Diversity of *Salmonella* Typhi-phages in Dhaka City through Whole Genome Sequencing

By

Shuborno Islam
ID: 22376001

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of
Master of Science in Biotechnology

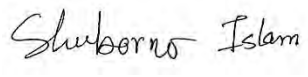
Department of Mathematics and Natural Sciences
BRAC University
June 2024

DECLARATION

I, Shuborno Islam, hereby declare that the dissertation/thesis entitled "Diversity of *Salmonella* Typhi-phages in Dhaka City through Whole Genome Sequencing" is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgment of collaborative research and discussions.

This thesis has been composed by me and has not been submitted in any previous application for any degree or qualification at this or any other institution. The work reported herein has been carried out by me under the supervision of Senjuti Saha, PhD at Child Health Research Foundation and Iftekhar Bin Naser, PhD at BRAC University

I have duly referred to and acknowledged the work of others and all sources of information and data used in the preparation of this thesis. This thesis complies with the ethical standards and guidelines set forth by Child Health Research Foundation and BRAC University.



Shuborno Islam
ID: 22376001
Department of Mathematics and Natural Sciences
BRAC University, Dhaka, Bangladesh

Approval

The thesis titled Diversity of *Salmonella* Typhi-phages in Dhaka City through Whole Genome Sequencing" submitted by

1. Shuborno Islam (22376001)

of Spring 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Master of Science in Biotechnology on 27th June 2024.

Examining Committee:

Supervisor:
(Internal)

_____
Iftekhar Bin Naser, PhD
Associate Professor, Department of Mathematics and Natural Sciences
BRAC University

Supervisor:
(External)

_____
Senjuti Saha, PhD
Deputy Executive Director
Child Health Research Foundation

Program Coordinator:
(Member)

_____
Munima Haque, PhD
Associate Professor, Department of Mathematics and Natural Sciences
BRAC University

External Expert Examiner:
(Member)

_____

Departmental Head:
(Chair)

_____
Dr. Firoze Haque
Associate Professor and Chairperson, Department of Mathematics and Natural Sciences
BRAC University

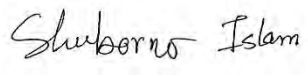
ETHICS STATEMENT

This research was conducted in accordance with the ethical standards and guidelines provided by Child Health Research Foundation and BRAC University. All procedures performed in this study involving environmental samples adhered to the ethical standards of the institutional and national research committees.

All environmental sampling methods were designed to be minimally invasive and to have no significant adverse effects on the ecosystems being studied. Efforts were made to ensure that the collection of water samples from various locations in Dhaka city did not disrupt local wildlife or contaminate the environment.

Although this study did not involve human subjects, the data collected was handled with the utmost care to ensure accuracy and integrity. Sample data were stored securely in a OneDrive file provided by Child Health Research Foundation, and all findings were reported in a manner that respects the privacy and confidentiality of any indirectly involved human communities.

In conducting this research, respect for local communities and their environments was a priority. Engagement with local stakeholders was conducted in a respectful and non-intrusive manner, ensuring that local customs and regulations were observed.



Shuborno Islam
ID: 22376001
Department of Mathematics and Natural Sciences
BRAC University, Dhaka, Bangladesh

“The more that you read, the more things you will know. The more that you learn, the more places you’ll go.”

-

Dr. Seuss

The Cat in the Hat

This page has been intentionally left blank

Table of Contents

List of Figures 8
List of Table 8
ABSTRACT 9
INTRODUCTION 10
Bacteriophage biology and its role in the ecosystem 10
Typhoid fever and its impact in a community 13
Bacteriophages and *Salmonella* Typhi in the environment 16
Objectives 18
METHODOLOGY 19
Overview 19
Sample Selection 20
Propagation and Purification of Phages 22
Extraction of Phage gDNA 22
Library Preparation of Phage gDNA 25
Computational Analysis for Phage gDNA Sequences 32
Conclusion 38
Data Availability 39
RESULTS 40
Overview 40
Genomic Characterization and Speciation of Typhi-phages 41
Genome Map and Plotting 43
Diversity of Typhi-phages 45
Conclusion 47
DISCUSSION 48
LIMITATIONS 53
REFERENCE 54

List of Figures

Figure 1: Structural overview of *Escherichia* T4 phages. 11

Figure 2: Types of phages and their phage infection cycles. 12

Figure 3: Global incidence rate of typhoid fever per 100 000 person-years in 2021,..... 15

Figure 4: Location of water sample collection and detection of *Salmonella* Typhi phages in water samples of urban Dhaka and rural Mirzapur, Bangladesh. 17

Figure 5: The workflow of this study. 20

Figure 6: The Infectivity Spectrum of the isolated phages collected from March 2023 to May 2023..... 21

Figure 7: Genome map of Bacteriophage Genus and species-wise. 44

Figure 8: Phylogenetic relationship of Typhi-phages over time and location. 46

List of Table

Table 1: Samples were selected for sequencing by studying the Infectivity Spectrum of the phages collected. 21

Table 2: Speciation of the 15 Typhi-phages sequenced in this study and their corresponding infectivity spectrum..... 41

Table 3: Genomic Characterization and Speciation of the 15 Typhi-phages sequenced in this study. 42

ABSTRACT

Bacteriophages and their hosts form a vast ecosystem, with bacteria evolving diverse defense mechanisms against phage infection while phages simultaneously develop strategies to overcome these defenses. Limited data on *Salmonella* Typhi phages hampers understanding of their genomics, evolution, and interactions with bacterial hosts. This dissertation aims to study *Salmonella* Typhi phages in the environmental water sources of Dhaka city, focusing on their genomic characterization and structure.

Phage DNA was sequenced using a combination of propagation experiments, genomic DNA extraction, and sequencing library preparation. Computational tools, including Pharokka and Kraken2, were employed to analyze the genetic makeup and taxonomical classification of these phages. A phylogenetic tree was generated using tail fiber protein sequences to explore evolutionary relationships with mid-20th-century phages.

Recent findings indicate the circulation of three phage genera, with *Kayfunavirus* being dominant, followed by *Teseptimavirus* and *Macdonaldcampvirus*. The generated genome map revealed gene structures and arrangements, highlighting annotated proteins required by the phages and further denoted large sequences of hypothetical proteins that can be crucial for phage survival and host interaction.

These insights contribute to understanding pathogen behavior, the spread of antibiotic-resistant genes, and environmental survival mechanisms, underscoring the importance of phages in microbial ecology.

Keywords: Bacteriophages, *Salmonella* Typhi, Whole Genome Sequencing, Phage diversity

INTRODUCTION

Bacteriophage biology and its role in the ecosystem

Bacteriophages are viruses that infect bacteria, and they shape the ecology and evolution of microbial communities. The interaction between phages and bacteria has been exploited to study the evolutionary mechanisms and composition of bacterial communities. A typical bacteriophage varies in size from 20nm to 200nm and the structure consists of several distinct elements that have important functions based on their life cycle. The head (capsid) of a phage is a key component composed of protein subunits that protect the phage's genetic material by forming a protective shell around it. This phage capsid has different structures; polyhedral, filamentous, and pleomorphic. The genomes of phages can either be DNA or RNA that may be double-stranded (ds) or single-stranded (ss). The tail structure of a phage is made of a tail tube, sheath, baseplates and tail fibers which are all connected to each other. The tail tube, connected to the phage capsid protein in one end and is surrounded by the sheath, is a hollow tube that allows the injection of phages genomic material inside the host bacteria, while the sheath contracts to fore the tail tube inside the bacterial cell wall. The baseplate that is connected to the other end of the tail tube is important is recognition and attachment as they have receptor-binding proteins to recognize certain molecules on the surface of bacterial cell. The tail fiber proteins are important as they facilitate the attachment to bacterial cell surface by recognizing specific receptors on the surface, that makes phages very specific to their hosts. These tail fibers are protein fiber extensions from the baseplate. Additionally, some phages have special structures that contribute to the stability and infectivity of phages, these structures include collar, neck, whiskers or spikes {Figure 1} (Dion et al., 2020; Yap & Rossmann, 2014). Most phages have a lytic infectious cycle (virulent phage) where they inject their genome into a phage-specific host and produce new viral progeny which kills the host cell and is released. There are some lysogenic phages (temperate phages) that incorporate their genome into the specific bacterial host genome and do not kill the host until the host undergoes stress which is when the lysogenic phages change their cycle (Runa et al., 2021). Recently, there have been certain diverse phages known as chronic phages that do not kill the host upon release. These phages bring certain changes within the host, such as by increasing the pathogenicity of the host cells {Figure 2} (Chevallereau et al., 2022).

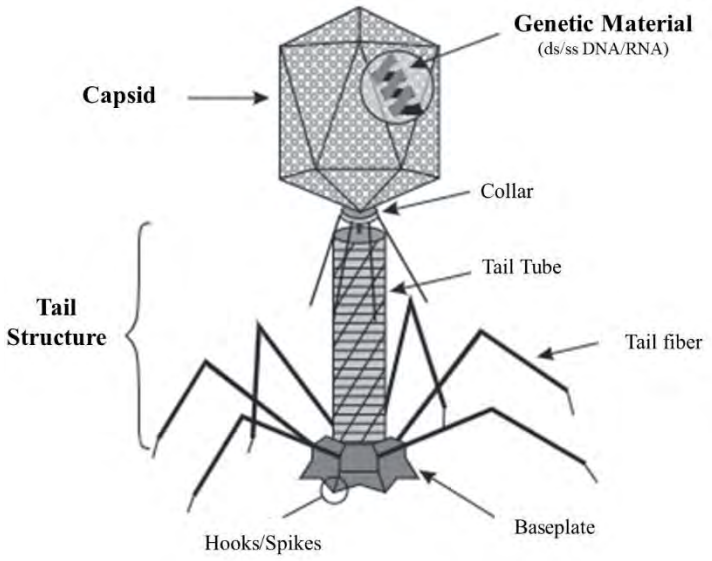


Figure 1: Structural overview of *Escherichia* T4 phages.
(Tokarz-Deptuła et al., 2011)

Bacterial populations are key players in the microbial world that shape many nutrient and biogeochemical cycles. It has also been realized through studies that outside the environmental microbial life, bacteria is present in medical applications, such as research on human microbiome gave insights on normal body functions. Until recently, attention has turned to the viruses of bacteria and through a series of studies, it was widely acknowledged that these bacterial viruses, bacteriophages, are ubiquitous and more diverse than any other microbial community. In spite of this, it is not clear how that phages shape the bacterial composition in the environment and it's a vital role in the environment by driving bacterial evolution and making a stability in the ecosystems. (Díaz-Muñoz & Koskella, 2014). Their interactive relationship with bacteria is complex as it also happens to influence bacterial virulence to eukaryotic fitness to the carbon cycle (Martiny et al., 2006).

In oceans and other aquatic bodies, these phages control the flow of carbon through viral infection and the killing of bacteria which is known as viral shunt (Suttle, 2005). Phages are also present in the human gut as they modulate the bacteria which has an impact on the physiology and metabolism of their hosts (Kim & Bae, 2018). In seawater, approximately 10^5 - 10^7 viral particles are present per millimeter while in soil the abundance lies between 10^3 - 10^9 viral particles per gram (Jansson & Wu, 2023) while the number increases to 10^{10} viral particles in the mammalian gut (Kim & Bae, 2018).

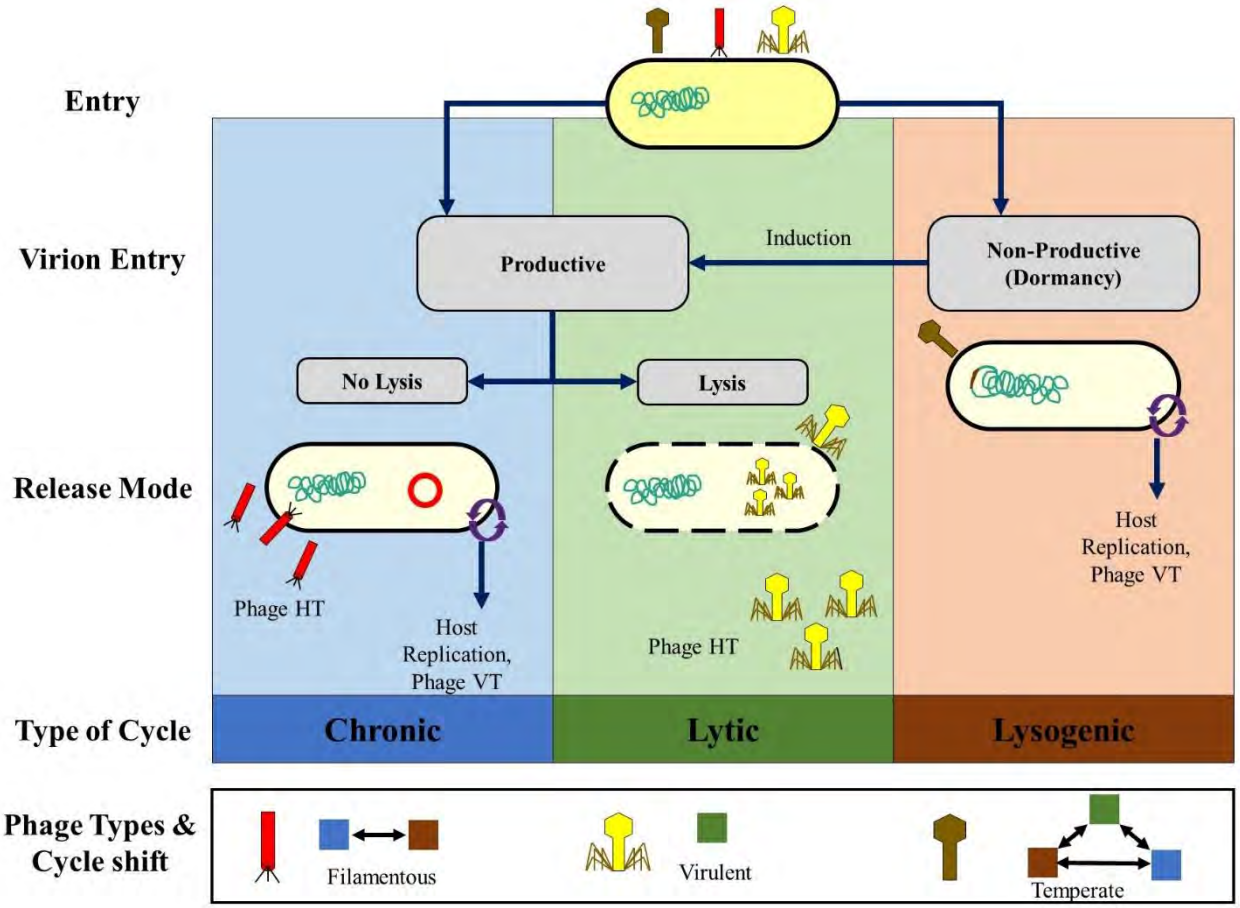


Figure 2: Types of phages and their phage infection cycles.
Adapted from: Chevallereau et al., 2022. Nature reviews Microbiology

A recent metagenomic study suggested that 96% of all phages that form the predominant biological entity on earth accounts to be tailed bacteriophages with double-stranded DNA genome; that is, the order of *Caudoviricete* and are subdivided into families based on their tail morphology while having the same major capsid protein (HK97 fold) (Zinke et al., 2022). The genome size of phages varies by a wide range and characteristics, such the recent discoveries of megaphages having over 200 kbp genome size and other unique features. Due to this diversity in phage size, structural morphology and genomic organization with unique genomic features, phages infecting different bacterial hosts generally show low genomic similarity among themselves. However, it is also known that phages infecting same hosts have substantial differences in their genome (Zhu et al., 2022). The exchange of genetic material between phages and bacteria depends dynamically on their environment, bringing in major changes in the population of both microbes over time. The study of phage genomic characteristics and their corresponding hosts has been going on for the

past two decades which resulted in several changes in phage phylogeny and reshaping it. Viral metagenomics provided insights of phage diversity in the ecosystem and global diversity and expanded our knowledge about the evolutionary perspective that several factors are counted in when it comes to high or low rate of gene exchange with its hosts' evolutional as well (Dion et al., 2020). Understanding the broader spectrum of phages and their interaction with bacteria can be achieved by studying more diverse phages of multiple hosts and adding more phage genome sequences to the databases, and the corresponding expansion of catalog of newly studied phage genomes can provide an in-depth knowledge about current viral taxonomical classification. This complex relationship among phages are visualized using a phylogenetic tree, which is suited to present new phage genera, subfamilies and families (Barylski et al., 2020).

Typhoid fever and its impact in a community

Typhoid fever is an invasive bacterial disease leading to bloodstream infection and fever caused by *Salmonella enterica* subspecies *enterica* serovar Typhi and is transmitted through contaminated food and water. This pathogen is a human-restricted pathogen that has no non-human animal reservoir and has evolved to become highly adapted to humans. This is an important cause of illness in areas with poor sanitation, having an estimate of 10.9 million cases and 116, 800 typhoid fever-related deaths annually. It poses a significant threat to many low- to middle- income countries (Stanaway et al., 2019), particularly countries in South Asia having the highest standardized incidence rate of enteric fever {Figure 3}. The Surveillance for Enteric Fever in Asia Project (SEAP) conducted from 2016 to 2019 revealed significant data on typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan. The study found high incidence rates of typhoid fever, particularly in Dhaka, Bangladesh, with an adjusted rate of over 913 cases per 100,000 person-years. The study underscores the substantial burden of enteric fever and the importance of targeted vaccination and healthcare interventions in these regions Infection of *S.* Typhi occurs via the fecal – to – oral system when an individual takes in contaminated food or water. During the early 1900s, typhoid fever was the first disease to be studied for asymptomatic carriage as a source of disease transmission, that includes the case of Mary Mallon. (Kirchhelle et al., 2019). This further led to understand the necessity of preventive measures for asymptomatic chronic carriers with antimicrobials to prevent the introduction of new cases of typhoid fever (Meiring et al., 2023).

The pathogen has earned its nickname as “stealth” pathogen as it evades the immune system and detection during acute infection to spread. Having 89% of shared genes with the related pathogen *Salmonella* Typhimurium which induces intestinal inflammation, the pathologies of typhoid fever caused by *S.* Typhi are very different (Brewer et al., 2021).

Most endemic countries lack the facilities and resources for proper diagnosis and antimicrobial susceptibility testing of *S.* Typhi through blood culture which leads to unauthorized use and overuse of unnecessary antimicrobials. This caused the emergence of multi-drug resistant MDR (multidrug resistance) *Salmonella* Typhi in Pakistan, India, and other Southeastern Asian countries in the 1980s which were resistant against three first-line drugs (ampicillin, chloramphenicol and co-trimoxazole) and by 1997, there were cases of ceftriaxone-resistant *S.* Typhi isolates circulating in Bangladesh, and other countries of South Asia. Primary treatment then focused on fluoroquinolones. By 2015, more than 90% of *Salmonella* Typhi isolates showed non-susceptibility to fluoroquinolones. With the increasing case reports of AMR in in-patient and out-patient departments and the community, third-generation cephalosporins became a primary drug of choice for treating suspected and confirmed cases of typhoid fever. The wide spread of antimicrobials caused an outbreak in Pakistan in 2016 by an extensively drug-resistant (XDR) strain of *Salmonella* Typhi that is resistant to chloramphenicol, ampicillin, co-trimoxazole, fluoroquinolones, and cephalosporins. By the next 4 years, >80% of *Salmonella* Typhi population consisted of XDR *Salmonella* Typhi. In 2020, 10 countries had detected cases of XDR *Salmonella* Typhi strains, and this initiated the use of azithromycin, the last option for oral antimicrobial. With the increasing use of azithromycin, it is a matter of time to detect cases of XDR *S.* Typhi isolates with azithromycin resistance.

The introduction of vaccines in communities is one solution to protect the vulnerable from typhoid fever and bring fast changes compared to substantial investments for infrastructural changes. Typhoid Conjugate Vaccines (TCV) are recommended for routine immunization programs and the World Health Organization granted prequalification of two TCVs: Typbar TCV and Typhibev, to prevent typhoid fever and the spread of drug-resistant *S.* Typhi (S. K. Saha et al., 2021).

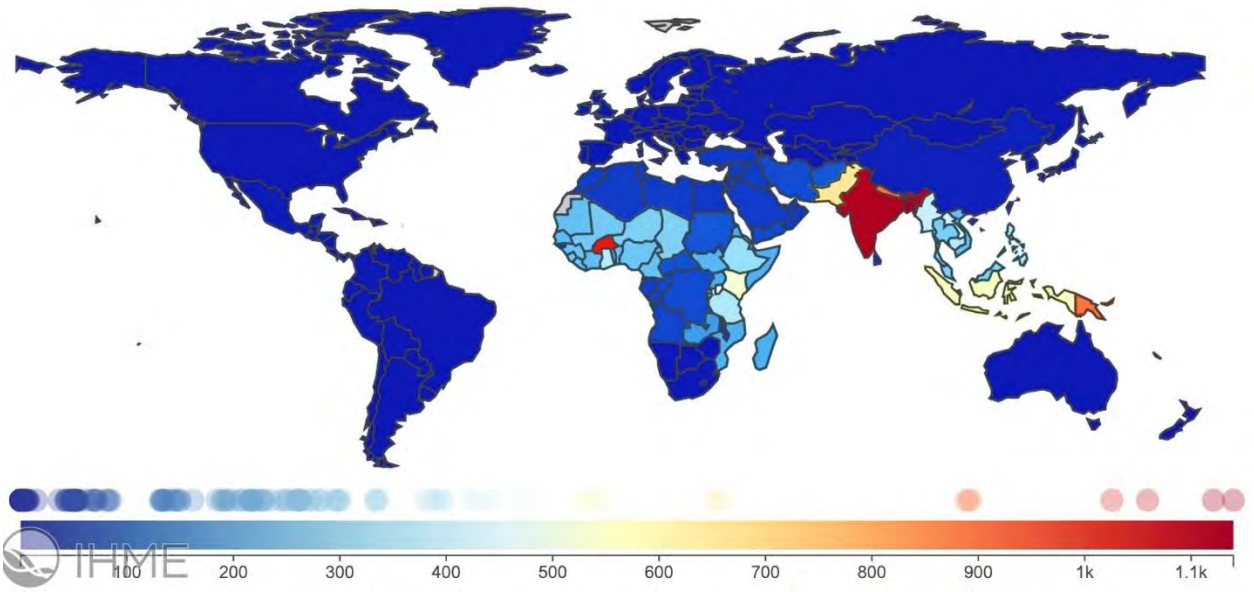


Figure 3: Global incidence rate of typhoid fever per 100 000 person-years in 2021,
Both sexes, <5 years. Areas with the highest incidence is shown in red, and areas with the lowest incidence in blue.
Source: GBD Compare, The Institute for Health Metrics and Evaluation

Through whole-genome sequencing and phylogenetic analysis, studies showed the transmission and emergence of antimicrobial resistance in *S.* Typhi that led to outbreaks. The International Typhoid Consortium analyzed over 2000 genomes across 69 countries and developed a genotyping scheme GenoTyphi, 68 single-nucleotide variants (SNV) markers were utilized that showed 4 primary clades, 16 clades, and 49 subclades (Wong et al., 2016). This GenoTyphi scheme was updated using the next five years of global pathogen population data after analyzing over 4500 published genomes. H58 (4.3.1) is spread out across most continents and is subdivided into 3 major lineages (H58 lineage 1: 4.3.1.1; H58 lineage 2: 4.3.1.2 and H58 lineage 3: 4.3.1.3). Bangladesh makes up one-third of the H58 lineage 1 population with genotype 4.3.1.3 is a monophyletic cluster being originated from Bangladesh and has another sublineage designated as 4.3.1.3. BDq, that is resistant to fluoroquinolones (MIC (Minimal Inhibitory Concentration): 4µg/mL) (Dyson & Holt, 2021). In Pakistan, along with Genotype 4.3.1.1 and derived genotypes, there is an emergence of the XDR sublineage (genotype 4.3.1.1 P1) that appeared in 2016 and became dominant across the nation. This led to the outbreak of XDR *Salmonella* Typhi strains. Non-H58 strains also circulate within Southern Asia with Bangladesh having the greatest diversity, having 3.2.2, 3.3.2 and 2.3.3. (Carey et al., 2023).

Bacteriophages and *Salmonella* Typhi in the environment

S. Typhi survives for a long period outside the human host in a viable, not-culturable state that eventually contributes to the survival and transmission of *S.* Typhi over long distances and prolonged time. This mechanism involves changes in the expression of certain metabolic and respiratory genes in aqueous and other environments (Cho & Kim, 1999) The presence of *S.* Typhi in the environment was observed through nucleic acid amplification which enhanced our knowledge about the contamination of *S.* Typhi in the community (S. Saha et al., 2018). Previously, studies were conducted to investigate and study *Salmonella* Typhi interaction and behavior within the human host, about pathogenicity and immunogenicity, but there is very little information available about the pathogen's life cycle in the environment. One such way to understand the pathogen through an environmental context is to study bacteriophages from the environment.

Bacteriophages have been observed during the early 1900s and they were initially used for controlling bacterial infections and many studies were done to investigate their efficacy. Over the past few decades, there have been increased interest in the bacteriophages of different bacteria and their interactions have been studied (Sharp, 2001). During this time, researchers started looking into the genomes of phages which further demonstrated that there is a huge diversity among different phages, and they have been evolving over billions of years Yet researchers conclude that they studied the surface of the phage genome exploration and future studies as well as studying phages from different hosts will reveal more (Hatfull & Hendrix, 2011). During the 1940s – 1980s, Typhi phages were used for bacterial typing until the advancement of molecular diagnostics methods (Farmer et al., 1975). Over the last 50 years, there is a lack of literature reviews on Typhi phages from typhoid-endemic countries that motivated researchers to investigate phage to understand the *Salmonella* Typhi in the environment.

In a recent study in Bangladesh by Child Heath Research Foundation (CHRF), Typhi-phages were isolated from environmental water sources that was compared with clinical data to correlate the presence of phages in water with the prevalence of local typhoid fever cases in Dhaka, an urbanized city and Mirzapur, rural district {Figure 3}. During this study, CHRF studied the diverse characterization of 86 phages by infecting them with different *Salmonella* Typhi genotypes. This further demonstrated that the Typhi strains that are circulating in Bangladesh are not equally susceptible to all phages circulating in the environment (Hooda et al., 2024).

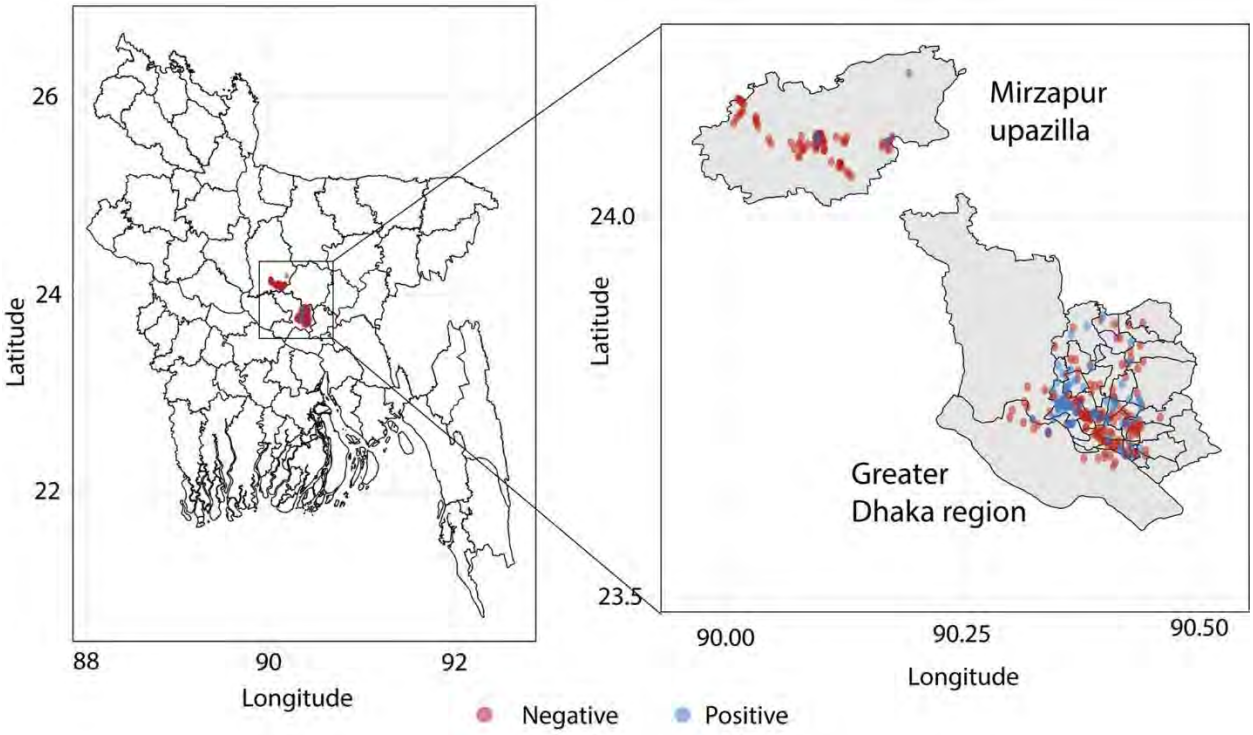


***Figure 4*:** Location of water sample collection and detection of *Salmonella* Typhi phages in water samples of urban Dhaka and rural Mirzapur, Bangladesh.
(Hooda et al., 2024)

CHRF has been collecting sewage samples for the last 2 years from 7 different thanas in Dhaka city where Typhi-phages are isolated and characterized, and their positivity is being correlated with typhoid fever positivity within these thanas. They are furthermore looking into the host range specificity and their interaction with 19 genotypes of *Salmonella* Typhi that circulate in Bangladesh.

Objectives

Over the years, the genome of Typhi phages has been studied on a small scale. Pickard *et al*., 2008, 2010 studied Typhi phages where they investigated tail fiber proteins. The phages they studied were from the mid-1900s. Till then, very few Typhi phages have been studied at a genomic level and as the phages evolve, there is a gap in understanding the current phages that circulate in the environment which interacts with *Salmonella* Typhi itself.

Phages live upon infecting bacterial host cells to produce new phage particles and releasing them. Phage tails have special molecular techniques that aid them to recognize bacterial host cells. These tail proteins, having different structure, help in infecting by penetration through the bacterial cell envelope and releasing the phage genome. Phage tails have been a keen interest in molecular biology as they determine the infection process and host specificity. These proteins have receptor-binding proteins (RBPs) at the end of the tail and interact with bacterial cell wall (Nobrega et al., 2018).

Through this research, we intend to use genomics to study the following objective:

Objective 1: To look for phages that are unique and kills the circulating the *Salmonella* Typhi genotypes. This is to be done by studying the Infectivity spectrum.

Objective 2: Using whole genome sequencing, to characterize the diverse Typhi-phage families at a genomic level

Objective 3: Using the tail-fiber protein to understand the evolutionary changes using a phylogenetic tree

METHODOLOGY

Overview

This chapter describes the methodological approaches that were employed in this dissertation with detailed research design, sample selection, and processing followed by data generation and data analysis methods. The following methodology is instrumental in achieving the objectives that were outlined in the research questions.

From a pilot study at the Child Health Research Foundation (CHRF), we learned about phage diversity and active interaction between phages and *S.* Typhi circulating in typhoid-endemic areas (Hooda et al., 2023). At present, phages are being isolated from an ongoing environmental surveillance study conducted by CHRF, which focuses on bacteriophages and the spread of drug-resistant *Salmonella* Typhi. The purpose of this environmental surveillance project is to study the impact of environmental niches on the spread of antibiotic resistance in *S.* Typhi.

From March 2023, the Child Health Research Foundation has been studying Typhi-phage dynamics in 7 thanas around Dhaka city. By focusing on sewage samples, 60 phages have been isolated from 204 sewage samples over 3 months. Upon collection, 15 mL of each sewage sample was centrifuged to pellet large soil particles, and the supernatant was filtered using a 0.22µm PES syringe filter. The filtered samples were enriched to increase phage number by incubating with the *S.* Typhi BRD948 strain. Upon enrichment, the samples were lawned for lysis and plaque observation by phages using the Double-Layer Agar Technique (DLA). Positive samples were further diluted and lawned to purify isolated phages. Isolated phages were diluted and spotted over *S.* Typhi BRD948 strain to calculate their Plaque-forming Unit (PFU), and the Infectivity Spectrum was generated by testing against 19 *Salmonella* Typhi genotypes to investigate their interacting diversity against the circulating *S.* Typhi genotypes in Bangladesh.

In addition, with plaque assays and plaque morphology, whole genome sequencing in important to learn more about phage genotypes, such as to understand which genes the phages carry and encode proteins and how the phages are different from other phages on a molecular level. To study their genomes using computational bioinformatics tools, a series of steps are required that involve

propagation and purification of isolated phages followed by gDNA extraction of phages and library preparation {Figure 4} (Islam et al., 2024).

Given the compatibility of research aims and the advantageous utilization of available resources, we opted to integrate these samples into this thesis to avoid duplication of samples and data collection.

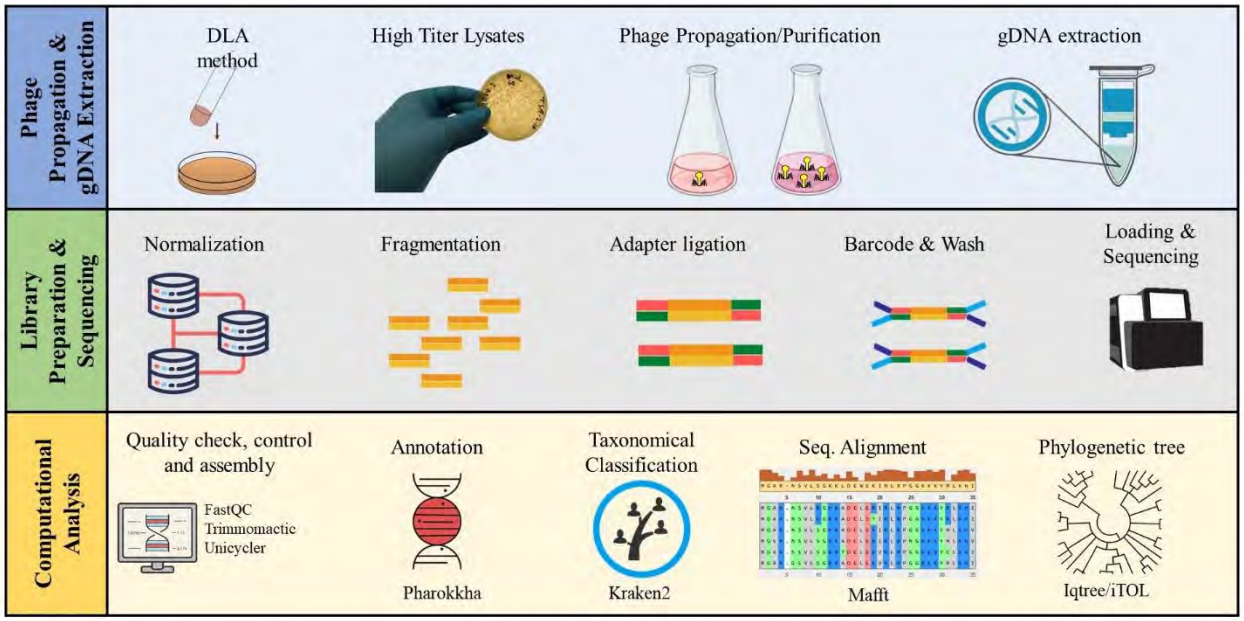


Figure 5: The workflow of this study.

Sample Selection

In this study, the Infectivity Spectrum is defined as how to characterize the phage interaction with the genotypes of *Salmonella* Typhi. By investigating the spectrum, 15 unique phage samples from clusters were selected for sequencing that kills the circulating Typhi genotypes (4.3.1.1 and 4.3.1.3) in Bangladesh. Moreover, certain genotypes (3.3.2 BD1, 3.3.2 BD2, and 4.3.1.3 BDq) have active phage defense systems in their genome and/or plasmids, and based on the interaction in the Infectivity Spectrum, phages have also been selected {Figure 5}.

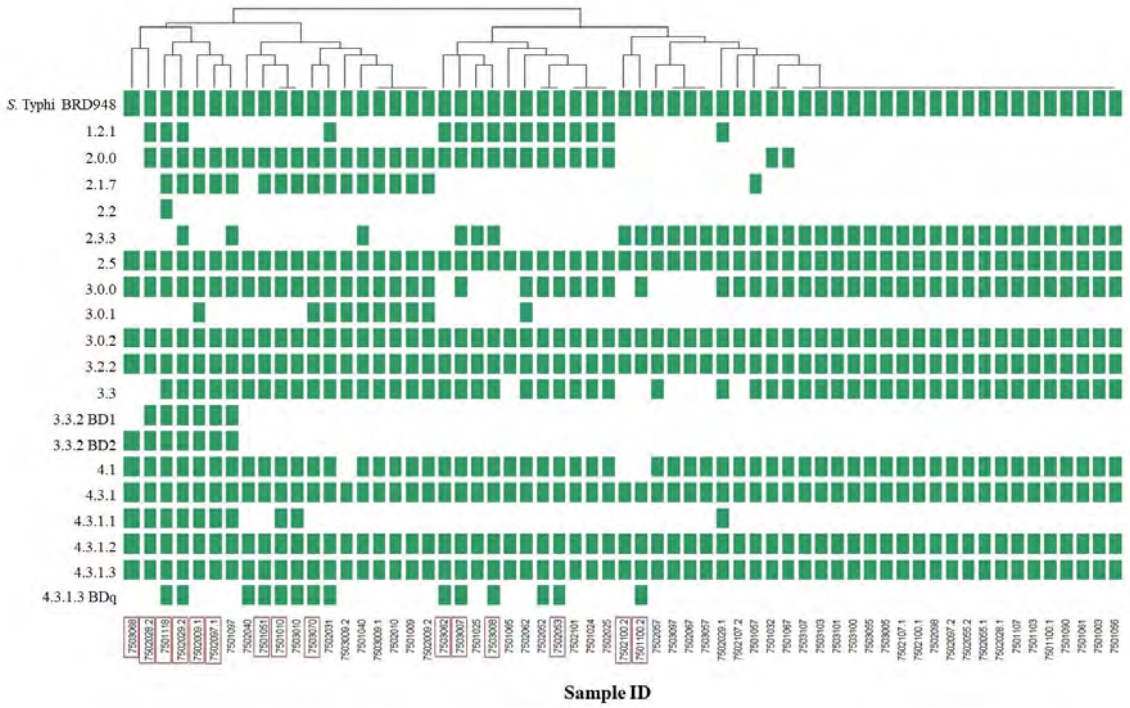
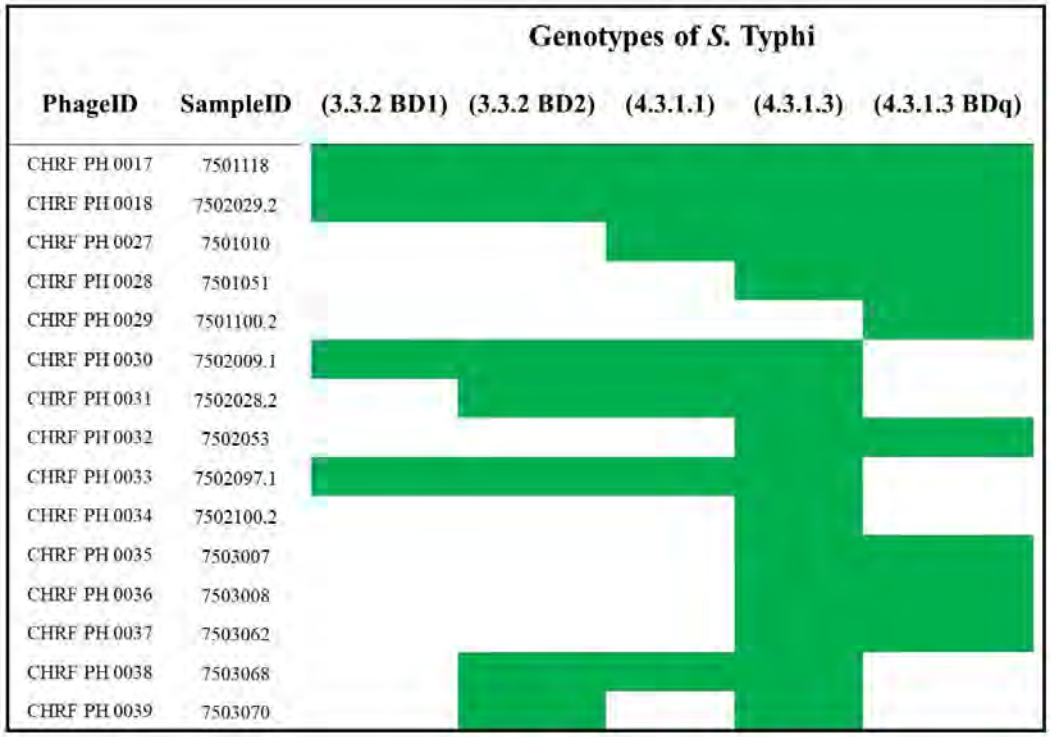


Figure 6: The Infectivity Spectrum of the isolated phages collected from March 2023 to May 2023.
Green indicates that phages infected the *S.* Typhi genotype. Red marked boxes indicate samples selected sequencing.

Table 1: Samples were selected for sequencing by studying the Infectivity Spectrum of the phages collected.
Green indicates that phages infected the *S.* Typhi genotype of interest.



Propagation and Purification of Phages

Phage propagation involves creating high-titer phage lysates which is achieved by growing living phage plaques over agar plates through Confluent Plate Lysis. Confluent Lysed Plates have densely packed confluent and living phages with a "web" of host bacteria living in between. In the pursuit of detecting phage activity, a systematic methodology was implemented.

For this experiment, *S.* Typhi BRD948 was used which is an attenuated *S.* Typhi Ty2 derivative that contains deletions in *aroC* and *aroD* and heat shock protein *HtrA*; making this strain to be used in containment level-2 laboratory.

To commence the experiment, co-culture preparations were initiated by adding 20 – 100 µL of each phage stock with at least 5000 pfu/mL of phage particles with 200 µL of overnight grown *Salmonella* Typhi BRD948 liquid culture and allowed to incubate for 20 minutes at room temperature.

Following the Double-Layer Agar Method, the incubated co-culture was added to 4 mL of molten 0.7% TSA (Tryptic Soy Agar) agar and subsequently poured over 1.5% TSA plates. These plates were then subjected to an incubation period of 16-18 hours at 37°C.

Post-incubation, the plates underwent observation for the presence of confluent lysis phage plaques. To elute phages, 4 mL of LB was methodically flooded over the plates and incubated at 4°C for 4 hours which allowed the phages to diffuse into the LB. In instances where plates exhibited no evidence of confluent lysis by phages, a repetition of the experiment was undertaken with varied volumes of 10^{-4} dilutions of phage stocks. Delicately using the edge of a pipette tip, the top agar layer was gently scraped off and subsequently transferred to a microcentrifuge tube. To this, 4 drops of chloroform were added, ensuring thorough mixing, followed by incubation at room temperature for 10 minutes. Post-incubation, the mixture underwent centrifugation at 10,000 rcf for 10 minutes, and the resulting supernatant was carefully transferred to a fresh microcentrifuge tube. This supernatant is the propagated phage lysates and is stored at 4°C for gDNA extraction.

Extraction of Phage gDNA

Genomic DNA extraction stands as a foundational process, laying the groundwork for comprehensive whole-genome sequencing and a deeper comprehension of genetic elements. In

this dissertation section, the protocol for extraction of genomic DNA from bacteriophages is explained in a step-by-step process along with all the crucial steps and modifications. The successful extraction of high-quality genomic DNA is pivotal, as it serves as a backbone for subsequent applications, including whole-genome sequencing, genotyping, and cloning. This section is crafted with the utmost precision, aiming to reproduce and obtain intact genomic material for downstream genomic analyses.

The extraction process unfolds through a well-defined protocol, utilizing a formulated kit coupled with an initial treatment step. During this treatment phase, enzymes selectively eliminate genetic material not intrinsic to the phage particles. Subsequently, the capsid protein of the phages undergoes digestion, and the remaining steps involve a series of washing using the QIAamp DNA mini kit [Cat: 51306, Qiagen, Germany]. This methodical approach ensures the extraction of pure and high-quality genomic DNA, establishing a robust foundation for the ensuing genomic analyses.

Removal of bacterial DNA/RNA:

In the preparation of the Digestion Enzyme Master mix, a solution was prepared using 50 µL of 10X DNA Digestion Buffer [Cat: E1010-1-4, Zymo Research, USA], 1 µL of DNase I (1 U/µL) [Cat: E1009-A, Zymo Research, USA], and 0.5 µL of RNase A (20 mg/mL) [Cat: 3018-2, New England Biolabs, USA]. This mixture underwent slow pipetting for thorough mixing and subsequently 51.5 µL of the master mix was added to 450 µL of phage lysates. Without vortexing or shaking, the resulting mixture was then subjected to an incubation period at 37°C for 90 minutes. Following the incubation period, an interruption of enzyme activity was introduced by adding of 20 µL of 0.5 M EDTA [Cat: 15575-038, GIBCO, Invitrogen Corporation, USA]. The tubes were allowed to incubate for an additional 10 minutes. This step minimizes the residual digestion activity of the enzymes, ensuring that subsequent analyses accurately reflect the initial DNA content in the phage lysates.

Digestion of Phage capsid:

In the preparation of 20 mg/mL Proteinase K, 1 mL of Proteinase K Storage Buffer [Cat: D3001-2-G, Zymo Research, USA] was added to 20 mg of lyophilized Proteinase K tubes [Cat: D3001-2-B, Zymo Research, USA] followed by thorough mixing. Subsequently, in the treatment of phage

lysates, 1.25 µL of Proteinase K solution was added. Without inducing any vortexing or shaking, the resultant mixture was incubated at 56°C for 90 minutes. This process yielded lysed phage lysates and is now ready for phage DNA extraction.

Phage DNA Extraction:

In the extraction and purification of phage genomic DNA, the QIAamp DNA minikit [Cat: 51306, Qiagen, Germany] was used. The reagents necessary for the extraction were prepared according to the user manual instructions.

In a microcentrifuge tube, 200 µL of the lysed phage lysate was combined with 200 µL of AL Buffer. The resulting mixture was vortexed and incubated at 70°C for 10 minutes. Following incubation, 200 µL of ice-cold 99.9% Ethanol [Merck, Germany] was added to the lysate-AL Buffer mixture, and vigorous vortexing was performed. This ethanol-containing mixture was subsequently transferred to a QIAamp DNA minikit spin column and centrifuged at 8,000 rcf for 1 minute. The filtrate was discarded, and the spin column was moved to a new 2 mL collection tube. To the spin column, 500 µL of AW1 (25 mL of 99% Ethanol added to concentrated 19 mL AW1 buffer) was added and centrifuged at 8,000 rcf for 1 minute. The filtrate was discarded, and the spin column was transferred to a new collection tube. This process was repeated with 500 µL of AW2 (30 mL of 99% Ethanol added to concentrated 13 mL AW1 buffer), centrifuging at 14,000 rcf for 1 minute. After discarding the filtrate, the spin column was transferred to a new collection tube and centrifuged at 14,000 rcf for 3 minutes to eliminate residual wash buffers and ethanol, which could potentially inhibit subsequent steps.

For the final elution step, the spin column was transferred to a 1.5 mL microcentrifuge tube. A total of 30 µL AE buffer was added directly over the spin column. After incubating the spin column for 5 minutes to ensure saturation of the filter, centrifugation at 8 000 rcf for 1 minute was performed to elute the DNA. To ensure complete elution, the eluate was then centrifuged at 8,000 rcf for 1 minute, passing through the same spin column.

Quality Check and Visualization

Using a Thermofisher Scientific Nanodrop One C, the A260/280 DNA purity quality was checked for each sample. Following the successful extraction of phage genomic, gel electrophoresis was applied to assess the quality of the DNA samples.

A 0.8% agarose gel was prepared by dissolving agarose powder in 1X TAE buffer. The mixture containing 0.8 grams of agarose powder in 100 mL of 1X TAE buffer was heated until completely dissolved, followed by cooling to approximately 60°C. To this, 2 µL of SYBR Safe was then added to the molten agarose gel for DNA visualization. DNA samples, obtained through the QIAamp DNA minikit extraction process, were mixed with 2.5X BlueJuice Gel Loading Buffer in a 1:1 ratio. The mixture was carefully loaded into the wells of the agarose gel using micropipettes. The gel was subjected to electrophoresis in 1X TAE buffer at 100 volts and 400 mAmp for approximately 55 minutes. During electrophoresis, DNA migrates through the gel based on size, resulting in distinct bands that can be visualized under UV light. After electrophoresis, the gel was placed in a Vilber UV transilluminator, and the DNA bands were visualized under UV light. SYBR Safe intercalates with DNA, and DNA bands can be seen under fluorescence. The resulting gel image was captured using gel documentation equipment. The gel electrophoresis results provided valuable insights into the quality of the extracted phage genomic DNA, that is ready for Whole Genome Sequencing.

Library Preparation of Phage gDNA

Through DNA sequencing, genetic information is revealed that is embedded within the phages. In this dissertation, the methodology employed for DNA sequencing is to achieve the objectives. The sequencing strategy adopted involves Next-Generation Sequencing (NGS) in the Illumina Platform and has been tailored to address the specific research questions posed in this study.

Quantification and Normalization using Qubit:

Normalization before library preparation ensures equal representation of DNA samples to reduce biases from sequencing depth and technical variations for accurate downstream genomic analysis. In this dissertation study, genomic DNA (gDNA) underwent a systematic serial dilution, achieving a 100-fold reduction in concentration. Subsequently, the concentrations of the samples were normalized to a final volume of 26 µL, ensuring a consistent DNA content of approximately 52 ng per sample. The quantification of double-stranded DNA (dsDNA) was executed through the utilization of the Qubit™ 1x dsDNA High Sensitivity HS and Broad Range BR Assay Kits [CAT: Q33231, Invitrogen, Thermofisher Scientific, USA].

To measure the concentrations of the 100-fold diluted samples, a working solution was aliquoted and added to both standards and individual samples. Before any experimental procedures, reagents and dye were equilibrated to room temperature, allowing for optimal conditions for 30 minutes.

For measuring standards, 190 µL of the Qubit™ 1X dsDNA HS BR working solution was dispensed into two separate Qubit™ Assay Tubes [CAT: Q32856, Invitrogen, Thermofisher Scientific, USA], each designated for a specific standard. Subsequently, 10 µL of each Qubit™ 1X dsDNA HS BR standard was introduced into the respective tube, followed by a brief vortexing for 2-3 seconds. On the other hand, for each sample, 198 µL of the Qubit™ 1X dsDNA HS BR working solution was allocated to separate tubes and 2 µL of the 100-fold diluted samples was added to each tube, followed by a vortexing step lasting 2-3 seconds. All the tubes were incubated in darkness for 2 minutes.

Following the incubation period, the concentrations of all tubes, beginning with the standards for calibration, were measured by the Qubit 3 Fluorometer (Thermofisher Scientific, USA). The measurements were conducted under the High Sensitivity Range to ensure accurate and precise quantification of DNA concentrations across all samples and standards. If the Qubit 3 Fluorometer read, "Too Low" for any sample, a new Qubit™ 1X dsDNA HS working solution was prepared using the 10-fold dilution of the respective sample and measured again. The concentrations of the diluted samples were recorded, and the concentration of the raw samples was calculated by the dilution factor. Using this, each of the samples was normalized up to 26 µL with approximately 52 ng of DNA.

Fragmentation of gDNA and End Repair

Library preparation for DNA sequencing was conducted following a protocol and reagents from the NEBNext® Ultra™ II FS DNA Library Prep (Cat: E7805L, New England Biolabs, USA). The library preparation involves the fragmentation of genomic DNA (gDNA) with a desired range of 300 bp – 700 bp. During the end of the fragmentation reaction, the end of the fragments was repaired by 5'-phosphorylation and extended 3'-dA tailing which is required for adapter ligation.

To initiate the fragmentation process, a master mix was prepared with 7 µL of NEBNext Ultra II FS Reaction buffer and 2 µL of NEBNext Ultra II FS Enzyme mix. Ensuring thorough homogenization, the master mix was vortexed and 9 µL aliquots of the master mix were dispensed

into each of the sample tubes. Each of the tubes contained normalized DNA, with a concentration of 52 ng in 26 µL.

The Master mix table and PCR profile are given below:

Component	Volume (µL)
NEBNext Ultra II FS Reaction Buffer (Yellow tube)	7
NEBNext Ultra II FS Enzyme mix (Yellow tube)	2
Normalized DNA (~52ng)	26
Total	**35**

Step	Time	Temperature
Fragmentation	7 minutes	37°C
Enzyme Deactivation	30 minutes	65°C
Hold	∞	4°C
Lid ON: 105°C		

Adapter Ligation:

During adapter ligation, adapters attached to DNA fragments help in systematic library preparation, cluster generation, and sequencing. The adapter sequences have primer binding regions, and index sequences, and have sites that can bind to the flow cell lawn. The adapter is a hairpin-loop structure with an Uracil (U) in the middle and binds to the 3`-dA tailing. It enables accurate template identification with signal generation and subsequent bioinformatics analysis for reliable NGS data generation. NEBNext® Multiplex Oligo for Illumina® Adapters [Cat: E6612A, New England Biolabs, USA] were pre-diluted 100-fold using NEBNext® Multiplex Oligo for Illumina® Buffer [Cat: E7762AA, New England Biolabs, USA] for DNA input >5ng.

In this master mix preparation, the pre-diluted adapter sequences were not added to avoid dimer formation within as that would lead to downstream anomalies. To initiate adapter ligation to each fragment, a master mix was prepared with 30 µL of NEBNext Ultra II Ligation Master mix and 1 µL of NEBNext Ligation Enhancer. Through homogenization, 9 µL of the master mix was aliquoted to 35 µL of the DNA Fragments for each sample of the previous reaction Upon

aliquoting, 2.5 µL of the 1:100 diluted NEDNext Illumina Adapter [Cat: E6612A] was added separately to each sample, making a total reaction volume of 68.5 µL*. The reaction mixture was then incubated at 20°C for 15 minutes in a thermocycler with its **LID OFF**.

The Master mix table:

Component	Volume (µL)	Remarks
NEBNext Ultra II Ligation Master mix (Red)	30	Master mix preparation
NEBNext Ligation Enhancer (Red)	1	
DNA Fragments	26	Previous reaction
NEDNext Illumina Adapter (1:100)	2.5	*Added Separately
Total	**68.5**	

Step	Time	Temperature
Fragmentation	15 minutes	20°C
Lid OFF		

Clean-up of Adapter-ligated DNA and Size Selection (0.9x wash)

For the libraries to be prepared for sequencing, 300 bp of library products are targeted for loading to avoid sequence data gaps during analysis. For this, unbounded DNA and adapters, and fragments less than 200 bp are eliminated through a series of wash. This process unfolds in three distinct stages, commencing with a 0.9X wash, where unbound DNA fragments are eliminated.

Preceding the commencement of the 0.9X wash, the SPRI Beckman Coulter AMPure XP beads [A63881, Beckman Coulter, USA] are brought to room temperature. To achieve a homogeneous mixture, vigorous vortexing was done. Additionally, a solution of 80% Ethanol is prepared.

To initiate the 0.9X wash on the adapter-ligated DNA, 61.65 µL of homogenized AMPure XP beads are added to the total volume of adapter-ligated products. The resulting mixture was incubated at room temperature for 5 minutes. Subsequently, the mixture in the tube was placed on a magnetic rack for 5 minutes, allowing for the separation of magnetic beads from the supernatant. The supernatant is removed, ensuring the beads remain untouched to prevent loss of adapter-

ligated DNA. Maintaining the tubes on the magnetic rack, 200 µL of 80% ethanol is introduced, followed by its removal after 30 seconds of incubation at room temperature. This ethanol washing step is repeated, and any residual ethanol is cautiously eliminated using a pipette tip.

Following the ethanol removal steps, the beads were left to air-dry for 5-10 minutes until they assumed a dark brown coloration, indicating the complete removal of residual ethanol. The sample tube was dislodged from the magnetic rack, followed by the addition of 17 µL of nuclease-free water. After careful pipetting, the tube was incubated for 2 minutes and then returned to the magnetic rack for an additional 2 minutes of incubation, during which the supernatant became clear. After this clarification, 15 µL of the supernatant was transferred to a fresh microcentrifuge tube. It is noteworthy that the eluted DNA is devoid of any unbound DNA fragments. After this step, the eluted DNA is suitable for storage at -20°C until subsequent procedures are undertaken.

USER Digestion and Barcoding

This step involves the USER enzyme to cleave the adapter at the Uracil region of the hairpin-loop of the adapters, creating a gap for IDT i5/i7 index primers (barcodes) to bind. During PCR enrichment each sample is tagged with a unique barcode that binds with 3`-dA tailing end along with the index primers for sample identification following amplification. PCR enrichment allows the barcode along with the adapters to increase the copy number of every fragment. and further helps in bioinformatical analysis while creating full contigs sequence.

To initiate USER End digestion, a master mix was prepared by aliquoting 3 µL of USER Enzyme [Cat: M5505L, New England Biolabs, USA] and 25 µL of NEBNext Ultra II Q5 master max [Cat: M0544L, New England Biolabs, USA] in a tube. Through homogenization, 28 uL of the master mix was added directly to the adapter-ligated washed tube. To this, 10 µL of pre-mixed forward and reverse Index primers (IDT i5/i7) were added separately**. A total of 53 µL of reaction mixture was vortexed briefly.

The Master mix table and PCR profile are given below:

Component	Volume (µL)	Remarks
USER Enzyme (Cat no. M5505L, 250 µL)	3	Master mix preparation
NEBNext Ultra II Q5 master mix	25	
Adapter-ligated DNA	15	Previous cleanup step

IDT i5/i7 pre-mixed primers	10	**Added Separately
Total	**53**	

Thermocycler steps	Cycle
37°C for 15 minutes	1
98°C for 30 seconds	1
98°C for 10 seconds	14
65°C for 75 seconds	
65°C for 5 minutes	1
Hold at 4°C	∞
Lid ON: 105°C	

Clean-up of Barcoded DNA and Size Selection (0.8x wash)

This 0.8x SPRI bead cleanup process washes the barcodes and eliminates products that are less than 100 bp.

Preceding the commencement of the 0.8X wash, the SPRI Beckman Coulter AMPure XP beads are brought to room temperature. To achieve a homogeneous mixture, vigorous vortexing is imperative. A solution of 80% Ethanol is pre-prepared was used.

To initiate the 0.8X wash for the barcoded products, 42.2 µL of homogenized AMPure XP beads are added to the total volume of barcoded products. The resulting mixture was incubated at room temperature for 5 minutes. Subsequently, the mixture in the tube was placed on a magnetic rack for 5 minutes, allowing for the separation of magnetic beads from the supernatant. The supernatant was removed, ensuring the beads remained untouched to prevent loss of adapter-ligated DNA. Maintaining the tubes on the magnetic rack, 200 µL of 80% ethanol was added, followed by its removal after 30 seconds of incubation at room temperature. This ethanol washing step was repeated, and any residual ethanol is cautiously eliminated using a pipette tip.

Following the ethanol removal steps, the beads were left to air-dry for 5-10 minutes until they assume a dark brown coloration, indicating the complete removal of residual ethanol. The sample

tube was dislodged from the magnetic rack, followed by the addition of 44 µL of nuclease-free water. After careful pipetting, the tube was incubated for 2 minutes and then returned to the magnetic rack for an additional 2 minutes of incubation, during which the supernatant becomes clear. After this clarification, 40 µL of the supernatant was transferred to a fresh microcentrifuge tube. It is noteworthy that the eluted DNA does not have unbounded DNA fragments and adapters or any DNA less than 100 bp. This product was further washed before being loaded into the sequencing machine.

Clean-up of Barcoded DNA and Size Selection (0.75x wash)

This 0.75x SPRI bead cleanup process washes the barcodes and eliminates products that are less than 200 bp.

Preceding the commencement of the 0.75X wash, the SPRI Beckman Coulter AMPure XP beads are brought to room temperature. To achieve a homogeneous mixture, vigorous vortexing is imperative. A solution of 80% Ethanol is pre-prepared was used.

To initiate the 0.75X wash for the barcoded products, 30 µL of homogenized AMPure XP beads are added to the total volume of barcoded products. The resulting mixture was incubated at room temperature for a duration of 5 minutes. Subsequently, the mixture in the tube was placed on a magnetic rack for a period of 5 minutes, allowing for the separation of magnetic beads from the supernatant. The supernatant was removed, ensuring the beads remain untouched to prevent loss of adapter-ligated DNA. Maintaining the tubes on the magnetic rack, 200 µL of 80% ethanol was added, followed by its removal after 30 seconds of incubation at room temperature. This ethanol washing step was repeated, and any residual ethanol is cautiously eliminated using a pipette tip.

Following the ethanol removal steps, the beads were left to air-dry for 5-10 minutes until they assume a dark brown coloration, indicating the complete removal of residual ethanol. The sample tube was dislodged from the magnetic rack, followed by the addition of 34 µL of nuclease-free water. After careful pipetting, the tube was incubated for 2 minutes and then returned to the magnetic rack for an additional 2 minutes of incubation, during which the supernatant becomes clear. After this clarification, 30 µL of the supernatant was transferred to a fresh microcentrifuge tube. It is noteworthy that the eluted DNA does not have unbounded DNA fragments and adapters or any DNA less than 200 bp.

Loading on Illumina iSeq100

Quantification of each library was conducted using the Qubit 3 Fluorometer. Subsequently, the individual libraries were combined into a single microcentrifuge tube and further diluted to achieve a final concentration of 110 pM. Additionally, a 5% PhiX [Illumina. Inc, USA] control at a concentration of 100 pM was prepared and incorporated into the pooled libraries.

Following the prescribed Illumina protocol for cartridge thawing and preparation on the iSeq100 platform, 20 uL of the diluted libraries containing 5% PhiX were loaded into the cartridge's loading well alongside the flow cell. The sequencing run was initiated following the manufacturer's instructions.

Upon completion of the iSeq100 run, the raw sequence data of respective samples were transferred and subjected to computational analysis using bioinformatic tools.

Computational Analysis for Phage gDNA Sequences

The bioinformatic section of this dissertation provides an elaborate methodology about the computational methods and analyses the biological raw data generated through this study. These analyses are instrumental in bringing new insights into Typhi phage genomics and diversity.

The bioinformatic analysis section starts with the preprocessing of the raw sequence data, which undergoes a quality control check of the respective files and, the removal of low-quality reads and adapter sequences through trimming. Cleaned reads are then assembled to form a full contig sequence file for respective samples.

Following the preprocessing steps, the contig sequence files are subjected to protein annotation, and hierarchy taxonomical classification and further studied for phylogenetic relationships among other known Typhi-phages.

Conda is an open-source package management system and environment system that is commonly used in the bioinformatics community. Developed by Continuum Analytics, part of Anaconda, Inc; Conda can simplify installation, runs, share, and update of packages and their dependencies among different computing platforms. Conda allows easy installation and update of packages for bioinformatics tools and further allows to create isolated environments with specific dependencies or packages.

Conda was downloaded from the Anaconda Download page and further installed using bash command that was provided. After installation, Conda was verified by checking its version and further updated to the latest version:

```
conda --version / conda -v
```

The Conda was further updated to the current version:

```
conda update conda -y
```

Managing Environment in Conda

A new environment was created in conda for each package to be installed.

```
conda create -n "env_name" biopython -y
conda activate "env_name"
```

To know the list of all the environments and their respective packages:

```
conda env list / conda list
```

To deactivate the current environment

```
conda deactivate
```

Quality Check

The raw data was evaluated for quality checks, through which the low-quality reads, short and duplicate reads, and sequences with high proportion of ambiguous bases were removed. For this step, FastQC tool (v#) was used to carry out quality control on the raw data.

Using conda repository, the tool was installed after creating an environment:

```
conda create -n fastqc
conda activate fastqc
conda install -c bioconda fastqc
```

The program was executed using the following command on the reads:

```
mkdir -p <QC_output_folder>
fastqc <Phage_##_read-1.fastqc.gz> <Phage_##_read-2.fastqc.gz> -o "QC_output_folder"
conda deactivate fastqc
```

Upon completion, an HTML format file and a zip file will be produced for each raw data. The HTML format file has basic statistical information about total sequence reads, read sequence length, poorly flagged sequences, and the percentage of G-C content.

The source code used for data analysis in this study is available on GitHub repository: https://github.com/s-andrews/FastQC

Removal of adapter sequences

Trimming of adapter sequences is an important preprocessing step during NGS analysis as it interferes with downstream analysis. Trimming is needed at the 3'-end of reads because adapter sequences are not found on the 5' ends. For this step, Trimmomatic tool was used where the fastq files and a universal adaptor sequence file (TruSeq3-PE-2.fa) provided by the user supplier will be used for trimming.

Using conda repository, the tool was installed after creating an environment:

```
conda create –n trimmomatic
conda activate trimmomatic
conda install -c bioconda trimmomatic
```

The program was executed using the following command on the reads:

```
trimmomatic PE -threads 4 <Phage_##_read-1.fastqc.gz>
<Phage_##_read-1.fastqc.gz> Phage_##_R-1_Trim_P Phage_##_R-1_Trim_S
Phage_##_R-2_Trim_P Phage_##_R-2_Trim_S
ILLUMINACLIP:adapter_folder/TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDER:4:20
MINLEN:36 LEADING:20 TRAILING:20
conda deactivate trimmomatic
```

The source code used for data analysis in this study in available on GitHub repository: https://github.com/usadellab/Trimmomatic

Sequence Assembly

After quality check of raw data files and removal of unwanted sequences, the files are finally assembled to produce highly accurate and complete genome assemblies or contig files. For this, Unicycler was used which is a bioinformatics tool that combines both *de novo assembly* and read mapping approaches (Wick et al., 2017).

Using conda repository, the tool was installed after creating an environment:

```
conda create –n unicycler
```

```
conda activate unicycler
conda install -c bioconda unicycler
```

The program was executed using the following command on the reads where the files selected are the trimmed files of the paired-reads:

```
mkdir -p <Unicycler_output_folder>
Unicycler -1 <Phage_##_R-1_Trim_P.fastq> -2 <Phage_##_R-2_Trim_P> -t 4 -o <Unicycler_output_folder>
conda deactivate unicycler
```

The source code used for data analysis in this study in available on GitHub repository: https://github.com/rrwick/Unicycler

Annotations of Phage Proteins

Understanding the function of proteins is important when it comes to learning about molecular pathways and cellular processes. For annotation of phage proteins, which has been an increasing interest over the years, Pharokka was used which is a rapid and consistent genomic annotation tool designed for bacteriophages (Bouras et al., 2022).

Using conda repository, the Pharokka tool and database was installed after creating an environment:

```
conda install mamba
conda create –n pharokka
conda activate pharokka
mamba install -c bioconda pharokka
```

The database was downloaded from GitHub repository which was installed following this command:

```
install_databases.py -d
```

To Update pharokka tools:

```
conda update -c bioconda pharokka
```

The program was executed using the following command:

```
mkdir -p <Annotated_output_folder>
pharokka.py -i <Phage_##.fasta> --prefix Phage_## -o <Annotated_output_folder> -d <installed_path/to/database_dir> -t 4
```

For gene mapping of each sample, `pharokka.gbk` file used by pharokka plotter using the following command:

```
pharokka_multiplotter.py -g Phage_##_pharokka.gbk -t Phage_## -o Phage_##_map
conda deactivate pharokka
```

The source code used for data analysis in this study in available on GitHub repository: https://github.com/gbouras13/pharokka

Hierarchical Analysis

Taxonomical analysis is employed for biological classification and ecological studies that provides a framework for loving organisms based on their evolutionary relationships. Taxonomical analysis aids researchers to classify the organisms of interest and further study their relationship with other closely related organisms. For phages, Kraken 2 was used for fast classification of taxonomical order (Wood et al., 2019).

Using conda repository, the Kraken 2 tool and database was installed after creating an environment:

```
conda install mamba
conda create -n kraken2
conda activate kraken2
mamba install kraken2
```

Kraken 2 update command:

```
conda update kraken2
```

The database was downloaded using the following command and will contain three files used for classification: `hash.k2d`, `opts.k2d`, `taxo.k2d`.

```
kraken2-build --standard --db <database_foldername>
```

The program was executed using the following command:

```
mkdir -p <kraken2_output_folder>
krakan2 -db <database_foldername> --threads 4 -output Phage_##.txt --report report_Phage_## Phage_##.fasta -o <kraken2_output_folder>
conda deactivate kraken2
```

Open the `report_Phage_##.txt` file for the classification.

The source code used for data analysis in this study in available on GitHub repository:
https://github.com/DerrickWood/kraken2
https://github.com/DerrickWood/kraken2/wiki/Manual

Genome Coverage

Genome coverage in sequencing data analysis directly related to the accuracy and reliability of downstream studies. Genome coverage provides sufficient reads for each nucleotide across the genome, that helps in detecting mutations or variants. The higher the coverage reads, the more accurate the sequencing depth while low coverage reads can lead to false variant data. The Quast tool was used here for calculating the genome coverage (Gurevich et al., 2013).

Using conda repository, the Quast tool and database was installed after creating an environment:

```
conda create –n quast
conda activate quast
conda install bioconda::quast
```

The program was executed using the following command:

```
mkdir -p <quast_output_folder>
quast <Phage_##.fasta> -g <Phage_##_pharokka.gff> -1 <Phage_##_read-1.fastqc.gz> -2 <Phage_##_read-2.fastqc.gz> -o <quast_output_folder>
conda deactivate quast
```

The source code used for data analysis in this study in available on GitHub repository:
https://anaconda.org/bioconda/quast

Phylogenetic tree

Studying the phylogenetic relationship between the phages is crucial as it talks about the evolutionary relationship with other phages. This helps in understanding the evolution in phage characterization that has happened over the years.

For studying the phylogenetic relationships of Typhi phages, previously studied phages from Pickard et al., 2008, 2010 and Shrestha et al., 2024 were collected from NCBI database. Using the

`pharokka.fnn` file generated through pharokka tools for protein annotation, the tail-fiber nucleotide sequence was extracted into separate fasta files that was followed by merging all the tail fiber fasta file into a single fasta file. Mafft and FigTree were used for phylogenetic tree file generation.

```
cat *.fasta > merged_tailfiber.fasta
conda create -n phylo
conda activate phylo
```

Using conda repository, the Mafft and Figtree tools were installed after creating an environment:

```
conda install -c conda-forge mafft iqtree
mafft merged_tailfiber.fasta > aligned_merged_tailfiber.fasta
Iqtree –s aligned_merged_tailfiber.fasta
conda deactivate phylo
```

The `aligned_merged_tailfiber.treefile` generated file was uploaded to iTOL website for visualization. The tree was re-rooted at midpoint and was further annotated using Colored Strips instructions dataset through iTol. For annotations and legends, Genus of respective phages, Year of Isolation and Location was included, and the color code was added from ColorBrewer: Code Advice for Maps.

Conclusion

To summarize, the methodology section above has provides a systematic outline methods and protocols used for the propagation and purification of phages from a single phage stock, genomic DNA extraction and library preparation for whole genome sequencing and further computational bioinformatical analysis. This chosen approach was followed to answer the objectives of this study. Key considerations, such as sample selection criteria, were managed to avoid sample duplications of the corresponding research project at the Child Health Research Foundation. The methodologies will provide a clear concept of the outlines to gain valuable insights from this study.

Data Availability

Genome sequencing data of the 14 Typhi phages from Bangladesh and 26 Typhi phages from Nepal is available in NCBI databases with the BioProject **PRJNA1081195** and BioProject **PRJNA933946** respectively.

The phages that have been studied by Pickard et al., 2008, 2010., are available online: ftp://ftp.sanger.ac.uk/pub/pathogens/Phage/.

RESULTS

Overview

This chapter of the dissertation section provides outcomes of the experimental investigations that have been conducted to answer the objectives of this study. These outcomes and findings provide information on the Typhi-phage population circulating in Dhaka city. Fifteen phage samples were selected based on their infectivity spectrum against circulating *S.* Typhi genotypes. The findings in this section are based on the computational bioinformatic experiments while the data have been generated through a series of laboratory experiments; namely phage purification from an isolated plaque that was propagated; and phage genomic DNA extraction followed by sequencing library preparation for Whole Genome Sequencing. The raw data generated was checked for any anomalies and was trimmed and assembled to make a full contig fasta format sequence file. Each of the assembled fasta files was utilized for locating the Open Reading Frames (ORFs) and annotation of phage proteins and analyzed for taxonomical classification. Subsequently, genome plots and phylogenetic tree were generated using the annotated protein files.

Genomic Characterization and Speciation of Typhi-phages

Between March 2023 and May 2023, 15 phages were selected from 61 isolated phages. Samples ID beginning with 7501 indicates phages collected in March 2023, 7502 indicates phages collected in April 2023 and 7503 indicates phages collected in May 2023. Some of the sewage samples have more than one phage plaque morphology, and these phages after isolation have been indicated with .1 or .2 at the end of the label.

As summarized in Table 2 and Table 3, three phage genera are circulating in Dhaka city, with 12 of them belonging to the *Kayfunavirus* genus with 57 – 62 ORFs, 2 were categorized under the *Teseptimavirus* genus with 53 ORFs and the remaining one is under the *Macdonaldcampvirus* genus containing 88 ORFs. On average, 40 – 50% of the ORFs are not annotated for protein functions and are denoted as "Hypothetical proteins". This section of the results was derived from the Pharokka and Kraken 2 tools. Using the Unicycler tool for genome assembly, both *Kayfunavirus* and *Teseptimavirus* have an average genome size of 38 000 base pairs while *Macdonaldcampvirus* has a genome size of over 46 000 base pairs. The Quast tool calculated the sequencing read coverage for each sample which ranged between 1000x to 1600x.

Table 2: Speciation of the 15 Typhi-phages sequenced in this study and their corresponding infectivity spectrum. Green indicates that phages infected the S. Typhi genotype of interest.

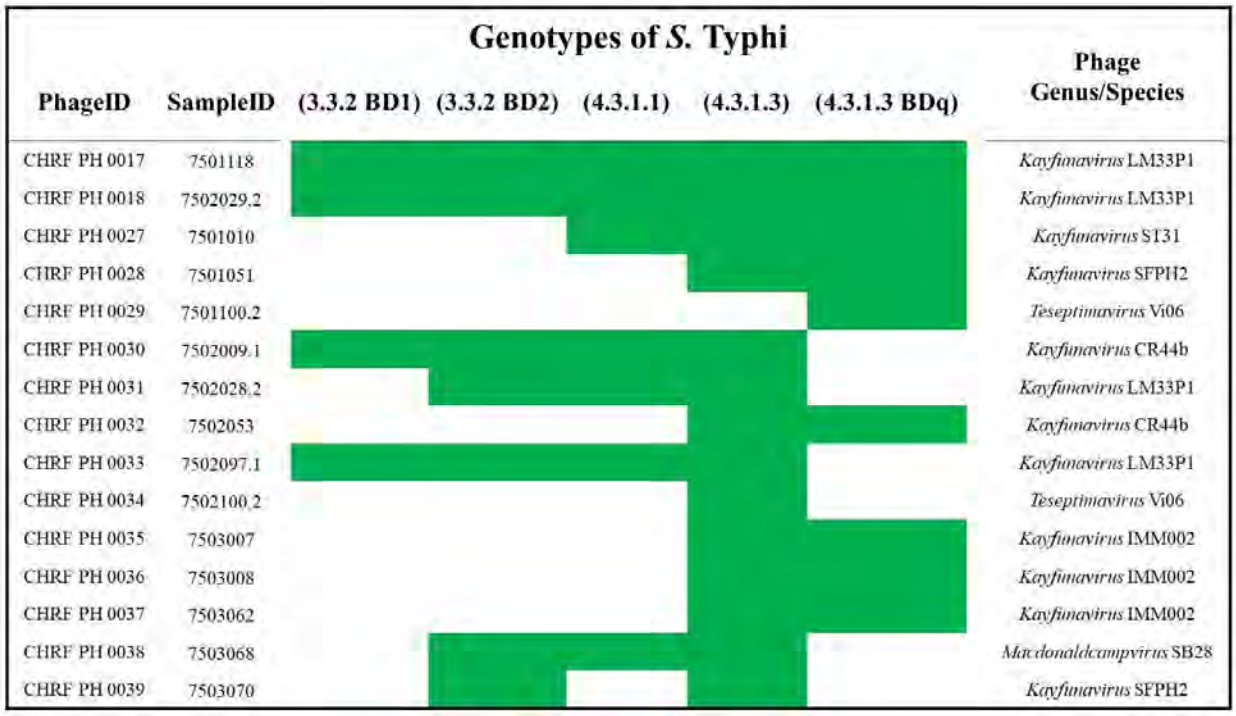


Table 3: Genomic Characterization and Speciation of the 15 Typhi-phages sequenced in this study.
The table contains information about the genome size, percentage of G-C content and the number of unknown proteins of respective phages. Additional information provides the taxonomical classification of these sequenced phages..

		Protein Annotation				Taxonomical Classification	
Phage ID	**Genome CoV**	**Genome Size**	**No. of ORFs**	**G+C %**	**Hypothetical protein**	**Family**	**Genus/ Species**
CHRF PH 0017	1107	38 244	58	50	28	*Autographiviridae*	*Kayfunavirus* LM33P1
CHRF PH 0018	1560	37 816	57	50	27	*Autographiviridae*	*Kayfunavirus* LM33P1
CHRF PH 0027	1329	37 201	57	50	27	*Autographiviridae*	*Kayfunavirus* ST31
CHRF PH 0028	1429	39 466	61	51	31	*Autographiviridae*	*Kayfunavirus* SFPH2
CHRF PH 0029	1175	38 161	52	49	19	*Autographiviridae*	*Teseptimavirus* Vi06
CHRF PH 0030	1210	38 698	59	51	26	*Autographiviridae*	*Kayfunavirus* CR44b
CHRF PH 0031	1073	36 169	58	50	26	*Autographiviridae*	*Kayfunavirus* LM33P1
CHRF PH 0032	1277	39 912	62	51	30	*Autographiviridae*	*Kayfunavirus* CR44b
CHRF PH 0033	1354	38 172	58	50	27	*Autographiviridae*	*Kayfunavirus* LM33P1
CHRF PH 0034	1345	38 162	53	49	19	*Autographiviridae*	*Teseptimavirus* Vi06
CHRF PH 0035	1200	39 568	63	51	30	*Autographiviridae*	*Kayfunavirus* IMM002
CHRF PH 0036	1455	39 568	61	51	29	*Autographiviridae*	*Kayfunavirus* IMM002
CHRF PH 0037	1286	39 568	62	51	30	*Autographiviridae*	*Kayfunavirus* IMM002
CHRF PH 0038	1245	46 277	88	46	54	Unclassified *Caudoviricete*	*Macdonaldcampvirus* SB28
CHRF PH 0039	1384	39 255	60	51	30	*Autographiviridae*	*Kayfunavirus* SFPH2

Genome Map and Plotting

Utilizing the outputs generated by the Pharokka and Multi Pharokka tools, a manual construction of linear phage protein sequence structures was conducted to visually depict the genome organization across respective genera {Figure 6}. Notably, within the *Autographiviridae* family, *Kayfunavirus* and *Teseptimavirus*, distinct features were observed. These phages, harboring RNA polymerase, were conspicuously absent from the "Unclassified *Caudoviricete*" family, particularly *Macdonaldcampvirus*.

Within the *Autographiviridae* family, a consistent arrangement of genes and their functionalities was discernible across various genera and species. Notably, genes associated with DNA/RNA and nucleotide metabolism tended to cluster together, indicative of a functional coherence within this family. Additionally, genes responsible for cellular lysis of bacterial cells, morons for prophage DNA, and host takeover were found clustered alongside DNA/RNA and nucleotide metabolism genes. Structural proteins, including connector proteins, head-packaging assembly proteins, and tail proteins, succeeded in this clustered arrangement.

Distinctive annotations were identified within *Macdonaldcampvirus*, presenting unique protein functionalities absent in other phages from the *Autographiviridae* family. Notably, the presence of the Anti-repressor (Transcriptional regulation) protein sets *Macdonaldcampvirus* apart, suggesting potential divergent evolutionary pathways.

Certain genes that have two different functions are observed to be close together in their arrangements, indicating to have close relations in their functions as well.

Overall, the structural proteins within these phages, notably those involved in tail protein formation and head assembly, exhibited a propensity for close genetic proximity, underscoring their functional interrelationships.

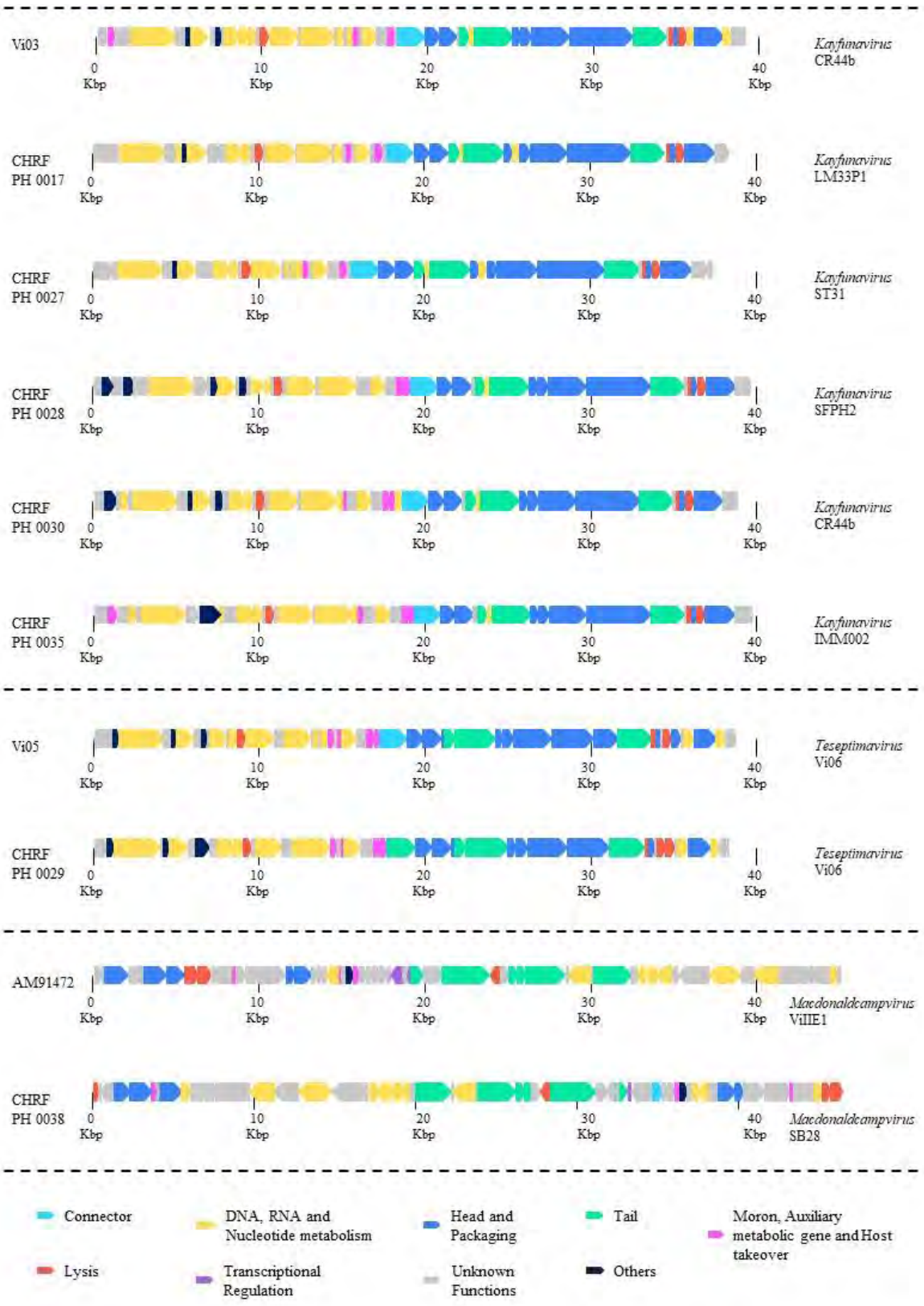


***Figure* 7**: Genome map of Bacteriophage Genus and species-wise.

Diversity of Typhi-phages

Phage diversity and changes over time were studied in this study to investigate the family of Typhi-phages and their evolutionary changes from the 1950s. A phylogenetic tree was generated using the tail fiber sequences of all the phages from which the Pharokka generated file. Along with the 15 phages sequenced in this study, 14 more phage sequences from Bangladesh have been included along with 26 phage sequences from Nepal and 7 phage sequences from the 1950s.

The phylogenetic tree was generated and established using the tail fiber nucleotide sequences, as tail fiber protein sequences are commonly used markers for phage phylogeny {Figure 8}. The analysis from Kraken 2 revealed three distinct genera under two different families: *Autographiviridae* (Order: *Caudoviricete*) and "Unclassified *Caudoviricete"*. The phages under *Autographiviridae* were assigned to *Kayfunavirus* and *Teseptimavirus* while the *Macdonaldcampvirus* was assigned to unclassified *Caudoviricete.* This generated using data from the GenBank tree showed five different genera of phages among which, *Kuttervirus* and *Vectrevirus* from the time 1930 to 1950 cannot be found in recent times while a large cluster is denoted as *Kayfunavirus* genus both in Bangladesh and Nepal followed by *Teseptimavirus* and *Macdonalcampvirus*.

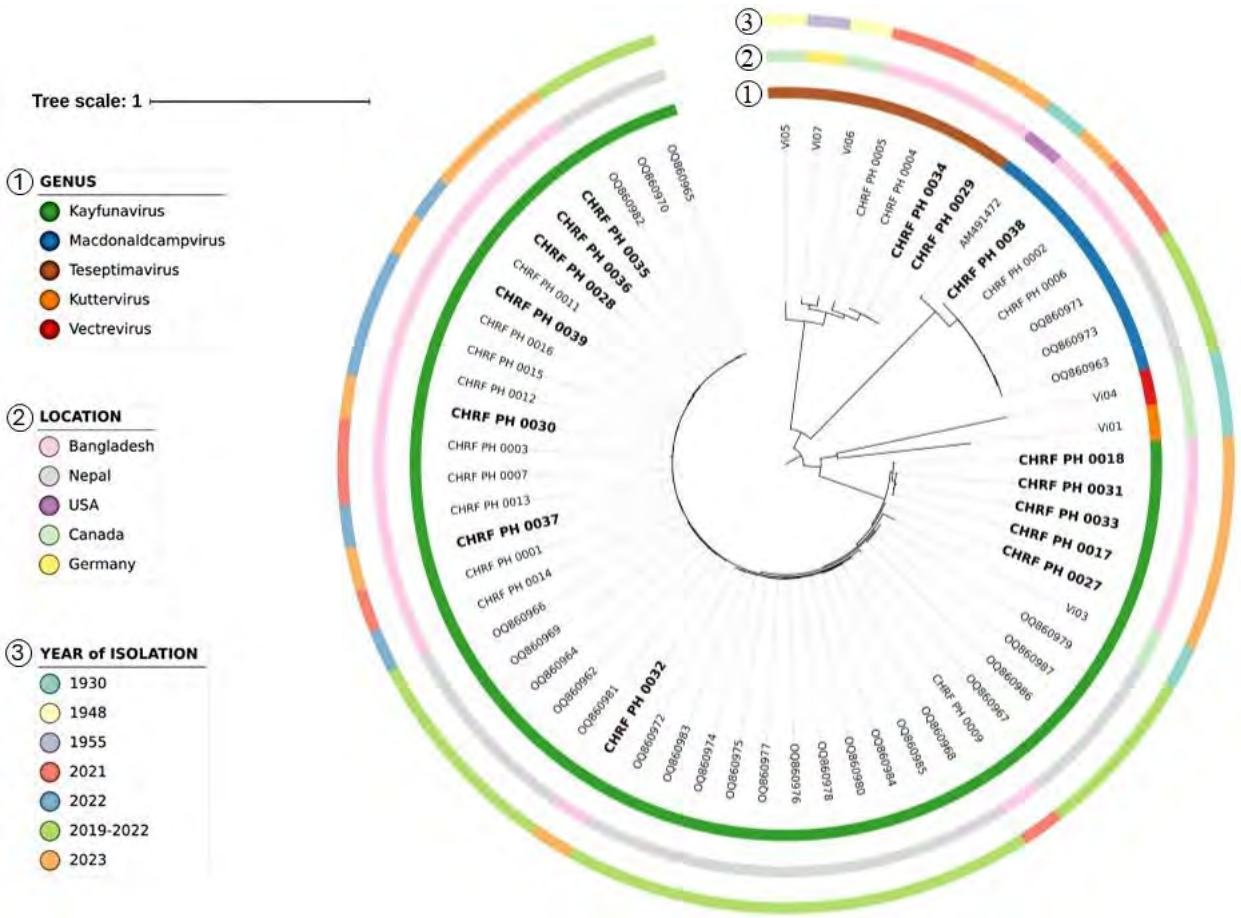


Figure 8: Phylogenetic relationship of Typhi-phages over time and location.
The branches with bold sample ID are the phages in this study. These 15 phages have been conceptualized against 14 phages from Bangladesh, 26 phages from Shrestha et al., 2024 and 7 phages from Pickard et al., 2008, 2010. The genus and location and year of isolation have also been shown.

Conclusion

In summary, the results section presents the genomic diversity of the Typhi-phages circulating in recent times in Bangladesh. Moreover, we observed the evolutionary changes of the phages over time. Through experimental assays, we selected phages with unique characteristics and those that interact with circulating *Salmonella* Typhi genotypes. The phage DNA was extracted, sequenced, and using the computational bioinformatic tools, the generated raw data were assembled and further analyzed for protein annotations and taxonomical classifications at the speciation level.

Our analysis briefly showed that *Kayfunavirus*, *Teseptimavirus,* and *Macdonalcampvirus* genera are circulating in the sewage system of Dhaka, Bangladesh while *Kuttervirus* and *Vectrevirus* cannot be found in recent times. *Kayfunavirus* and *Teseptimavirus* are classified under the same family, *Autographiviridae* while *Macdonaldcampvirus* is under "Unclassified *Caudoviricete"*. Importantly, gene plots were generated shedding light about the gene structure and organization and locating hypothetical proteins, which can suggest potential avenues for future research explorations.

DISCUSSION

This study addresses the significant gap in the current understanding of the genomic characteristics and evolutionary adaptations of Typhi phages. Previous research, such as the studies by Pickard et al., 2008, 2010, primarily examined mid-20th century Typhi phages, leaving modern variants largely unexplored. Given the molecular studies related to *S.* Typhi, phages also undergo continuous evolution. Therefore, it is critical to investigate their present-day counterparts.

Using Whole Genome Sequencing (WGS), this study explores the diverse Typhi-phages circulating in Bangladesh. Our objectives include identifying dominant phage families, assessing potential extinction risks, and examining the speciation and behaviors of these phages through detailed taxonomic classification and genome plotting. A key focus is on tail fiber proteins, which are essential for bacterial host recognition and cell envelope penetration. These proteins, particularly receptor-binding proteins (RBPs), play a crucial role in the infection process and host specificity. This study utilizing the tail fiber proteins provides much-needed genomic data on Typhi phages, enhancing our comprehension of their evolutionary dynamics and structural proteins. This research contributes significantly to molecular biology by improving our understanding of Typhi phage evolution and interactions with *Salmonella* Typhi. The findings presented in this dissertation chapter expand upon the key points and observations outlined in the results section.

This study focused on sequencing Typhi phages isolated from Dhaka city using the Illumina Platform. Samples were selected based on their infectivity spectrum across an assay of 19 *S.* Typhi genotypes, particularly targeting the H58 lineage, which is currently the dominant circulating genotype, as well as specific genotypes resistant to phage infection.

Our analysis revealed the presence of three distinct genera of phages within two families. *Kayfunavirus*, found in this region, and *Teseptimavirus* belong to the *Autographiviridae* family (Order: *Caudoviricetes*). Additionally, *Macdonaldcampvirus*, categorized as "Unclassified *Caudoviricetes*," was identified. This analysis showed some consistency with previous phage sequence data from Nepal, where they also observed the *Kayfunavirus* circulating followed by *Macdonaldcampvirus*, although the presence of *Teseptimavirus* was not observed. These compiled results were then compared with historical Typhi-phages isolated from Canada, the USA, and Germany between 1938 and 1955. Notably, genera such as *Kuttervirus* and *Vectrevirus* from this

period are absent in contemporary samples, due to the decreased prevalence of typhoid fever in Europe and North America, suggesting these phages may have gone extinct.

All sequenced phages in this study are tailed dsDNA (Order: *Caudoviricetes*), which aligns with their known environmental abundance (Ackermann, 2007). The phylogenetic tree analysis, constructed based on the tail-fiber protein, offers significant insights into the evolutionary relationships among these phages. The distinct and independent clusters observed highlight the genus-specific evolutionary pathways. *Macdonaldcampvirus* phages, particularly specimen CHRF_PH_0038, form a cluster under the same node with other contemporary phages within the same taxonomic genus, that involves Bangladesh and Nepal of recent times. However, Phage AM491472, dating back to 1930, stands apart from this node, indicating a potential temporal or evolutionary distinction. This suggests differential genetic diversification or environmental pressures over time, necessitating further *Macdonaldcampvirus* phage sequences to elucidate these distinctions fully. In the case of *Teseptimavirus*, recent Bangladeshi phages grouped under a single node, while historical phages from 1948 and 1955 formed a separate node. This pattern underscores the temporal and geographical influences on the genetic diversification of these phages. These findings also shed light on the variability of phage susceptibility among different S. Typhi genotypes and lineages. This variability may be attributed to natural variations in the Vi capsular polysaccharide and/or phage defense mechanisms, influencing the efficacy of phage infection and cell lysis. Additionally, *S.* Typhi has evolved mechanisms such as expressing phage inhibitory proteins or utilizing the CRISPR-Cas system to evade phage attacks (Shabbir et al., 2016). Future experimental studies could investigate these bacterial defense mechanisms by cloning them into phage-susceptible bacteria, thereby enhancing our understanding of bacterial evolution and lineage extinction dynamics.

Moreover, the close genetic proximity observed between the terminase gene and the HNH endonuclease gene suggests a functional association. The terminase enzyme in phages is crucial for DNA packaging, while the highly conserved HNH endonuclease gene plays a significant role in the homologous recombination of DNA (Zhang et al., 2017). Understanding these genetic relationships could provide deeper insights into the molecular mechanisms underlying phage biology and their evolutionary strategies. The sequential arrangement and clustering of structural proteins, such as Connector proteins (Head-tail adapter), Head and packing proteins, and Tail proteins, appear to play a critical role in determining the shape and structural integrity of the phage.

This organization is essential for the coordinated transcription and translation of phage proteins, which are necessary for the assembly and release of virions from bacterial cells (Osadchy & Kolodny, 2011). The strategic positioning of these structural proteins facilitates efficient assembly processes, ensuring that the necessary components are readily available for the formation of new virions. Additionally, the presence of lysis proteins, including Amidase, Holin, and Rz-like Spanin, is vital for the phage infection cycle. These proteins also accumulate lysozymes on the bacterial cell envelope, creating lesions or holes in the inner membrane, peptidoglycan, and outer membrane, thereby facilitating the release of newly assembled virions (Young, 2014). The spatial proximity of membrane-penetrating lysis proteins is close to tail fiber proteins, which are responsible for recognizing bacterial host cells, suggesting a highly coordinated system that enhances the efficiency of infection and lysis (Nobrega et al., 2018). Analysis of protein annotations and gene map plotting reveals that the genes responsible for bacterial cell wall lysis are located downstream of the tail genes. This downstream positioning indicates a functional interdependence between the translation and activity of tail proteins and lysis proteins. The tail proteins attach to the cytoplasmic membrane, after which the lysis proteins act to breach the bacterial cell envelope, facilitating the release of progeny phages. This alignment of genes underscores a tightly regulated process where structural and lysis functions are closely linked, ensuring the successful propagation of the phage. The findings highlight the sophisticated organization within phages that not only supports their structural integrity but also optimizes their infectious cycle. The interplay between structural and lysis proteins is a crucial aspect of phage biology, emphasizing the evolutionary adaptations that enable efficient infection, replication, and dissemination of phages within bacterial populations.

Throughout the gene map plot, hypothetical proteins comprise 40% - 50% of the entire genome. These proteins are clustered in similar patterns and locations across the genomes of phages within the *Autographiviridae* family. Notably, large sequences of hypothetical proteins are incorporated into the genome structure of the *Macdonaldcampvirus*. These hypothetical proteins have been identified through computational biology; however, their functions remain unannotated due to a lack of empirical evidence. It is hypothesized that these genes may play roles in phage infection processes and in overcoming bacterial phage-defense mechanisms. Recent advancements in bioinformatics tools and databases, such as Glimmer and PHAST, facilitate the prediction of genes in phage genomes. Additionally, databases like UniProt and SWISS-MODEL assist in annotating

these genes and predicting their protein structures. To elucidate the functions of hypothetical proteins, experimental approaches can be employed. For instance, gene knockouts or disruptions using transposons can reveal the impact of these proteins on phage biology and corresponding phenotypes. Another approach involves recombinant expression and purification of these proteins. By cloning the genes into bacterial expression systems, the resulting proteins can be studied in host cells or phage infection models. This allows for a deeper understanding of their roles and mechanisms of action during the phage infection cycle. These combined computational and experimental methods are crucial for uncovering the functions of hypothetical proteins, which could provide significant insights into phage biology and their interactions with bacterial hosts.

A study reveals the role of phages-plasmids (P-Ps) in gene exchange between plasmids and phages that drives evolution and emergences of certain bacterial population or genotypes within a species. Analyzing the sequence similarity network among mobile genetic elements, such as phages and plasmids and P-Ps, can provide insights into genetic recombination. critical for rapid evolution process. Studying the mobile genetic elements can highlight the proportion of recombing genes compared to phages and plasmids that will indicate the role if horizontal gene transfer and genetic diversity as well (Pfeifer & Rocha, 2024). A different study published in *The EMBO Journal* described a novel phage resistance mechanism, the Bacteriophages Exclusion system (BREX), that is found in many bacterial genomes The BREX system allows the phages to bind to the bacteria but blocks the phage DNA replication system via methylation-dependent process to stop the production of progeny phages. The BREX is characterized by six core genes that help bacteria to recognize themselves from the foreign particles, highlighting the complexity and diversity of bacterial defense strategies changes over time against phage infections. (Goldfarb et al., 2015). The BREX anti-phage defense system was studied in a plasmid in *Escherichia fergusonnii* (Picton et al., 2021).

Among the dominant *Salmonella* Typhi H58 linages, 4.3.1.3 BDq is a drug-resistant strain and has a plasmid of over 91 kbp possess the Bacteriophage Exclusion (BREX) phage defense system (Lima et al., 2019). Looking into the infectivity spectrum {Figure 6}, this 4.3.1.3 BDq genotype is resistant to many of the phages that have been isolated compared to another closely related genotype 4.3.1.3 which is susceptible to phages. This shows that the BREX system harboring in the pk91 plasmid has a role in defending from phage attacks. On the other hand, some phages are also overcoming the BREX phage defense system. Studying and characterizing the phages and the

plasmids separately, can unveil the interaction phages with bacteria and further inform the scientific community about the other potential novel phage defense system as well as other anti-phage defense systems.

In conclusion, this study not only enhances our understanding of the current Typhi phage diversity in Dhaka but also provides a historical perspective on phage evolution, revealing significant insights into their genome structure and the interactions between phages and *S*. Typhi genotypes. This knowledge could inform future strategies for utilizing phages in therapeutic applications and for understanding bacterial resistance mechanisms.

LIMITATIONS

This study provides insights into the genomic characterization of Typhi-phages in Dhaka's water sources across seven distinct thanas. However, there are certain limitations that impact the interpretation of the findings.

Firstly, the sample size and area of collection were limited, as water samples were collected from only certain parts of Dhaka over a three-month period. This does not fully represent the Typhi-phage population across the entire city. To study the diversity of phages, their population structure and family of these phages, understanding how phage populations change over time and how phages can be different over long distances and in places where typhoid case is low, continuous surveillance is necessary to observe behavioral changes against the genotypes of *S.* Typhi. Moreover, the gap of phage data between mid-20th century and now further led to difficulties in interpreting and analyzing how extinction of certain phages took place. The results show that certain genera of phages are not circulating in recent years but have limited data about how and why this extinction took place. As a result, more sequencing of the phages over long time can help in investigating the diversity over a longer period and can also reveal if there are seasonal variations in phage distribution alongside typhoid cases in different time of the year, which was not possible in this 3-months.

Additionally, there are limitations in analyzing phage sequencing data. Although tools like Pharokka and Kraken2, are constantly get updated but can have certain biaseness in genome annotations and taxonomical classifications, which might indirectly cause effect in our understanding of the function and origin of hypothetical proteins. With information gap of Typhi-phages sequences worldwide, these tools often depend on limited resources.

Despite these limitations, the study provides valuable insights into the genomic characteristics of *S.* Typhi phages. Future research should aim to include a larger, more diverse sample, employ multiple sequencing techniques, and consider longitudinal studies to better understand the dynamics of phage populations over time.

REFERENCE

- Ackermann, H.-W. (2007). 5500 Phages examined in the electron microscope. *Archives of Virology*, *152*(2), 227–243. https://doi.org/10.1007/s00705-006-0849-1

- Barylski, J., Enault, F., Dutilh, B. E., Schuller, M. B., Edwards, R. A., Gillis, A., Klumpp, J., Knezevic, P., Krupovic, M., Kuhn, J. H., Lavigne, R., Oksanen, H. M., Sullivan, M. B., Jang, H. B., Simmonds, P., Aiewsakun, P., Wittmann, J., Tolstoy, I., Brister, J. R., … Adriaenssens, E. M. (2020). Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages. *Systematic Biology*, *69*(1), 110–123. https://doi.org/10.1093/sysbio/syz036

- Bouras, G., Nepal, R., Houtak, G., Psaltis, A. J., Wormald, P.-J., & Vreugde, S. (2022). Pharokka: A fast scalable bacteriophage annotation tool. *Bioinformatics*, *39*(1), btac776. https://doi.org/10.1093/bioinformatics/btac776

- Brewer, S. M., Twittenhoff, C., Kortmann, J., Brubaker, S. W., Honeycutt, J., Massis, L. M., Pham, T. H. M., Narberhaus, F., & Monack, D. M. (2021). A *Salmonella* Typhi RNA thermosensor regulates virulence factors and innate immune evasion in response to host temperature. *PLOS Pathogens*, *17*(3), e1009345. https://doi.org/10.1371/journal.ppat.1009345

- Carey, M. E., Dyson, Z. A., Ingle, D. J., Amir, A., Aworh, M. K., Chattaway, M. A., Chew, K. L., Crump, J. A., Feasey, N. A., Howden, B. P., Keddy, K. H., Maes, M., Parry, C. M., Van Puyvelde, S., Webb, H. E., Afolayan, A. O., Alexander, A. P., Anandan, S., Andrews, J. R., … Global Typhoid Genomics Consortium Group Authorship. (2023). Global diversity and antimicrobial resistance of typhoid fever pathogens: Insights from a meta-analysis of 13,000 *Salmonella* Typhi genomes. *eLife*, *12*, e85867. https://doi.org/10.7554/eLife.85867

- Chevallereau, A., Pons, B. J., van Houte, S., & Westra, E. R. (2022). Interactions between bacterial and phage communities in natural environments. *Nature Reviews Microbiology*, *20*(1), Article 1. https://doi.org/10.1038/s41579-021-00602-y

- Cho, J.-C., & Kim, S.-J. (1999). Viable, but non-culturable, state of a green fluorescence protein-tagged environmental isolate of *Salmonella* typhi in groundwater and pond water. *FEMS Microbiology Letters*, *170*(1), 257–264. https://doi.org/10.1111/j.1574-6968.1999.tb13382.x

- Díaz-Muñoz, S. L., & Koskella, B. (2014). Chapter Four—Bacteria–Phage Interactions in Natural Environments. In S. Sariaslani & G. M. Gadd (Eds.), *Advances in Applied Microbiology* (Vol. 89, pp. 135–183). Academic Press. https://doi.org/10.1016/B978-0-12-800259-9.00004-4

- Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology*, *18*(3), Article 3. https://doi.org/10.1038/s41579-019-0311-5

- Dyson, Z. A., & Holt, K. E. (2021). Five Years of GenoTyphi: Updates to the Global *Salmonella* Typhi Genotyping Framework. *The Journal of Infectious Diseases*, *224*(Supplement_7), S775–S780. https://doi.org/10.1093/infdis/jiab414

- Farmer, J. J., Hickman, F. W., & Sikes, J. V. (1975). Automation of salmonella typhi phage typing. *The Lancet*, *306*(7939), 787–790. https://doi.org/10.1016/S0140-6736(75)80004-3

- Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., & Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO Journal*, *34*(2), 169–183. https://doi.org/10.15252/embj.201489455

- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

- Hatfull, G. F., & Hendrix, R. W. (2011). Bacteriophages and their genomes. *Current Opinion in Virology*, *1*(4), 298–303. https://doi.org/10.1016/j.coviro.2011.06.009

- Hooda, Y., Islam, S., Kabiraj, R., Rahman, H., Sarkar, H., Silva, K. E. da, Raju, R. S., Luby, S. P., Andrews, J. R., Saha, S. K., & Saha, S. (2024). Old tools, new applications: Use of environmental bacteriophages for typhoid surveillance and evaluating vaccine impact. *PLOS Neglected Tropical Diseases*, *18*(2), e0011822. https://doi.org/10.1371/journal.pntd.0011822

- Hooda, Y., Islam, S., Kabiraj, R., Rahman, H., Silva, K. E. da, Raju, R. S., Luby, S. P., Andrews, J. R., Saha, S. K., & Saha, S. (2023). *Old tools, new applications: Use of environmental bacteriophages for typhoid surveillance and evaluating vaccine impact* (p. 2023.02.14.23285884). medRxiv. https://doi.org/10.1101/2023.02.14.23285884

- Islam, S., Kabiraj, R., Amin, A., Pranto, S. H., Dey, S., Malaker, A. R., Keya, D. P., Tanmoy, A. M., Saha, S. K., Hooda, Y., & Saha, S. (2024). *Protocol for Genomic DNA Extraction and Sequencing Library Preparation from Phage Stock*. https://www.protocols.io/view/protocol-for-genomic-dna-extraction-and-sequencing-dbkk2kuw

- Jansson, J. K., & Wu, R. (2023). Soil viral diversity, ecology and climate change. *Nature Reviews Microbiology*, *21*(5), Article 5. https://doi.org/10.1038/s41579-022-00811-z

- Kim, M.-S., & Bae, J.-W. (2018). Lysogeny is prevalent and widely distributed in the murine gut microbiota. *The ISME Journal*, *12*(4), Article 4. https://doi.org/10.1038/s41396-018-0061-9

- Kirchhelle, C., Pollard, A. J., & Vanderslott, S. (2019). Typhoid—From Past to Future. *Clinical Infectious Diseases*, *69*(Supplement_5), S375–S376. https://doi.org/10.1093/cid/ciz551

- Lima, N. C. B., Tanmoy, A. M., Westeel, E., de Almeida, L. G. P., Rajoharison, A., Islam, M., Endtz, H. P., Saha, S. K., de Vasconcelos, A. T. R., & Komurian-Pradel, F. (2019). Analysis of isolates from Bangladesh highlights multiple ways to carry resistance genes in *Salmonella* Typhi. *BMC Genomics*, *20*(1), 530. https://doi.org/10.1186/s12864-019-5916-6

- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., Reysenbach, A.-L., Smith, V. H., & Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, *4*(2), Article 2. https://doi.org/10.1038/nrmicro1341

- Meiring, J. E., Khanam, F., Basnyat, B., Charles, R. C., Crump, J. A., Debellut, F., Holt, K. E., Kariuki, S., Mugisha, E., Neuzil, K. M., Parry, C. M., Pitzer, V. E., Pollard, A. J., Qadri, F., & Gordon, M. A. (2023). Typhoid fever. *Nature Reviews Disease Primers*, *9*(1), 1–18. https://doi.org/10.1038/s41572-023-00480-z

- Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., Dutilh, B. E., & Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology*, *16*(12), 760–773. https://doi.org/10.1038/s41579-018-0070-8

- Osadchy, M., & Kolodny, R. (2011). Maps of protein structure space reveal a fundamental relationship between protein structure and function. *Proceedings of the National Academy of Sciences*, *108*(30), 12301–12306. https://doi.org/10.1073/pnas.1102727108

- Pfeifer, E., & Rocha, E. P. C. (2024). Phage-plasmids promote recombination and emergence of phages and plasmids. *Nature Communications*, *15*(1), 1545. https://doi.org/10.1038/s41467-024-45757-3

- Pickard, D., Thomson, N. R., Baker, S., Wain, J., Pardo, M., Goulding, D., Hamlin, N., Choudhary, J., Threfall, J., & Dougan, G. (2008). Molecular Characterization of the *Salmonella* enterica Serovar Typhi Vi-Typing Bacteriophage E1. *Journal of Bacteriology*, *190*(7), 2580–2587. https://doi.org/10.1128/jb.01654-07

- Pickard, D., Toribio, A. L., Petty, N. K., van Tonder, A., Yu, L., Goulding, D., Barrell, B., Rance, R., Harris, D., Wetter, M., Wain, J., Choudhary, J., Thomson, N., & Dougan, G. (2010). A Conserved Acetyl Esterase Domain Targets Diverse Bacteriophages to the Vi

Capsular Receptor of *Salmonella* enterica Serovar Typhi. *Journal of Bacteriology*, *192*(21), 5746–5754. https://doi.org/10.1128/jb.00659-10

- Picton, D. M., Luyten, Y. A., Morgan, R. D., Nelson, A., Smith, D. L., Dryden, D. T. F., Hinton, J. C. D., & Blower, T. R. (2021). The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses. *Nucleic Acids Research*, *49*(19), 11257–11273. https://doi.org/10.1093/nar/gkab906

- Runa, V., Wenk, J., Bengtsson, S., Jones, B. V., & Lanham, A. B. (2021). Bacteriophages in Biological Wastewater Treatment Systems: Occurrence, Characterization, and Function. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/article/10.3389/fmicb.2021.730071

- Saha, S. K., Tabassum, N., & Saha, S. (2021). Typhoid Conjugate Vaccine: An Urgent Tool to Combat Typhoid and Tackle Antimicrobial Resistance. *The Journal of Infectious Diseases*, *224*(Supplement_7), S788–S791. https://doi.org/10.1093/infdis/jiab443

- Saha, S., Tanmoy, A. M., Andrews, J. R., Sajib, M. S. I., Yu, A. T., Baker, S., Luby, S. P., & Saha, S. K. (2018). Evaluating PCR-Based Detection of *Salmonella* Typhi and Paratyphi A in the Environment as an Enteric Fever Surveillance Tool. *The American Journal of Tropical Medicine and Hygiene*, *100*(1), 43–46. https://doi.org/10.4269/ajtmh.18-0428

- Shabbir, M. A. B., Hao, H., Shabbir, M. Z., Wu, Q., Sattar, A., & Yuan, Z. (2016). Bacteria vs. Bacteriophages: Parallel Evolution of Immune Arsenals. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.01292

- Sharp, R. (2001). Bacteriophages: Biology and history. *Journal of Chemical Technology & Biotechnology*, *76*(7), 667–672. https://doi.org/10.1002/jctb.434

- Shrestha, S., Silva, K. E. D., Shakya, J., Yu, A. T., Katuwal, N., Shrestha, R., Shakya, M., Shahi, S. B., Naga, S. R., LeBoa, C., Aiemjoy, K., Bogoch, I. I., Saha, S., Tamrakar, D., & Andrews, J. R. (2024). Detection of *Salmonella* Typhi bacteriophages in surface waters as

a scalable approach to environmental surveillance. *PLOS Neglected Tropical Diseases*, *18*(2), e0011912. https://doi.org/10.1371/journal.pntd.0011912

- Stanaway, J. D., Reiner, R. C., Blacker, B. F., Goldberg, E. M., Khalil, I. A., Troeger, C. E., Andrews, J. R., Bhutta, Z. A., Crump, J. A., Im, J., Marks, F., Mintz, E., Park, S. E., Zaidi, A. K. M., Abebe, Z., Abejie, A. N., Adedeji, I. A., Ali, B. A., Amare, A. T., … Hay, S. I. (2019). The global burden of typhoid and paratyphoid fevers: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Infectious Diseases*, *19*(4), 369–381. https://doi.org/10.1016/S1473-3099(18)30685-6

- Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), Article 7057. https://doi.org/10.1038/nature04160

- Tokarz-Deptuła, B., Sliwa-Dominiak, J., Adamiak, M., Deptuła, T., Nowak, M., & Deptuła, W. (2011). Application of bacteriophages—Selected data. *Central-European Journal of Immunology*, *36*, 87–91.

- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, *13*(6), e1005595. https://doi.org/10.1371/journal.pcbi.1005595

- Wong, V. K., Baker, S., Connor, T. R., Pickard, D., Page, A. J., Dave, J., Murphy, N., Holliman, R., Sefton, A., Millar, M., Dyson, Z. A., Dougan, G., & Holt, K. E. (2016). An extended genotyping framework for *Salmonella* enterica serovar Typhi, the cause of human typhoid. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms12827

- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

- Yap, M. L., & Rossmann, M. G. (2014). Structure and Function of Bacteriophage T4. *Future Microbiology*. https://doi.org/10.2217/fmb.14.91

- Young, R. (2014). Phage lysis: Three steps, three choices, one outcome. *Journal of Microbiology*, *52*(3), 243–258. https://doi.org/10.1007/s12275-014-4087-z

- Zhang, L., Xu, D., Huang, Y., Zhu, X., Rui, M., Wan, T., Zheng, X., Shen, Y., Chen, X., Ma, K., & Gong, Y. (2017). Structural and functional characterization of deep-sea thermophilic bacteriophage GVE2 HNH endonuclease. *Scientific Reports*, *7*(1), 42542. https://doi.org/10.1038/srep42542

- Zhu, Y., Shang, J., Peng, C., & Sun, Y. (2022). Phage family classification under Caudoviricetes: A review of current tools using the latest ICTV classification framework. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.1032186

- Zinke, M., Schröder, G. F., & Lange, A. (2022). Major tail proteins of bacteriophages of the order *Caudovirales*. *Journal of Biological Chemistry*, *298*(1), 101472. https://doi.org/10.1016/j.jbc.2021.101472
