobacterium originally isolated from humans that is mostly non-pathogenic[1,4].

Transcriptome studies are limited in their ability to measure direct changes in the functional activity of a cell[27]. Many regulatory steps are between mRNA abundance and protein abundance including pre, co- and post-translational modifications[28]. Additionally, alternative splicing of a single mRNA can result expression of hundreds of different proteins. Thus, it is critical to examine the interaction of phage and bacteria at the level of the protein through quantitative proteomics to obtain a comprehensive assessment of the interaction[23]. In this study, a global proteomic analysis was conducted to investigate *M. smegmatis* host proteome response during infection with the temperate phage Ochi17 at the mid-lytic phase (3.5 h), as well as phage Ochi17 protein expression during infection of the host. Given that the doubling time for *Mycobacterium smegmatis* is 3-3.5 h, we proposed that Ochi17 will primarily hinder host macromolecular synthesis at this infection time point. Results reveal upregulation of key cellular pathways involved in macromolecular synthesis, such as homologous recombination, amino acid metabolism, ribosome, and vitamin metabolism. Ochi17 could be utilizing the host *M. smegmatis* molecular machinery for its own propagation. Results from this study suggest possible survival mechanisms of the mycobacterium host and might help identify previously elusive mechanisms involved in mycobacteriophage-mycobacteria interactions.

3.3 Results

3.3.1 *M. smegmatis* infection and Ochi17 One-step Growth Curves

Following initial incubation of *M. smegmatis* in 7H9 media up to mid-exponential stage, the cells were infected with phage Ochi17 in triplicates at a multiplicity of infection (MOI) of 10 at 37^{o}C in order to achieve approximately 100% infection efficiency[29].

For the first 90 min (1.5 h) of phage Ochi17 infection, both the phage-infected and un-infected host (control) showed a similar lag in growth. However, the growth rate from 1.5 h and 6 h post-infection (p.i.) in infected and uninfected cells differed by 10-20% (Fig. 3.1a). Infected cells grew more slowly than uninfected controls and had a reduced carrying capacity, likely due to partial host cell lysis (~70-80%).

At the same infection times, we observed the replication cycle of phage Ochi17 (Fig. 3.1a). There was no change in the number of phage particles within the first 30 min. This is probably the latent period of phage Ochi17 infection where adsorption to the cell wall of *M. smegmatis* takes place after insertion of Ochi17 genome. Following 30 min (0.5 h) p.i., we began observing increase in phage replication. Early lytic stage was depicted as the period between 0.5 and 1.5 h p.i. Between 1.5 h and 6 h was considered the middle stage of Ochi17 infection, where increased rate of replication and synthesis of phage constituents takes place. Between 6 h and 30 h, denoted as the late stage of Ochi17 infection, is the period where phage assembly occurs.

In order to understand the response of *M. smegmatis* to phage Ochi17 infection, we decided to capture the Ochi17-*M. smegmatis* interaction at the middle stage where Ochi17 infection was beginning to show some observable effects on *M. smegmatis* (Fig. 3.1a). Infected (phage-treated) and un-infected (control) cultures were harvested and prepared for proteomics analysis around 3.5 h p.i. (Fig. 3.1a-b).

3.3.2 LC-MS Reproducibility

A prerequisite for accurate intensity-based label free quantitation is reproducible LC retention times of peptides[30]. To validate the precision of the label-free method used for quantifying protein abundance, the reproducibility of peptide retention times of the common peptides having more than one MS/MS count for the three phage-infected replicates were evaluated. The coefficient of determination (R squared) of 0.99 indicated a very high retention time consistency among the replicates (Fig. 3.2a). Average and median coefficient of variation of the peptide retention time was < 1.0%, which indicated good reproducibility for the intensity-based label-free quantification. Correlation analysis of LFQ intensities of the replicates for each phage-treated and control sample also showed high Pearson correlation coefficients and further confirmed high reproducibility among the replicates (Fig. 3.2b).

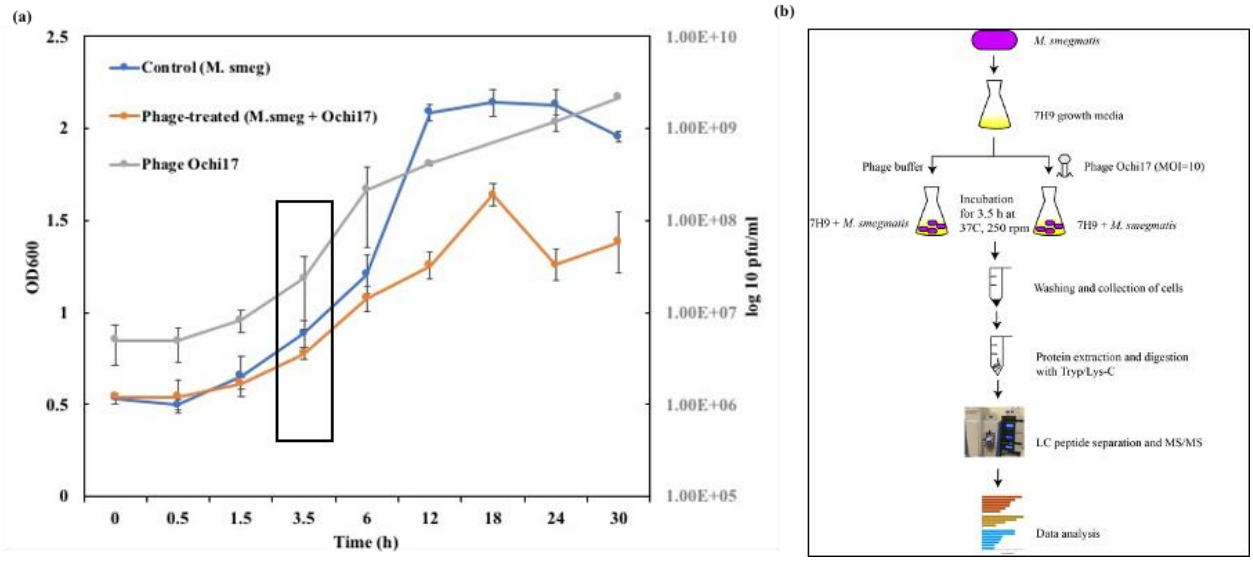


Figure 3.1. ***M.* smegmatis infection curve, Ochi17 one-step growth curve and experimental design.** (**a**) Growth curves of non-infected *M. smegmatis* (control) and *M. smegmatis* infected with phage Ochi17 (phage-treated), alongside one-step growth curve of phage Ochi17. The data represent the mean ± SD of three replicates for each treatment. Insert box represent time period of sample collection for proteomic analysis. (**b**) Experimental workflow for the proteomic analysis of phage Ochi17-infected *M. smegmatis. M. smegmatis* was infected with phage Ochi17 for ~ 3.5 h (see insert in Fig. 1A) at Multiplicity of Infection (MOI) of 10 with shaking at 250 rpm, 37ₒC. Experiments were done in triplicates. See the materials and methods for detailed procedures. LC- Liquid Chromatography; MS/MS- tandem mass spectrometry.

3.3.3 Global Proteome Analysis

A total of 2095 and 1896 proteins were identified in at least two of the three replicates in the phage-infected and control samples respectively, accounting for an overall total of 2188 proteins (Fig. 3.2c). While 1803 proteins were common between both samples, 93 proteins were found only in the control and 233 proteins only in the phage-infected (Fig. 3.2c).

Of the 1803 common proteins, 162 were significantly regulated at $p \leq 0.05$, with 108 of the proteins having absolute log2 fold change ≥ 0.5 (Fig. 3.2c; Appendix A) compared to the control. Principal Component Analysis (PCA) showed that these differentially regulated proteins clustered into two separate groups of control and phage-infected, accounting for 89.2% of the total variation among proteins between the two groups (Fig. 3.3a). PC1 accounts for about 82.3% of the difference, while PC2 accounts for only 6.95%, indicating that the difference between the groups is larger than the difference within the groups (Fig. 3a). These differentially regulated proteins were hierarchically clustered into a heatmap, which shows a clear distinction between the phage-infected and control groups, as well as consistent expression patterns among the different replicates

for both phage-treated and control (Fig. 3.3b). Proteins that were exclusively found in either the control group (93 proteins) or the phage-infected group (233 proteins) were also analyzed as part of the significantly regulated proteins. In total, there were 488 significantly differentially expressed proteins: 331 upregulated and 157 downregulated proteins relative to untreated controls.

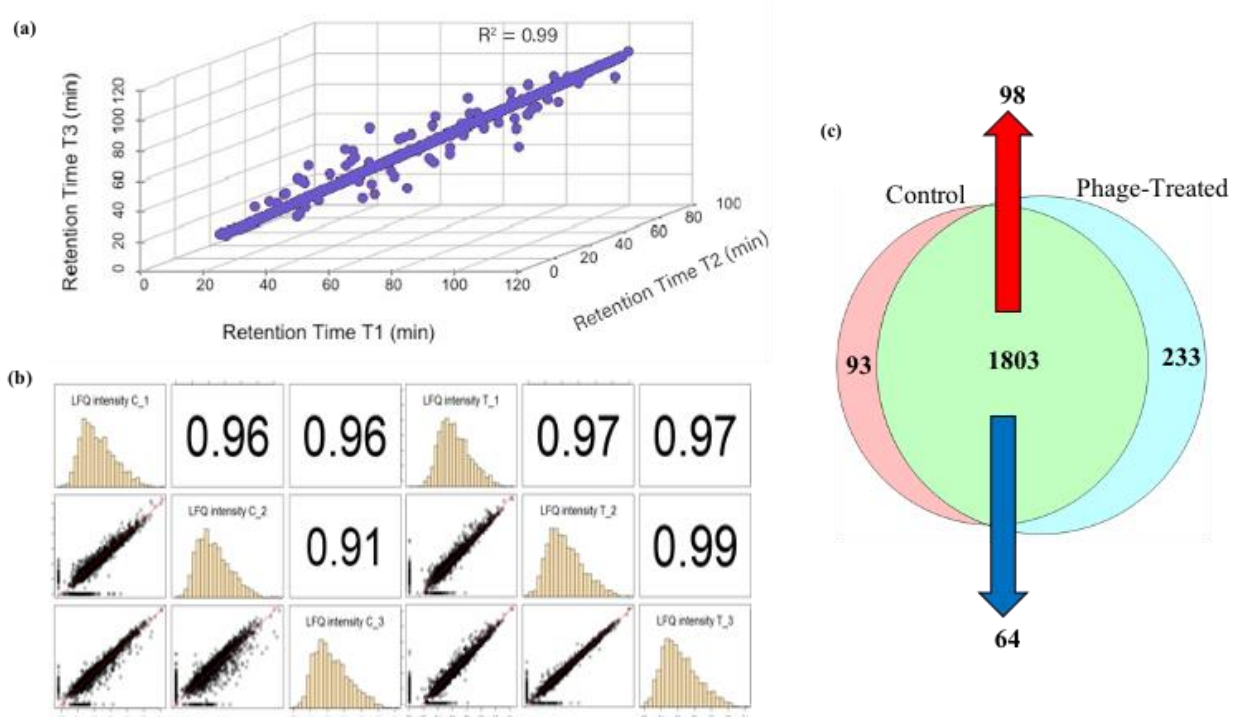


Figure 3.2. **Quantitative proteome data analysis.** (**a**) Correlation of the peptide retention times among the three phage-infected samples showing LC-MS reproducibility. (**b**) Correlation analysis of LFQ intensities of the replicates for control and phage-infected samples. (**c**) Total *M. smegmatis* proteins identified from three biological replicates in each control and phage-infected samples. Upward orange arrow denotes significant upregulation, and downward blue arrow denotes significant downregulation. The numbers the arrows are pointing to represent the number of differentially significantly regulated proteins.

3.3.4 Analysis of differentially expressed proteins

The top 10 upregulated proteins were primarily DNA binding proteins involved in homologous recombination and DNA repair such as replicative DNA helicase (dnaB), recombinase A (recA), Holliday junction ATP-dependent DNA helicase (ruvA), exodeoxyribonuclease III (Xth) and uracil-DNA glycosylase (ung or UDG). Others include L-lysine-epsilon aminotransferase (lat), phoH-like protein (phoH2), integral membrane protein, piperideine-6-carboxylic acid dehydrogenase (Pcd), and one unknown protein, MSMEI_4295. The top 10 downregulated proteins were mostly energy-related, such as iron-containing alcohol dehydrogenase (MSMEI_1354), hydroxymethylglutaryl-coA lyase (MSMEI_2027), glycerol

dehydratase (MSMEI_1511) and malonyl CoA-acyl carrier protein transacylase (fabD). Others include coenzyme pqq synthesis protein E (pqqE), probable conserved transmembrane protein, DXP reductoisomerase (dxr), and major facilitator superfamily protein. Two proteins with unknown functions (MSMEI_5553 and MSMEG_2261) were also among the top downregulated proteins (Fig. 3.3c).

A total of identified 488 differentially expressed proteins were functionally mapped using DAVID 6.8[31]. Of these, 436 (89%) were mapped to their functional pathways and groups using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) terms (Fig. 3.4a-c). Among upregulated proteins, the most represented KEGG pathways included biosynthesis of secondary metabolites, amino acid metabolism, microbial metabolism in diverse environments, biosynthesis of antibiotics, homologous recombination/DNA repair, and ABC transporters. The most represented pathways among downregulated proteins included biosynthesis of secondary metabolites, amino acid metabolism, microbial metabolism in diverse environments, biosynthesis of antibiotics, metabolism of starch and sugar, ABC transporters, and glycero(phospho)lipid metabolism (Appendix B.1). Homologous recombination, vitamin metabolism, base (nucleotide) excision repair, aminobenzoate degradation, and pantothenate and CoA biosynthesis pathways were only represented in upregulated proteins, while starch and sugar metabolism were only represented in downregulated proteins (Appendix B.1). The biological processes that were most represented among upregulated proteins include transcription regulation, amino acid metabolism, DNA (homologous) recombination, transport and translation. Among downregulated proteins, the most represented biological processes are transport, cell wall/cell shape organization, transcription regulation, and amino acid metabolism. DNA or homologous recombination, DNA repair regulation, DNA replication, carbohydrate metabolic process, antibiotic biosynthesis, and protein folding/unfolding were exclusively represented in upregulated proteins. Response to stress and antibiotic catabolism, however, were uniquely represented in downregulated proteins (Appendix B.2). The most represented molecular functions among upregulated and downregulated proteins included transcription factor activity/DNA binding, ATP binding, oxidoreductase activity, and metal ion binding. Exonuclease activity was only represented in the upregulated proteins, while glycerol kinase and hydrogenase activities were represented only in downregulated proteins (Appendix B.3).

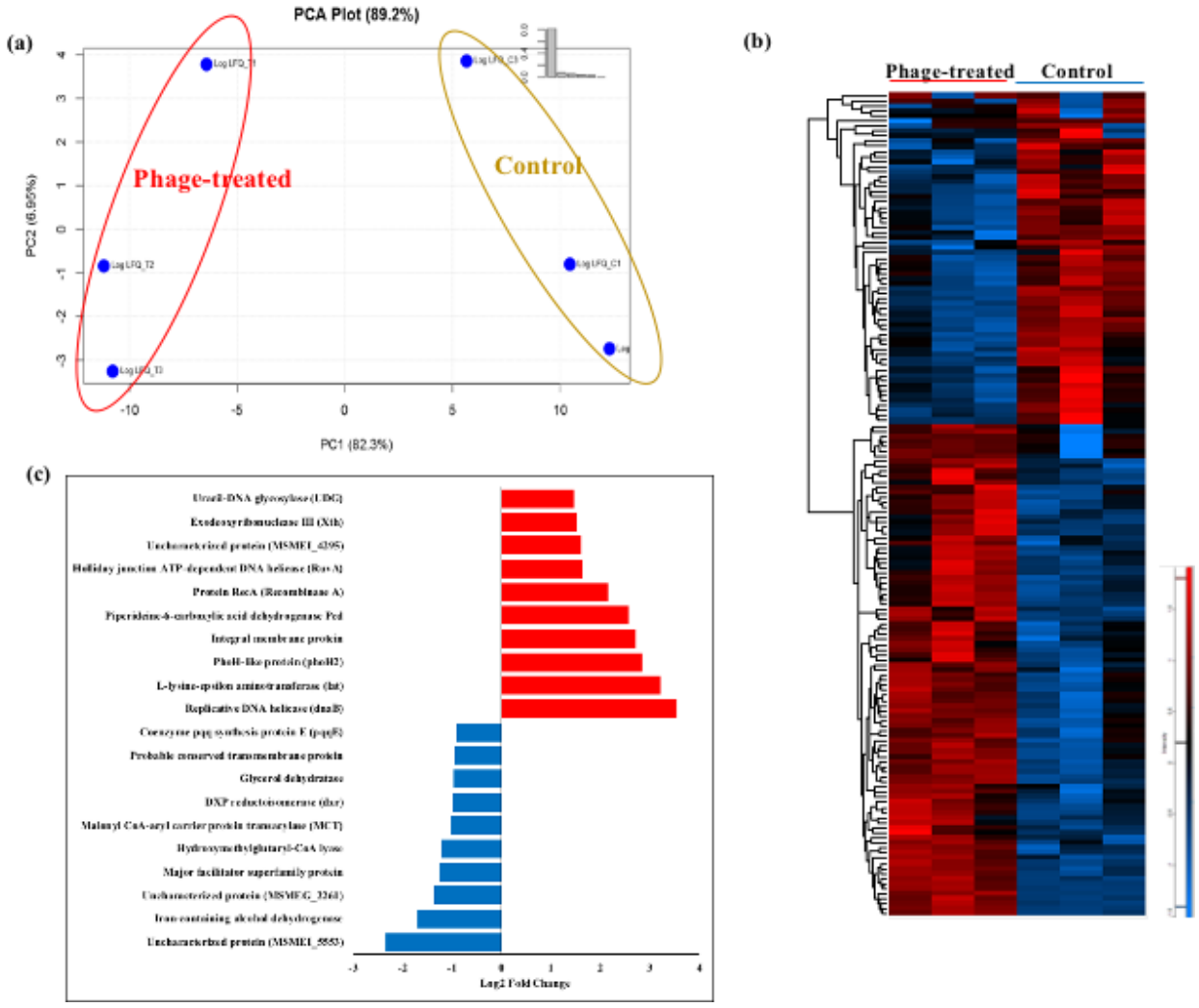


Figure 3.3. Proteomic profiles in control and phage-treated samples. (**a**) PCA plot of significantly regulated proteins from control and phage-treated samples showing difference in both groups. (**b**) Heatmap of significantly regulated proteins from control and phage-infected samples showing difference in both groups. Each experimental group comprosed three replicates. LFQ- Label-free Quantitation. (**c**) The top 10 upregulated and downregulated proteins based on log2 fold change. T- Phage-infected; C- Control.

3.3.5 Impact of phage Ochi17 on host molecular mechanisms

To investigate which host pathway and/or functional groups were significantly impacted by Ochi17 infection at the onset of phage infection, differential enrichment of KEGG pathways and GO terms with fold change ≥ 0.5 among the significantly expressed proteins were also analyzed (Fig. 3.4a-c). Among significantly upregulated KEGG pathways were homologous recombination, purine and pyrimidine metabolism, ribosome (translation), fatty acid degradation, base/nucleotide excision repair, pyruvate and glycolysis/gluconeogenesis metabolism, vitamin metabolism, and

amino acid metabolism (Fig. 3.4a). Significantly downregulated pathways included starch and sugar metabolism, fatty acid biosynthesis, glycerolipid/glycerophospholipid metabolism, carbon metabolism, and microbial metabolism in diverse environment (Fig. 3.4a). Differentially upregulated biological processes included DNA repair, replication, transcription, cell wall organization, among others. The downregulated biological processes were glycerol-3-phosphate metabolism, fatty acid biosynthesis, and stress response (Fig. 3.4b). RNA binding, exonuclease, ATPase, reductase and oxidoreductase activities were some of the molecular functions differentially upregulated. while, dehydrogenase, hydrogenase, and glycerol kinase activities were downregulated (Fig. 3.4c).

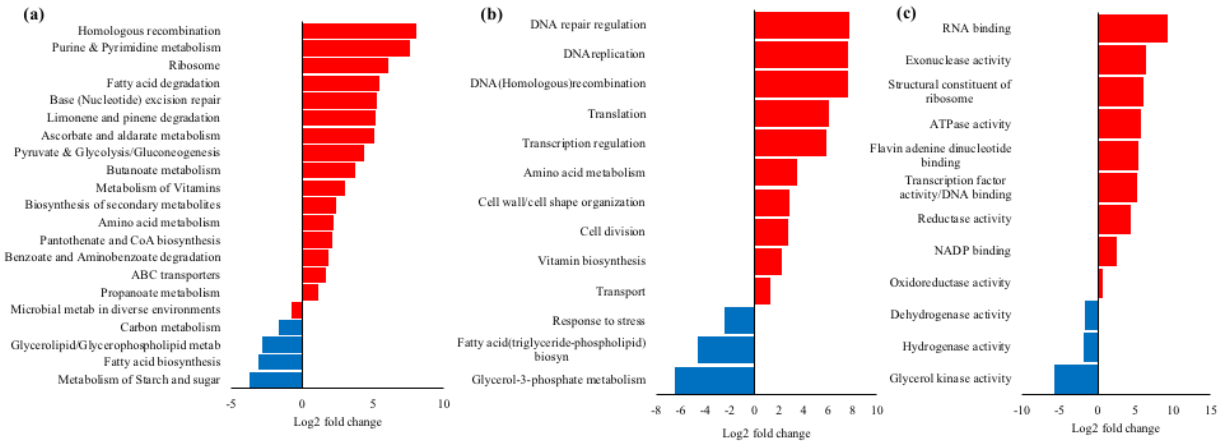


Figure 3.4. Differential regulation of KEGG pathways and functional groups. (**a**) Pathway classification according to KEGG database (**b**) Biological process classification according to GO terms. (**c**) Molecular function classification according to GO terms. Amino acid metabolism includes arginine and proline, histidine, tryptophan, phenylalanine, tyrosine, alanine, aspartate and glutamate, cysteine and methionine, glycine, serine and threonine metabolism; lysine, valine, leucine and isoleucine degradation and biosynthesis. For a complete list of the proteins and their functional categories, see Tables S3 and S4 in the Supplemental material.

Following the functional enrichment, we then examined the possible interactions and molecular actions of the significantly identified proteins. Using STRING v.11.0, which generates a protein-protein interaction network based on computational and database annotation, the significantly regulated proteins were clustered into four groups (Fig. 3.5a). Two of these clusters were involved in cell wall organization while the remaining two contained proteins associated with homologous recombination and ribosome/translation, respectively (Fig. 3.5a-b). Counterintuitively, recX, which inhibits recA, is upregulated with recA and other rec proteins (Fig. 3.5a-b). Additionally, two proteins with unknown functions, such as MSMEG_0758 and MSMEG_4752 were found to cluster with the cell wall related proteins (Fig. 3.5a).

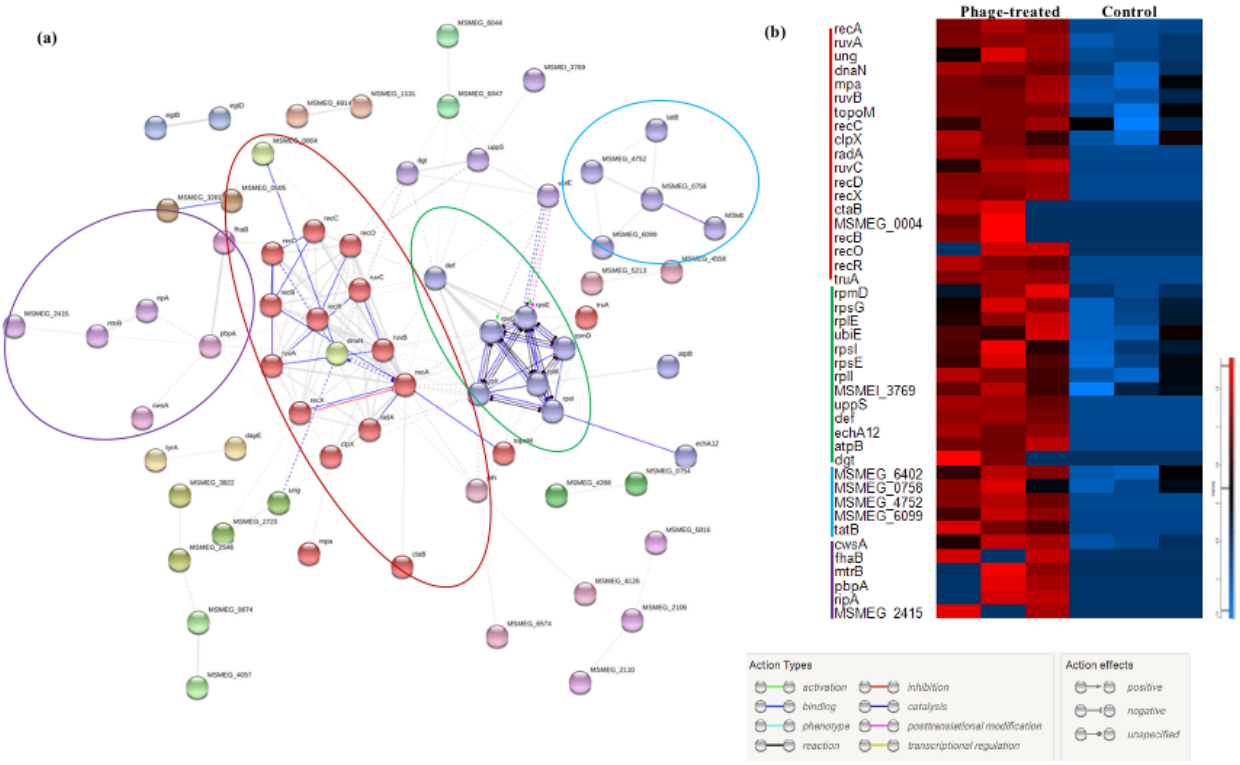


Figure 3.5. Interactions of differentially regulated proteins. (**a**) STRING interaction of differentially regulated proteins from Ochi17-*M. smegmatis* interaction with minimum required interaction score as medium confidence (0.4). The edges represent the different molecular actions existing among the proteins. Colored circles represent different clusters, which were determined by an MCL clustering analysis with an inflation parameter of 3. Red, homologous recombination; green, protein synthesis; purple and cyan, cell wall-related. Unconnected nodes were hidden. (**b**) Heatmap showing the protein expressions of the proteins shown in the STRING map. Vertical colored lines match colored circles in the STRING map.

Fatty acids and lipids have been reported to be vital in the organization of mycobacterial cell walls, particularly mycolic acids[32]. The functional enrichment analysis showed that fatty acid metabolism was significantly affected. We observed upregulation of fatty acid degradation and downregulation of fatty acid biosynthesis, glycerol-3-phosphate metabolism and glycerol kinase activity (Fig. 3.4a-c). Using STRING (v. 11.0) analysis, interactions among proteins involved in fatty acid metabolism pathway (biosynthesis and degradation) and glycerol-3-phosphate metabolism were depicted and grouped into three clusters (Fig. 3.6a). For the glycerol-3-phosphate metabolism, MSMEG_6229 and MSMEG_6759 were downregulated, while MSMEG_1736 was upregulated. All three proteins show evidence of a binding interaction (Fig. 3.6a-b). The proteins (MSMEG_2227, echA12 and fadA2) associated with fatty acid degradation were all upregulated, and bind to each other (Fig. 3.6a-b). Out of the four fatty acid biosynthesis proteins in the

interaction, two (MSMEG_0269 and MSMEG_5642) were upregulated, while the other two (fabD and ino1) were downregulated (Fig. 3.6b).

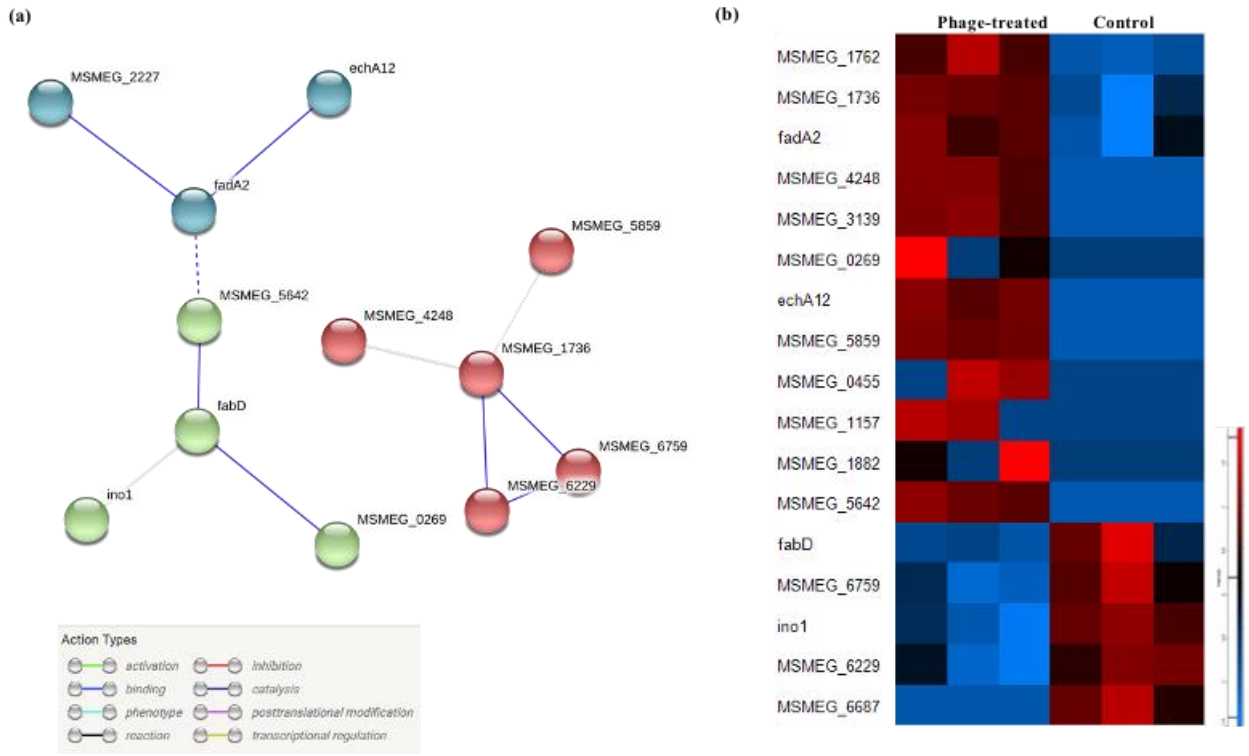


Figure 3.6. Interactions of differentially regulated proteins involved in fatty acid and lipid metabolism. (**a**) STRING interaction of differentially regulated proteins involved in fatty acid metabolism with minimum required interaction score as medium confidence (0.4). The edges represent the different molecular actions existing among the proteins. Clusters were determined by an MCL clustering analysis with an inflation parameter of 3. Unconnected nodes were hidden. (**b**) Heatmap showing the expressions of the fatty acid metabolism proteins shown in the STRING map. Vertical colored lines match colored circles in the STRING map.

3.3.6 Validation of selected proteins

Real-time quantitative PCR (RT-qPCR) analysis was conducted to determine if mRNA transcript expression levels were aligned with the observed proteome expression after 3-3.5 h post infection with phage Ochi17. The following proteins were selected as representatives of significantly regulated pathways: Aldehyde dehydrogenase (AldA) in the fatty acid metabolism pathway, 30S ribosomal protein S7 (rpsG), 30S ribosomal protein S9 (rpsI), and 50S ribosomal protein L5 (rplE) in the translational (ribosomal) pathway collectively; recA from the homologous recombination (HR) pathway; and a ribosome hibernation promoting factor (hpf). Our results showed that only rpsG and AldA showed any significant change in mRNA expression. They were both reduced by almost 10-fold relative to the control (Fig. 3.7). This is inconsistent with the expression we reported

at the proteome level where both proteins were upregulated (Fig. 3.5b; Fig. 3.6b). Although, the downregulation of hpf and upregulation of recA transcripts were reflected in the proteome analysis, they were not significant (Fig. 3.7).

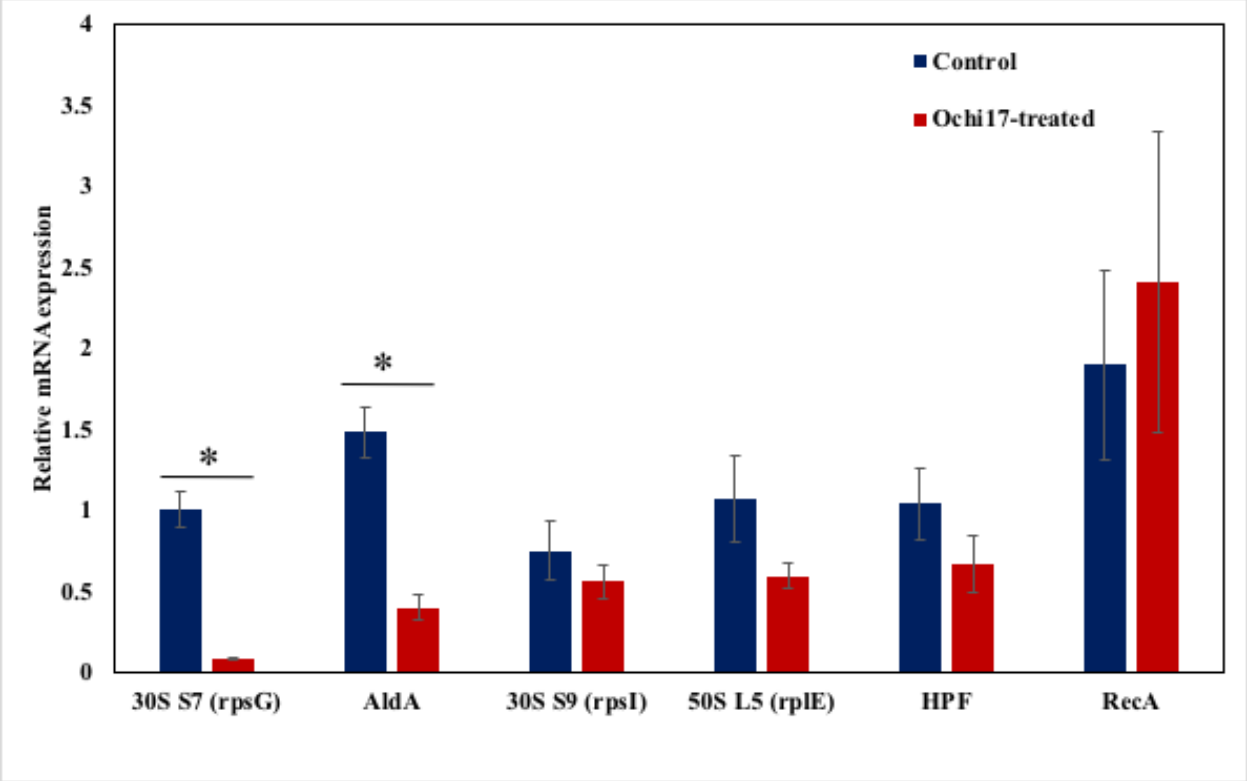


Figure 3.7. mRNA fold change of selected proteins. RT-qPCR ΔΔCq was used to determine relative fold change of mRNA expression level after phage Ochi17 infection of *M. smegmatis* for ~ 3.5 h. The qPCR results were normalized using 16S rRNA. rpsG = 30S ribosomal protein S7, HPF= hibernation promoting factor, rplE = 50S ribosomal protein L5, rpsI = 30S ribosomal protein S9, rpmD = 50S ribosomal protein L30, RecA = Protein RecA (Recombinase A). Significant difference represented by unpaired two-tailed Student's t-test *$p \leq 0.05$.

3.3.7 Proteome expression of phage Ochi17

A total of 68 phage proteins were identified in Ochi17-infected *M. smeg*matis (Appendix C; Table 3.1). This represents 62% expression of the predicted proteins from its genome. Out of this, 54 proteins were common to all three replicates, and 14 were found in any two replicates. The top 10 abundant proteins include major capsid, major tail subunit, lysin A, tail assembly chaperone, tape measure, portal, galactosyltransferase, lysin B, and two proteins of unknown function (Table 3.1). Majority of these top abundant proteins were classified as particle formation proteins, followed by host-associated proteins.

Based on the total number of Ochi17 proteins identified, 41% proteins are of unknown function followed by particle formation proteins (30%), DNA metabolism and replication (14%), host interaction (10%), and host lysis (4%). In terms of LFQ abundance, however, proteins involved in particle formation appear to be the most represented (72%), followed by proteins with unknown function (11%), host lysis (7%), DNA metabolism and replication (5%), and host interaction (4%) (Fig. 3.8a-b; Table 3.1).

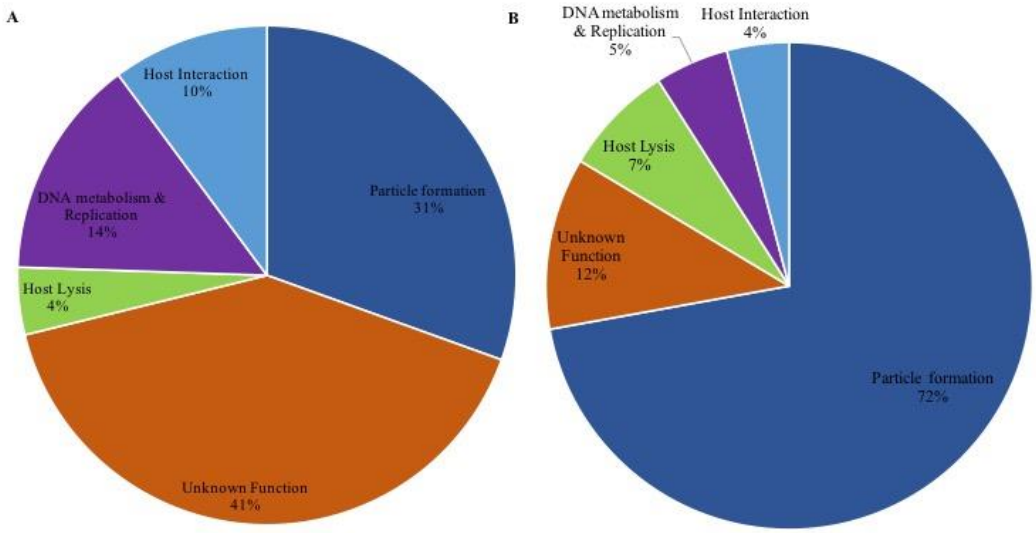


Figure 3.8. Functional classification of identified phage Ochi17 proteins. (**a**) represented as percentage of total label-free intensity quantitation (LFQ) of identified proteins. (**b**) represented as percentage of the total number of proteins identified. For a full list of all the Ochi17 proteins and their respective groups, see Table 3.1.

The particle formation proteins include major capsid protein, major and minor tail proteins, tape measure protein, portal protein, tail assembly chaperone, head-to-tail connector protein, scaffolding protein, terminase large and small subunit, and capsid maturation protease (Table 3.1). Host lysis includes proteins such as lysin A, lysin B, and holin. DNA metabolism and replication proteins include GIY-YIG endonuclease, HTH DNA-binding, DnaQ-like exonuclease, AAA ATPase, HNH endonuclease, WhiB, and MPME 1. Host interaction comprises proteins such as glycosyltransferase, galactosyltransferase, DNA methylase, integrase, antirepressor, and trigger factor.

Table 3.1. Ochi17 proteins identified at 3.5 h post-infection. LFQ and MS/MS counts represent average of three replicates. LFQ = Label-free Quantitation. DNA metabolism & rep = DNA metabolism & replication.

Protein start	Putative function	LFQ intensity	MS/MS count	Classification
5027	Major capsid	11598700000	395	Particle formation
7534	Major tail subunit	3472100000	63	Particle formation
25598	Lysin A	1307520000	60	Host Lysis
8423	Tail assembly chaperone	980306666.7	76	Particle formation
25692	Hypothetical protein	740956666.7	40	Unknown
9449	Tape measure	728790000	72	Particle formation
2227	Portal protein	598850000	45	Particle formation
39399	Hypothetical protein	557713333.3	38	Unknown
56258	Galactosyltransferase	513880000	33	Host Interaction
26464	Lysin B	512013333.3	45	Host Lysis
17618	Minor tail protein	423746666.7	35	Particle formation
16791	Minor tail protein	404153333.3	16	Particle formation
13247	Minor tail protein	371513333.3	32	Particle formation
52530	GIY-YIG endonuclease	328403333.3	27	DNA metabolism & rep
42710	Hypothetical protein	271263333.3	5	Unknown
35174	HTH DNA-binding protein	264756666.7	14	DNA metabolism & rep
20071	Minor tail protein (D-ala-D-ala carboxypeptidase)	261346666.7	29	Particle formation
29006	DNA polymerase III subunit	255895666.7	28	DNA metabolism & rep
43287	AAA ATPase	248300000	23	DNA metabolism & rep
27954	Holin	240139666.7	8	Host Lysis
15037	Minor tail protein	224696666.7	25	Particle formation
23559	Minor tail protein	213873333.3	16	Particle formation
52982	Hypothetical protein	213773333.3	27	Unknown
5854	Head-to-tail adaptor	204686666.7	22	Particle formation
50740	Hypothetical protein	184526666.7	12	Unknown
54873	Glycosyltransferase/methyltransferase	180680000	21	Host Interaction
24405	Hypothetical protein SEA FRANKIE 25	179153333.3	22	Unknown
37018	Hypothetical protein PBI Llama 51	166922666.7	14	Unknown
23210	Hypothetical protein PBI_LLAMA_21	152453333.3	6	Unknown
44264	DNA methylase	121695333.3	12	Host Interaction
22008	Minor tail protein	119728666.7	16	Particle formation
4349	Scaffolding protein	113080000	16	Particle formation
3566	Capsid maturation protease	112938333.3	9	Particle formation
32570	Integrase	112732000	10	Host Interaction

42253	HTH DNA binding domain protein	104943666.7	14	DNA metabolism & rep
36134	Putative antirepresser	96663333.33	4	Host Interaction
23779	Minor tail protein	93219000	13	Particle formation
559	Terminase large subunit	84981333.33	8	Particle formation
49194	Hypothetical protein PBI_LLAMA_88	84576666.67	9	Unknown
44768	HNH endonuclease	78258333.33	6	DNA metabolism & rep
25327	Hypothetical protein PBI_SQUIRTY_23	75797000	5	Unknown
37473	Hypothetical protein PBI_LLAMA_52	74558666.67	8	Unknown
54248	Trigger factor	73196666.67	7	Host Interaction
51709	Hypothetical protein PBI_LLAMA_97	65688000	7	Unknown
54054	Hypothetical protein PBI_LLAMA_104	61192000	4	Unknown
45109	Hypothetical protein PBI_LLAMA_71	61081333.33	7	Unknown
31113	Hypothetical protein PBI Llama 40	60175333.33	6	Unknown
6955	Head-to-tail connector	54124000	11	Particle formation
56821	Glycosyltransferase	53569000	5	Host Interaction
32131	GIY-YIG endonuclease	50931333.33	11	DNA metabolism & rep
38331	Hypothetical protein PBI_LLAMA_56	47946666.67	8	Unknown
38818	Hypothetical protein PBI_LLAMA_58	34226666.67	2	Unknown
28409	Hypothetical protein SEAGREEN_34	33555666.67	1	Unknown
8829	Tail assembly chaperone	32541000	4	Particle formation
53071	Hypothetical protein PBI_LLAMA_103	31730666.67	3	Unknown
40990	Whib transcription factor	31126333.33	6	DNA metabolism & rep
38570	Hypothetical protein PBI_LLAMA_57	29559333.33	3	Unknown
28666	Hypothetical protein FLORINDA_34	23739500	3	Unknown
87	Terminase small subunit	12847000	1	Particle formation
46357	Hypothetical protein PBI_LLAMA_76	8299600	1	Unknown
48872	Hypothetical protein PBI_CHE9D_93	8115000	3	Unknown
50145	Hypothetical protein PBI_LLAMA_92	7522000	1	Unknown
6147	Head-to-tail stopper	6402333.333	2	Particle formation
43066	Hypothetical protein SEA_KIMBERLIUM_66	3622000	1	Unknown
50312	Hypothetical protein PBI_LLAMA_93	3321533.333	1	Unknown
35682	Hypothetical protein PBI_LLAMA_49	2061500	1	Unknown
32226	GIY-YIG nuclease family protein	1520000	1	DNA metabolism & rep
48269	MPME 1 protein	1230766.667	1	DNA metabolism & rep

3.4 Discussions

Global proteomic study of phage-bacteria interactions can be vital in elucidating the dynamic relationship between phages and host bacteria. In this study, we described the response of *M. smegmatis* to middle stage lytic infection by mycobacteriophage Ochi17 (Fig. 3.1a). Growth of the infected host did not follow the usual lytic pattern which would be (almost) complete cell lysis. Instead we observed a reduction in host cell multiplication.

Proteomics results showed more host proteins significantly upregulated than downregulated (Fig. 3.2c). This is in contrast to some previously reported transcriptomic studies where more genes were downregulated than upregulated[12, 33,34]. This could be because of the slower cell growth in ochi17-infected *M. smegmatis* resulting in protein upregulation, despite transcript reduction as seen in gene expression growth-rate dependent studies[35]. There is also the possibility that differences in infection times, phage or host type used in the various studies could impact the corresponding changes in transcript and protein expression levels.

The exclusive upregulation of proteins involved in homologous recombination and base excision/DNA repair pathways could be the result of phage Ochi17 integrating its DNA into the host chromosome through homologous recombination, or a global host response to DNA damage caused by phage Ochi17 where the host cell cycle is arrested and DNA repair is induced. Temperate phages integrate into the host chromosome via homologous recombination and depend heavily upon the host replication machinery for replication and development[36-38]. Mycobacteria use recA-dependent HR to repair double strand breaks and avoid mutation in the host genome[39]. This could also be employed as a phage resistance mechanism by *M. smegmatis* to combat integration of the temperate phage into the host chromosome[40]. STRING analysis showed the negative interaction between recX and recA which are both involved in HR and were both upregulated (Fig. 3.5a-b). By binding to recA, recX inhibits recA activity and subsequently the HR process[41]. The upregulation of recX could be a response to a passive upregulation of recA, arising from the reduced growth of *M. smegmatis*, to prevent Ochi17 integration into its DNA.

Proteins involved in transcriptional factor activities, translational activities (ribosomal proteins), amino acid metabolism, and metabolism of vitamins (Fig. 3.4a-c) were significantly regulated, with more proteins upregulated than downregulated in processes associated with the synthesis of macromolecules. More specifically, outcomes from enrichment analysis revealed that these macromolecular processes, including purine and pyrimidine metabolism were significantly

upregulated (Fig. 3.4a-c). Previous studies with phage infection have shown similar results. Enrichment of amino acid metabolism and ribosome pathways were observed after infection of *Acinetobacter baumannii* with phage ϕAbp1 at the transcript level for 10 min[18] and an elevation in the concentration of amino acids in the late phase of a lytic phage-infected *Sulfitobacter sp.* 2047 at the metabolite level was reported[42]. A recent proteome study investigating *Lactococcus lactic MG1363* response to phage p2 infection observed an increase in pyrimidine biosynthesis and protein synthesis proteins[23]. Although others observed the cessation of host macromolecular synthesis with phage infection[43], the upregulation of the metabolism of macromolecules could indicate either some form of hijacking of *M. smegmatis* macromolecular synthesis by phage Ochi17 for its own propagation (Fig. 3.1a) or a reaction from host *M. smegmatis* in the form of rapid cell multiplication.

The significant downregulation of starch and sugar metabolism, fatty acid/lipid biosynthesis, glycerol kinase activity and the upregulation of fatty acid degradation (Fig. 3.4a-c), may reflect efforts of the phage Ochi17 to weaken host metabolism and cease the synthesis of host macromolecules[14,32]. Also, more proteins involved in glycerol(phospho)lipid metabolism, and glycerol-3-phosphate metabolism were downregulated than upregulated while proteins involved in glycerol kinase activity were exclusively downregulated as well. Of all the fatty acid biosynthesis proteins identified in this study, fabD serves as the central connector (Fig. 3.6a). fabD is a critical protein in the biosynthesis of fatty acids and it is known to be actinobacterium-specific[44]. Our results showed that it was significantly downregulated (Fig. 3.6b), which could end up limiting fatty acid biosynthesis. The attack of a host by a phage usually begins at the surface of the host's cell wall and ends with the lysis of the host cell[45]. Glycerolipids are a critical component of cellular membranes and thus it is not surprising that glycerolipid metabolism was targeted. Also, mycobacterial cell walls are primarily made up of mycolic acids which are produced from the biosynthesis of fatty acids[44]. It is also possible that the cessation of the host macromolecular synthesis could result from a host abortive infection anti-phage mechanism to protect the general mycobacteria population[46].

Only two genes (rpsG and AldA) showed significant regulation of mRNA expression (Fig. 3.7). Both were at variance with the proteomics results, which showed upregulation of the proteins (Fig. 3.5b, 3.6b). Although other genes examined were not significantly changed, their expression did not match the proteomics results, with the exception of recA, rpmD and hpf (Fig. 3.7). Some

studies have reported downregulation of the expression of ribosomal proteins. More specifically, 30S and 50S ribosomal protein transcripts were reported to be downregulated after 15, 30 and 50 min of phage infection[16] and the downregulated expression of several ribosomal proteins up to 80 min after phage infection were previously reported[12]. Downregulation at the RNA transcript level suggests that only a small fraction of the host protein synthesis capacity is required for virion production in late infection stage[16]; however, this is not supported by studies that report coat protein synthesis begins after 15 min post infection[34,47]. A recent RNA-seq study revealed upregulated genes of *Acinetobacter baumannii* (AB1) were significantly enriched in ribosome pathways after phage infection[18]. Given that mRNAs are short-lived, it is possible that transcript abundance will vary periodically, and is thus insufficient in estimating a cell's metabolic activity[27]. Protein abundance studies are essential to accurately understand cellular metabolism, especially given the high stability of proteins.

The proteomic profile of phage Ochi17 in infected *M. smegmatis* revealed significant expression of 68 proteins (Table 3.1) out of 110 genes previously annotated. The high abundance of structural proteins after a 3.5-h infection of the host suggests that Ochi17 particle assembly may be happening during this infection period. We also observed that phage proteins, such as Rec and Ruv, that will assist phage Ochi17 in integrating into the host were not expressed. It could be that these proteins were of low abundance, and thus were not identified or they were not required by Ochi17 for host integration. These proteins were found in another study on the proteome expression of a lytic phage, Patience in *M. smegmatis* at both early and late stages[24]. Phages overcome the restriction-modification systems of host bacteria by acquiring cognate methylase[48,49]. Interestingly, our results also showed high expression of DNA methylase and methyltransferase Ochi17 proteins in the infected host *M.smegmatis* (Table 3.1).

In conclusion, the aim of this study was to understand how mycobacterium species, using *M. smegmatis* as a model, will respond to a mycobacteriophage infection. To the best of our knowledge, this will be the first proteome study investigating a mycobacterium response to a phage attack. We showed that phage Ochi17 works to utilize host *M. smegmatis* macromolecular synthesis for its own propagation. Also, we reported that fatty acid and lipid metabolism of *M. smegmatis* is a primary target for Ochi17 attack, to weaken host defense and probably compete in host energy utilization. Understanding phage-bacteria interactions will likely depend on several

factors such as the type of phage (lytic or temperate), the host bacterium, and the infection time points employed during experimental analysis.

3.5 Materials and Methods

3.5.1 *M. smegmatis* strain and growth conditions

Following slight modification of protocol[50], *M. smegmatis* strain mc^{2}155 colonies were inoculated in two flasks of 7H9 liquid medium containing 50 mL 7H9 Middlebrook broth supplemented with 0.05% tween-80, 1 mM calcium chloride, 10% AD supplement, 40% glycerol, 50 ug/ml carbenicillin, and 10 ug/ml cycloheximide at 37^{o}C with constant agitation at 250 rpm. After two days, 50 ul of this culture was added to a fresh 50 mL 7H9 liquid medium without tween-80, with constant agitation at 250 rpm until it reaches the exponential stage of growth (OD_{600} of 0.5- 0.7).

3.5.2 Phage infection, *M. smegmatis* growth curve and Ochi17 one-step growth curve

On reaching the exponential growth, *M. smegmatis* culture in each flask was aliquoted into three separate 15 mL conical tubes of 12 mL each and then centrifuged at 500 xg for 10 min after which 10 ml of supernatant was removed and the pellet resuspended in the remaining broth. For the phage-treated samples, the concentrated culture was then infected with phage Ochi17 at a MOI of 10 and mixed gently. For the control samples, the concentrated culture was infected with phage buffer. The mixture was incubated for 15 min at 37 ^{o}C to initiate phage adsorption to the host's cell wall, and then the volume of the conical tube was brought back to 12 ml with pre-warmed broth culture media. The contents were then transferred to sterile 50 ml flask containing several cut pipette tips. The flasks were later incubated at 37 ^{o}C with agitation at 250 rpm.

For *M. smegmatis* growth curve, 2 ml of cells was used to measure the OD600 at 0, 0.5, 1.5, 3.5, 6, 12, 18, and 24 h post-infection with a nanophotometer NP80 (IMPLEN, Westlake Village, CA, USA). 7H9 media was used as the standard.

Ochi17 one-step growth curve was obtained by collecting 10 ul of each Ochi17-infected replicate sample and making serial dilutions for each sample up to 10^{-7} according to a modified protocol[61]. Each dilution was then used for spot titer plaque assay using L-agar plates. After 24 h

of incubation, plaques were counted and used to calculate the phage forming unit per ml (pfu/ml) using the formula: (Number of plaques/volume used) * 10_3 * dilution $factor_{50}$.

3.5.3 Sample preparation, protein extraction and digestion

After incubation for ~ 3.5 h (at the point of early effect of phage Ochi17), 1.5 ml of the cells was transferred from the flask into two groups of three sterile micro centrifuge tubes. The cells were eventually pelleted at 14,000 xg for 5 min and the supernatant removed. Each pellet was later resuspended in 750 ul of phage buffer and centrifuged at 14,000 xg for 8 min. Pelleted *M. smegmatis* cells were washed with Phosphate Buffer Saline (PBS) pH 7.4 and spun at 14, 000 xg for 5 min. Supernatant was removed and pellets were solubilized in 50 ul 1X LDS sample buffer containing SDS and sonicated for 3 min and heated at 90^oC for 5 min. After cooling, 30 ul of each sample was run on 10% Bis-Tris gel (Life Technologies, Carlsbad, CA, USA) for 20 min at 120 V. The gel was then rinsed with milliQ water and stained with commassie blue for about 2 h. The gel was further destained by rinsing three times in milliQ water. Protein bands were cut from the gel and 25 mM ABC/100% ACN was added and then vortexed several times until staining solution is removed, and 100% ACN was eventually added and then removed after vortexing. The gel pieces were dried in a vacuum centrifuge for 15 min. Reduction and alkylation of cysteines were carried out on the gel pieces using 10 mM DTT in 25 mM ABC at 55^oC for 1 h, and using 55 mM IAA in 25 mM ABC at room temperature in the dark for 45 min. For the trypsin digestion, 20 µg of Lys- C/trypsin (Promega) was dissolved in 400 µL of 25 mM ABC and 50 µL of enzyme mixture was added to each sample to achieve an enzyme-to-substrate ratio of 1:25 or until the gels are below liquid level. Gel samples containing enzyme solution were then placed in a temperature-controlled shaker at 37^oC overnight. After digestion, supernatants were removed, and peptides were extracted using 60% ACN/5% trifluoroacetic acid (TFA) and were dried in vacuum centrifuge for ~2 h to prepare them for LC-MS/MS.

3.5.4 RNA extraction

Total RNA was isolated using a RNAspin Mini kit (GE Healthcare, Pittsburgh, PA, USA). Briefly, pelleted *M. smegmatis* cells were resuspended in 100 µl TE buffer containing 2 mg/ml lysozyme and lysed in 350 µl lysis solution. The lysate was then filtered to reduce viscosity and

de-salted using the RNAspin mini filter and 350 µl desalting buffer respectively. To digest DNA present in the lysate, DNase I mixture was added to the lysate and then incubated at room temperature for 15 min. After two washes, RNA was then eluted in 100 µl RNase-free water. RNA concentration was measured using a nanophotometer NP80 (IMPLEN, Westlake Village, CA, USA).

3.5.5 Liquid chromatography and tandem mass spectrometry (LC-MS/MS) data acquisition

Phage Ochi17-infected and uninfected samples were analyzed by reverse-phase high-performance liquid chromatography−electrospray ionization−tandem mass spectrometry (HPLC−ESI−MS/MS) using the Dionex UltiMate 3000 RSLC nano System (Thermo Fisher Scientific, Waltham, MA, USA) coupled to the Q-Exactive High-Field (HF) Hybrid Quadrupole Orbitrap MS (Thermo Fisher Scientific, Waltham, MA, USA) and a Nano- electrospray Flex ion source (Thermo Fisher Scientific, Waltham, MA, USA). Peptides were resuspended in 3% ACN/0.1% FA/97% MilliQ formic acid and loaded onto a trap column (300 µm ID × 5 mm) packed with 5 µm 100 Å PepMap C18 medium and washed using a flow rate of 5 µL/min with 98% purified water/2% acetonitrile (ACN)/0.01% formic acid (FA). The trap column was then switched in-line with the analytical column after 5 min. Peptides were separated using a reverse-phase Acclaim PepMap RSLC C18 (75 µm × 15 cm) analytical column using a 120 min method at a flow rate of 300 nL/min. The analytical column was packed with 2 µm of 100 Å PepMap C18 medium (Thermo Fisher Scientific, Waltham, MA, USA). Mobile phase A consisted of 0.01% FA in water, and mobile phase B consisted of 0.01% FA in 80% ACN. The linear gradient started at 5% B and reached 30% B in 80 min, 45% B in 91 min, and 100% B in 93 min. The column was held at 100% B for the next 5 min before being brought back to 5% B and held for 20 min to equilibrate the column. Sample was injected into the QE HF through the Nanospray Flex Ion Source fitted with an emission tip from Thermo Scientific. The column temperature was maintained at 35 °C. MS data were acquired with a Top 20 data-dependent MS/MS scan method. The full scan MS spectra were collected over 300−1650 m/z range with a maximum injection time of 100 ms, a resolution of 120 000 at 200 m/z, spray voltage of 2, and an AGC target of 1 × 106. Fragmentation of precursor ions was performed by high-energy C-trap dissociation (HCD) with the normalized collision energy of 27 eV. MS/MS scans were acquired at a resolution of 15 000 at m/z 200. The

dynamic exclusion was set at 20 s to avoid repeated scanning of identical peptides. Three biological sample replicates from each treatment were utilized for LC−MS/MS, which was sufficient for good statistical power. Instrument optimization and recalibration was carried out at the start of each batch run using the Pierce calibration solution.

3.5.6 Real-time quantitative polymerase chain reaction (RT-qPCR)

RT-qPCR was performed on eight differentially expressed genes from three biological and technical replicates, as listed in supplemental table S6. The cDNA was synthesized from 100 ng of the total RNA using QuantiTect® reverse transcription kit (QIAGEN, Germantown, MD, USA). RT-qPCR reactions were done using QuantStudio5™ Pro Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, USA) with SYBR Green Universal Master (Applied Biosystems, Foster City, CA, USA) in a 96-well plate. The primers used for this experiment are listed in table S6 in the supplemental material. The relative expression levels were normalized to the expression of 16S rRNA, which has been shown to be a valid and stable marker for RT-qPCR assays[12,18,51].

3.5.7 Bioinformatics and data analysis

The raw MS/MS data (.raw files) were processed using MaxQuant (v1.6.0.16) with its integrated Andromeda search engine[52,53] with the spectra matched against the Uniprot *M. smegmatis* fasta (http://www.uniprot.org) concatenated with a common contaminants database and a reverse-decoy database. The cleavage enzyme used was set as trypsin/P and LysC while allowing for up to 2 missed cleavages. MaxQuant search was done at 1% FDR at both the peptide and protein levels. The minimum peptide length required for database search was set to seven amino acids. Precursor mass tolerance of ± 10 ppm, MS/MS fragment ions tolerance of ± 0.5 Da, maximum missed cleavage for tryptic digestion was set to two, methionine oxidation and protein N-terminal acetylation (K) were set as the variable, while Carbamidomethyl (C) was set as a fixed modification. The “unique plus razor peptides” were used for peptide quantitation. Razor peptides are non-unique peptides assigned to the protein group with most other peptides.

Removing proteins with reverse identification and those identified as contaminants were the first steps in filtering the MaxQuant data. Then proteins without any LFQ intensity and without any MS/MS counts were filtered out. Also, proteins with negative score were removed. MaxQuant

results were exported to Data Analysis and Extension Tool (DAnTE) and analyzed for Pearson correlations coefficients to determine reproducibility among replicates in both control and phage-infected samples. In order to generate high confidence data, the following criteria were further used to filter the MaxQuant data: protein quantified by LFQ intensity in less than two replicates was removed; protein quantified by MS/MS counts in less than two replicates was also removed, except the protein has a total MS/MS count > 5. The LFQ intensity, used as the primary quantitative measure of abundance, was converted to log2 values and averaged across replicates. Fold change was calculated by the difference in the average log2 values between phage-infected proteins and control proteins. Two-tailed unpaired T-test was used to calculate significance values. Proteins with fold change values of < -0.5 or > 0.5 and $p < 0.05$ were considered significantly regulated by phage Ochi17 infection.

Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of significantly regulated proteins was performed using DAVID 6.8 (the Database for Annotation, Visualization and Integrated Discovery; https://david.ncifcrf.gov/). Enrichment was considered significant at a maximum modified Fisher Exact P-value (EASE Score) of 0.1.

3.6 References

1. Fu, X., Ding, M., Zhang, N. & Li, J. Mycobacteriophages: an important tool for the diagnosis of *Mycobacterium tuberculosis* (review). *Mol Med Rep* **12**, 13-19 (2015).
2. Bloom, B.R. Tuberculosis: pathogenesis, protection and control. *Nat Med* **2**, 244-245, doi: 10.1038/nm0296-244b (1995).
3. Gordon, S.V. & Parish, T. Microbe profile: *Mycobacterium tuberculosis*: humanity's deadly microbial foe. *Microbiol* **164**, 437-439 (2018).
4. Hatfull, G. The secret lives of mycobacteriophages. *Adv Virus Res* **82**, 179-288 (2012).
5. Broxmeyer, L. *et al*. Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a mycobacteriophage delivered by a nonvirulent mycobacterium: a model for phage therapy of intracellular bacterial pathogens. *The J Infectious Disease* **186**, 1155-1160 (2002).
6. World Health Organization. Global tuberculosis report. *WHO* 5-147 (2017).
7. World Health Organization. Global tuberculosis report. *WHO* **15**, 1-261 (2019).

8. Satish, R. & Desouza, A. Study of characteristics of mycobacteriophage- a novel tool to treat *Mycobacterium spp. Int J Mycobacteriol* **8**, 170-174 (2019).
9. Hatfull, G. Mycobacteriophages: windows into tuberculosis. *PLoS Pathogens* **10**, 12-14 (2014).
10. Brussow, H. & Hendrix, R.W. Phage genomics: small is beautiful. *Cell* **108**, 13-16 (2002).
11. Nechaev, S. & Severinov, K. Bacteriopage-induced modifications of host RNA polymerase. *Annu Rev Microbiol* **57**, 301-322 (2003).
12. Zhao, X., *et al*. Global transcriptomic analysis of interactions between *Pseudomonas aeruginosa* and bacteriophage PaP3. *Sci Reports* **6**, 19237, doi: 10.1038/srep19237 (2016).
13. Samson, J.E., Magadan, A.H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defenses. *Nature Rev Microbiol* **10**, 675-687 (2013).
14. De Sordi, L., Lourenco, M. & Debarbieux, L. The battle within: interactions of bacteriophages and bacteria in the gastrointestinal tract. *Cell Host Microbe* **25**, 210–218 (2019).
15. Poranen, M.M., *et al*. Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* **80**, 8081–8088 (2007).
16. Ravantti, J.J., Ruokoranta, T.M., Alapuranen, A.M. & Bamford, D.H. Global transcriptional responses of Pseudomonas aeruginosa to phage PRR1 infection. *J Virol* **82**, 2324–2329 (2008).
17. Leskinen, K., Blasdel, B.G., Lavigne, R. & Skurnik, M. RNA-sequencing reveals the progression of phage-host interactions between ϕR1-37 and Yersinia enterocolitica. *Viruses* **8**, 111 doi: 10.3390/v8040111 (2016).
18. Yang, Z. *et al.* global transcriptomic analysis of the interactions between phage φAbp1 and extensively drug-resistant *Acinetobacter baumannii*. *mSystems* **4**, e00068-19 (2019).
19. Mojardin, L., & Salas, M. Global transcriptional analysis of virus-host interactions between phage ϕ29 and *Bacillis subtilis*. *J Virol* **90**, 9293-9304 (2016).
20. Sacher, J.C. *et al*. Transcriptomic analysis of the *Campylobacter jejuni* response to T4-like phage NCTC 12673 infection. *Viruses* **10**, 332, doi:10.3390/v10060332 (2018).
21. Lavigne, R. *et al.* A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *mBio* **4**, e00061-13 (2013).
22. Weintraub, S.T. *et al*. Global proteomic profiling of *Salmonella* infection by a giant phage. *J Virol* **93**, e01833-18, doi: 10.1128/JVI.01833-18 (2019).
23. Lemay, M-L. *et al*. Investigating Lactococcus lactis MG1363 response to phage p2 infection at the proteome level. *Mol Cel Prot* **18**, 704-714 (2019).

24. Pope, W.H. *et al*. Genomics and proteomics of Mycobacteriophage Patience, an accidental tourist in the *Mycobacterium* neighborhood. *mBio* **5**, 1-11 (2014).
25. Maynard, N.D. *et al*. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. *PLoS Genet* **6**, e1001017, doi: 10.1371/journal.pgen.1001017 (2010).
26. Cisek, A.A., Dabrowska, I., Gregorczyk, K.P. & Wyzewski, Z. Phage therapy in bacterial infections treatment: one hundred years after the discovery of bacteriophages. *Curr Microbiol* **74**, 277-283 (2017).
27. Schwanhäusser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
28. Feder, M.E. & Walser, J.C. The biological limitations of transcriptomics in elucidating stress and stress responses. *J Evol Biol* **18**, 901-910 (2005).
29. Knipe, D.M. & Howley, P.M. Fields virology: vol 1. 5th ed. *Lippincott Williams & Wilkins* ISBN 9780781760607 OCLC 71812790 (2007).
30. Aryal, U.K., McBride, Z., Chen, D., Xie, J. & Szymanski, D.B. Analysis of protein complexes in Arabidopsis leaves using size exclusion chromatography and label-free protein correlation profiling. *J Proteomics*, **166**, 8-18 (2017).
31. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc* **4**, 44-57 (2009).
32. Kinsella, R.J., Fitzpatrick, D.A., Creevey, C.J. & Mclnerney, J.O. Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution, and gene duplication. *Proc Natl Acad Sci USA* **100**, 10320-10325 (2003).
33. Sacher, J.C. *et al.* Transcriptomic analysis of the *Campylobacter jejuni* response to T4-like phage NCTC 12673 infection. *Viruses* **10**, 332, doi:10.3390/v10060332 (2018).
34. Zhao, X. *et al.* Transcriptomics and metabolomics profiling of phage-host interactions between phage PaP1 and *Pseudomonas aeruginosa Front Microbiol* **8**, 548 (2017).
35. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366-1375 (2009).
36. Friedman, D.I. *et al*. Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. *Microbiol Rev* **48**, 299-325 (1984).

37. Worobey, M. & Holmes, E.C. Evolutionary aspects of recombination in RNA viruses. *J Gen Virol* **80**, 2535–2543 (1999).
38. Osterhout, R.E., Figueroa, I.A., Keasling, J.D. & Arkin, A.P. Global analysis of host response to induction of a latent bacteriophage. *BMC Microbiol* **7**, 82, doi: 10.1186/1471-2180-13-183 (2007).
39. Gupta, D.K., Patra, A.T., Zhu, L., Gupta, A.P. & Bozdech, Z. DNA damage regulation and its role in drug-related phenotypes in the malaria parasites. *Sci Rep* **6**, 23603, doi: 10.1038/srep23603 (2016).
40. Cirz, R.T. *et al*. Inhibition of Mutation and Combating the Evolution of Antibiotic Resistance. *PLoS Biol* **3**, e176 (2005).
41. Stohl, E.A. *et al.* Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo. *J Biol Chem* **278**, 2278-2285 (2003).
42. Ankrah, N.Y.D. *et al*. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *The ISME J* **8**, 1089-1100 (2014).
43. Koerner, J.F. & Snustad, D.P. Shutoff of host macromolecular synthesis after T-even bacteriophage infection. *Microbiol Rev* **43**, 199-223 (1979).
44. Seed, K. Battling phages: how bacteria defend against viral attack. *PLoS Pathogens* **11**, e1004847, doi:101371/journalppat1004847 (2015).
45. Labrie, S.J., Samson, J.E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Rev Microbiol* **8**, 317-327 (2010).
46. Beremand, M.N., Blumenthal, T. Overlapping genes in RNA phage: a new protein implicated in lysis. *Cell* **18**, 257-266 (1979).
47. Vogel, C. & Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**, 227-232 (2012).
48. McGrath, S., Seegers, J.F.M. L., Fitzgerald, G.F. & van Sinderen, D. Molecular characterization of a phage-encoded resistance system in *Lactococcus lactis*. *Appl Environ Microbiol* **65**, 1891–1899 (1999).
49. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**, 43–51 (2011).
50. Poxleitner, M., Pope, W., Jacobs-Sera, D., Sivanathan, V. & Hatfull, G. Phage discovery guide for use with Mycobacterium smegmatis. *HHMI* 1-118 (2016).

51. Shi, L., Jung, Y.J., Tyagi, S., Gennaro, M.L. & North, R.J. Expression of Th1-mediated immunity in mouse lungs induces a Mycobacterium tuberculosis transcription pattern characteristic of nonreplicating persistence. *Proc Natl Acad Sci USA* **100**, 241–246, doi: 10.1073/pnas.0136863100 (2003).
52. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein identification. *Nat Biotechnol* **26**, 1367-1372 (2008).
53. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526 (2014).

4. TEMPORAL PROTEOMICS ANALYSIS OF MYCOBACTERIOPHAGE-MYCOBACTERIA INTERACTION

4.1 Abstract

For mycobacteriophages to serve as effective therapeutic agents in the treatment of mycobacterial infections, a robust understanding of mycobacteriophage-mycobacteria interaction is vital in elucidating how mycobacteria attempt to evade mycobacteriophage attack throughout its life cycle. In chapter 3, the study on mycobacteriophage-mycobacteria interaction focused on a single infection time point, where for the first time we reported the molecular impact of a mycobacteriophage on any mycobacterial host. However, we discovered that the physiological effects of Ochi17 on the mycobacterium host varied over the entire life cycle of the host. In this study, we investigated the temporal mycobacteriophage-mycobacteria interaction by applying label-free quantitative proteomics on the host, *Mycobacterium smegmatis*, infected with the temperate phage, Ochi17 for 48 h. Proteomics analysis was done at 0h, 6 h, 12 h and 24 h post infection time points. A total of 2,181 host proteins and 56 Ochi17 proteins were identified across all the time points. We showed that phage Ochi17 proteome expression is time-dependent and the proteins were grouped into three clusters based on expression time. Our results showed that host lysis and lysis-promoting Ochi17 proteins were majorly expressed together, especially during the middle stage, and aggressively at the late stage of infection. Proteomic results suggest that Ochi17 suppresses the growth of *Mycobacterium smegmatis* not just by hijacking the macromolecular synthesis of the host, but also by suppressing host transcription, two-component system and degrading fatty acids. This study may contribute in the application of mycobacteriophages as therapeutic agents on *Mycobacterium tuberculosis* focusing on specific targets.

4.2 Introduction

Mycobacterium tuberculosis, a human pathogenic bacterium, is the causative agent of tuberculosis[1]. Despite the global use of several antibiotics and vaccines, tuberculosis (TB) is still one of the deadliest infectious diseases in the world, killing about 1.6 million people yearly[2,3]. Due to the overuse of antibiotics, TB has become resistant to drugs such as rifampicin that were previously effective against it[3]. As a result of this evolution and emergence of multi-drug resistant

TB, it is important to consider alternative therapeutic agents that could be used either in place of antibiotics or in conjunction with antibiotics in treating TB[4]. One of such option is the use of mycobacteriophages, viruses that specifically attack mycobacterial hosts[4,5].

Since the 1940s, many mycobacteriophages have been isolated and characterized[6]. Integrated research and education programs like the Phage Hunters Integrating Research and Education (PHIRE) and the Howard Hughes Medical Institute Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (HHMI SEA- PHAGES), have focused on isolating and characterizing mycobacteriophages using mycobacterial hosts, especially *Mycobacterium smegmatis*[5]. We have previously shown that a novel temperate mycobacteriophage that was isolated at Purdue University in 2017 had the ability to infect and cause molecular disruptions in *Mycobacterium smegmatis* at the translational level (chapter 3). We also showed the necessity of using proteomics to study phage-bacteria interactions to give a comprehensive understanding of this interaction in addition to previous transcriptomic and metabolomic studies[7-10].

Initial growth studies of Ochi17-infected *Mycobacterium smegmatis* revealed significant suppression of cell replication after 6 h of infection (Fig. 4.1). Therefore, we decided to investigate the proteome level interactions between both organisms to gain an understanding of the temporal impact of Ochi17 on *Mycobacterium smegmatis* as well as the temporal expression of Ochi17 proteins. We hypothesized that the significant suppression of the growth of *Mycobacterium smegmatis* with Ochi17 after 6 h is also due to the downregulation of proteins involved in macromolecular synthesis as was previously reported at 3.5 h (chapter 3). Our results showed that phage Ochi17 proteome expression is cluster-based and dependent on the stage of infection. Consequently, this produces a time-dependent response from the host. We also suggested that Ochi17 suppresses growth of *Mycobacterium smegmatis* by downregulating the host proteins involved in DNA transcription, while promoting host protein synthesis for its own propagation. This temporal proteomic analysis gave a broader and full picture of the interactions between Ochi17 and *Mycobacterium smegmatis*.

4.3 Results

4.3.1 *M. smegmatis* infection and growth curve

In chapter 3, I reported the growth dynamics of phage Ochi17-infected *M. smegmatis* cells for the first 30 h, and consequently studied their early interactions at a single time point. In this study, I decided to investigate the Ochi17- *M. smegmatis* interactions over the course of the life cycle of the host *M. smegmatis*. To start with the investigation, the growth dynamics of phage Ochi17-infected *M. smegmatis* for an extended period of up to 48 h was first examined (Fig. 4.1). I observed the same growth dynamics as previously reported in chapter 3 for the first 30 h of Ochi17 infection (Fig. 4.1). Not much difference was observed until 6 h post infection, showing that phage Ochi17 had relatively little or no effects in the first 6 h of infection. Following 6 h post infection, a huge difference in cell replication between the control and Ochi17-infected *M. smegmatis* was observed. At 24 h post infection, a reduction in cell replication relative to previous time points was observed. However, this reduction was not observed after 30 h post infection (Fig. 4.1).

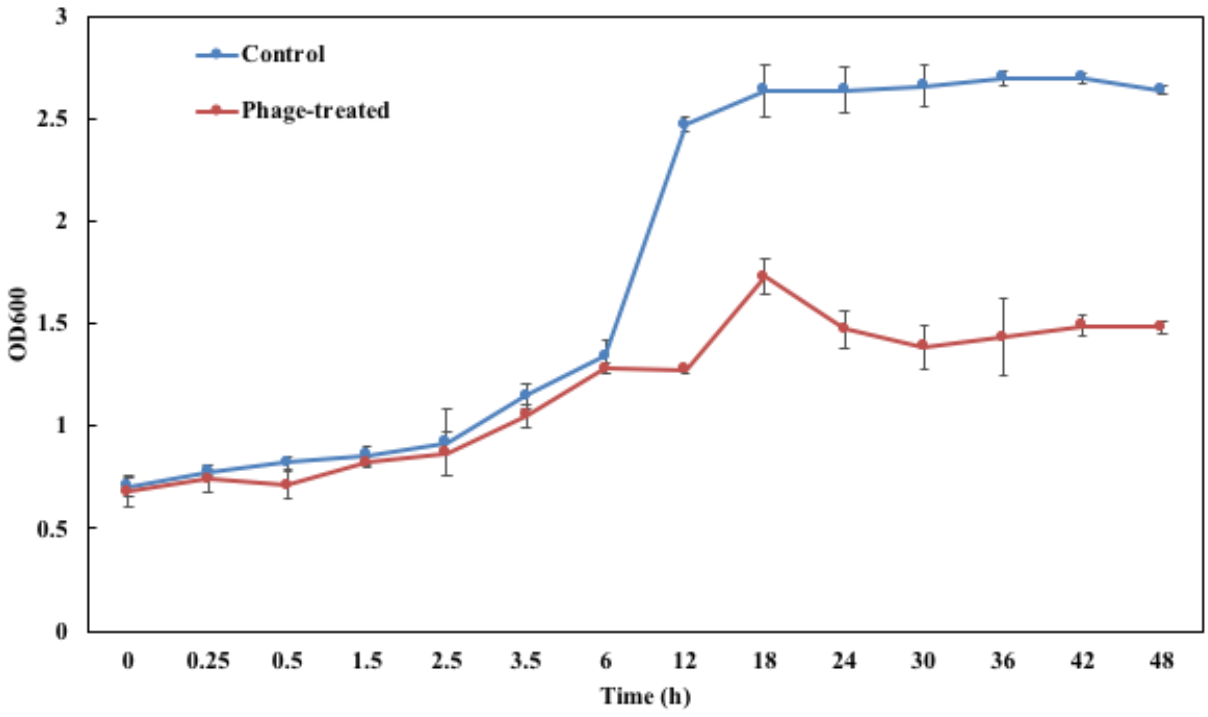


Figure 4.1. *M. smegmatis* infection growth curve. Growth curves of non-infected *M. smegmatis* (control) and *M. smegmatis* infected with phage Ochi17 (phage-treated). The data represent the mean ± SD of three replicates for each treatment. Phage-treated and control samples for proteomics analysis were collected in triplicates at 0, 6, 12, and 24 h. Time point 0 h for the treated samples (T0) represents time after pre-attachment incubation with Ochi17 but before actual infection incubation.

4.3.2 Experimental design and MS/MS reproducibility

In order to investigate the global proteomics responses during the interaction over the life cycle of *M. smegmatis*, samples in triplicates at 0 h, 6 h, 12 h, and 24 h post infection time points based on the growth dynamics were collected (Fig. 4.1; 4.2a). 0 h post infection represents initial phage attachment on host cell wall before infection. I picked 6 h to examine the difference in interaction relative to 0 h. Time point of 12 h was used since it was the next time point following 6 h where the large difference in growth was observed. I also chose 24 h time point since it was the first time point where growth reduction began (Fig. 4.1).

Correlation analysis of the label-free quantitation (LFQ) intensities of the control and Ochi17-treated replicates for each of the four time points (0 h, 6 h, 12 h and 24 h) showed very high Pearson correlation coefficients. This confirmed high reproducibility among the replicates of each treatment, and variations among the different treatments (control v. Ochi17-treated) for each time point with the exception of 0 h (Fig. 4.2b).

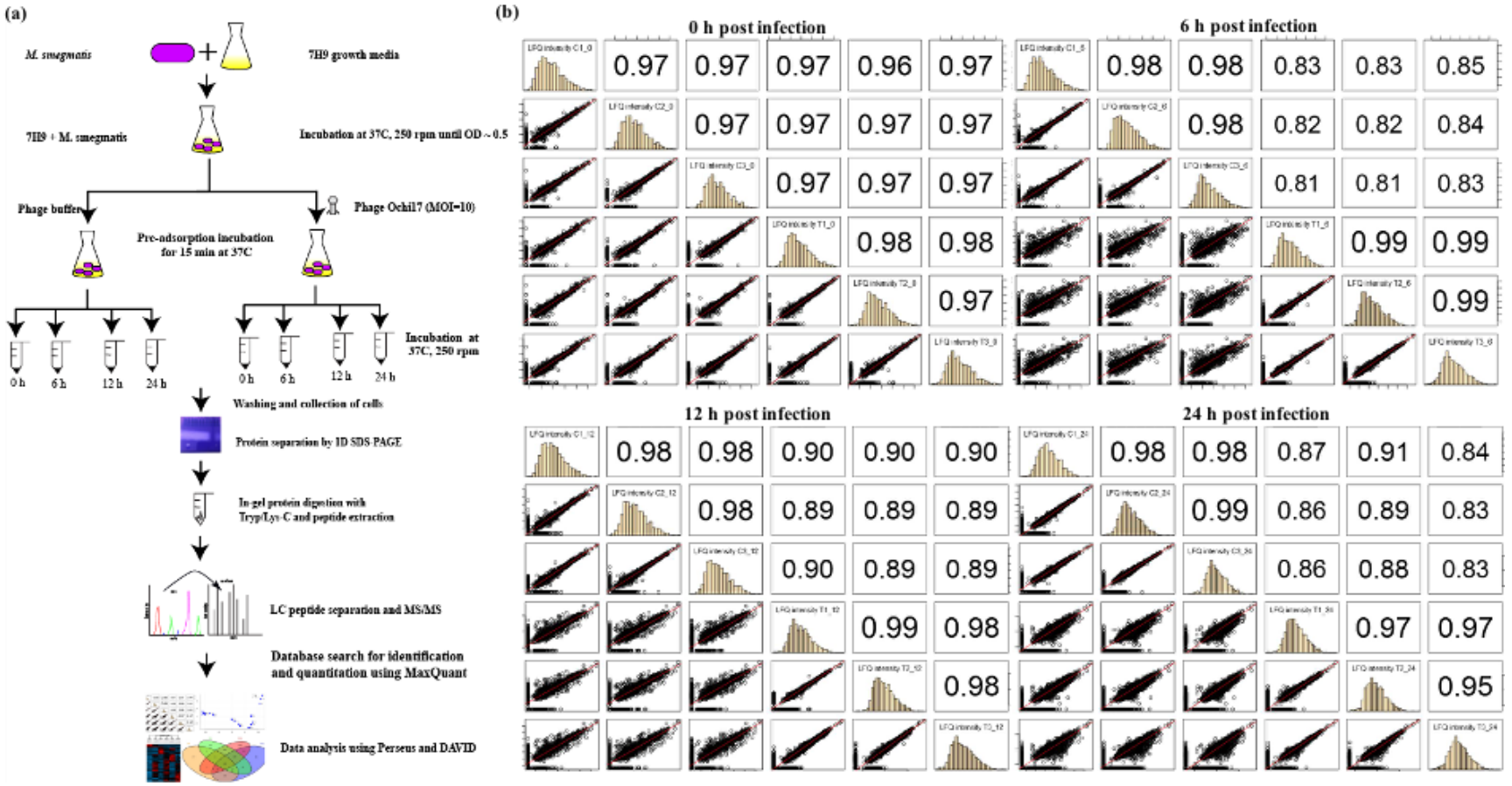


Figure 4.2. Experimental design for proteomics analysis. (**a**) Workflow for sample collection and proteomics analysis of the collected samples at the designated time points. MOI- multiplicity of infection; LC- Liquid Chromatography; MS/MS- tandem mass spectrometry (**b**) Correlation analysis of the LFQ intensities of the replicates for the control and Ochi17-infected *M. smegmatis* samples for 0 h (top left), 6 h (top right), 12 h (bottom left) and 24 h (bottom right) time points.

4.3.3 Phage Ochi17 global proteome analysis

The variation among the different timepoints was represented using principal component analysis (PCA) plot (Fig. 4.3a). The PCA plot showed that the four timepoints separated into four distinct clusters, confirming that proteomic expression evolves with infection stage (Fig. 4.3a). A total of 56 phage Ochi17 proteins were identified across all timepoints (Fig. 4.3b). Out of these, 30 proteins were common in all timepoints. Four proteins (integrase, tail assembly chaperone, and two uncharacterized proteins) were found only in 24 h post infection (T24) but not at any other time points. These proteins have multiple MS/MS counts in each replicate (Fig. 4.3b; Table 1). Also, five proteins were found in all time points except at T24. They include two proteins found in only 6 h and 12 h post infection (T6 and T12)- DNA methylase and a protein of unknown function, two proteins among T0, T6 and T12- minor tail and terminase small subunit proteins, and one protein between 0 h and 6 h- immunity repressor (Fig. 4.3b; Table 1). Further investigations showed that these five proteins also have multiple MS/MS counts in each replicate. The top three most abundant proteins across all the time points were lysin B, minor tail, and D-ala-D-ala carboxypeptidase proteins (Table 4.1).

All 56 proteins were later classified into functional groups based on the annotated functions of their genes. Based on the number of proteins in each group, particle formation made up 32% of the proteins followed by host interaction (16%), and then DNA metabolism and replication (11%) and host lysis (4%). The remaining 37% of the proteins were proteins of unknown functions (Fig. 4.3c). In order to understand the effects of infection time on phage protein expression, we examined how each of these functional groups were represented at each time point. The same number of host lysis proteins, and almost equal number of DNA metabolism and replication

proteins were present at all time points (Fig. 4.3d). However, proteins involved in host interaction and particle formation, and also unknown proteins were observed to be more abundant in 6 h, 12 h, and 24 h time points. Host interaction proteins were more abundant at 6 h, while particle formation proteins were more abundant at 24 h (Fig. 4.3d).

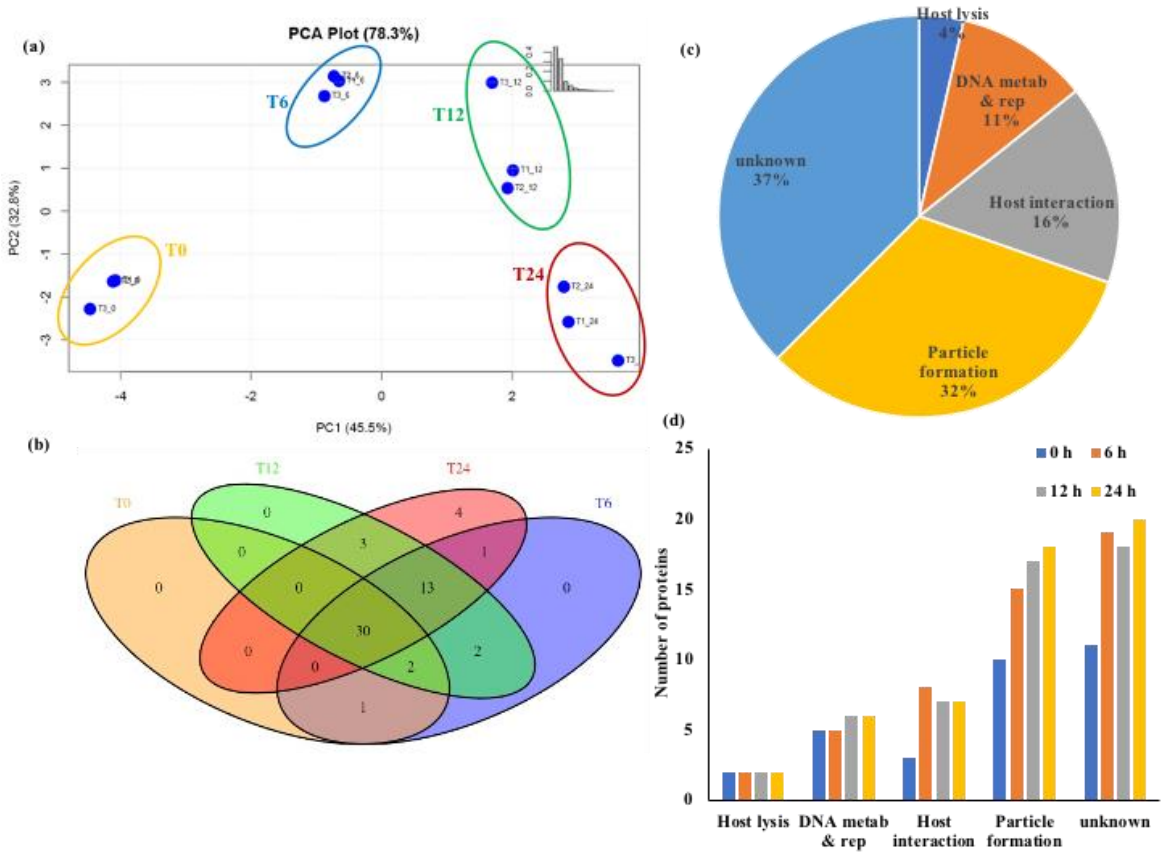


Figure 4.3. Phage Ochi17 proteome analysis. (**a**) PCA plot of phage Ochi17 proteins identified at the different time points in the phage-treated samples. T0: 0 h post infection; T6: 6 h post infection; T12: 12 h post infection; T24: 24 h post infection (**b**) Venn diagram showing the distribution of all the phage Ochi17 proteins identified in the study among the different timepoints (**c**) Classification of all 56 Ochi17 proteins identified at all the four timepoints. DNA metab & rep: DNA metabolism & replication (**d**) Bar chart showing the number of proteins represented in each functional classification at each timepoint.

Table 4.1. LFQ intensities of phage Ochi17 proteins identified across all infection time points. Intensities are average of three replicates. DNA metab & rep = DNA metabolism & replication.

Phage Ochi17 proteins	0 h	6 h	12 h	24 h	Classification
Lysin B	462170000	886073333	943583333	648190000	Host lysis
D-ala-carboxypeptidase	549063333	1372000000	834526667	464673333	Host interaction
Minor tail protein	321453333	552016667	511663333	429253333	Particle formation
Galactosyltransferase	284200000	309826667	343930000	330633333	Host interaction
Uncharacterized protein	8343867	239966667	371536667	328596667	Unknown
DNA polymerase III sub	150423333	118326667	193896667	262350000	DNA metab & rep
Uncharacterized protein	76533333	83256000	156313333	194953333	Unknown
Uncharacterized protein	9778167	118956000	185373333	161470000	Unknown
AAA-ATPase	99587667	142336667	146990000	160340000	DNA metab & rep
Uncharacterized protein	11204000	65079333	102743667	123083333	Unknown
Uncharacterized protein	47776000	48713333	54342667	114973333	Unknown
Terminase large subunit	139273333	299286667	322660000	103840000	Particle formation
Minor tail protein	48096667	58819000	64449667	89217000	Particle formation
Uncharacterized protein		68495667	84747667	83942333	Unknown
Antirepressor		31216000	46416000	80955000	Host interaction
Tail assembly chaperone		11084533	60038000	79447000	Particle formation
Head-to-tail adaptor	57322000	57348000	50700333	75804333	Particle formation
Glycosyltransferase	188510000	33783333	37194000	75645667	Host interaction
Minor tail protein		15630667	40677333	74620000	Particle formation
Minor tail protein		22807000	43395667	74508667	Particle formation
Uncharacterized protein	9035000	46665667	81256667	69519333	Unknown
Portal protein	98289667	83411333	68865000	68788000	Particle formation
Lysin A	11283000	56593000	66438000	67321000	Host lysis
Uncharacterized protein	24823333	26401333	44497333	62750667	Unknown
Uncharacterized protein	4631633	16257000	40065333	56385667	Unknown
DUF1508 domain-containing protein	13730333	38570333	64455667	51210000	Unknown
G-I-YY-I-G endonuclease	358083333	198230000	187640000	47341967	DNA metab & rep
Capsid maturation protease	107186667	98793333	83843000	47140333	Particle formation
Major tail protein	19264667	119396667	108653333	44998000	Particle formation
Uncharacterized protein		21155667	34359333	42971000	Unknown
Glycosyltransferase		14957000	27835333	38746000	Host interaction
Uncharacterized protein	256180000	21724000	29248667	35545000	Unknown
Major capsid protein			34707333	34520667	Particle formation
Uncharacterized protein		21335333	27701000	33973667	Unknown

Tail terminator	55871667	54392333	39119333	25947333	Particle formation
Uncharacterized protein		7960600	15404333	25744667	Unknown
WhiB family transcription factor	153213333	265413333	191523333	20228067	DNA metab & rep
Uncharacterized protein	26776333	63691333	50888333	17070167	Unknown
Tape measure protein	19277000	36446333	10751667	16918967	Particle formation
Scaffolding protein			12253033	15142000	Particle formation
HNH endonuclease			10721833	14396767	DNA metab & rep
Cro protein		31986667	25854667	10313000	Host interaction
Uncharacterized protein				9775667	Unknown
Uncharacterized protein		4148267		9544700	Unknown
Minor tail protein		26728667	10351133	9470333	Particle formation
Uncharacterized protein		5752300	7551733	7357333	Unknown
Uncharacterized protein				7269333	Unknown
Integrase				7148667	Host interaction
Tail assembly chaperone				6769367	Particle formation
HNH endonuclease	24374000	13066667	9254833	5862133	DNA metab & rep
Uncharacterized protein		4371333	6224867	4445267	Unknown
Minor tail protein	580123333	144058000	261627333		Particle formation
DNA methylase		11530333	9675467		Host interaction
Uncharacterized protein		4141433	6904033		Unknown
Terminase small subunit	52031667	9373800	4515200		Particle formation
Immunity repressor	6859200	4201700			Host interaction

4.3.4 Proteomics analysis revealed time-dependent expression of Ochi17 proteome

All the identified phage Ochi17 proteins were represented in a heatmap and showed the clustering of the phage proteins into three groups of early, middle and late expression proteins (Fig. 4.4). The early proteins were those highly expressed at 0 h and 6h but had little or no expression at 12 and 24 h. some of these proteins include glycosyltransferase, terminase small subunit, G-I-Y Y-I-G endonuclease, HNH endonuclease, and immunity repressor, among others (Fig. 4.4). The middle-expressed proteins were those highly expressed at 6 and 12 h, with little or no expression at 0 and 24 h. They include tape measure, minor tail, DNA methylase, D-ala-D-ala carboxypeptidase, lysin A, lysin B, WhiB family transcription factor, among others (Fig. 4.4). Most of the proteins were classified as late expressed proteins due to their high expression at 24 h with little or no expression at the other time points. Some of them include integrase, tail assembly

chaperone, antirepressor, galactosyltransferase, DNA polymerase subunit III subunit, and many uncharacterized proteins (Fig. 4.4). The genes located on the 3' end of Ochi17 genome were mostly clustered as late expression genes (Fig. 4.4). MS/MS spectra of selected high scoring peptides used for the identification of some of these Ochi17 expressed proteins (lysin A, lysin, integrase, and tail assembly chaperone) showed a good match between the theoretical and experimental masses, thus leading to efficient protein identification (Appendix D)

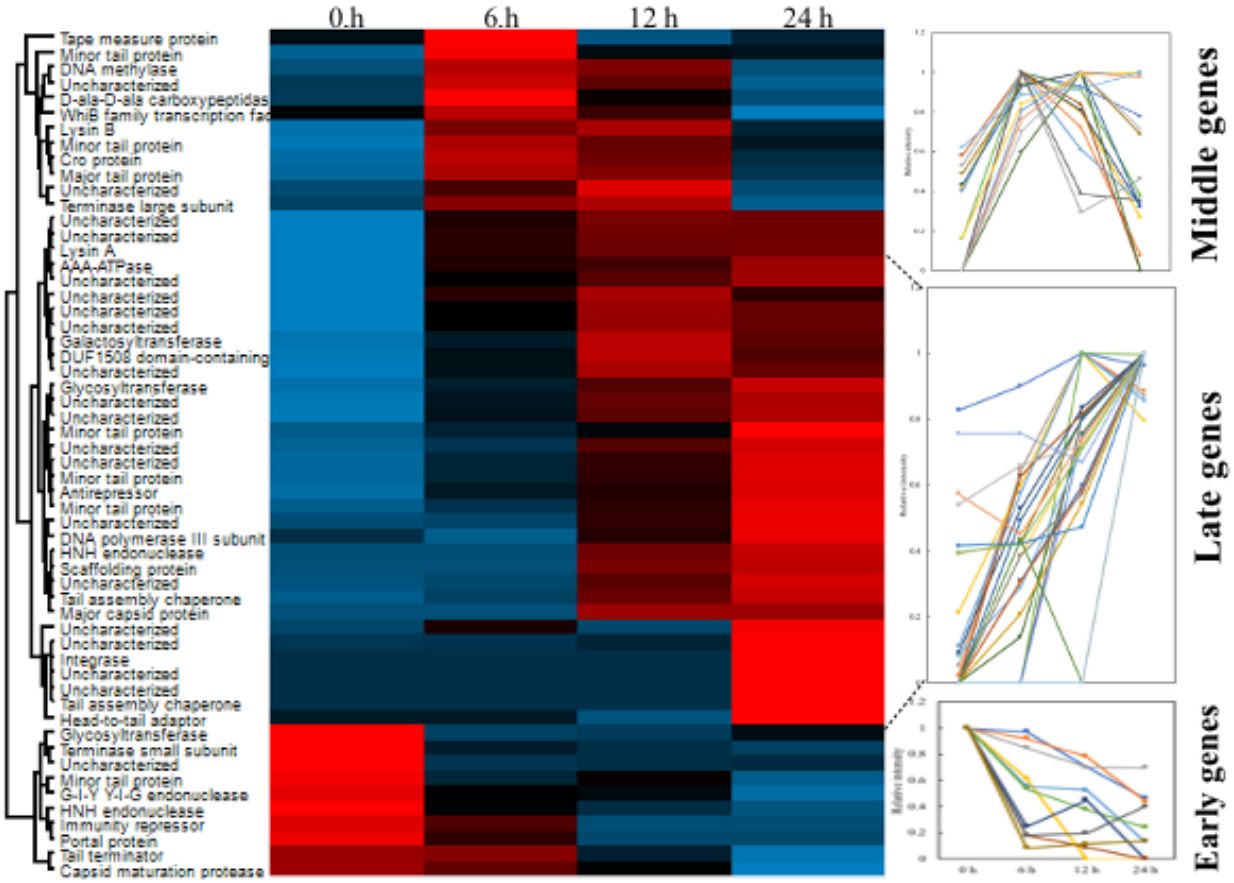


Figure 4.4. Profile of all the identified phage Ochi17 proteins. Heatmap showing the average LFQ intensities as a representation of protein expression at each of the four time points- 0 h, 6 h, 12 h and 24 h. Profile charts on the right side depict relative quantitative abundance of each protein. The phage proteins were clustered into three clusters (early, middle and late) based on time of expression.

4.3.5 Host *M. smegmatis* global proteome analysis

Proteins were counted as identified if they were found in at least two replicates for each control and treated sample. PCA clustering showed the variations among the treatment groups for each time point. The control and treated groups at 0 h showed little or no difference with the control

group at 6 h (Fig. 4.5a). A total of 2181 unique *M. smegmatis* proteins were identified across all timepoints. Out of these, 924 proteins were common in all four time points. Specifically, 1017 proteins were identified in 0 h, 1712 proteins in 6 h, 1849 proteins in 12 h, and 1959 proteins in 24 h (Fig. 4.5b; Table S2).

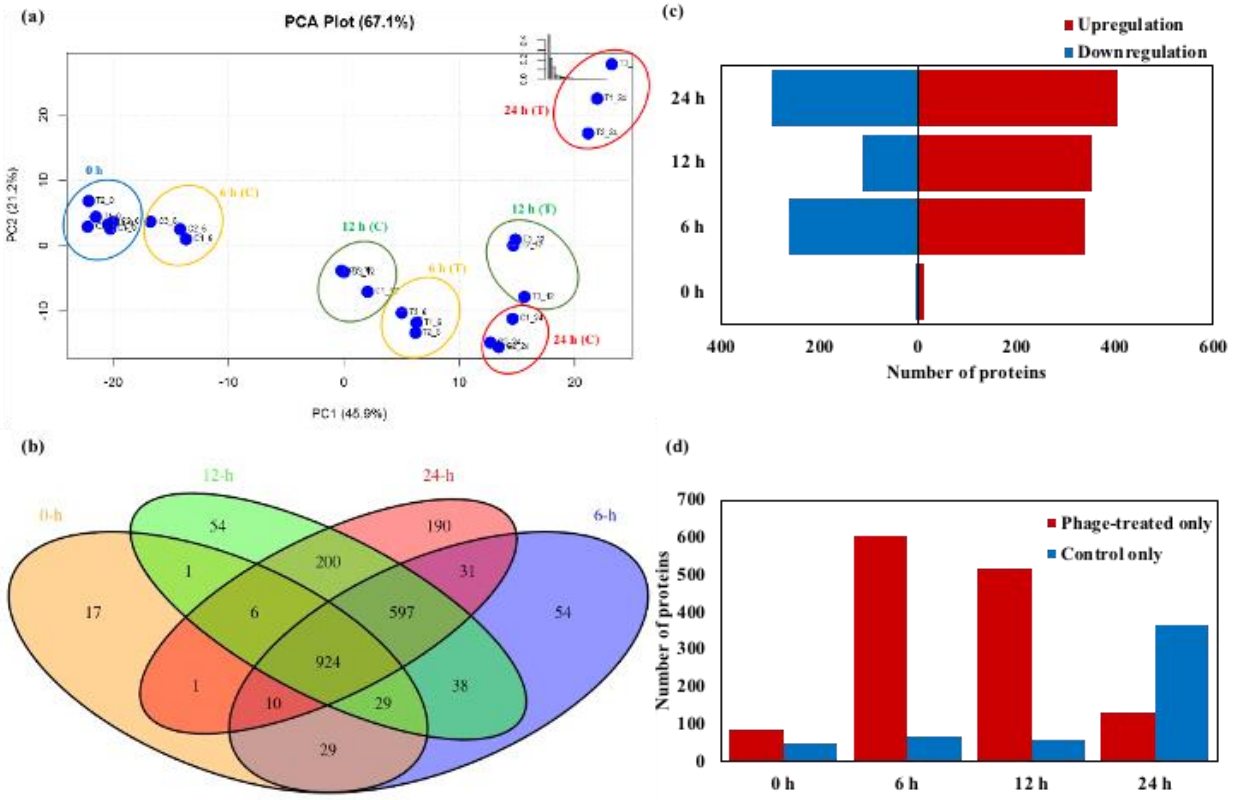


Figure 4.5. Global proteome composition of phage Ochi17-infected *M. smegmatis*. (**a**) PCA plot of *M. smegmatis* proteins identified at the different time points showing variation in treatment C: control; T: Ochi17-infected/-treated (**b**) Venn diagram showing the total number of proteins identified across the four different time points of 0 h, 6 h, 12 h and 24 h (**c**) Number of significantly differentially regulated proteins identified after Ochi17 infection of *M. smegmatis* at the four different time points ($p \leq 0.05$) (**d**) Number of unique proteins identified exclusively in phage Ochi17-treated and control *M. smegmatis* at each time point.

At 0 h post infection, only 17 proteins were differentially regulated- 12 upregulated and 5 downregulated, while 85 proteins and 47 proteins were only found in Ochi17-treated and control respectively (Fig. 4.5c, 4.5d). At 6 h post infection, 340 proteins were upregulated and 261 were downregulated, while 604 proteins were only found in the treated samples and 65 proteins only in the control samples. At 12 h, 353 and 114 proteins were upregulated and downregulated respectively, while 518 and 56 proteins were exclusively identified in the treated and control samples respectively (Fig. 4.5c, 4.5d). Meanwhile, 407 proteins were differentially upregulated

and 297 were differentially downregulated at 24 h (Fig. 4.5c). Compared to other time points, more proteins were exclusively found in the control samples than in the treated samples- 364 and 132 proteins respectively (Fig. 4.5d)

4.3.6 Functional analysis of differentially expressed proteins

The proteins that were differentially regulated and exclusively found in either the control or phage-treated samples for each time point were represented in a heatmap (Fig. 4.6). Overall, we observed a consistent expression of identified proteins across the replicates of the different experimental groups (control and treated). Using DAVID 6.8[11], significantly differential proteins at each of the infection time point were classified into Gene Ontology terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Fig. 4.7a-d). At 0 h post infection, KEGG pathways ABC transporters, ribosome, and valine, leucine and isoleucine degradation, were the most represented pathways among upregulated proteins. Most represented KEGG pathways among downregulated proteins include glycolysis/gluconeogenesis, and purine metabolism (Fig. 4.7a). More proteins were upregulated in all the pathways with the exception of the glycolysis/gluconeogenesis pathway.

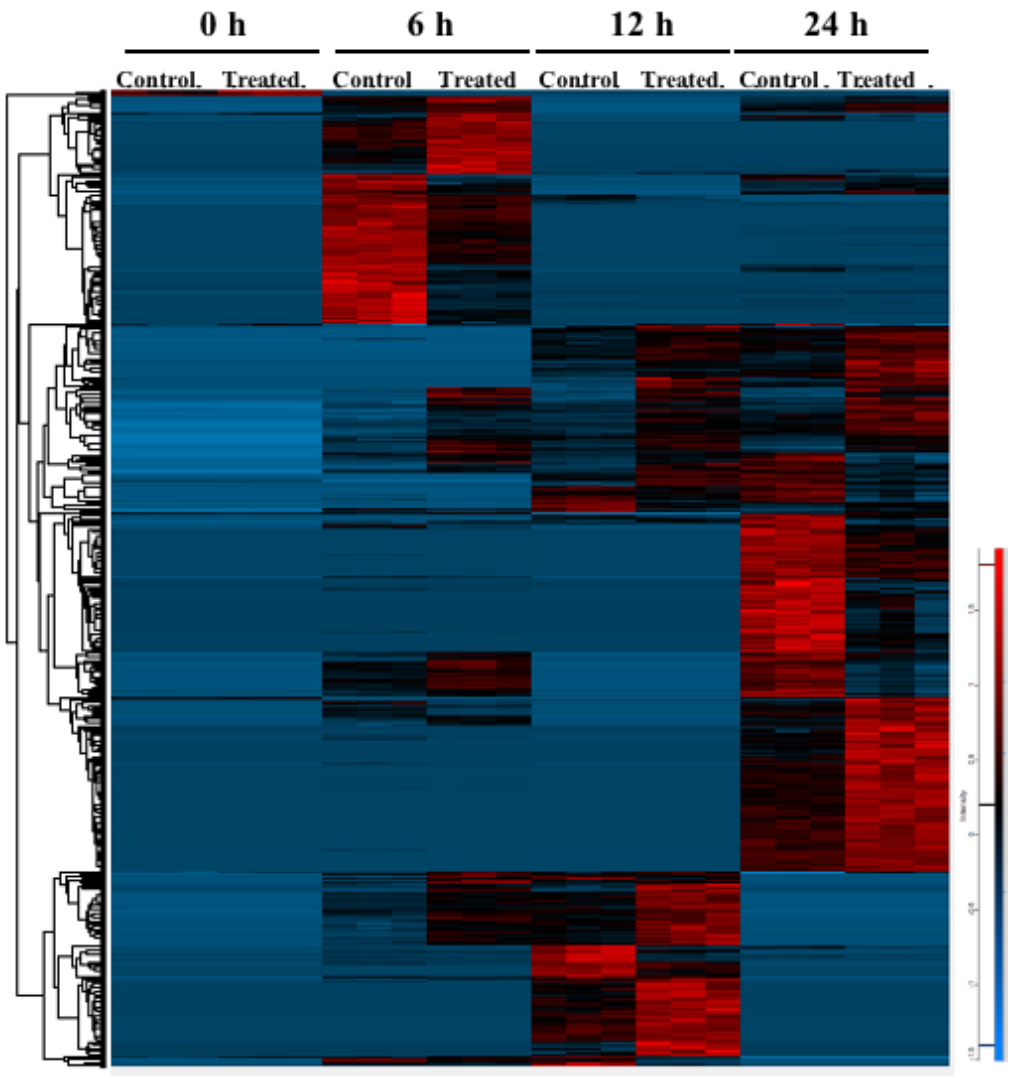


Figure 4.6. Heatmap showing significantly regulated proteins at each time point after phage Ochi17 infection of *M. smegmatis*. Each control and treated group under each time point comprised three replicates. Blue color represents proteins either not detected or with little expression in a replicate.

At 6 h post infection, the top upregulated and downregulated KEGG pathways include ABC transporters, ribosome, glycolysis/gluconeogenesis, purine and pyrimidine metabolism, among others. More proteins were found in all the upregulated pathways than in downregulated pathways. Two component system, DNA replication, nucleotide/base excision repair, streptomycin biosynthesis were only represented in the upregulated proteins (Fig. 4.7b). The same top regulated pathways identified at 6 h were also represented at 12 h post infection (Fig. 4.7c). However, at 12 h post infection, only two ribosomal proteins were downregulated, while 24 were upregulated. Also, fatty acid degradation, pentose phosphate pathway, two-component system,

pyrimidine metabolism, folate biosynthesis, glycerophospholipid metabolism, pantothenate and CoA biosynthesis and metabolic pathways were only represented in the upregulated proteins.

Examination of proteins regulated at 24 h post infection showed that downregulation of top KEGG pathways represented by the proteins identified were reported to be upregulated at 6 and 12 h post infection. They included ABC transporters, purine metabolism, glycolysis/gluconeogenesis, two-component system, among others (Fig. 4.7d). We also reported upregulation of more proteins involved in ribosome and fatty acid degradation. Proteins involved in glycerophospholipid metabolism, amino sugar and nucleotide sugar metabolism, and ubiquinone and other terpenoid-quinone biosynthesis were only represented in downregulated proteins (Fig. 4.7d).

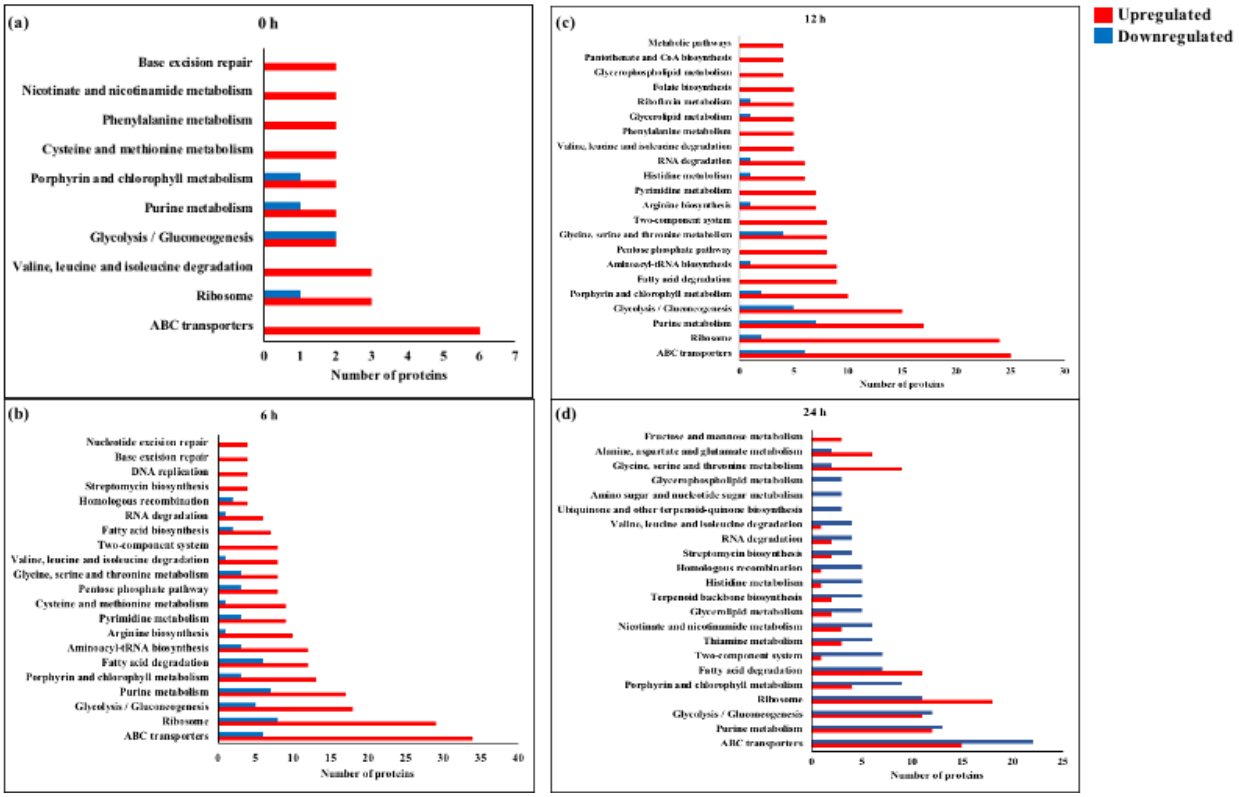


Figure 4.7. Top KEGG pathways represented at different infection time points based on number of proteins (**a**) 0 h (**b**) 6 h (**c**) 12 h (**d**) 24 h.

4.3.7 Time-dependent effect of phage Ochi17 on host *M. smegmatis*

The number of proteins represented in a functional category does not necessarily reflect the effect of a treatment on the functional group. In order to gain an understanding of the temporal impact of phage Ochi17 on host *M. smegmatis*, we analyzed the enrichment of KEGG pathways

and GO terms using DAVID 6.8 at the recommended EASE score of 0.1[11] (Fig. 4.8a-d). Our results showed that at 0 h molecular functions of heme binding and transcription factor activity were upregulated, while transporter activity and small protein activating enzyme activity were downregulated (Fig. 4.8a). At 6 h post infection, ABC transporters and ribosome pathways were upregulated, along with molecular functions- rRNA binding and structural constituent of ribosome, and biological processes transcription and translation. Meanwhile, cell redox homeostasis, flavin adenine dinucleotide binding and cell were downregulated (Fig. 4.8b). At 12 h post infection, more functional categories were downregulated than upregulated. Only integral component of membrane of the cellular component GO term and the molecular function, transporter activity were upregulated. Zinc ion binding, magnesium ion binding, DNA repair and intracellular were all downregulated (Fig. 4.8c).

The effect of Ochi17 on *M. smegmatis* at 24 h post infection saw more upregulated functional groups than downregulated. Extracellular region (cellular component), RNA binding, acyl-CoA dehydrogenase activity and transcription regulation were all downregulated. Meanwhile, biosynthesis of antibiotics, amino acids and secondary metabolites, structural constituent of ribosome, ribosome, and metabolism of nitrogen, fructose and mannose, and glycine, serine and threonine were all upregulated (Fig. 4.8d).

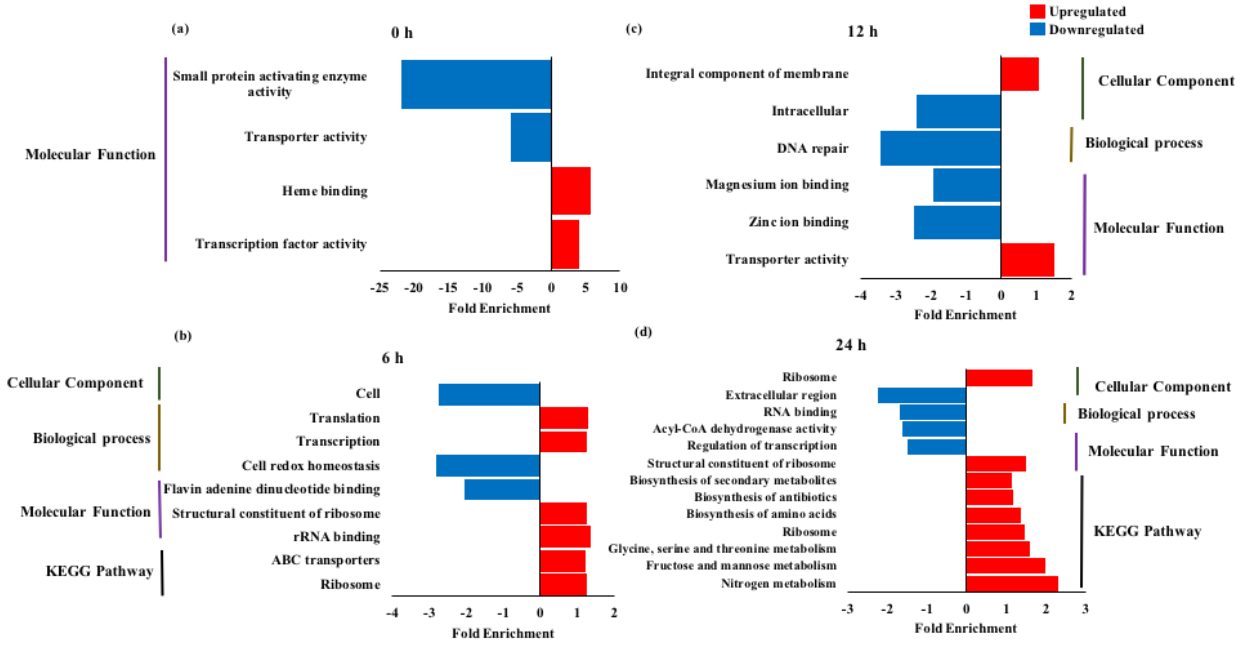


Figure 4.8. Enrichment of GO terms and KEGG pathways at (**a**) 0 h (**b**) 6 h (**c**) 12 h and (**d**) 24 h

4.3.8 Phage Ochi17 impact on *M. smegmatis* two component system (TCS)

TCS in bacteria play important signaling roles such as in pathogenesis and cell-to-cell communication[12]. Their exclusive presence in bacteria provides an opportunity to use them as targets for developing antibiotics and other therapeutic agents[12]. Proteomic results showed that all the proteins involved in TCS pathway were exclusively upregulated at 6 and 12 h post infection, while they were significantly downregulated at 24 h post infection (Fig. 4.7b-d). In order to understand how the proteins involved in the TCS pathway interact at 24 h post infection, STRING v 11.0 was used to examine the molecular actions occurring among the proteins. A key finding was the downregulation of the two-component regulatory systems mtrA/mtrB and mprA/mprB only at 24 h post infection. The STRING analysis showed a positive activation of mtrA and mprA by mtrB. mtrB phosphorylates mtrA, which plays a role in cell division[13,14] (Fig. 4.9).

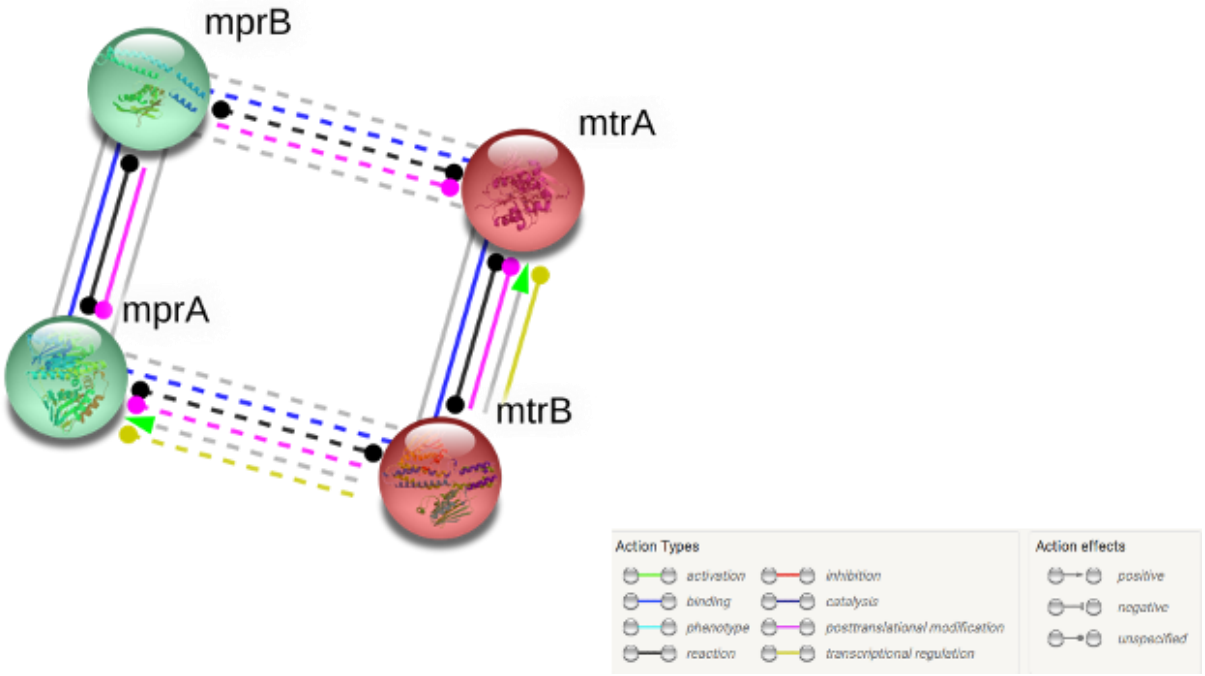


Figure 4.9. Proteins involved in TCS pathway. STRING analysis showing interactions and molecular actions of the four proteins involved in the mtrA/mtrB and mprA/mprB regulatory systems.

4.4 Discussion

The use of proteomics in studying phage-bacteria interactions is becoming prominent[15-17]. In the previous study on phage Ochi17-*M. smegmatis* interaction (chapter 3), proteomics was also applied in studying the response of *M. smegmatis* to phage Ochi17 infection at a single time point

of 3.5 h. In order to get a deeper understanding into Ochi17-*M. smegmatis* interaction beyond one time point, I carried out a time-series investigation of the impact of Ochi17 on the proteome of *M. smegmatis* as well as the temporal expression of phage Ochi17 proteins across the infection timeline.

The *M. smegmatis* growth curve showed that a significant impact of Ochi17 on *M. smegmatis* was not observed until after 6 h post infection (Fig. 4.1). This is relatively late compared to the reported responses of other bacteria. For example, a study reported that the host *Bacillus subtilis* was quick to respond to lytic phage ϕ29 infection at 16 min[7]. Using another lytic phage PRD1, another study reported initial *E. coli* response at 30 min after infection[18]. A proteome study of the interaction between *Lactococcus lactis* MG1363 and the virulent phage p2 reported an initial response of about 3 h after infection[16]. This difference in the response time of the host could be attributed to the kind of host used, the life cycle of the phage used (lytic or temperate), and/or the MOI used. Complete cell lysis of the host up to 48 h post infection was not observed (Fig. 4.1), unlike other observations that used lytic phages[7,16,18]. However, the observations in this study were in line with a previous study that showed that the lytic activities of two temperate phages, MP22 and D3112 only caused a minimal reduction in cell replication within the first five hours of infection[19].

To gain insights into the molecular interactions happening throughout the different infection times, the phage proteins were clustered into three groups (early, middle and late stage genes) based on their degree of expression (Fig. 4.4), since protein clustering can give some insights into specific cell activities associated with a class of proteins[20]. Proteins in the early stage were majorly particle formation and host interaction proteins. These proteins were dominantly expressed at 0 h. At this time point, phage attachment to host cell has just taken place and infection is yet to begin[21]. Therefore, it was not surprising to observe structural proteins at this time point.

The high expression of Ochi17 proteins involved in host lysis activities between 6 and 12 h post infection (middle genes) corresponds to the significantly reduced bacterial growth observed after 6 h post infection (Fig. 4.1; 4.4). After penetration into the host cell and subsequent multiplication, phages typically burst out of the host cell using lysin A and lysin B proteins to lyse the mycolic acid-rich cell wall[22]. The high abundance of these two proteins, coupled with other lysis-promoting proteins like Cro and D-ala-D-ala-carboxypeptidase, at 6 h and 12 h post infection suggests that this time period is likely when host cell lysis occurs. Cro protein, a repressor protein

is involved in the switch from lysogeny to lytic phase[23], while D-ala-D-ala-carboxypeptidase is involved in cell proteolysis[24]. This partial cell lysis is also supported by the high expression of Ochi17 Whib family transcription factor, which is required for cell differentiation[25] and DNA methylase (Fig. 4.4). DNA methylase prevents viral DNA degradation by the host restriction-modification antiviral defense system[26]. During this period, Ochi17 continues to undergo rapid replication (Fig. 3.1 in chapter 3).

Concurrently, host proteome response within this same time period (6 h and 12 h post infection) showed significant upregulation of RNA degradation, and amino acid (valine, leucine and isoleucine) degradation, suggesting these pathways were targeted to cause reduced growth rate of Ochi17-infected *M. smegmatis* compared to the non-infected control (Fig. 4.1). Valine, leucine and isoleucine, also known as branched-chain amino acids are known targets of antibacterial drug development. Deletions of their genes in *M. tuberculosis* significantly reduced its virulence[27]. Also, the biosynthetic pathway of all three amino acids has been shown to be necessary for the growth and survival of *M. tuberculosis*[28,29]. Further examination showed that at 12 h post infection, proteins involved in fatty acid degradation were exclusively upregulated (Fig. 4.7c). Bacterial fatty acids are major part of the membrane lipids and important source of energy[30]. Degradation of fatty acids suggests another form of attack by phage Ochi17 on *M. smegmatis* as was reported in chapter 3.

Although the carrying capacity of Ochi17-infected *M. smegmatis* was significantly reduced when compared to the non-infected control, the *M. smegmatis* cell replication kept increasing with time until 24 h post infection (Fig. 4.1). Proteome analysis revealed that the vast majority of proteins and pathways involved in macromolecular synthesis were upregulated at 6 h and 12 h post infection (Fig. 4.7a-d; Fig. 4.8b-c). This could explain the increase in the growth rate of Ochi17-infected *M. smegmatis.* The upregulation of macromolecular synthesis proteins could also suggest that phage Ochi17 is utilizing the host macromolecular synthesis for its own cell replication as has been previously reported[10].

Continuous Ochi17 infection of *M. smegmatis* past 12 h started showing a decline in host cell replication was at 24 h post infection (Fig. 4.1). Protein clustering showed that phage Ochi17 lytic activities progressed towards the late stage of infection (12 h – 24 h) due to the relatively high expression of antirepressor protein and DNA polymerase III at 24 h post infection. Antirepressor prevents the phage c2 repressor from binding to its operators to maintain lysogeny, therefore

promoting lysis[31], while DNA polymerase III promotes cell replication[32]. *M. smegmatis* proteome response at 24 h post infection showed that proteins and functional groups involved in macromolecular synthesis, such as transcription, were mostly downregulated, with the exception of ribosomal proteins and pathway (Fig. 4.7d; Fig. 4.8d). This observation suggests that the growth reduction detected at 24 h could be as a result of DNA transcription inhibition from Ochi17 lysis, which ultimately hinders cell multiplication[10,33]. The upregulation of ribosomal pathway could be the result of Ochi17 usage of the host protein synthesis machinery for its own propagation[10]. Ochi17 one-step curve showed increased cell multiplication at this time point (chapter 3).

The sudden decline of cell replication at 24 h post infection makes this time point vital in understanding the effect on the host. Two Component System (TCS) are essential targets in anti-bacterial drug development due to their signaling roles in a variety of events such as pathogenesis[12]. At 6h and 12 h post infection, TCS was exclusively upregulated (Fig. 4.7b-c). However, at 24 h post infection, all the proteins were downregulated with the exception of one (DNA-binding response regulator). The only four proteins that showed interactions were those of the mtrA/mtrB and the mprA/mprB TC systems. MtrA is a DNA-binding response regulator that plays a role in cell division and controls the expression of certain genes such as ripA, dnaA, fbpB, among others, while mtrB phosphorylates mtrA[14]. mprA and mprB contribute in maintaining a balance among several systems involved in stress resistance. mprB also phosphorylate mprA[13]. The differential regulation of TCS proteins suggests TCS as a temporal target in Ochi17-*M. smegmatis* interaction.

In conclusion, this study reports a time-dependent interaction between Ochi17 and its host, *M. smegmatis*. The results presented here showed that Ochi17 proteome expression is dynamic over the infection period, which in turn revealed time-dependent expression of certain *M. smegmatis* functional groups and pathways that are proven targets for anti-bacterial therapeutic agents. Overall, this study provides the basis for further investigation of singular targets at specific time periods in using mycobacteriophages as therapeutic agents for the treatment of tuberculosis.

4.5 Methods

4.5.1 *M. smegmatis* strain and growth conditions and phage infection

Following slight modification of protocol[21] *M. smegmatis* strain mc[2]155 colonies were inoculated in two flasks of 7H9 liquid medium containing 50 mL 7H9 Middlebrook broth

supplemented with 0.05% tween-80, 1 mM calcium chloride, 10% AD supplement, 40% glycerol, 50 ug/ml carbenicillin, and 10 ug/ml cycloheximide at 37$_{o}$C with constant agitation at 250 rpm. After two days, 50 ul of this culture was added to a fresh 50 mL 7H9 liquid medium without tween-80, with constant agitation at 250 rpm until it reaches the exponential stage of growth (OD_{600} of 0.5- 0.7).

On reaching the exponential growth, a modified protocol[34] was followed to infect *M. smegmatis* with phage Ochi17. *M. smegmatis* culture in each flask was aliquoted into three separate 15 mL conical tubes of 12 mL each and then centrifuged at 500 xg for 10 min after which 10 ml of supernatant was removed and the pellet resuspended in the remaining broth. For the phage-treated samples, the concentrated culture was then infected with phage Ochi17 at an MOI of 10 and mixed gently. For the control samples, the concentrated culture was infected with phage buffer. The mixture was incubated for 15 min at 37 $_{o}$C to initiate phage adsorption to the host's cell wall, and then the volume of the conical tube was brought back to 12 ml with pre-warmed broth culture media. The contents were then transferred to sterile 50 ml flask containing several cut pipette tips. The flasks were later incubated at 37 $_{o}$C with agitation at 250 rpm. For *M. smegmatis* growth curve, 2 ml of cells was used to measure the OD600 at 0, 0.5, 1.5, 3.5, 6, 12, 18, and 24 h post-infection with a nanophotometer NP80 (IMPLEN, Westlake Village, CA, USA). 7H9 media was used as the standard.

4.5.2 Sample preparation, protein extraction and digestion

After incubation for 0, 6, 12, and 24 h, 1.5 ml of the cells was transferred from the flask into two groups of three sterile micro centrifuge tubes. The cells were eventually pelleted at 14,000 xg for 5 min and the supernatant removed. Each pellet was later resuspended in 750 ul of phage buffer and centrifuged at 14,000 xg for 8 min. Pelleted *M. smegmatis* cells were washed with Phosphate Buffer Saline (PBS) pH 7.4 and spun at 14, 000 xg for 5 min. Supernatant was removed and pellets were solubilized in 50 ul 1X LDS sample buffer containing SDS and sonicated for 3 min and heated at 90$_{o}$C for 5 min. After cooling, 30 ul of each sample was run on 10% Bis-Tris gel (Life Technologies, Carlsbad, CA, USA) for 20 min at 120 V. The gel was then rinsed with milliQ water and stained with commassie blue for about 2 h. The gel was further destained by rinsing three times in milliQ water. Protein bands were cut from the gel and 25 mM ABC/100% ACN was added and then vortexed several times until staining solution is removed, and 100%

ACN was eventually added and then removed after vortexing. The gel pieces were dried in a vacuum centrifuge for 15 min. Reduction and alkylation of cysteines were carried out on the gel pieces using 10 mM DTT in 25 mM ABC at 55$_{o}$C for 1 h, and using 55 mM IAA in 25 mM ABC at room temperature in the dark for 45 min. For the trypsin digestion, 20 μg of Lys- C/trypsin (Promega) was dissolved in 400 μL of 25 mM ABC and 50 μL of enzyme mixture was added to each sample to achieve an enzyme-to-substrate ratio of 1:25 or until the gels are below liquid level. Gel samples containing enzyme solution were then placed in a temperature-controlled shaker at 37$_{o}$C overnight. After digestion, supernatants were removed, and peptides were extracted using 60% ACN/5% trifluoroacetic acid (TFA) and were dried in vacuum centrifuge for ~2 h to prepare them for LC-MS/MS.

4.5.3 Liquid chromatography and tandem mass spectrometry (LC-MS/MS) data acquisition

Samples were analyzed by reverse-phase high-performance liquid chromatography−electrospray ionization−tandem mass spectrometry (HPLC−ESI−MS/MS) using the Dionex UltiMate 3000 RSLC nano System (Thermo Fisher Scientific, Waltham, MA, USA) coupled to the Orbitrap Fusion Lumos (Thermo Fisher Scientific, Waltham, MA, USA) and a Nano- electrospray Flex ion source (Thermo Fisher Scientific, Waltham, MA, USA). Peptides were resuspended in 3% ACN/0.1% FA/ 97% MilliQ formic acid and loaded onto a trap column (300 μm ID × 5 mm) packed with 5 μm 100 Å PepMap C18 medium and washed using a flow rate of 5 μL/min with 98% purified water/ 2% acetonitrile (ACN)/0.01% formic acid (FA). The trap column was then switched in-line with the analytical column after 5 min. Peptides were separated using a reverse-phase Acclaim PepMap RSLC C18 (75 μm × 15 cm) analytical column using a 120 min method at a flow rate of 300 nL/min. The analytical column was packed with 2 μm of 100 Å PepMap C18 medium (Thermo Fisher Scientific, Waltham, MA, USA). Mobile phase A consisted of 0.01% FA in water, and mobile phase B consisted of 0.01% FA in 80% ACN. The linear gradient started at 5% B and reached 30% B in 80 min, 45% B in 91 min, and 100% B in 93 min. The column was held at 100% B for the next 5 min before being brought back to 5% B and held for 20 min to equilibrate the column. Sample was injected into the QE HF through the Nanospray Flex Ion Source fitted with an emission tip from Thermo Scientific. The column temperature was maintained at 35 °C. MS data were acquired with a Top 20 data-dependent

MS/MS scan method. The full scan MS spectra were collected over 300−1650 m/z range with a maximum injection time of 100 ms, a resolution of 120 000 at 200 m/z, spray voltage of 2, and an AGC target of 1 × 106. Fragmentation of precursor ions was performed by high-energy C-trap dissociation (HCD) with the normalized collision energy of 27 eV. MS/MS scans were acquired at a resolution of 15 000 at m/z 200. The dynamic exclusion was set at 20 s to avoid repeated scanning of identical peptides. Three biological sample replicates from each treatment were utilized for LC−MS/MS, which was sufficient for good statistical power. Instrument optimization and recalibration was carried out at the start of each batch run using the Pierce calibration solution.

4.5.4 Bioinformatics and data analysis

The raw MS/MS data (.raw files) were processed using MaxQuant (v1.6.0.16) with its integrated Andromeda search engine[35,36] with the spectra matched against the Uniprot *M. smegmatis* fasta (http://www.uniprot.org) concatenated with a common contaminants database and a reverse-decoy database. The cleavage enzyme used was set as trypsin/P and LysC while allowing for up to 2 missed cleavages. MaxQuant search was done at 1% FDR at both the peptide and protein levels. The minimum peptide length required for database search was set to seven amino acids. Precursor mass tolerance of ± 10 ppm, MS/MS fragment ions tolerance of ± 0.5 Da, maximum missed cleavage for tryptic digestion was set to two, methionine oxidation and protein N-terminal acetylation (K) were set as the variable, while Carbamidomethyl (C) was set as a fixed modification. The “unique plus razor peptides” were used for peptide quantitation. Razor peptides are non-unique peptides assigned to the protein group with most other peptides.

Removing proteins with reverse identification and those identified as contaminants were the first steps in filtering the MaxQuant data. Then proteins without any LFQ intensity and without any MS/MS counts were filtered out. Also, proteins with negative score were removed. MaxQuant results were exported to Data Analysis and Extension Tool (DAnTE) and analyzed for Pearson correlations coefficients to determine reproducibility among replicates in both control and phage-infected samples. In order to generate high confidence data, the following criteria were further used to filter the MaxQuant data: protein quantified by LFQ intensity in less than two replicates was removed; protein quantified by MS/MS counts in less than two replicates was also removed, except the protein has a total MS/MS count > 5. The LFQ intensity, used as the primary quantitative measure of abundance, was converted to log2 values and averaged across replicates.

Fold change was calculated by the difference in the average log2 values between phage-infected proteins and control proteins. Two-tailed unpaired T-test was used to calculate significance values. Proteins with fold change values of < -0.5 or > 0.5 and $p < 0.05$ were considered significantly regulated by phage Ochi17 infection.

Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of significantly regulated proteins was performed using DAVID 6.8 (the Database for Annotation, Visualization and Integrated Discovery; https://david.ncifcrf.gov/). Enrichment was considered significant at a maximum modified Fisher Exact P-value (EASE Score) of 0.1.

4.6 References

1. Ryan, K.J. & Ray, C.G. Mycobacteria. In: Sherris Medical Microbiology : an introduction to infectious diseases (4th ed.). *McGraw-Hill*, NY p. 439 (2004).
2. Smith, I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev* **16**, 463-496 (2003).
3. World Health Organization. Global tuberculosis report. *WHO* 5-147 (2017).
4. Fu, X., Ding, M., Zhang, N. & Li, J. Mycobacteriophages: an important tool for the diagnosis of *Mycobacterium tuberculosis* (review). *Mol Med Rep* **12**, 13-19 (2015).
5. Hatfull, G. Mycobacteriophages: windows into tuberculosis. *PLoS Pathogens* **10**, 12-14 (2014).
6. Hatfull, G. The secret lives of mycobacteriophages. *Adv Virus Res* **82**, 179-288 (2012).
7. Mojardin, L., & Salas, M. Global transcriptional analysis of virus-host interactions between phage ϕ29 and *Bacillis subtilis*. *J Virol* **90**, 9293-9304 (2016).
8. Sacher, J.C. *et al*. Transcriptomic analysis of the *Campylobacter jejuni* response to T4-like phage NCTC 12673 infection. *Viruses* **10**, 332, doi:10.3390/v10060332 (2018).
9. Leskinen, K., Blasdel, B.G., Lavigne, R. & Skurnik, M. RNA-sequencing reveals the progression of phage-host interactions between ωR1-37 and Yersinia enterocolitica. *Viruses* **8**, 111 doi: 10.3390/v8040111 (2016).
10. Zhao, X., *et al*. Global transcriptomic analysis of interactions between *Pseudomonas aeruginosa* and bacteriophage PaP3. *Sci Reports* **6**, 19237, doi: 10.1038/srep19237 (2016).
11. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc* **4**, 44-57 (2009).

12. Wang, S. Bacterial Two-Component Systems: Structures and Signaling Mechanisms. In Protein phosphorylation in human health (Ed. Huang, C). *IntechOpen* (2012).
13. Fol, M. *et al.* Modulation of *Mycobacterium tuberculosis* proliferation by MtrA, an essential two-component response regulator. *Mol Microbiol* **60**, 643-657 (2006).
14. Plocinska, R. *et al.* Septal localization of the Mycobacterium tuberculosis MtrB sensor kinase promotes MtrA regulon expression. *J Biol Chem* **287**, 23887-23899 (2012).
15. Lavigne, R. *et al.* A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *mBio* **4**, e00061-13 (2013).
16. Lemay, M-L. *et al*. Investigating Lactococcus lactis MG1363 response to phage p2 infection at the proteome level. *Mol Cel Prot* **18**, 704-714 (2019).
17. Weintraub, S.T. *et al*. Global proteomic profiling of *Salmonella* infection by a giant phage. *J Virol* **93**, e01833-18, doi: 10.1128/JVI.01833-18 (2019).
18. Poranen, M.M., *et al*. Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* **80**, 8081–8088 (2007).
19. Chung, I.Y., Sim, N. & Cho, Y.H. Antibacterial Efficacy of Temperate Phage-Mediated Inhibition of Bacterial Group Motilities. Antimicrob Agents Chemother. **56**, 5612–5617 (2012).
20. Okekeogbu, I.O. *et al.* Differential distributions of trafficking and signaling proteins of the maize ER-Golgi apparatus. *Plant Sig Behav* **14**, e1672513, doi:10.1080/15592324.2019.1672513 (2019).
21. Poxleitner, M., Pope, W., Jacobs-Sera, D., Sivanathan, V. & Hatfull, G. Phage discovery guide for use with Mycobacterium smegmatis. *HHMI* 1-118 (2016).
22. Payne, K., Sun, Q., Sacchettini, J. & Hatfull, G.F. Mycobacteriophage lysin B is a novel mycolylarabinogalactan esterase. *Mol Microbiol* **73**, 367-381 (2009).
23. Lee, S., Lewis, D.E.A., Adhya, S. The Developmental Switch in Bacteriophage lambda: A Critical Role of the Cro Protein. *J Mol Biol* **430**, 58-68 (2018).
24. Rawlings, N.D. & Barrett, A.J. Families of serine peptidases. *Meth Enzymol* **244**, 19-61 (1994).
25. Kormanec, J. & Homerova, D. Streptomyces aureofaciens whiB gene encoding putative transcription factor essential for differentiation. *Nucleic Acid Res* **21**, 2512-2512 (1993).

26. Zinoviev, V.V., Evdokimov, A.A., Malygin, E.G., Schlagman, S.L. & Hattman S. Bacteriophage T4 Dam DNA-(N6-adenine)-methyltransferase. Processivity and orientation to the methylation target. *J Biol Chem* **278**, 7829-7833 (2003).
27. Tathyana, M., Franco, A. & Blanchard, J.S. Bacterial Branched-Chain Amino Acid Biosynthesis: Structures, Mechanisms, and Drugability. *Biochemistry* **56**, 5849–5865 (2017).
28. Sassetti, C.M., Boyd, D.H. & Rubin, E.J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77–84 (2003).
29. Grandoni, J.A., Marta, P.T. & Schloss J.V. Inhibitors of branched-chain amino acid biosynthesis as potential antituberculosis agents. *J Antimicrob Chemother* **42**, 475–482 (1998).
30. Fujita, Y., Matsuoka, H. & Hirooka, K. Regulation of fatty acid metabolism in bacteria. *Mol Microbiol* **66**, 829–839 (2007).
31. Kim, M. & Ryu, S. Antirepression system associated with the life cycle switch in the temperate *Podoviridae* phage SPC32H. *J Virol* **87**, 11775-11786 (2013).
32. Glover, B.P. & McHenry, C.S. An asymmetric dimeric replicative complex with leading and lagging strand polymerases. *Cell* **105**, 925-934 (2001).
33. Nechaev, S. & Severinov, K. Bacteriophage-induced modifications of host RNA polymerase. *Annu Rev Microbiol* **57**, 301-322 (2003).
34. Wolyniak, M.J., Reyna, N.S., Plymale, R., Pope, W.H. & Westholm, D.E. Mass spectrometry as a tool to enhance "-omics" education. *J Microbiol Biol Edu* 19, doi:10.1128/jmbe.v19i.1459 (2018).
35. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein identification. *Nat Biotechnol* **26**, 1367-1372 (2008).
36. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526 (2014).

5. CONCLUSIONS AND FUTURE DIRECTIONS

There have been renewed research interests in phage-bacteria interactions due to the bactericidal property of phages, which can be utilized in various applications ranging from therapeutic uses to food preservation, in addition to using them as molecular tools for manipulating bacteria[1,2]. With the advancement in molecular technologies, such as microarray, RNAseq, and mass spectrometry, recent studies have largely focused on global molecular interactions at the omics level. Most global omics studies on phage-bacteria interactions have been transcriptomics with very few on proteomics and metabolomics, even though transcriptomics only give limited understanding to molecular activities[3].

Despite the progress on the studies on phage-bacteria interactions, none has been reported for a mycobacterium host of which *M. tuberculosis* belongs to. *M. tuberculosis* is the deadliest human pathogen in the world[4] and are quickly becoming multi-drug resistant and extensively drug-resistant[5]. Hence, the widespread research on the potential of phages to control and treat bacterial infections and diseases.

This study applied global proteomics to study the interactions between a novel mycobacteriophage, Ochi17 and a mycobacterium host, *M. smegmatis*. Understanding how *M. smegmatis* responds to a mycobacteriophage attack will help in engineering and exploiting mycobacterial hosts for various purposes, especially in therapeutics. This study looked into answering two major research questions: how does *M. smegmatis* respond to a mycobacteriophage attack at the molecular (proteome) level? And does the molecular response of *M. smegmatis* and phage molecular expression change over the entire infection period? (chapter 4). To answer the first question, a novel phage was first isolated and characterized (chapter 2), and then followed by a global proteomics study of *M. smegmatis* infected with the isolated phage (ochi17) at a specific time point of 3.5 h (chapter 3). Global proteomics study of phage Ochi17-*M. smegmatis* interaction at various infection time points (0 h, 6 h, 12 h and 24 h) was also applied to answer the second research question (chapter 4). Below is a summary of the findings for each research question, and future directions emanating from the observations.

5.1 Research question I: how does *M. smegmatis* respond to phage Ochi17 infection at the molecular level?

In order to understand the response of *M. smegmatis* to phage ochi17 infection, phage Ochi17 was first isolated by a direct method according to already established protocol[6]. Phage Ochi17 was isolated directly from the soil, purified and then amplified using the plaque assay to get a single clonal population. The structure and genome of the isolated phage was then characterized using Transmission Electron Microscopy (TEM) at Purdue University and Illumina sequencing at the Pittsburgh Bacteriophage Institute. TEM revealed that Ochi17 has a capsid head (56.67 nm) and a non-contractile tail that is 200 nm in length. Based on this structure, it was classified into the Siphoviridae morphotype (chapter 2). Genome annotation using DNAMaster[7] and Phamerator[8] showed ochi17 possesses an integrase, making it a temperate phage (Fig. 2.3 in chapter 2). Genome organization showed that genes were arranged in functional sections. Genes with related functions are located close to one another (Fig. 2.3).

Following the characterization of Ochi17, label-free quantitative proteomics was used to study the molecular response of *Mycobacteria smegmatis* infected with the phage Ochi17 at 3.5 h post infection (chapter 3). More proteins were upregulated than downregulated. HR and processes of macromolecular synthesis such as amino acid metabolism, DNA replication, transcription, translation, and vitamin metabolism were significantly upregulated. Based on this, it was suggested that phage Ochi17 probably utilizes host *M. smegmatis* macromolecular synthesis for its own propagation. RecX, an inhibitor of RecA was found to also be upregulated, despite the upregulation of HR. Chapter 3 also showed that fatty acid and lipid metabolism were primary targets of phage Ochi17. Specifically, fabD, a critical protein in fatty acid biosynthesis was significantly downregulated, making it a prime target for therapeutic purposes.

5.2 Research question II: does the molecular response of *M. smegmatis* and phage molecular expression change over the entire infection period?

This question was addressed using global label-free quantitative proteomics to investigate the temporal mycobacteriophage-mycobacteria interaction over 24 h of Ochi17 infection in chapter 4. The growth curve of *M. smegmatis* showed that significant reduction in the carrying capacity of Ochi17-infected *M. smegmatis* by as much as 70-80% after 6 h of infection (Fig. 4.1). The physiological effects of Ochi17 on the mycobacterium host varied over the entire life cycle of the

host. Proteomics analysis was done at 0h, 6 h, 12 h and 24 h post infection time points. A total of 2,181 host proteins and 56 Ochi17 proteins were identified across all the time points. Phage Ochi17 proteome expression is time-dependent and the proteins were grouped into three clusters based on expression time. The results showed that host lysis and lysis-promoting Ochi17 proteins were majorly expressed together, especially during the middle stage of infection and into the late infection stage. Proteomic results suggest that Ochi17 suppresses the growth of *Mycobacterium smegmatis* not just by utilizing the macromolecular synthesis of the host, but also by suppressing host transcription, two-component system and degrading fatty acids (chapter 4).

5.3 Future directions

This study focused on a mycobacterial host, and results obtained may not be applicable to other classes of bacteria. However, the insights may contribute in the application of mycobacteriophages as therapeutic agents on *Mycobacterium tuberculosis* focusing on specific targets.

Future work will look into farther timeline beyond 24 h phage infection to understand the extent and durability of phage Ochi17 infection. To the best of my knowledge, this is the only omics study that has investigated up to 24 h of phage-bacteria interaction. Another future study based on the clustering profile in chapter 4, will be protein interaction study to investigate which host proteins are interacting with each phage Ochi17 phage cluster. This can be done using co-immunoprecipitation[9] of the whole cell extracts at the three different infection time periods representing the early, middle, and late infection stage. Additional future work will focus on the interactions between highly expressed individual host proteins such as fabD, recX, and proteins involved in TCS and potential Ochi17 proteins using yeast 2 hybrid[9]. Further validation of specific proteins of interest (macromolecular synthesis proteins, fabD, mtrA, and mtrB) using qPCR[10].

Finally, a complete understanding of phage-bacteria interactions will likely depend on several factors such as the type of phage used (lytic or temperate), the host bacterium, and the infection time points employed during experimental analysis. Overall, this study provides the basis for further investigation of singular targets at specific time periods in using mycobacteriophages as therapeutic agents for the treatment of tuberculosis.

5.4 References

1. Hatfull, G. F. *The Secret Lives of Mycobacteriophages*. *Advances in Virus Research* vol. 82 (Elsevier Inc., 2012).
2. Fu, X., Ding, M., Zhang, N. & Li, J. Mycobacteriophages: An important tool for the diagnosis of Mycobacterium tuberculosis (Review). *Molecular Medicine Reports* **12**, 13–19 (2015).
3. Schwanhüusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
4. Gordon, S. v. & Parish, T. Microbe profile: Mycobacterium tuberculosis: Humanity's deadly microbial foe. *Microbiology (United Kingdom)* **164**, 437–439 (2018).
5. Broxmeyer, L. *et al.* Killing of Mycobacterium avium and Mycobacterium tuberculosis by a mycobacteriophage delivered by a nonvirulent Mycobacterium: A model for phage therapy of intracellular bacterial pathogens. *Journal of Infectious Diseases* **186**, 1155–1160 (2002).
6. Poxleitner, M., Pope, W. H., Jacobs-Sera, D., Sivanathan, V. & Hatfull, G. F. Phage Discovery Guide. 1–19 (2017).
7. Dedrick, R. M. *et al.* HHS Public Access. 1–26 (2017) doi:10.1038/nmicrobiol.2016.251.Prophage-mediated.
8. Cresawn, S. G. *et al.* Phamerator: A bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* **12**, (2011).
9. Kumar, P., Dwivedi, M., Patel, M. B., Prakash, C. & Prasad, B. D. Advances in molecular techniques to study diversity. *Plant Biotechnology, Volume 1: Principles, Techniques, and Applications* **2014**, 341–365 (2017).
10. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nature Protocols* **1**, 1559–1582 (2006).

APPENDIX A. SIGNIFICANTLY DIFFERENTIAL PROTEINS

All significant proteins ($p \leq 0.05$) and unique proteins identified in phage-treated and control samples. Red font = upregulated proteins; green font = downregulated proteins.

Protein names	Gene names	Avg Log LFQ_T	Avg Log LFQ_C	Log2 Fold Change
Replicative DNA helicase (EC 3.6.4.12)	dnaB MSMEI_6709	27.8842	24.3310	3.55
L-lysine-epsilon aminotransferase (EC 2.6.1.36)	lat MSMEI_1721	28.8811	25.6734	3.21
PhoH-like protein phoH2 (Phosphate starvation-inducible protein psiH)	phoH2 MSMEI_5109	27.4267	24.5989	2.83
Integral membrane protein	MSMEI_6600	30.2327	27.5186	2.71
Piperideine-6-carboxylic acid dehydrogenase Pcd (EC 1.2.1.3)	pcd MSMEI_1719	28.9888	26.4210	2.57
Protein RecA (Recombinase A)	recA MSMEG_2723 MSMEI_2656	31.1272	28.9800	2.15
Holliday junction ATP-dependent DNA helicase RuvA (EC 3.6.4.12)	ruvA MSMEG_2944 MSMEI_2870	26.9331	25.3016	1.63
Uncharacterized protein	MSMEI_4295	27.2539	25.6469	1.61
Exodeoxyribonuclease III Xth	MSMEI_1617	27.3015	25.7682	1.53
Uracil-DNA glycosylase (UDG) (EC 3.2.2.27)	ung MSMEG_2399 MSMEI_2339	27.7612	26.2867	1.47
Uncharacterized protein	MSMEI_1209	28.0665	26.6138	1.45
Uncharacterized protein	MSMEI_4957	25.1697	23.7189	1.45
Beta sliding clamp (Beta clamp) (Sliding clamp) (Beta-clamp processivity factor) (DNA polymerase III beta sliding clamp subunit) (DNA polymerase III subunit beta)	dnaN MSMEG_0001 MSMEI_0003	30.6804	29.4230	1.26
Putative decaprenylphosphoryl-5-phosphoribose phosphatase MSMEG_6402 (DPPR phosphatase) (EC 3.1.3.-) (Phospholipid phosphatase)	MSMEG_6402 MSMEI_6234	27.2712	26.0341	1.24
Proteasome-associated ATPase (AAA ATPase forming ring-shaped complexes) (ARC) (Mycobacterial proteasome ATPase)	mpa MSMEG_3902 MSMEI_3813	29.4392	28.2075	1.23
Phosphoadenosine phosphosulfate reductase	MSMEG_1245	26.5101	25.3132	1.20
50S ribosomal protein L30	rpmD MSMEG_1473 MSMEI_1437	29.7785	28.5899	1.19
Glutaredoxin electron transport component of NRDEF (Glutaredoxin-like protein) NrdH (EC 1.17.4.1)	nrdH MSMEI_0988 MSMEI_2239	26.6300	25.4727	1.16
Holliday junction ATP-dependent DNA helicase RuvB (EC 3.6.4.12)	ruvB MSMEG_2945 MSMEI_2871	26.0145	24.8620	1.15
Cell wall synthesis protein CwsA (Cell wall synthesis and cell shape protein A)	cwsA MSMEG_0023 MSMEI_0025	28.8748	27.7711	1.10

30S ribosomal protein S7	rpsG MSMEG_1399 MSMEI_1361	32.5771	31.4920	1.09
Ribonucleoside-diphosphate reductase subunit alpha 2 (EC 1.17.4.1) (Ribonucleotide reductase R1 subunit 2)	nrdE2 MSMEG_2299 MSMEI_2241	31.6958	30.6230	1.07
Uncharacterized protein	MSMEI_3193	25.6181	24.5498	1.07
NAD-dependent epimerase/dehydratase (EC 5.1.3.2)	MSMEI_0922	25.1366	24.0706	1.07
DNA polymerase III, delta (EC 2.7.7.7)	MSMEI_4460	26.5937	25.5705	1.02
Topoisomerase subunit TopoM (EC 5.99.1.3)	topoM MSMEG_0456 MSMEI_0443	28.0836	27.0902	0.99
50S ribosomal protein L5	rplE MSMEG_1467 MSMEI_1431	31.8248	30.8370	0.99
GntR family transcriptional regulator	MSMEI_4022	26.0851	25.0997	0.99
Demethylmenaquinone methyltransferase (EC 2.1.1.163)	menG MSMEG_1115 MSMEI_1083	26.5353	25.5595	0.98
6-phosphogluconate dehydrogenase (EC 1.1.1.44)	MSMEI_0004	28.5028	27.5464	0.96
DNA repair exonuclease SbcD	MSMEI_4877	24.7299	23.7919	0.94
Uncharacterized protein	MSMEI_1208	26.2444	25.3137	0.93
RNA polymerase sigma factor	MSMEG_0219	26.0715	25.1690	0.90
Cyclase/dehydrase	MSMEI_5540	26.9895	26.0909	0.90
Tryptophan-rich sensory protein	MSMEG_1131	27.5496	26.6668	0.88
Ferredoxin-dependent glutamate synthase (EC 1.4.1.13)	MSMEI_6101	24.6542	23.8516	0.80
Putative flavin-containing monoamine oxidase aofH (EC 1.4.3.4)	aofH MSMEI_1990	24.4748	23.6831	0.79
DNA-directed DNA polymerase (EC 2.7.7.7)	dnaE1 MSMEI_3096	28.1114	27.3326	0.78
Uncharacterized protein	MSMEI_1937	26.5144	25.7466	0.77
Ribonucleoside-diphosphate reductase subunit beta (EC 1.17.4.1)	nrdF2 MSMEI_1002 MSMEI_2253	30.2948	29.5348	0.76
RecBCD enzyme subunit RecC (EC 3.1.11.5) (Exonuclease V subunit RecC) (ExoV subunit RecC) (Helicase/nuclease RecBCD subunit RecC)	recC MSMEG_1328	25.0098	24.3028	0.71
Uncharacterized protein	MSMEG_3204	25.7109	25.0355	0.68
Type III restriction enzyme (EC 3.1.21.3)	MSMEI_1204	26.2324	25.5740	0.66
Putative chromosome-partitioning protein parB	parB MSMEI_6746	28.1400	27.4913	0.65
NifS-like class-V aminotransferase, probable cysteine desulfurase (EC 2.8.1.7)	MSMEI_1207	27.0857	26.4384	0.65
Delta-1-pyrroline-5-carboxylate dehydrogenase (EC 1.2.1.88)	rocA MSMEI_4990	26.1819	25.5355	0.65
Transcriptional regulator, TetR family	MSMEI_0999 MSMEI_2250	27.3459	26.7062	0.64
Spermidine/putrescine import ATP-binding protein PotA (EC 3.6.3.31)	potA MSMEG_3281	25.9456	25.3420	0.60
6-phosphogluconate dehydrogenase NAD-binding (3-hydroxyisobutyrate dehydrogenase) (EC 1.1.1.31)	MSMEI_4024	24.4754	23.8774	0.60
Transcriptional regulator, TetR family	MSMEI_5871	24.9299	24.3372	0.59
Secreted protein	MSMEI_2065	25.9597	25.3862	0.57

Putative membrane protein	MSMEI_4921	27.1066	26.5491	0.56
Isocitrate lyase (EC 4.1.3.1)	icl MSMEI_0889	29.9177	29.3611	0.56
TetR-family transcriptional regulator	MSMEI_3675	23.6662	23.1110	0.56
Uncharacterized protein	MSMEG_6914	25.3598	24.8170	0.54
UvrABC system protein B (Protein UvrB) (Excinuclease ABC subunit B)	uvrB MSMEI_3727	26.7558	26.2158	0.54
Conserved transmembrane protein	MSMEI_4186	28.0983	27.5647	0.53
Crp/Fnr familytranscriptional regulator	MSMEI_6029	31.1479	30.6167	0.53
Carboxymuconolactone decarboxylase (EC 4.1.1.44)	MSMEI_0884	25.1280	24.6020	0.53
GCN5-related N-acetyltransferase	MSMEI_5462	26.6303	26.1048	0.53
Glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	glpD2 MSMEI_1696	26.6988	26.1947	0.50
Transcriptional regulator, TetR family	MSMEI_4723	28.0533	27.5552	0.50
30S ribosomal protein S9	rpsI MSMEG_1557 MSMEI_1520	32.4758	31.9805	0.50
Cyclohexanone monooxygenase (EC 1.14.13.22)	MSMEI_2251	29.8313	29.3535	0.48
LppI	MSMEI_3759	24.7778	24.3032	0.47
ATP-dependent Clp protease ATP-binding subunit ClpX	clpX MSMEG_4671 MSMEI_4553	28.0325	27.5767	0.46
Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site) lyase)	nth MSMEI_6027	25.3068	24.8580	0.45
NAD dependent epimerase/dehydratase	MSMEI_4183	25.6767	25.2331	0.44
Cyclase/dehydrase family protein	MSMEG_0129	24.2206	23.7906	0.43
Uncharacterized protein	MSMEG_0758	25.0652	24.6360	0.43
Short-chain dehydrogenase/reductase SDR	MSMEI_0721	25.9181	25.5049	0.41
Uncharacterized protein	MSMEG_1643	24.4689	24.0809	0.39
Tryptophan--tRNA ligase (EC 6.1.1.2) (Tryptophanyl-tRNA synthetase) (TrpRS)	trpS MSMEI_1618	28.2649	27.8865	0.38
30S ribosomal protein S5	rpsE MSMEG_1472 MSMEI_1436	31.2295	30.8573	0.37
ABC transporter ATP-binding protein	MSMEI_0623	28.8578	28.4876	0.37
ABC transporter related protein	MSMEI_2465	26.2918	25.9338	0.36
50S ribosomal protein L9	rplI MSMEG_6894 MSMEI_6710	30.3245	29.9777	0.35
Conserved alanine valine and glycine rich protein	MSMEI_2704	26.3226	25.9817	0.34
MscS Mechanosensitive ion channel	MSMEI_2041	24.9503	24.6206	0.33
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_5726	26.1980	25.8693	0.33
Conserved hypothetical hydrolase	MSMEI_0778	25.6647	25.3373	0.33
Putative acetyl-COA acyltransferase Fada2 (3-ketoacyl-COA thiolase) (Beta-ketothiolase) (EC 2.3.1.16)	fadA2 MSMEI_0366	29.0574	28.7607	0.30
Geranylgeranyl reductase	MSMEI_0998 MSMEI_2249	30.3615	30.0657	0.30
Peptidoglycolipid exporter Gap (GPLs addressing protein)	gap MSMEG_0403 MSMEI_0393 LJ00_02005	24.5709	24.2781	0.29
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.50)	MSMEI_3616	26.5736	26.2900	0.28

Branched-chain amino acid ABC transporter (LivG)	MSMEI_3167	27.0850	26.8155	0.27
NUDIX hydrolase	MSMEI_3116	24.4613	24.1981	0.26
Uncharacterized protein	MSMEI_3067	27.6235	27.3636	0.26
Uncharacterized protein	MSMEI_1218	24.7005	24.4433	0.26
Amidohydrolase 3	MSMEI_3771	24.1540	23.9022	0.25
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.243)	MSMEI_4090	27.5564	27.3127	0.24
1-aminocyclopropane-1-carboxylate deaminase (EC 3.5.99.7)	MSMEI_6558	27.9067	27.6713	0.24
Short-chain dehydrogenase/reductase SDR	MSMEI_0843	27.3324	27.1027	0.23
Polyprenol monophosphomannose synthase (PPM synthase) (Polyprenol-P-Man synthase) (Ppm1) (EC 2.4.1.-) (Dolichol-phosphate mannose synthase) (EC 2.4.1.83)	ppm1 MSMEG_3859 MSMEI_3769	28.0504	27.8293	0.22
FAD dependent oxidoreductase	MSMEI_0621	24.7849	24.5751	0.21
Short-chain dehydrogenase/reductase SDR	MSMEI_4605	29.2562	29.0517	0.20
Conserved transmembrane proteinm, MmpS5	mmpS5 MSMEI_3415	26.0809	25.9075	0.17
Beta-lactamase	MSMEI_2966	25.3768	25.2064	0.17
DNA repair protein RadA (EC 3.6.4.-) (Branch migration protein RadA)	radA MSMEG_6079 MSMEI_5919	26.6128		26.61
Uncharacterized protein	MSMEI_3147	25.7859		25.79
Hydrolase, isochorismatase family protein	MSMEG_4126	25.3771		25.38
Conserved hydrolase	MSMEI_4876	25.9750		25.98
PE-PPE, C-terminal domain protein	MSMEI_1145	25.9482		25.95
Crossover junction endodeoxyribonuclease RuvC (EC 3.1.22.4) (Holliday junction nuclease RuvC) (Holliday junction resolvase RuvC)	ruvC MSMEG_2943 MSMEI_2869	25.3895		25.39
Conserved MCE associated membrane protein	MSMEI_0140	24.4781		24.48
RecBCD enzyme subunit RecD (EC 3.1.11.5) (Exonuclease V subunit RecD) (ExoV subunit RecD) (Helicase/nuclease RecBCD subunit RecD)	recD MSMEG_1325 MSMEI_1288	24.4859		24.49
Histidine triad (HIT) protein (EC 3.6.1.17)	MSMEI_4900	24.2928		24.29
Cation ABC transporter, periplasmic cation-binding protein, putative	MSMEG_6047	25.6245		25.62
Gluconate permease	MSMEI_0439	23.5418		23.54
Uncharacterized protein	MSMEG_4752	27.4220		27.42
DNA helicase (EC 3.6.4.12)	uvrD MSMEI_1909	26.1849		26.18
Decaprenyl diphosphate synthase (DecaPP) (EC 2.5.1.86) (EC 2.5.1.87) (Decaprenyl pyrophosphate synthase) (Long-chain isoprenyl diphosphate synthase) (Trans,polycis-decaprenyl diphosphate synthase)	uppS MSMEG_4490 MSMEI_4379	24.8844		24.88
Protein lsr2	lsr2 MSMEI_5934	26.6819		26.68

Conserved integral membrane alanine and leucine rich protein	MSMEI_2700	23.5744		23.57
Transcriptional regulator (Bacterial regulatory proteins, IclR family)	MSMEI_2326	24.4254		24.43
Uncharacterized protein	MSMEI_2614	24.7360		24.74
Uncharacterized protein	MSMEI_3339	25.1594		25.16
Uncharacterized protein	MSMEG_4558	24.6704		24.67
Putative succinate-semialdehyde dehydrogenase [NADP(+)] (SSADH) (SSDH) (EC 1.2.1.79)	gabD2 MSMEG_5912 MSMEI_5752	24.4501		24.45
Gtf2	gtf2	24.8985		24.90
Peptide deformylase (PDF) (EC 3.5.1.88) (Polypeptide deformylase)	def MSMEG_0832 MSMEI_0813	25.3782		25.38
Phospholipid/glycerol acyltransferase (EC 2.-.-.-)	MSMEI_0923	24.7162		24.72
Biotin synthase (EC 2.8.1.6)	bioB MSMEG_3194 MSMEI_3112	23.8581		23.86
MOSC domain protein	MSMEG_3411	23.9227		23.92
Transcriptional regulator, GntR family	MSMEI_3887	24.6286		24.63
2,3-dihydroxybenzoate-AMP ligase (EC 2.7.7.58)	MSMEI_4013	24.0126		24.01
1-acylglycerol-3-phosphate O-acyltransferase (EC 2.3.1.51)	MSMEI_4148	23.5401		23.54
UvrD/REP helicase	MSMEI_1900	24.2973		24.30
Uncharacterized protein	MSMEI_0359	23.6582		23.66
Conserved hypothetical membrane protein	MSMEI_0522	24.6486		24.65
Uncharacterized protein	MSMEI_0564	23.9252		23.93
Transcriptional regulator, Fis family	MSMEI_0814	23.8325		23.83
AsnC-family transcriptional regulator	MSMEI_0819	25.3070		25.31
Uncharacterized protein	MSMEG_2197	24.2689		24.27
Enoyl-CoA hydratase/isomerase	MSMEG_3139	25.0321		25.03
Transmembrane ATP-binding protein ABC transporter cydD (EC 3.6.3.25)	cydD MSMEI_3149	24.5479		24.55
Uncharacterized protein	MSMEI_4152	21.0208		21.02
von Willebrand factor type A	MSMEI_5359	24.9461		24.95
Uncharacterized protein	MSMEG_5816	23.5460		23.55
Cobalamin B12-binding protein	MSMEI_6367	23.8710		23.87
Methyltransferase type 11 (EC 2.1.1.-)	MSMEI_0095	24.2871		24.29
Uncharacterized protein	MSMEI_0911	23.8754		23.88
Uncharacterized protein	MSMEI_0950	24.4993		24.50
Uncharacterized protein	MSMEG_1823	24.8709		24.87
Aminoglycoside phosphotransferase	MSMEI_2494	24.1521		24.15
Uncharacterized protein	MSMEI_2523	22.7565		22.76
Transcriptional regulator, PadR family	MSMEI_2794	22.4876		22.49
NADH-dependent glutamate synthase small subunit (EC 1.4.1.-)	gltD MSMEI_3144	24.7904		24.79
Transmembrane transport protein MmpL1	mmpL5 MSMEI_3416	25.0562		25.06
1,4-Dihydroxy-2-naphthoate synthase (EC 4.1.3.36)	MSMEI_4015	24.8737		24.87
Uncharacterized protein	MSMEI_4576	24.7043		24.70
Uncharacterized protein	MSMEI_5253	23.3005		23.30

Probable conserved transmembrane protein rich in alanine	MSMEG_6099	24.4275		24.43
DNA polymerase III subunit delta (EC 2.7.7.7)	holB MSMEI_5996	25.1574		25.16
Uncharacterized protein	MSMEG_6228	24.7331		24.73
Enoyl-CoA hydratase/isomerase (EC 4.2.1.17)	MSMEI_6380	23.7858		23.79
Uncharacterized protein	MSMEI_6568	24.0160		24.02
Pirin domain protein	MSMEI_0051	28.2674		28.27
Uncharacterized protein	MSMEG_0840	24.3723		24.37
Aminoglycoside phosphotransferase	MSMEI_0977 MSMEI_2228	23.7991		23.80
Luciferase-like protein	MSMEI_2248	24.0846		24.08
Urocanate hydratase (Urocanase) (EC 4.2.1.49) (Imidazolonepropionate hydrolase)	hutU MSMEG_1179 MSMEI_1148	23.7129		23.71
Cutinase (EC 3.1.1.74)	cut2 MSMEI_1491	23.8920		23.89
Alternative RNA polymerase sigma-D factor, SigD	sigD MSMEI_1560	23.3141		23.31
Amidase (EC 3.5.1.4)	MSMEI_2437	23.9168		23.92
Isovaleryl-CoA dehydrogenase (EC 1.3.8.4)	MSMEI_4019	24.1315		24.13
Succinyl-diaminopimelate desuccinylase (SDAP desuccinylase) (EC 3.5.1.18) (N-succinyl-LL-2,6-diaminoheptanedioate amidohydrolase)	dapE MSMEI_4975	24.4091		24.41
Uncharacterized protein	MSMEG_5213	21.9078		21.91
Fatty-acid-CoA ligase FadD18	fadD19 MSMEI_5754	24.2223		24.22
Epoxide hydrolase EphE (EC 3.3.2.-)	ephE MSMEI_6023	23.7190		23.72
ArsR family transcriptional regulator	MSMEI_4375	24.6519		24.65
FHA domain-containing protein FhaB (FtsZ-interacting protein A)	fhaB fipA MSMEG_0034 MSMEI_0036	25.3026		25.30
PPE family protein	ppe68 MSMEI_0065	23.5128		23.51
ABC transporter related protein	MSMEI_0111	23.8679		23.87
S-methyl-5'-thioadenosine phosphorylase (EC 2.4.2.28) (5'-methylthioadenosine phosphorylase) (MTA phosphorylase) (MTAP)	mtnP pnp MSMEG_0990 MSMEI_0963	23.4832		23.48
Membrane-flanked domain protein	MSMEI_1774	23.4235		23.42
MmpS3 protein	MSMEG_1932	24.3337		24.33
Urea amidolyase, allophanate hydrolase subunit (EC 3.5.1.-)	MSMEI_2135	24.8252		24.83
Uncharacterized protein	MSMEI_4581	25.7862		25.79
Uncharacterized protein	MSMEI_5873	24.4957		24.50
Glyoxalase/bleomycin resistance protein/dioxygenase (EC 1.13.11.39)	bphC MSMEI_5875	23.1203		23.12
ABC-type cobalt transport system	MSMEG_6724	23.8481		23.85
Regulatory protein RecX	recX MSMEG_2724 MSMEI_2657	22.5834		22.58
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_0262	23.6867		23.69
Transcriptional regulator, TetR family	MSMEI_0696	23.8025		23.80
Uncharacterized protein	MSMEI_2084	22.8030		22.80

Enoyl-CoA hydratase EchA1 (EC 4.2.1.17)	echA1 MSMEI_2173	23.6545		23.65
Putative methyltransferase	MSMEI_2352	24.6630		24.66
Succinate-semialdehyde dehydrogenase (NAD(P)+) (EC 1.2.1.16)	MSMEI_2428	23.3073		23.31
Alpha/beta hydrolase fold protein (EC 3.-.-.-)	bphD MSMEI_2457	23.4599		23.46
Putative amidase amiC (EC 3.5.1.4)	amiC MSMEI_2462	24.5640		24.56
Transcriptional Regulator, GntR family protein	MSMEG_2546	23.1676		23.17
Conserved protein	MSMEG_2999	24.2120		24.21
Transcriptional regulator, GntR family protein	MSMEG_4057	23.0688		23.07
Putative amidase amiA2 (EC 3.5.1.4)	amiA2 MSMEI_4381	23.6351		23.64
Sec-independent protein translocase protein TatB	tatB MSMEG_5069 MSMEI_4942	24.6120		24.61
TetR family transcriptional regulator	MSMEI_5680	23.8416		23.84
Putative aldehyde dehydrogenase AldA	aldA MSMEI_5700	24.6218		24.62
Periplasmic binding proteins and sugar binding domain of the LacI family protein, putative	MSMEG_6044	22.8817		22.88
Prephenate dehydrogenase (EC 1.3.1.12)	MSMEG_6330	25.2817		25.28
Transcriptional regulator, TetR family	MSMEI_6548	23.6348		23.63
Uncharacterized protein	MSMEI_6368	24.5745		24.57
Uncharacterized protein	MSMEG_0240	24.0093		24.01
Uncharacterized protein	MSMEG_0754	23.9418		23.94
Conserved secreted protein	MSMEI_0871	23.4143		23.41
TetR family transcriptional regulator	MSMEI_0959	24.4039		24.40
SAM-dependent methyltransferase	MSMEI_1020 MSMEI_2271	23.7606		23.76
Conserved integral membrane protein	MSMEI_1619	24.1748		24.17
TetR-family protein transcriptional regulator	MSMEG_1741	24.0797		24.08
Sensor histidine kinase MtrB (EC 2.7.13.3)	mtrB MSMEG_1875 MSMEI_1836	23.5491		23.55
Putative SOS response-associated peptidase (EC 3.4.-.-)	MSMEI_1851	25.0606		25.06
Flavin-binding monooxygenase	MSMEI_1992	23.9213		23.92
Uncharacterized protein	MSMEG_2582	24.3284		24.33
Glycine betaine/L-proline ABC transporter (EC 3.6.3.32)	MSMEI_2853	24.7661		24.77
Protoheme IX farnesyltransferase (EC 2.5.1.-) (Heme B farnesyltransferase) (Heme O synthase)	ctaB MSMEG_3105 MSMEI_3026	22.7467		22.75
Uncharacterized protein	MSMEG_3855	23.7197		23.72
Uncharacterized protein	MSMEG_3918	22.8727		22.87
Putative monooxygenase yxeK	MSMEI_3986	24.6169		24.62
Uncharacterized protein	MSMEI_4088	23.8343		23.83
Acetyl-/propionyl-CoA carboxylase (Beta subunit) AccD1 (EC 6.4.1.3)	MSMEI_4600	23.0650		23.07
Pyruvate ferredoxin/flavodoxin oxidoreductase (EC 1.2.7.8)	MSMEI_4973	23.1820		23.18
Uncharacterized protein	MSMEI_5071	24.9467		24.95

Septum formation initiator subfamily protein, putative	MSMEG_5414	23.5421		23.54
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_5417	24.6747		24.67
Secreted protein	MSMEG_6049	23.8936		23.89
Transcriptional regulator, XRE family	MSMEI_6103	23.4973		23.50
Uncharacterized protein	MSMEG_6464	23.7413		23.74
UPF0232 protein MSMEG_0004/MSMEI_0006	MSMEG_0004 MSMEI_0006	24.5123		24.51
Penicillin-binding protein A (PBPA)	pbpA MSMEG_0031 MSMEI_0033	23.5810		23.58
SMP-30/Gluconolaconase/LRE-like region	MSMEI_0207	22.3474		22.35
Uncharacterized protein	MSMEI_0547	22.6391		22.64
NUDIX hydrolase	MSMEI_1232	23.8818		23.88
RecBCD enzyme subunit RecB (EC 3.1.11.5) (Exonuclease V subunit RecB) (ExoV subunit RecB) (Helicase/nuclease RecBCD subunit RecB)	recB MSMEG_1327 MSMEI_1289	24.8031		24.80
Methyltransferase type 11 (EC 2.1.1.137)	MSMEI_1853	22.7851		22.79
Uncharacterized protein	MSMEI_1894	24.2107		24.21
Aminoglycoside phosphotransferase	MSMEI_1996	25.2951		25.30
Protoporphyrinogen oxidase (EC 1.3.3.4)	hemY MSMEI_2712	23.0840		23.08
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.53)	fabG3 MSMEI_3435	24.0837		24.08
Cytochrome B561	MSMEI_3621	23.5491		23.55
23S rRNA methyltransferase TsnR (EC 2.1.1.34)	MSMEI_3702	23.1498		23.15
Uncharacterized protein	MSMEI_3706	22.4712		22.47
Regulatory protein GntR, HTH	MSMEG_3822	24.1364		24.14
Alpha/beta hydrolase fold protein	MSMEI_3980	22.6932		22.69
Thiamine pyrophosphate enzyme-like TPP-binding protein (EC 2.2.1.6)	ilvX MSMEI_4247	23.4588		23.46
DNA repair protein RecO (Recombination protein O)	recO MSMEG_4491 MSMEI_4380	23.2532		23.25
Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.193)	MSMEI_4391	24.0508		24.05
ATP synthase subunit a (ATP synthase F0 sector subunit a) (F-ATPase subunit 6)	atpB MSMEG_4942	23.6849		23.68
L-proline dehydrogenase (EC 1.5.99.8)	MSMEI_4989	23.6964		23.70
Peptidyl-tRNA hydrolase (PTH) (EC 3.1.1.29)	pth MSMEG_5432 MSMEI_5283	22.8407		22.84
Peptidase S9, prolyl oligopeptidase active site region (EC 3.4.19.1)	MSMEI_5371	23.3055		23.31
Sensor histidine kinase PhoR	MSMEG_5870	24.3740		24.37
Putative enoyl-CoA hydratase echA13 (EC 4.2.1.17)	echA13 MSMEI_5727	22.9992		23.00
Pe family protein	MSMEG_6144	23.5130		23.51
Histidine N-alpha-methyltransferase (EC 2.1.1.44) (Histidine trimethyltransferase)	egtD MSMEG_6247 MSMEI_6086	23.6132		23.61
Uncharacterized protein	MSMEI_6470	24.3132		24.31
Putative ESX-1 scaffolding and assembly protein SaeA	saeA MSMEG_0044 MSMEI_0046	24.1809		24.18

Aldehyde Dehydrogenase (EC 1.2.1.-)	MSMEI_0442	25.3556		25.36
S-adenosyl-L-methionine-dependent methyltransferase (EC 2.1.1.-)	MSMEG_0607	22.8449		22.84
Transcriptional regulator, GntR family protein	MSMEG_0874	22.4689		22.47
Dicarboxylic acid transport integral membrane protein KgtP	kgtP MSMEI_0958	23.6672		23.67
Major facilitator superfamily MFS_1	MSMEI_0994 MSMEI_2245	22.8240		22.82
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_1125	22.5343		22.53
DEAD/DEAH box helicase-like protein	MSMEI_1219	23.2828		23.28
ABC transporter sugar-binding protein (EC 3.6.3.17)	MSMEI_1664	23.9620		23.96
Transcriptional regulatory protein (Probably TetR-family)	MSMEI_1800	23.5540		23.55
Diacylglycerol O-acyltransferase (EC 2.3.1.20)	MSMEI_1842	23.9470		23.95
Conserved alanine and valine rich protein	MSMEI_1888	23.4474		23.45
Uncharacterized protein	MSMEG_2110	24.1287		24.13
Ketopantoate reductase ApbA/PanE-like protein (EC 1.1.1.169)	MSMEI_2846	23.6556		23.66
Ethyl tert-butyl ether degradation EthD	MSMEI_2879	22.3405		22.34
Xylulose kinase	MSMEG_3263	24.5036		24.50
Integral membrane transporter with CBS domains (EC 1.1.1.205)	MSMEI_3552	26.8382		26.84
2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125)	MSMEI_4018	24.5270		24.53
CAIB/BAIF family protein	MSMEG_4120	22.7767		22.78
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_4230	23.5978		23.60
Conserved membrane protein	MSMEI_4772	25.2486		25.25
Acyl-CoA synthase	MSMEG_5291	24.3203		24.32
Glutamine-fructose-6-phosphate transaminase (Isomerizing) (EC 2.6.1.16)	MSMEI_5185	23.3191		23.32
AMP-dependent synthetase and ligase (EC 6.2.1.1)	MSMEI_5500	23.0159		23.02
Uncharacterized protein	MSMEG_5682	22.3480		22.35
Uncharacterized protein	MSMEG_5726	25.0858		25.09
Uncharacterized protein	MSMEG_5741	23.3403		23.34
D-amino acid aminohydrolase (EC 3.5.1.82)	MSMEI_5982	22.3136		22.31
Recombination protein RecR	recR MSMEG_6279 MSMEI_6115	23.9657		23.97
2-dehydropantoate 2-reductase (EC 1.1.1.169) (Ketopantoate reductase)	MSMEI_6206	22.0493		22.05
ABC transporter	MSMEI_6347	24.3180		24.32
UPF0678 fatty acid-binding protein-like protein MSMEG_6574/MSMEI_6396	MSMEG_6574 MSMEI_6396	23.4184		23.42
O-methyltransferase Omt	omt MSMEI_6481	24.1100		24.11
Uncharacterized protein	MSMEI_0935	22.4769		22.48
Deoxyguanosinetriphosphate triphosphohydrolase-like protein	dgt MSMEG_4483 MSMEI_4372	24.8294		24.83

Probable sugar ABC transporter, substrate-binding protein, putative	MSMEG_0505	23.6888		23.69
Methyltransferase type 11	MSMEI_0531	23.9998		24.00
Aminopeptidase Y Metallo peptidase MEROPS family M28A (EC 3.4.11.15)	MSMEI_0788	22.4037		22.40
Uncharacterized protein	MSMEI_0984 MSMEI_2235	22.9457		22.95
tRNA pseudouridine synthase A (EC 5.4.99.12) (tRNA pseudouridine(38-40) synthase) (tRNA pseudouridylate synthase I) (tRNA-uridine isomerase I)	truA MSMEG_1527	23.0296		23.03
Putative transcriptional regulator	MSMEG_2106	22.8582		22.86
Acyl-CoA dehydrogenase (EC 1.3.8.1)	MSMEI_2156	22.9498		22.95
L-lactate 2-monooxygenase	MSMEI_2453	25.1070		25.11
Glutamate transporter permease protein GluD	MSMEG_2725	24.4717		24.47
Diguanylate cyclase (Ggdef) domain protein	MSMEG_2774	23.4828		23.48
Putative ribonuclease D (EC 3.1.-.-)	MSMEI_2709	23.4457		23.45
Peptidoglycan endopeptidase RipA (EC 3.4.-.-) (Resuscitation-promoting factor interaction partner A) (Rpf-interacting protein A)	ripA MSMEG_3145 MSMEI_3064	24.1615		24.16
Precorrin-6x reductase cobK (EC 1.3.1.54)	cobK MSMEI_3786	23.4910		23.49
Integral membrane protein	MSMEI_4110	22.3681		22.37
Peptidase S58 DmpA	MSMEI_4781	22.1945		22.19
Dihydropteroate synthase (DHPS) (EC 2.5.1.15) (Dihydropteroate pyrophosphorylase)	folP2 MSMEI_4959	22.6304		22.63
Putative carboxyltransferase subunit of acetyl-CoA carboxylase (EC 6.4.1.2)	accD3 MSMEI_5492	24.8315		24.83
Uncharacterized protein	MSMEI_5495	22.4706		22.47
Uncharacterized protein	MSMEI_5554	23.0776		23.08
Polysaccharide biosynthesis protein	MSMEG_5968	23.0067		23.01
Transcriptional regulator, LysR family	MSMEI_5967	24.1327		24.13
Transcriptional regulatory protein (Probably TetR-family)	MSMEI_5981	22.2004		22.20
Hercynine oxygenase (EC 1.14.99.50) (Gamma-glutamyl hercynylcysteine S-oxide synthase)	egtB MSMEG_6249 MSMEI_6088	23.8084		23.81
MarR family transcriptional regulator	MSMEI_6336	22.1799		22.18
UDP-glycosyltransferase, MGT (EC 2.4.1.-)	MSMEI_5187	23.7884		23.79
Transcriptional regulator, TetR family	MSMEI_0451	22.1851		22.19
Intracellular protease, PfpI family	MSMEI_0521	23.0365		23.04
Aldehyde oxidase and xanthine dehydrogenase molybdopterin binding protein (EC 1.17.1.4)	MSMEI_1253	22.7767		22.78
NmrA-like protein	MSMEI_2104	22.8076		22.81
Hydrolase MutT1	mutT1 MSMEI_2330	22.8869		22.89
Hemerythrin HHE cation binding region	MSMEG_2415	23.7161		23.72
Dienelactone hydrolase (EC 3.1.1.45)	MSMEI_2383	23.5067		23.51

Fatty-acid-CoA ligase FadD9 (EC 6.2.1.3)	fadD9 MSMEI_2882	24.0277		24.03
Integral membrane protein	MSMEG_4288	24.3369		24.34
Cysteine desulfurase (EC 2.8.1.7)	MSMEI_4425	22.1738		22.17
Putative conserved lipoprotein LppH (EC 2.7.11.1)	MSMEI_4482	24.6317		24.63
Uncharacterized protein	MSMEI_4795	22.2735		22.27
Alpha/beta hydrolase fold-3 domain protein	MSMEI_5083	23.5796		23.58
Sensor-type histidine kinase prrB (EC 2.7.13.3)	prrB MSMEI_5514	22.7458		22.75
Cysteine synthase/cystathionine beta-synthase (EC 2.5.1.47)	MSMEI_6042	24.1747		24.17
Transcriptional regulator, MarR family	MSMEI_6194	23.5161		23.52
Uncharacterized protein	MSMEI_5553	23.2669	25.6065	-2.34
Iron-containing alcohol dehydrogenase (EC 1.1.1.1)	MSMEI_1354	27.6732	29.3632	-1.69
Uncharacterized protein	MSMEG_2261	24.0929	25.4597	-1.37
Major facilitator superfamily protein	MSMEG_1477	23.0847	24.3512	-1.27
Hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4)	MSMEI_2027	24.4340	25.6487	-1.21
Malonyl CoA-acyl carrier protein transacylase (MCT) (EC 2.3.1.39)	fabD MSMEG_4325 MSMEI_4225	28.3663	29.3875	-1.02
1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXP reductoisomerase) (EC 1.1.1.267) (1-deoxyxylulose-5-phosphate reductoisomerase) (2-C-methyl-D-erythritol 4-phosphate synthase)	dxr MSMEG_2578 MSMEI_2516	25.5064	26.4902	-0.98
Glycerol dehydratase (EC 4.2.1.30)	MSMEI_1511	26.4490	27.4192	-0.97
Probable conserved transmembrane protein	MSMEG_0068	23.4950	24.4491	-0.95
Coenzyme pqq synthesis protein E pqqE	pqqE MSMEI_1388	26.4896	27.4035	-0.91
Iron-containing alcohol dehydrogenase (EC 1.1.1.202)	MSMEI_6081	31.1741	32.0476	-0.87
Nickel-dependent hydrogenase large subunit (EC 1.12.5.1)	MSMEI_2207	24.2487	25.1063	-0.86
Protein glcG	MSMEI_1507	25.0372	25.8693	-0.83
Peroxisomal hydratase-dehydrogenase-epimerase (EC 1.1.1.-) (EC 4.2.1.-)	MSMEG_0096	26.4283	27.2502	-0.82
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60) (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP cytidylyltransferase) (MCT)	ispD MSMEI_5916	26.0387	26.8352	-0.80
Glutathione S-transferase (EC 2.5.1.18)	MSMEI_5544	24.0391	24.8088	-0.77
Peptidase M20 domain-containing protein 2	amiB1 MSMEI_1638	27.4492	28.2142	-0.76
TnpR	tnpR	25.0957	25.8408	-0.75
Pyridoxamine 5'-phosphate oxidase-related, FMN-binding protein	MSMEI_5004	27.7580	28.4756	-0.72
Phosphoadenosine phosphosulfate reductase (EC 1.8.4.8) (3'-phosphoadenylylsulfate reductase) (PAPS reductase, thioredoxin dependent)	cysH MSMEG_4528	27.0780	27.7833	-0.71

(PAPS sulfotransferase) (PAdoPS reductase)				
Aldehyde dehydrogenase (NAD+) (EC 1.2.1.3)	MSMEI_1506	29.5148	30.2080	-0.69
Ribosome hibernation promoting factor (HPF)	hpf MSMEI_1839	25.2608	25.9531	-0.69
UspA	MSMEI_3849	24.4900	25.1807	-0.69
Cytoplasmic peptidase PepQ (EC 3.4.13.9)	pepQ MSMEI_2959	25.4586	26.1436	-0.68
Uncharacterized protein	MSMEI_3845	24.8306	25.5117	-0.68
UspA domain protein	MSMEI_3854	29.1580	29.8109	-0.65
Hydrogenase expression/formation protein HypD	MSMEI_2218	24.0584	24.7099	-0.65
Uncharacterized protein	MSMEI_3860	26.6948	27.3447	-0.65
Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)	glpK MSMEI_6577	32.8064	33.4083	-0.60
Quinolinate synthase (EC 2.5.1.72)	nadA MSMEG_3199	27.2601	27.8581	-0.60
Type I antifreeze protein	MSMEG_5479	23.8724	24.4577	-0.59
UspA	MSMEI_5107	29.4082	29.9804	-0.57
Lipolytic enzyme, G-D-S-L	MSMEI_5133	23.1774	23.7213	-0.54
Integral membrane efflux protein EfpA	efpA MSMEI_2557	24.5393	25.0609	-0.52
3-oxoacyl-(Acyl-carrier-protein) synthase 1 KasA (EC 2.3.1.41)	kasA MSMEI_4227	30.1243	30.6148	-0.49
Isochorismate synthase/isochorismate-pyruvate lyase mbtI	MSMEI_4412	23.8216	24.3090	-0.49
Carbonic anhydrase (EC 4.2.1.1)	MSMEI_4858	27.7278	28.2100	-0.48
Inositol-3-phosphate synthase (IPS) (EC 5.5.1.4) (Myo-inositol 1-phosphate synthase) (MI-1-P synthase) (MIP synthase)	ino1 MSMEG_6904 MSMEI_6720	30.6359	31.1163	-0.48
Uncharacterized protein	MSMEI_4205	25.1965	25.6658	-0.47
Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase (NN:DBI PRT) (EC 2.4.2.21) (N(1)-alpha-phosphoribosyltransferase)	cobT MSMEI_4174	25.0598	25.5201	-0.46
Uracil phosphoribosyltransferase (EC 2.4.2.9) (UMP pyrophosphorylase) (UPRTase)	upp MSMEI_1654	25.2948	25.7468	-0.45
Uncharacterized protein	MSMEI_6418	27.5421	27.9926	-0.45
ABC Fe3+-siderophores transporter, periplasmic binding protein	MSMEI_4447	26.1477	26.5962	-0.45
Alpha/beta hydrolase fold protein	MSMEI_6721	24.8268	25.2744	-0.45
Acetamidase/Formamidase	MSMEI_4278	26.1102	26.5463	-0.44
Alcohol dehydrogenase, zinc-binding protein	MSMEI_2034	30.5800	30.9971	-0.42
4-hydroxy-2-oxovalerate aldolase 2 (HOA 2) (EC 4.1.3.39) (4-hydroxy-2-keto-pentanoic acid aldolase 2) (4-hydroxy-2-oxopentanoate aldolase 2)	bphI-2 MSMEG_5937 MSMEI_5778	27.0856	27.4998	-0.41
L-cysteine:1D-myo-inositol 2-amino-2-deoxy-alpha-D-glucopyranoside ligase (L-Cys:GlcN-Ins ligase) (EC 6.3.1.13) (Mycothiol ligase) (MSH ligase)	mshC cysS2 MSMEG_4189 MSMEI_4091	26.8026	27.2097	-0.41

Putative neutral zinc metallopeptidase	MSMEG_4893	28.0438	28.4179	-0.37
Thiopurine S-methyltransferase (Tpmt) superfamily protein	MSMEG_6235	25.8255	26.1934	-0.37
tRNA-specific 2-thiouridylase MnmA (EC 2.8.1.13)	mnmA trmU MSMEG_2358 MSMEI_2298	25.0996	25.4396	-0.34
Oligopeptide/dipeptide ABC transporter	MSMEI_4870	24.2348	24.5702	-0.34
Putative S-adenosyl-L-methionine-dependent methyltransferase MSMEG_1482/MSMEI_1446 (EC 2.1.1.-)	MSMEG_1482 MSMEI_1446	23.2518	23.5708	-0.32
Glutamine-binding periplasmic protein/glutamine transport system permease protein	MSMEG_6307	27.3402	27.6506	-0.31
GTP cyclohydrolase 1 type 2 homolog	MSMEI_4206	24.2923	24.5970	-0.30
Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)	glpK MSMEG_6229	27.9117	28.2109	-0.30
Amidohydrolase	MSMEG_1703	25.5796	25.8558	-0.28
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_3347	25.6746	25.9208	-0.25
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7) (Enoylpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl transferase) (EPT)	murA MSMEI_4805	27.5325	27.7604	-0.23
Cystathionine gamma-synthase	metB MSMEI_5127	28.7999	28.9909	-0.19
Putative cytochrome P450 135B1 (EC 1.14.-.-)	MSMEG_6478	23.4487	23.6386	-0.19
Cob(I)alamin adenosyltransferase (EC 2.5.1.17)	cobO MSMEG_2616	24.3178	24.5029	-0.19
Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (5'-phosphoribosylglycinamide transformylase) (GAR transformylase) (GART)	purN MSMEG_5516	24.5075	24.6720	-0.16
Putative UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	galU MSMEI_5321	27.8386	27.9589	-0.12
Antibiotic biosynthesis monooxygenase	MSMEI_5928		25.6157	-25.62
Conserved hypothetical membrane protein	MSMEI_0236		27.2043	-27.20
PE	pE MSMEG_0412		25.4197	-25.42
Uncharacterized protein	MSMEI_2688		25.7335	-25.73
Oxidoreductase FAD-binding region	MSMEI_1702		24.8425	-24.84
Probable conserved membrane protein	MSMEG_0235		25.1555	-25.16
YCII-related protein	MSMEI_1008 MSMEI_2259		25.2993	-25.30
Beta-lactamase (EC 3.5.2.6)	MSMEI_3884		25.4626	-25.46
Conserved regulatory protein	MSMEI_4142		25.3494	-25.35
Bifunctional enzyme mbtA: salicyl-AMP ligase + salicyl-S-ArCP synthetase	mbtA MSMEI_4404		25.1570	-25.16
Phosphohistidine phosphatase SixA	MSMEI_4878		24.6136	-24.61
Uncharacterized protein	MSMEI_5617		25.3092	-25.31
Glyoxalase II, GloB (EC 3.1.2.6)	MSMEI_1296		26.0212	-26.02

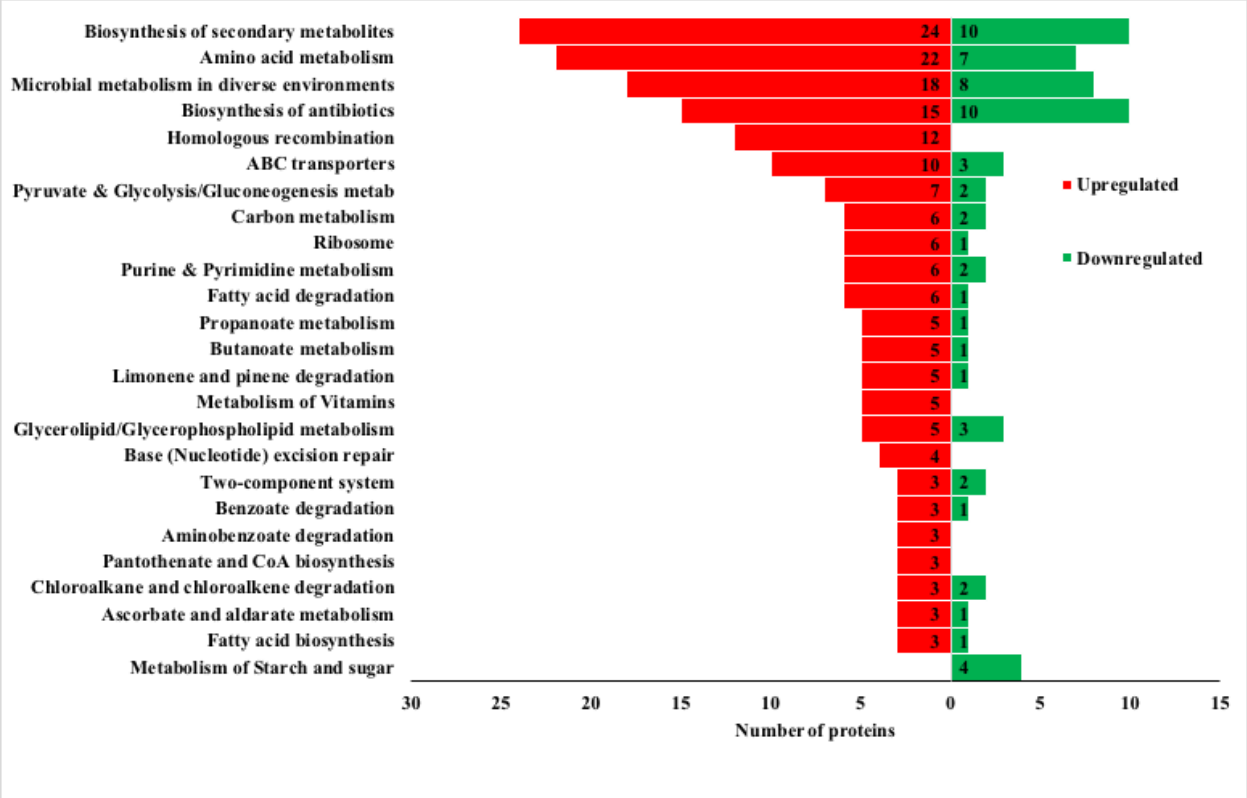
Fructose-bisphosphate aldolase (EC 4.1.2.13)	MSMEI_1336		24.3199	-24.32
ABC transporter ATP-binding protein (EC 3.6.3.-)	modC MSMEI_1969		24.4573	-24.46
Hemerythrin HHE cation binding domain protein	MSMEI_2929		23.4007	-23.40
Uncharacterized protein	MSMEG_5501		25.2345	-25.23
Cutinase (EC 3.1.1.74)	MSMEG_5878		23.7491	-23.75
Uncharacterized protein	MSMEI_5730		24.6378	-24.64
Histidine kinase (EC 2.7.13.3)	MSMEI_1598		24.4549	-24.45
Uncharacterized protein	MSMEG_1422		25.0973	-25.10
Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	MSMEI_3267		24.0168	-24.02
Sugar-transport integral membrane protein ABC transporter SugB	sugB MSMEI_4932		25.5766	-25.58
FMN-dependent NADH-azoreductase (EC 1.7.-.-) (Azo-dye reductase) (FMN-dependent NADH-azo compound oxidoreductase)	azoR MSMEI_3301		24.3754	-24.38
Putative ESX-1 scaffolding and assembly protein SaeB	saeB MSMEG_0045 MSMEI_0047		23.1726	-23.17
ESAT-6-like protein EsxA	esxA MSMEG_0066 MSMEI_0067		25.5184	-25.52
50S ribosomal protein L28	rpmB-3 rpmB MSMEI_2340		27.4361	-27.44
Aldo/keto reductase	MSMEI_2415		23.3652	-23.37
Shikimate 5-dehydrogenase AroE (EC 1.1.1.25)	aroE MSMEI_2953		24.7891	-24.79
Conserved hypothetical membrane protein	MSMEI_5932		26.7414	-26.74
Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB) (1-phosphofructokinase protein) (EC 2.7.1.56)	MSMEI_0084		23.6179	-23.62
Amino acid permease-associated region	MSMEI_0271		24.5684	-24.57
Putative ethyl tert-butyl ether degradation protein EthD	MSMEI_0275		23.7362	-23.74
Putative S-adenosyl-L-methionine-dependent methyltransferase MSMEG_1479/MSMEI_1443 (EC 2.1.1.-)	MSMEG_1479 MSMEI_1443		24.5721	-24.57
Peptidase M22, glycoprotease	MSMEG_1578		24.7280	-24.73
Aspartate ammonia-lyase (EC 4.3.1.1)	MSMEI_1637		24.4326	-24.43
Ribosome maturation factor RimP	rimP MSMEG_2624 MSMEI_2562		24.6569	-24.66
Conserved membrane glycine rich protein	MSMEI_2925		24.6963	-24.70
NAD-dependent protein deacylase Sir2 (EC 3.5.1.-) (Regulatory protein SIR2 homolog)	sir2 cobB npdA MSMEG_5175 MSMEI_5041		23.8532	-23.85
Uncharacterized protein	MSMEI_6093		24.3424	-24.34
Galactan 5-O-arabinofuranosyltransferase (EC 2.4.2.46) (Arabinofuranosyltransferase AftA)	aftA MSMEG_6386 MSMEI_6218		24.3194	-24.32

GCN5-related N-acetyltransferase (EC 2.3.1.128)	MSMEI_0183		23.4903	-23.49
Uncharacterized protein	MSMEG_0251		26.1409	-26.14
ATP-dependent RNA helicase	MSMEG_1540		24.7942	-24.79
NADH ubiquinone oxidoreductase 20 kDa subunit (EC 1.12.99.6)	MSMEI_2206		23.6137	-23.61
Uncharacterized protein	MSMEI_2561		23.7475	-23.75
Putative metal-sulfur cluster biosynthetic enzyme	MSMEI_3048		25.3265	-25.33
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.-)	MSMEI_3520		24.1986	-24.20
Uncharacterized protein	MSMEI_4246		23.3535	-23.35
Secreted protein	MSMEG_6180		24.6975	-24.70
Uncharacterized protein	MSMEI_6032		24.8499	-24.85
Iron-containing alcohol dehydrogenase (EC 1.1.1.202)	MSMEI_6078		23.8094	-23.81
Nitroreductase	MSMEI_6096		24.2417	-24.24
Urea amidolyase related protein (EC 3.5.1.54)	MSMEI_0424		24.5754	-24.58
Uncharacterized protein	MSMEG_0672		27.3493	-27.35
UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.3.1.98) (UDP-N-acetylmuramate dehydrogenase)	murB MSMEG_0928 MSMEI_0906		23.3955	-23.40
Alanine dehydrogenase (EC 1.4.1.1)	ald MSMEG_2659 MSMEI_2596		24.6082	-24.61
Short-chain dehydrogenase/reductase SDR (EC 1.1.1.100)	MSMEI_3216		23.0703	-23.07
Virginiamycin B lyase (EC 4.2.99.-) (Streptogramin B lyase)	vgb MSMEG_4914 MSMEI_4787		25.0768	-25.08
Trehalose 6-phosphate phosphatase (EC 3.1.3.12)	MSMEI_5882		22.7912	-22.79
Transcriptional regulator, LacI family	MSMEI_0478		24.0398	-24.04
Uncharacterized protein	MSMEI_1397		24.8742	-24.87
Purine catabolism PurC-like protein	MSMEI_2386		22.7587	-22.76
Sugar ABC transporter, ATP-binding protein	MSMEI_3186		24.3194	-24.32
Uncharacterized protein	MSMEI_3948		24.5594	-24.56
Methyltransferase type 11 (EC 2.1.1.80)	MSMEI_4202		24.3809	-24.38
Putative citrate lyase (Beta subunit) (EC 4.1.3.6)	MSMEI_4596		24.0047	-24.00
Short-chain dehydrogenase/reductase SDR	MSMEI_5703		24.5637	-24.56
Uncharacterized protein	MSMEI_5764		23.5267	-23.53
Aldehyde dehydrogenase (EC 1.2.1.-)	MSMEI_6507		23.7906	-23.79
Aminotransferase class-III (EC 4.1.1.64)	MSMEI_6659		23.6393	-23.64
Lipid transfer protein or keto acyl-CoA thiolase Ltp4 (EC 2.3.1.-)	MSMEI_5762		23.7044	-23.70
Glyoxalase/bleomycin resistance protein/dioxygenase	MSMEI_0205		22.1058	-22.11
Glyoxalase/bleomycin resistance protein/dioxygenase	MSMEI_0591		24.6420	-24.64
Peptidase S8 and S53, subtilisin, kexin, sedolisin	MSMEI_0608		22.7769	-22.78

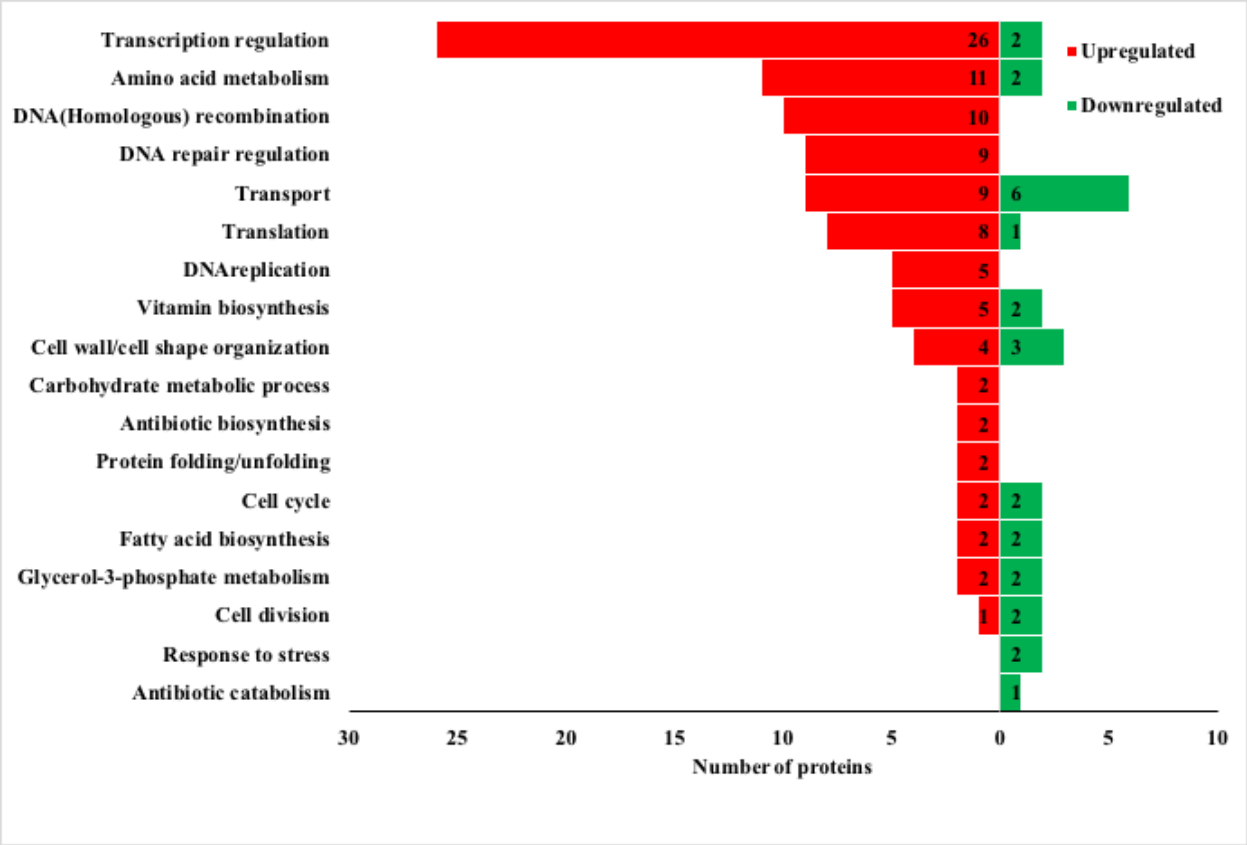
Putative conserved transmembrane protein	MSMEI_0933		24.5003	-24.50
OsmC family protein	MSMEI_2360		24.2808	-24.28
tRNA dimethylallyltransferase (EC 2.5.1.75) (Dimethylallyl diphosphate:tRNA dimethylallyltransferase) (DMAPP:tRNA dimethylallyltransferase) (DMATase) (Isopentenyl-diphosphate:tRNA isopentenyltransferase) (IPP transferase) (IPPT) (IPTase)	miaA MSMEG_2734 MSMEI_2667		25.4000	-25.40
Unidentified antibiotic-transport integral membrane ABC transporter	MSMEI_3038		23.6508	-23.65
Uncharacterized protein	MSMEG_3674		24.0981	-24.10
Alpha/beta hydrolase	MSMEI_4340		24.5374	-24.54
Polyketide synthase MbtD	mbtD MSMEI_4400		22.3510	-22.35
Glycine cleavage T protein (Aminomethyl transferase) (EC 2.1.2.10)	MSMEI_5642		24.5149	-24.51
NADP-dependent fatty aldehyde dehydrogenase (EC 1.2.1.4)	MSMEG_6134		24.6050	-24.60
Oxidoreductase, FAD/FMN-binding	MSMEG_6486		23.3380	-23.34
Uncharacterized protein	MSMEI_6596		25.1892	-25.19
Class II aldolase/adducin	MSMEI_6680		22.8196	-22.82
Short-chain dehydrogenase/reductase SDR	MSMEI_1699		24.4727	-24.47
Transcriptional regulator, GntR family	MSMEI_2429		23.6522	-23.65
Glyoxalase/bleomycin resistance protein/dioxygenase	MSMEI_2868		24.0545	-24.05
Putative conserved transmembrane protein	MSMEI_4120		24.9316	-24.93
Putative ATP-dependent DNA helicase	MSMEI_4785		22.0612	-22.06
Zinc-binding oxidoreductase (EC 1.6.5.5)	MSMEI_6565		23.0061	-23.01

APPENDIX B. PATHWAY AND FUNCTION CLASSIFICATIONS OF ALL DIFFERENTIALLY REGULATED PROTEINS

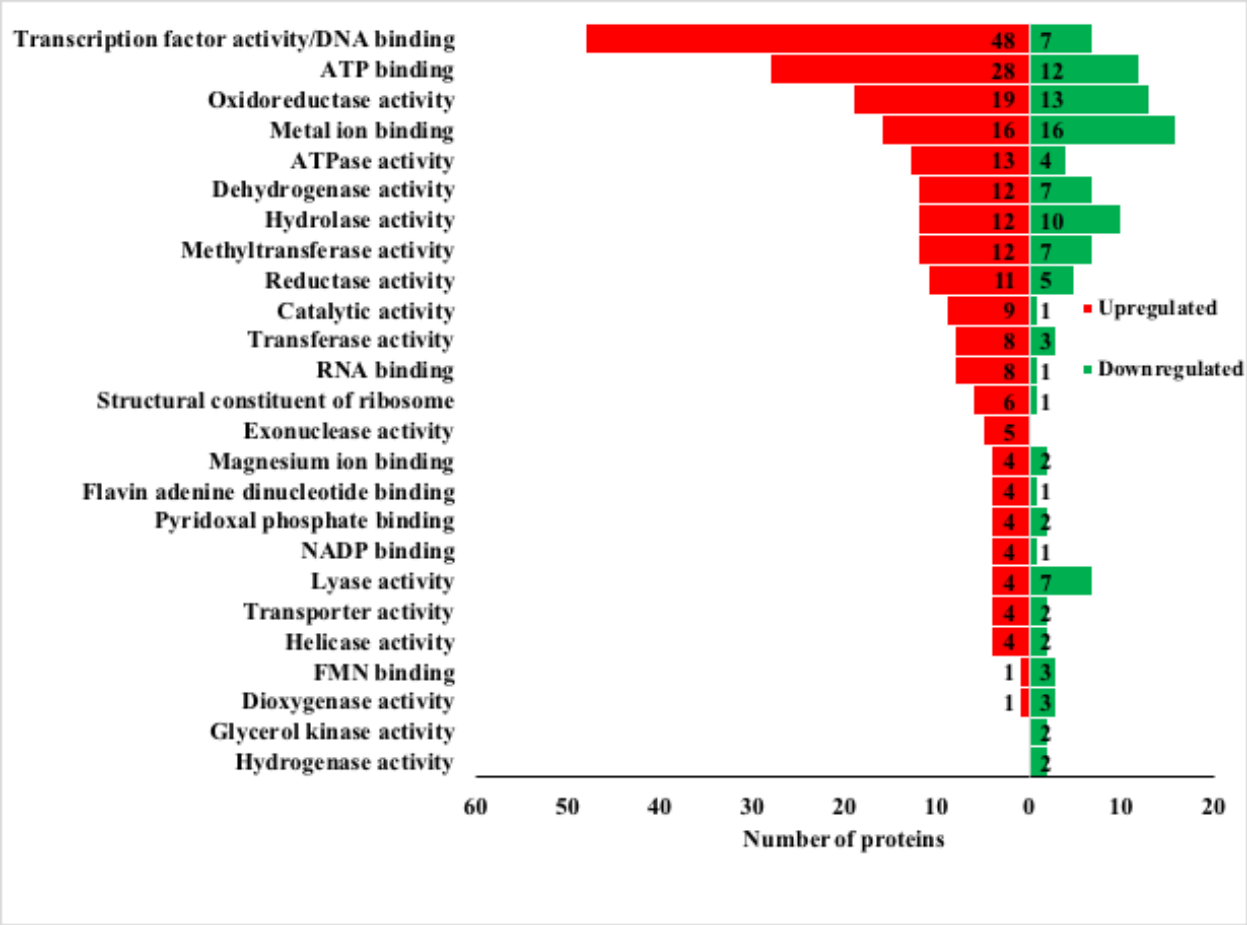
I. Pathway classification according to KEGG database



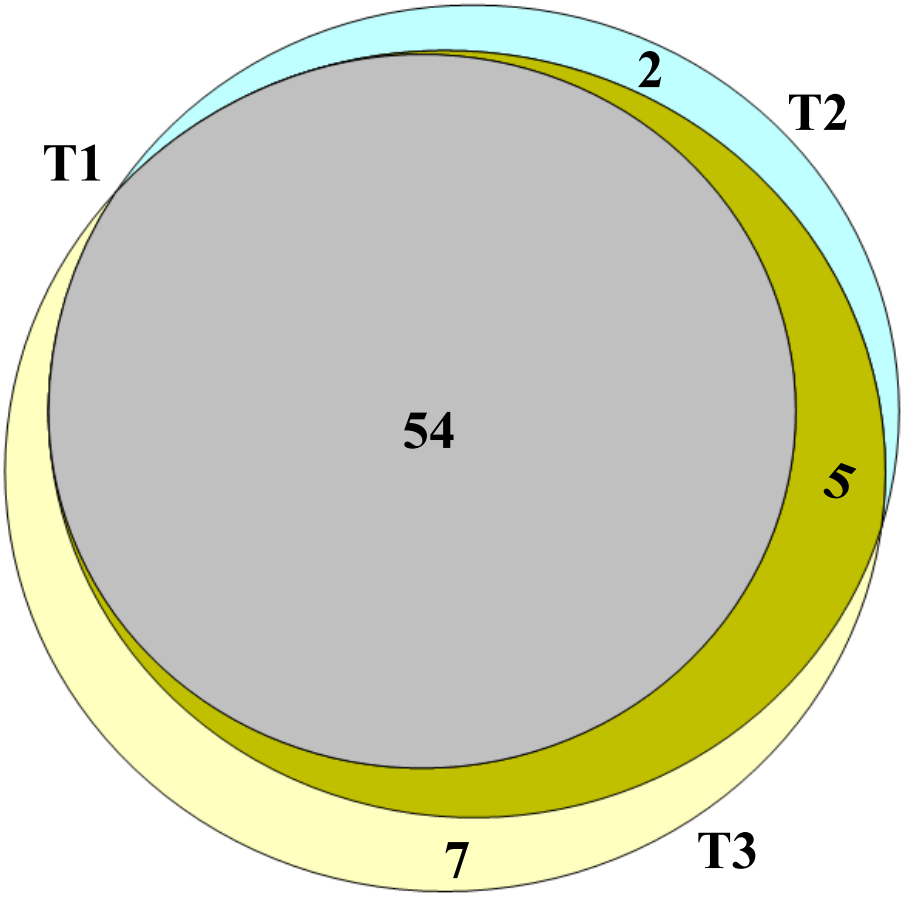
II. Biological process classification according to GO terms



III. Molecular function classification according to GO terms

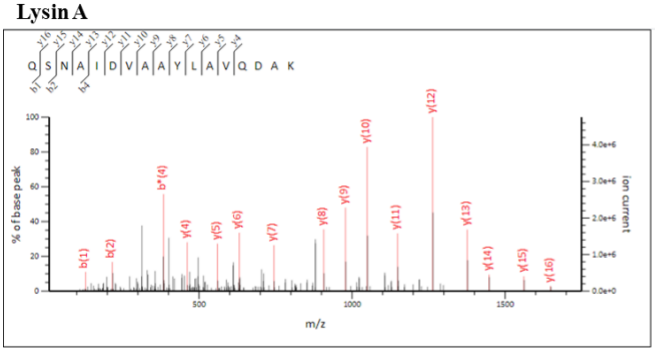


APPENDIX C. NUMBER OF OCHI17 PROTEINS IDENTIFIED IN ALL THREE OCHI17-TREATED (T) REPLICATES

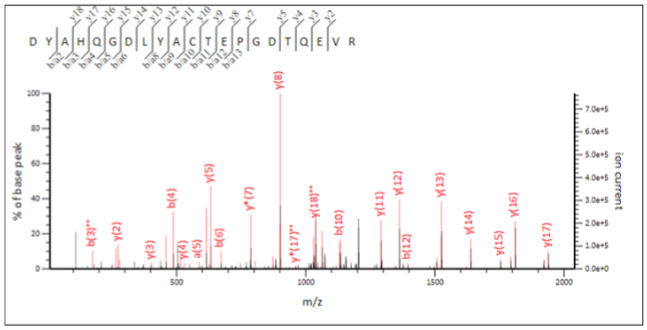


APPENDIX D. MS/MS FRAGMENTATION OF HIGHEST SCORING REPRESENTATIVE PEPTIDES FROM SELECTED PHAGE PROTEINS. SPECTRA WERE COLLECTED FROM MASCOT.

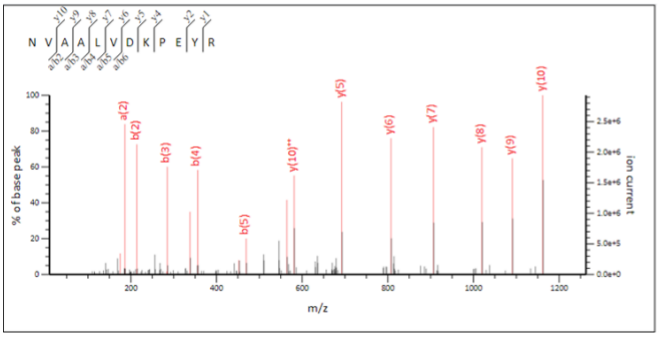
Lysin A



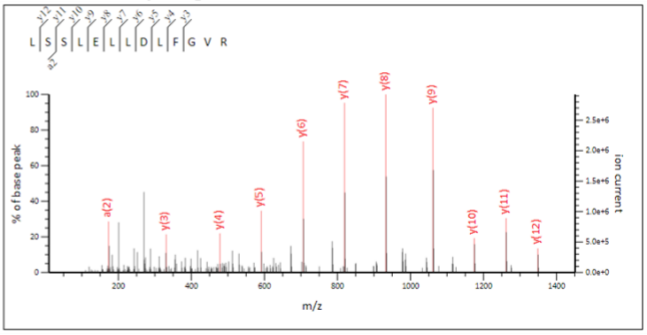
Lysin B



Integrase



Tail assembly chaperone



The red spectra represent the experimental masses, while the black spectra are the theoretical masses.