COMPARATIVE ANALYSIS REVEALS THE CONTRIBUTION OF PHAGE-DERIVED GENOMIC ISLANDS TO THE INCREASING PATHOGENICITY OF FACULTATIVE BACTERIAL PATHOGENS

By

Israt Jahan Khan Poly
16136009

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

Department of Mathematics and Natural Sciences
Brac University
September 2020

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Israt Jahan Khan

Israt Jahan Khan Poly

16136009

Approval

The thesis titled "Comparative analysis reveals the contribution of phage-derived genomic islands to the increasing pathogenicity of facultative bacterial pathogens" submitted by

Israt Jahan Khan Poly (16136009)

of Spring, 2016 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of BSc on September of 2020.

Examining Committee:

Supervisor:
(Member)

_____

Dr. Iftekhar Bin Naser
Assistant Professor, Mathematics and Natural Sciences
BRAC University

Supervisor:
(Member)

_____

Tokee Mohammad Tareq
Lecturer, Mathematics and Natural Sciences
BRAC University

Program Coordinator:
(Member)

_____

Dr. Iftekhar Bin Naser
Assistant Professor, Mathematics and Natural Sciences
BRAC University

Departmental Head:
(Chair)

_____

Prof. A. F. M. Yusuf Haider
Chairperson, Mathematics and Natural Sciences
BRAC University

Ethics Statement

All data collection, storage and analysis were conducted at the computer laboratory of the department of Mathematics and Natural Sciences, located on the 19th floor of building two in Brac University. All of the data were collected from the appropriate databases and analyzed with precision and integrity with the help of suitable tools and programs. The contents of this study would prove to be an essential component in further wet laboratory research regarding bacterial increased pathogenicity.

Abstract

Bacteriophages are the most abundant entity in the biosphere outnumbering bacteria by a factor of at least 10. Other than infecting bacteria, phages also influence the evolution of bacteria in an unexpected way to provide more pathogenicity to the species in the course of time. It has thereby become very significant to establish an equation between bacterial increased pathogenicity and phage derived genome islands.

Computational analysis of vast number of bacterial pathogenic genomes divided by two different time periods (before and after 1990) yielded the virulence regions emerging from prophages. Notable addition of pathogenic phage proteins was observed in bacteria of the latter time periods indicating increase in the pathogenesis of already pathogenic bacteria. As we have few information on phage diversity and their contribution on bacterial evolution, hence this study would create an opportunity to understand their relationship to a great extent.

Keywords: Bacterial pathogenicity; Phage derived genomic islands; Virulence factors

Acknowledgement

First and foremost, I am grateful to my parents for always supporting me in every aspect of my life. They have always encouraged me to pursue what I really loved to do. Without their support my passion to work with microorganisms would never get materialized into this thesis.

I am grateful to the Chairperson of the Department of Mathematics and Natural Sciences, Brac University, **Professor A F M Yusuf Haider** to allow me to work in the computer laboratory of the department and to provide me with immense opportunities.

My profound gratitude goes to my supervisor **Dr. Iftekhar Bin Naser**, Assistant Professor, and co-supervisor **Tokee Mohammad Tareq**, Lecturer, Department of Mathematics and Natural Sciences, BRAC University to accept me as their thesis student and allowing me to finish my degree.

I cannot express enough gratitude to my mentor **Tushar Ahmed Shishir**, Graduate Teaching Assistant, Department of Mathematics and Natural Sciences, BRAC University for providing me insight and expertise into this work. Even when he was under immense pressure handling his masters research, he never stopped reaching out to me.

I am grateful to all the members of computer laboratory, Department of Mathematics and Natural Sciences, Brac University who kindly helped me throughout the entire period and made it feel like home to me.

Lastly, I would like to thank all my friends for their constant support mentally and academically. I am lucky to have them for always reaching out to me whenever I needed a helping hand.

Israt Jahan Khan Poly

Table of Contents

List of Tables

List of Figures

List of Acronyms

HGT	Horizontal Gene Transfer
MIGE	Mobile and Integrative Genetic Element
PAI	Pathogenicity island
pVOG	Prokaryotic Virus Orthologous Groups
NGS	Next Generation Sequencing
EP	Effector proteins
T3SS	Type three secretion system
NCBI	National Center for Biotechnology Information
PATRIC	Pathosystems Resource Integration Center
VFDB	Virulence Factors Database
RAST	Rapid Annotation using Subsystem Technology
PPI	Protein-Protein Interactions
PHIDIAS	Pathogen-Host Interaction Data Integration and Analysis System
BLAST	Basic Local Alignment Search Tool
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool)
PSSM	position-specific scoring matrix

Chapter 1

Introduction

1.1 Introduction to phage provided bacterial virulence

In 2016, Yong stated that “I was struck by how different everything seemed with microbes in mind”. The smaller an organism gets, the complex it becomes. Talking about microbes’ bacteria and bacteriophage take the center stage. Phages are the most abundant biological entity in the environment. They outnumber bacteria in the biosphere on a huge scale where the ratio is 10 phages for 1.1 bacteria, making them more dominant (Brüssow & Hendrix, 2002, p. 14). The mammalian gut also represents similar pictures with 10^{13}-10^{14} bacteria in per gram of fecal matter where the human gut is densely populated by 10^{8} viral particles in per gram of feces (Fortier & Sekulovic, 2013, p. 357). Hence undeniably creating a vast playground for phages in all the ecosystems.

Usually, bacteriophage is considered to infect and replicate within bacteria. Though it would be unjust to simplify the interaction since bacteriophage does not only infect a bacterium but also provides virulence in several pathogen bacteria. Phages even play a role in the emergence and evolution of those pathogens through horizontal gene transfer and acquisition of virulence factors (Boyd & Brüssow, 2002, p. 522). When movement of genetic material occurs between unicellular and (or) multicellular organisms not via reproduction, the transfer is then called horizontal gene transfer (HGT) or lateral gene transfer (LGT) (Keeling & Palmer, 2008, p. 607) and molecules that are produced by unicellular organisms to enhance their effectivity and to achieve immunoevasion, immunosuppression, inhibition of the host's immune response and entry into and exit out of host cells is known as virulence factors (Cross, 2008, p. 197). Virulence factors

for a variety of human pathogens get encoded on mobile and integrative genetic elements (MIGEs) which are associated with different stages of bacterial pathogenesis (The development of a disease and the chain of events leading to that disease) (E. F. Boyd, 2012, p. 98). Their contribution to the pathogenicity towards its host began to uncover in 1927 through an experiment conducted via Frobisher and Brown (Wagner & Waldor, 2002, p. 3988). They discovered that nontoxigenic streptococci produced scarlatinal toxin (The strep bacteria produces a toxin that generates a rash which appears 1 to 2 days after the onset of illness, Scarlet Fever) when exposed to the filtered supernatants of toxigenic streptococci (*Definition of Scarlatina*, 2017). Later, phage encoded scarlatinal toxin was found in the supernatant causing the toxicogenicity (The ability of a pathogenic organism to produce injurious substances that damage the host) (Frobisher & Brown, 1927). From then on, acquired virulence of bacteria or in classical term known as phage conversion (a process whereby phage-encoded virulence factors convert their bacterial host from a nonvirulent strain to a virulent strain) became widely accepted (E. F. Boyd, 2012, p. 98). Bacteria can obtain virulence properties by the transfer of virulent genes from phages known as transduction or via transferring mobile and integrative genetic elements (MIGEs) such as insertion sequences, plasmids, bacteriophages, conjugative transposons, integrative and conjugative elements (ICEs), and pathogenicity islands (PAIs) (A large region of chromosomal DNA (35–200 kb) that encodes several virulence factors and that is present in all pathogenic isolates and usually absent from non-pathogenic isolates of a species) known as conjugation. Consequently, it has also been observed that toxic genes are not the only medium of pathogenicity. Instead it is only a subset of the diverse virulence factors encoded by bacteriophages. In addition to that, phage encoded regulatory factors increase the expression of virulent genes (Spanier & Cleary, 1980c, p. 1394). Whereas, alteration of bacterial components also happens due to phage encoded enzymes (Guan et al., 1999, p. 1271).

Besides, the unique properties of bacteriophages also contribute to bacterial pathogenicity. Therefore, the interaction between a bacteriophage and a bacterium cannot simply be called a parasite-host interaction. It is indeed a captivating case of co-evolution (Desiere et al., 2001, p. 331).

1.2 Phage-bacteria interaction

Generally, there are two main categories of phages based on their lifecycle. Strictly lytic phages (or virulent) and temperate phages. Some phages also embrace a pseudo temperate lifestyle, though they refrain from integrating their chromosome into the host (Guttman, Raya, Kutter, & Sulakvelidze, 2005) (er por Guttman et al., 2005). Virulent phages only replicate via lytic cycle, identified by the lysis of the infected bacterial host observed at the end of the phage replication cycle. Adsorption of the phage to a specific bacterial cell wall receptor, along with the injection of the phage DNA into the bacterial cytoplasm, transcription of phage genes, and replication of the phage genome inside the bacterial genome indicates the lytic cycle of DNA phages (Fig. 01). Afterwards, viral proteins are synthesized and followed by the congregation of it into capsids and tails that gets packaged into the capsids to form virions.

The phage holin (a small protein which forms oligomers) creates pores within the cytoplasmic membrane (Wang et al., 2000, p. 813) that causes hydrolysis of the peptidoglycan layer as it allows the phage-encoded endolysin to gain access, resulting in cell lysis and the release of thousands of infectious particles into the neighboring habitat. These particles are capable enough to re-infect susceptible bacteria and start the cycle over again. Temperate phages are also capable of lysis, similar to virulent phages. On top of that, they can as well opt for a nonlytic lifestyle (Guttman, Raya, Kutter, & Sulakvelidze, 2005) (er por Guttman et al., 2005).

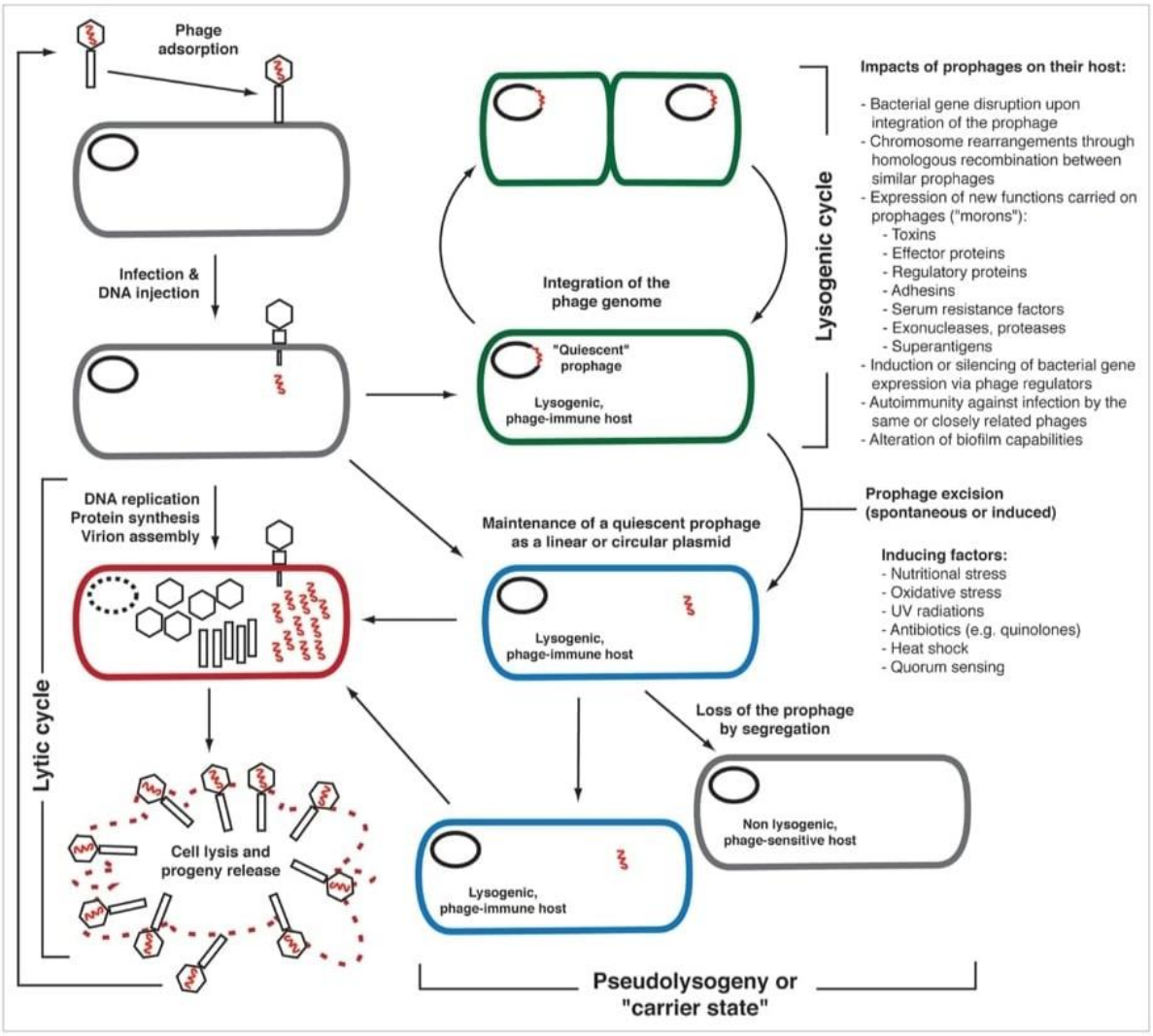


Figure 01: Different lifestyles embraced by phages. Virulent phages follow the lytic cycle for their replication whereas temperate phages may replicate through the lytic cycle like virulent phages, or may opt for the lysogenic cycle (Fortier & Sekulovic, 2013, p. 357).

The phage can decide on whether to initiate a lytic cycle or to inject its DNA into the bacterial chromosome to become a prophage on the way, as soon as it enters the DNA into the bacterial cytoplasm. If the second option is selected, a specific phage repressor (CI in phage λ) comes into play through inhibiting transcription of most of the phage genes and genes required for the lytic

cycle. Alongside the bacterial chromosome during cell division, phage DNA is replicated synchronously which can last for an infinite period of time. A lysogen, a bacterial cell carrying one or more prophages, typically is immune toward lytic or lysogenic infection by other phages (Guttman, Raya, Kutter, & Sulakvelidze, 2005) (er por Guttman et al., 2005). As temperate phages can establish a stable relationship with their host, so it may have profound consequences on the virulence of the lysogens (Brüssow et al., 2004b, p. 566). Besides, through prophage integration, many clusters of phage genes as well as genes encoding virulence factors that are associated with the phage genome gets integrated into the bacterial genome. These phage-associated virulence factors or pVOGs have an immense impact on bacterial pathogenicity. pVOGs (Prokaryotic Virus Orthologous Groups) are a set of phage-associated virulence factors who encode proteins. The differential expression of these proteins also adds to the pathogenicity. The expressed proteins from these clusters has a minimal effect on growth in other environments but mostly associated with affecting pathogenicity of bacteria while most of the time increasing virulence is the case (Busby et al., 2012, p. 311).

The primary sources of genetic diversity and strain variation affiliated with the virulence of many bacterial pathogens including *Pseudomonas*, *Neisseria* and *Klebsiella* are embedded in prophage. Phages linked with virulent strains primarily encode effector proteins, powerful extracellular toxins, enzymes such as phospholipase, DNase, superoxide dismutase, staphylokinase, proteins affecting resistance to serum and altering antigenicity, superantigens, adhesion factors, proteinases, and mitogenic factors which leads towards added pathogenicity (Brüssow et al., 2004b, p. 566). Due to the help of Next generation sequencing (NGS), the number of complete and draft bacterial genome sequences are increasing day-by-day in public databases. The genomic sequencing of pathogenic bacterial strains made it abundantly clear that free-living bacteria can

convert from being benign to pathogenic by the acquisition of prophages, a major evolutionary route (Busby et al., 2012, p. 311). Though the huge collection of genomic data available in public repositories help people access more information, it also imposes one of the greatest challenges in phage biology to link the presence of prophages in bacterial chromosomes to apparent bacterial phenotypes. This task becomes way more complex when prophages becomes critical to be linked with virulence and do not encode known virulence factors on the basis of bioinformatics analyses. It is a matter of relief that there are other ways in phage biology to connect prophage with the bacterial virulence.

1.3 Phage moron: A fitness advantage for bacterial cell

In company with the suppression of phage replication, the expression of most of the phage genes are repressed too in the lysogenic state. Still, a subset of prophage genes is expressed. Those are related to evolutionarily adaptive for the phage and they present themselves like a fitness advantage for the bacterium by simultaneously functioning as both friend and foe to their host (Tsao et al., 2018, p. 12). These extra genes in the context of bacteriophage genetics are called “Phage Moron”. These additional genes signify that these bacteriophage genomes have "more on" them, coining the term moron (Cumby et al., 2012). The GC contents of morons differ from those of the surrounding genes, implying an acquisition by horizontal gene transfer, and consist of their own promoter and terminator sequences to enable expression from the prophage.

Morons may enhance bacterial survival via several mechanisms such as encoding superinfection exclusion proteins which inhibit further phage infection by altering the bacterial cell envelope (Cumby, Edwards, et al., 2012, p. 5017). Similarly, morons also increase the pathogenicity of bacterial strains infecting humans, hence recognizing as virulence factors. By minimizing interference with adjacent prophage genes, morons allow expression of the virulence genes

(Hendrix et al., 2000, p. 507). For example, *Pseudomonas aeruginosa* FIZ15 prophage expresses genes that enhance adherence to epithelial cells and provide resistance to phagocytosis and human serum (Vaca-Pacheco et al., 1999, p. 241). To gain further information by which prophages alter bacterial behavior, (Tsao et al., 2018, p. 12) characterized the effects of expressing 14 different moron genes in two different *P. aeruginosa* clinical strain backgrounds. They found a noticeable alteration in at least one measurable bacterial phenotype with the expression of 85% of these genes. (Tsao et al., 2018, p. 12) They demonstrated that morons could modulate bacterial phenotypes and highlighted the complex impacts that prophages might have exerted on bacterial behavior and human disease.

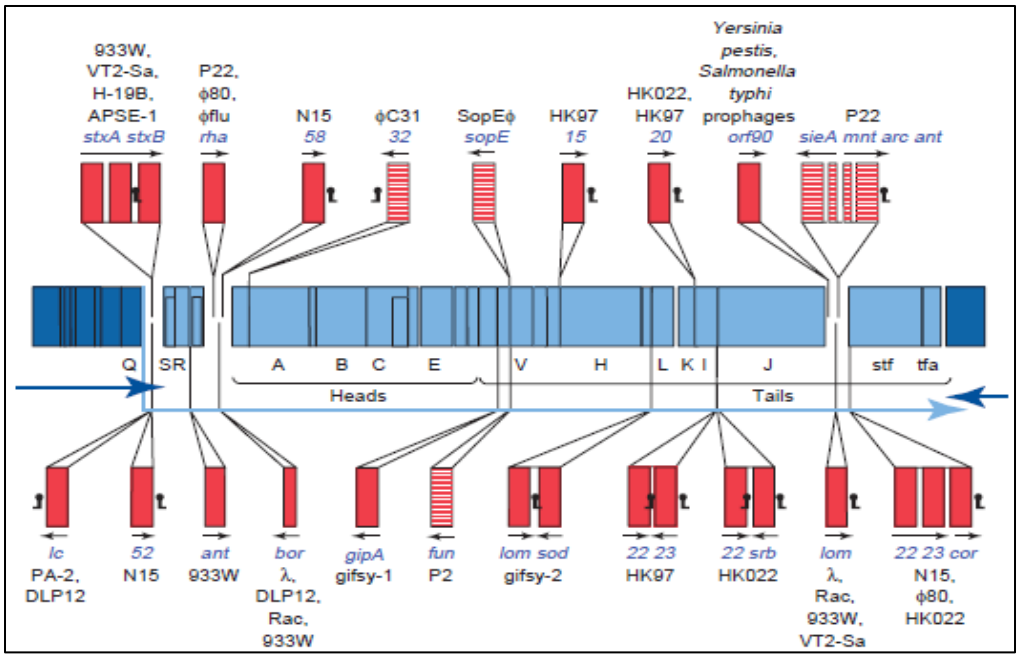


Figure 02: Temperate bacteriophage late-operon morons. Light blue rectangles indicate late-operons' structure and lysis genes. Early genes in other operons are showed as dark blue rectangles. Transcription across these regions during lambda infection is represented by light blue and dark blue arrows, respectively. Each red rectangle indicates the location of a moron, and the striped red rectangles denote morons present in non-homologous known phages and prophages that carry it are given in black (Hendrix et al., 2000, p. 507).

Comparative phage genomics enables the quick identification of morons and the function of the virulence factors. A structural demonstration was done by (Hendrix et al., 2000, p. 507) which located the position of the moron of temperate phage l (Fig. 02) and their respective functions (Table. 01) in that prophage.

Table 01. Functions of temperate phage late-operon morons (Hendrix et al., 2000, p. 507)

Gene	Function
52	DNA adenine methylase
58	DNA cytosine methylase
ant	Anti-repressor
bor	Lysogen more resistant to mammalian serum
cor	Lysogen resistance to phage attack
fun	Lysogen resistance to phage attack
gipA	Lysogen survives better in Peyer's patches
lom	Lysogen has improved binding to mammalian cells
rha	Affects ability of phage to propagate on certain hosts
mnt-arc-ant	Anti-repressor and attendant regulation apparatus
sieA	Lysogen resistance to phage attack
sod	Superoxide dismutase; lysogen more virulent in mice
sopE	Stimulates mammalian cell ruffling and cell invasion
stx	Might suppress late-operon expression in lysogen Shiga-like toxin; kills mammalian cells by damaging rRNA

Hendrix et al. (2000, p. 506) created a compelling case that recent evolution of phages has included the addition to their genomes of transcriptionally autonomous genes, the morons which provide selective benefits to the phage genome.

1.4 Bacterial entry into its host and pathogenicity

Bacteria must find entry points to enter the host. The entrance could occur by ingestion of contaminated food or water in the case of gastrointestinal pathogens, by inhalation of airborne infectious droplets in the case of respiratory tract pathogens, sexual transmission among urinary and reproduction tract pathogens or cuts or abrasions on the natural physical barrier of the skin for blood borne infections (E. F. Boyd, 2012, p. 98).

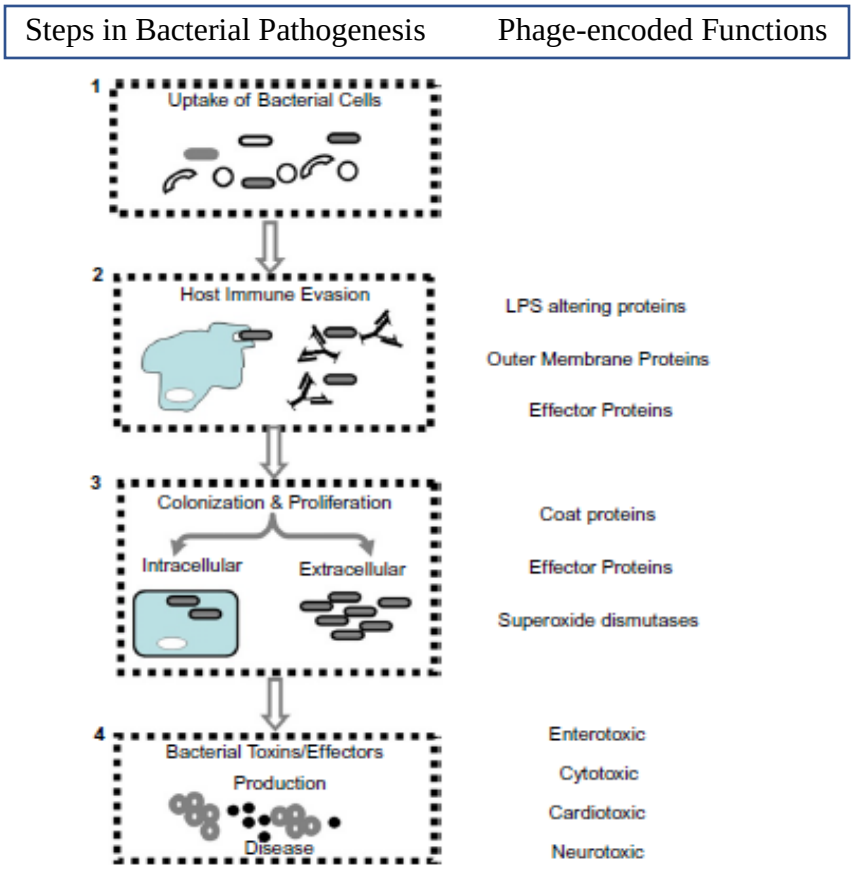


Figure 03: Stages in bacterial pathogenesis and virulence factors (E. F. Boyd, 2012, p. 98)

Upon entry, the pathogen must reach to the site of attachment and proliferation, toxin elaboration, or its specific intracellular host target. To accomplish many of these steps (Fig. 03), the required functions and the toxins are produced by phage DNA (residing in the chromosome of the bacterium as temperate, helper, or satellite prophages). As bacteria has adapted to live in different niches such as human host, hence it is very important to discuss the traits encoded by lysogenic phages that enable bacterial colonization in the eukaryotic host cell and survival in the human host.

1.5 Phage derived effector proteins (EPs) and pathogenicity island

Intracellular or temperate pathogens must gain entry into the eukaryotic host cell such as human cell to inject and infect. To facilitate the entry of bacteria into the target cell, phage encoded virulence genes play a significant role. Apart from converting a non-virulent bacterium to a virulent bacterium, bacteriophage-encoded virulence genes also provide novel mechanisms to bacteria. These virulence genes produce proteins which enable attachment, invasion and survival within the eukaryotic host (Tobe et al., 2006, p. 14944).

Phage-encoded exotoxins are found in both Gram-negative and Gram-positive bacteria, including *Corynebacterium diphtheriae, Escherichia coli, Shigella spp, Pseudomonas aeruginosa, Klebsiella pneumoniae, Vibrio cholerae, Neisseria meningitidis Staphylococcus aureus and Streptococcus pyogenes* (Tobe et al., 2006, p. 14944). These exotoxins may be cytotoxic, enterotoxic or neurotoxic that may create variety of diseases starting from mild gastrointestinal disease to life-threatening sepsis and toxemia. These phage-encoded virulence factors are commonly known as effector proteins (EPs) or toxins. Three secretion systems (T3SSs) found in Gram-negative animal and plant pathogens secret these specialized types of proteins. These are large proteins that translocate directly from the bacterial cytosol into the target cell cytosol using the contact dependent bacterial secretion systems (E. F. Boyd et al., 2012, p. 141). These effector

proteins cause a variety of effects once they are inside their eukaryotic host cell due to their multifunctionality with multi-domain that appears to have arisen through recombination reassortment events (Stavrinides et al., 2006). Similarity can be observed on the N-terminal region of one effector to the C-terminal region of another effector but lack any other homology. Effector proteins mostly require a chaperone too. Among 39 different EPs that are transported by the T3SS in *Escherichia coli* (EHEC), 7 EPs are encoded on the locus of the enterocyte effacement (LEE) pathogenicity island (PAIs), genomic islands or gene clusters incorporated in the genome, chromosomally or extra-chromosomally, of pathogenic bacteria acquired from bacteriophage by HGT and the rest are encoded throughout the genome, mostly in prophage genomes (Ogura et al., 2009, p. 17940). Different EPs among strains explain differences in pathogenesis and cell tropism (Ogura et al., 2009, p. 17940). These may target actin filaments, microtubules, mitochrondria, lysosomes, the nucleus, the inner cell membrane and tight junctions of their eukaryotic cell and disrupt many different cells signaling pathways such as MAPK, cyclic AMP and GTPases Rho, Rac and Cdc42 (E. F. Boyd et al., 2012, p. 141).

To begin the approach into host cell, the pathogen has to attach to the host cell surface and manipulate the surface to achieve internalization (Fig. 04) (Tobe et al., 2006, p. 14944). Three secretion systems T3SS of *Escherichia coli* (EHEC) inject an array of EPs into epithelial cells of their human host, so that cell attachment and colonization can occur (Ogura et al., 2009, p. 17940). Subsequent secretion of EPs into target cell by T3SS allows EPs to reach the cytosol of target cell, which results in actin microfilament rearrangement with pseudopod or pedestal formation, that is followed by internalization within a phagosome (fig: 04). Some EPs on the other hand, prevent lysosome fusion with the pathogen containing vacuole while other EPs act as either ubiquitin ligases or deubiquitinases to prevent autophagy. A number of EPs are finally believed to activate

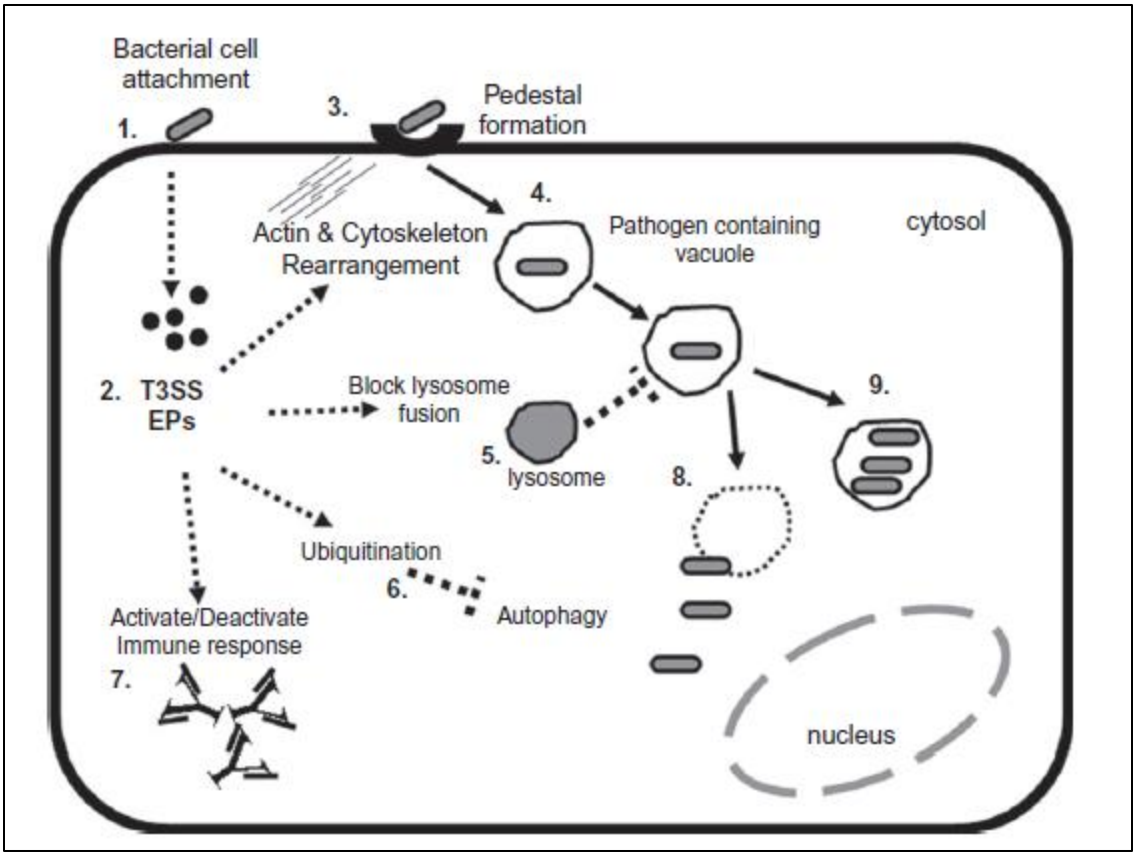


Figure 04: A simplified schematic of some of the roles of T3SS effector proteins (EPs) in pathogen uptake and survival within its target eukaryotic cell (dotted arrows) (E. F. Boyd, 2012, p. 98).

and deactivate the host cell immune response. Pathogens either exit the vacuole and duplicate within the cell cytosol or remain within the vacuole and replicate there. However, to aid in each of these processes' EPs are supposedly identified (E. F. Boyd, 2012, p. 98).

1.6 Phages, Pathogenicity islands, mobile and integrative genetic elements (MIGEs)

Pathogenicity islands are large chromosomal clustered genomic regions believed to encode several virulence factors which can be easily seen in pathogenic isolates (E. F. Boyd, 2012, p. 98). Phages

and PAIs are not closely related to one another as pathogenicity islands were considered to be degenerate forms of phages or other MIGEs. This was hypothesized mainly due to the presence of phage-like integrases and some regions associations with other MIGEs (Tobe et al., 2006, p. 14944). As phage integrases were the first integrases to be identified, many different MIGEs that encode a ‘‘phage-like integrase’’ is characterized by BLAST analysis. They encode an integrase, and insert at a tRNA locus. Most PAIs’ guanine-cytosine and dinucleotide frequency compared to that of their host genome are found almost similar (Hacker & Kaper, 2000, p. 663). Their ability to integrate and excise from host chromosome using an integrase, an attachment site, and a recombination directionality factor excisionase is also noteworthy (Napolitano et al., 2011, p. 428). By site-specific recombination of a circular intermediate, the recombination module allows many MIGEs such as phages and ICEs to insert into the host genome (E. F. Boyd, 2012, p. 98).

Based on a phylogenetic analysis of their encoded integrase, evolutionary relationships have been stablished among PAIs and prophages from *E. coli* strains (Napolitano et al., 2011, p. 428). Analyzed PAIs were evolutionarily distinct from prophages from BLAST analysis (E. F. Boyd, 2012, p. 98). T3SSs encoded within PAIs are required for the transport of specific effectors encoded within diverse phages (E. F. Boyd, 2012, p. 98). Evolutionary necessities indeed altered the prophage genome to its suitable way which created different PAIs and significant for the invasion, attachment and survival of diverse pathogenic bacteria to a variety of target eukaryotic cell.

Thus, it can be inarguably stated that phage is not only considered to be a foe towards bacteria, instead temperate phages or prophages often tend to act like a friend in need of pathogens to make them more pathogenic or to allow non pathogens to gradually become pathogens via phage encoded virulent factors or phage mediated horizontal gene transfer.

1.7 Commensal bacteria and their pathogenicity

Pathogens are those bacteria who can cause damage of the host cell by specific virulence factors. and those virulence factors can either come from phages or from their PAIs. Although the above definition is used to describe pathogens, it would be unfair to define them in such a simple manner due to the transition of harmless commensals into lethal pathogens. Commensal bacteria colonize inside humans by co-existing and without harming humans, maintaining a mutualistic relationships which is often found in the skin, saliva, mammary glands, seminal fluid, placenta, uterus, ovarian follicles, lung, oral mucosa, conjunctiva, biliary tract, and gastrointestinal tract and can be called microbiota as well (Marchesi & Ravel, 2015, p. 21). While they are normally considered to be symbiont, good for the host, they can also become deadly which is frequently happening nowadays. Therefore, the term "pathobiont" was surfaced to describe commensal, harmless bacteria that under specific circumstances, can turn hostile (Rath et al., 2018, p. 12). Methodological advancements in computational biology enabled detailed insights into whole bacterial assemblages and created another prominent term "dysbiosis", describing the altered structures of host microbiota that are associated with disease (Tamboli, 2004, p. 2). With those broader concepts and terms, borders between the so-called commensals, opportunistic pathogens, and pathobionts are increasingly getting blurred and elusive (Rath et al., 2018, p. 12).

The transitions of these commensals could be described by the term "pathogenic function (pathofunction)" coined by (Rath et al., 2018, p. 12) which represents specific features of host bacterial communities to cause non-communicable disease. Pathofunctions have various modes of action including the production of harmful metabolites, extracellular enzymes to immunostimulatory surface structures (Table. 02). Host damage is a direct result of pathofunction activity which usually requires excessive concentrations or longer-term exposure to create a

disease (Rath et al., 2018, p. 12). Pathofunctions, the functions that are shared by various, taxonomically distinct organisms can be distinguished from traditional virulence factors encoded by phages, as they are restrictedly used for characterizing particular strains as pathogens. The transition can also occur if there is any breach in the intestinal epithelium, indicating pathogenic infiltration which alarms the immune system to mount a response. Particularly in the context of epithelial cells, a commensal bacterium can become pathogenic if there is a damage to the epithelial barrier (Ribet & Cossart, 2015, p. 175). For example, the pathogenic bacterium *C. difficile* can manipulate the epithelial tight junction barrier functions and can invade tissues.

Table 02. Selected (putative) pathofunctions of gut microbiota and associated diseases (Rath et al., 2018, p. 12)

Pathofunction	Mode	Associated disease
TMA(O)	Metabolite	CVD, T2D, kidney disease
LCA/DCA	Metabolite	CRC, liver cancer
Hydrogen sulfide	Metabolite	IBD, pouchitis, CRC
Indole/phenol/p-cresol	Metabolite	IBD, CVD, renal failure
N-Nitrosamine	Metabolite	Stomach cancer
Ammonia	Metabolite	Several conditions (e.g., hepatic encephalopathy)
Branched-chain amino acids	Metabolite	Obesity-associated insulin resistance
4-Ethylphenylsulfate	Metabolite	Neurodevelopmental disorders
Uric acid	Metabolite	Gout
Bacterial proteases	Enzyme	IBD

Broadly, there is no intrinsic property that defines a microbe as either a commensal or pathogen and largely it seems to be dictated by immune state of the host or danger signals of tissue damage. So, it can be predicted that these two groups do not share a common invasion pathway. Whereas, the interesting thing is the ability of the innate immune system to recognize and mount an appropriate response against commensals (tolerogenic response) vs pathogens (inflammatory response). For instance, LPS component is a pathogen associated molecular pattern (PAMP) that is present in both commensals as well as pathogens and is recognized by the pattern recognition receptors (PRRs) (such as TLRs) expressed by the host. Using the term "microbe associated molecular pattern" (MAMP) would be more appropriate than PAMP which could ultimately trigger an immune response. To comply with the response, as the environment changes, commensals can also turn hostile for their survival as they were continuously confronted with macrophages due to the mixture of pathogen and commensals at the same space (Miskinyte et al., 2013). Commensal bacteria that had been exposed to macrophages started to show changes in their phenotype. The selective pressure imposed by the presence of the macrophages prompted changes in the bacteria (*From Friend to Foe: How Benign Bacteria Evolve into Virulent Pathogens*, 2013). The changes affected the phenotype of the bacteria ("small colonies of pathogens" or "mucoid colonies") and their genetic make-up. Therefore, having increased ability to cause disease, indicating changes in morphologies and traits similar to those of pathogenic bacteria to rapidly alter commensals to emerge as pathogens (Miskinyte et al., 2013).

1.8 *Pseudomonas, Klebsiella,* and *Neisseria*: Commensalism to pathogenesis

Commensals are certainly turning into deadly pathogens. While most of the commensals are turning steadily or not turning at all, there are few who are breaking the mold from being an opportunistic pathogen, causing treatable infections to emerging as superbugs by being as fatal as

possible. According to a list given by WHO in 2017, *Pseudomonas, Klebsiella*, and *Neisseria* are considered in the critical and high priority level in pathogen list due to their recent contribution in effectively causing diseases and killing people at a highly alarming rate (World Health Organization: WHO, 2017). Thus, these three pathogens are the topic of experiment in this thesis paper, making it very important to know how these three functions as effective killing machines.

Pseudomonas is gram-negative, belonging to the family Pseudomonadaceae, primarily residing in soil, water, vegetation, in the skin of healthy persons and also in the throat and stool of non-hospitalized patients (EUZéBY, 1997, p. 590). They usually do not cause infections in healthy people. If an infection does occur in a healthy person, it is generally mild. More severe infections occur in people who are already hospitalized with another illness or condition, or people who have a weak immune system, causing nosocomial infections which are also considered as opportunistic infections. Although the infections were supposed to be mild and gone within few days with proper antibiotic treatment, in the recent years it has soon become, a powerful resistant superbug who is spreading death in human patients unprecedently. According to Health Protection Agency figures, the number of cases of *Pseudomonas* rose by 41 per cent from 2,605 in 2002 to 3,663 in 2007 in England alone (Campbell, 2017). The 2012 *Pseudomonas* outbreak in Ireland made the situation more critical (Campbell, 2017). The infection specially from P*seudomonas aeruginosa* has become easy to catch on in a hospital while being extremely hard to get rid of through antibiotic treatment, hence being very similar to the potentially fatal MRSA and *C. difficile* infections (Campbell, 2017).

Commonly being found in nature, *Klebsiella,* in humans, are present in parts of the digestive tract where they do not generally cause problems, yet *Klebsiella pneumoniae* and *Klebsiella oxytoca* are the two strains responsible for most human illnesses (*Klebsiella Infection | Genetic and Rare*

Diseases Information Center (GARD) – an NCATS Program, 2018). Like *Pseudomonas*, many *Klebsiella* infections are also acquired in the hospital setting among people with a compromised immune system or people who have a medical device implanted in him (*Klebsiella Infection | Genetic and Rare Diseases Information Center (GARD) – an NCATS Program*, 2018). These infections are becoming more aggressive and difficult to treat due to antibiotic-resistant strains of *Klebsiella* (*Klebsiella Infection | Genetic and Rare Diseases Information Center (GARD) – an NCATS Program,* 2018). In 2011, a sudden *Klebsiella pnuemomiae* Carbapenamase (KPC) outbreak happened at the National Institutes of Health in Bethesda, Maryland, known as the NIH, one of the most established and renowned hospitals (*Klebsiella Pneumoniae Pathogenesis - Microbewiki,* 2016). A 43-year-old woman was transferred to the NIH just after lung transplant and with a KPC infection. As there was no information of the outbreak, weeks later, a new patient tested positive for KPC and over the next six months, several patients were showing up positive too, eventually quarantining the hospital. Known combination of antibiotic treatment seemed to not work and ultimately, a total of 18 patients became infected with 6 patients died from KPC (*Klebsiella Pneumoniae Pathogenesis - Microbewiki,* 2016). After genetic analysis, the bacteria are thought to be home grown as *K. pneumoniae* living in our digestive systems consists a gene that can spread resistance to different bacteria. This incident is just one notable among other examples, leaving a possibility of KPC outbreak which is still lurking in hospitals with no known drugs or treatments.

Neisseria are considered to be commensals as they colonize the host without causing disease that are found at the mucosal surfaces or normal flora in the throat. They also appear to be opportunistic pathogens because of the infections caused by commensal *Neisseria* in persons with deficient immune systems (*Pathogenicity - Gonorrhea - STD Information from CDC,* 2017). Only *N.*

gonorrhoeae is not considered to be at the normal flora, as it is considered always to be pathogenic, infecting the mucosal surfaces of urogenital sites (cervix, urethra, rectum) and the oro- and nasopharynx (throat) and causing symptomatic or asymptomatic infections (*Pathogenicity - Gonorrhea - STD Information from CDC*, 2017). On the contrary, strains of *N. meningitidis* may be carried at normal flora in the throat under some circumstances, even though *N. meningitidis* is also pathogenic, known to cause sporadic cases or epidemics of meningitis. Between 3% and 30% of healthy persons carrying *N. meningitidis* are asymptomatic (*Pathogenicity - Gonorrhea - STD Information from CDC*, 2017). From the cervix or urethra of men or women, *N. meningitidis* has been isolated indicating that it is not a normal flora in genitourinary sites. *Neisseria* was considered to be creating mild infections till 2002. Because in that year at NJ-Conference, among a gathering of around 500 people, a 45 years old woman, Joanne Hemstreet, was rushed to hospital after suffering from a fever, headache, vomiting, and shortness of breath. Doctors thought of this as anthrax but after her death, test results showed Hemstreet died of *Neisseria meningitidis* and infecting 80 other people (*Woman at NJ Conference Dies of Meningitis*, 2002). This type of strange and sudden death also happened to a boy who went to visit Egypt in 2017. As soon as he returned home, he became sick, and within few days he died after many failed treatments (Tawfeek, 2017). Thus, hypervirulence of *N. meningitidis* is definitely able to kill humans within days, while the main evil-doer remains at large.

1.9 Objective of the study

Virulent becoming more virulent and non-virulent achieving virulence or commensals' journey to pathogenesis – all of these clarify one statement that the line distinguishing between pathogen and non-pathogen are getting blurrier with the formers' increased tendency to turn towards the latter with each day. Among a variety of reasons mentioned above, "hypervirulence" is one of the most

important reasons to fit the scenario. Hypervirulence generally originates from virulenence factors which is inherent in bacterial genome or integrated to the genome via prophage. Although the bacterial genome has gone through a plenty of mutations, hypervirulence due to bacterial intrinsic genome does not quite fit. It had to come from outside sources, otherwise bacterial characteristics would have altered way more than that what observed.

Therefore, in this study, the prime concern is to establish a relationship between bacterial increased pathogenicity with rapid inclusion of virulence genes from prophage via computational analysis of genetics.

Chapter 2

Materials and Method

2.1 Databases

Recent advancement in computational biology has made genome analysis way easier than before. Hundreds of databases have been established worldwide to function as repositories of genes. Three things were required to figure out the increased bacterial pathogenicity (*Pseudomonas*, *Klebsiella*, and *Neisseria*) alongside acquisition of increased prophage genes. Those are- a. virulence bacterial genome, b. virulence protein sequence of bacterial pathogens, and c. prophage derived virulence protein sequence. While the database NCBI is considered the father of genome repositories, it is also a quite generalized one, thus making it hard to select only the pathogenic genomes. Virulent bacterial genomes were found in PATRIC database while virulence protein sequence of bacterial pathogens and phages were collected from VFDB and PHIDIAS databases respectively.

The Pathosystems Resource Integration Center (PATRIC), funded by the National Institute of Allergy and Infectious Diseases, is a genomics-centric relational database and bioinformatics

resource, assisting scientists in infectious-disease research (Gillespie et al., 2011, p. 4288). PATRIC provides scientists with a comprehensive bacterial genomics database, an extensive suite of computational tools and platforms for bioinformatics analysis, and a plethora of associated data relevant to genomic analysis. All publicly available genome-scale data for bacteria are compiled and continually updated, enabling comparative analyses to reveal the differences between infectious free-living and commensal species. PATRIC, divides the resources into two major categories: (i) organisms, genomes, and comparative genomics and (ii) recurrent integration of community-derived associated data. All genomes, more than 10000 in PATRIC, are consistently annotated using RAST, the Rapid Annotations using Subsystems Technology (Wattam et al., 2013). PATRIC also includes information related to disease and (Protein-protein interaction) PPIs which are freely available.

Set up in 2004, Virulence factor database (VFDB) is dedicated for providing current knowledge of virulence factors (VFs) from various medical significant bacterial pathogens to facilitate pathogenomic research (Yang et al., 2007). VFDB was therefore upgraded to present the enormous diversity of bacterial genomes in terms of their virulence genes and their organization and also provide (i) detailed tabular comparison of virulence composition of a given genome with other genomes, (ii) multiple alignments and statistical analysis of homologous VFs and (iii) graphical comparison of virulence genes (Yang et al., 2007).

The Pathogen-Host Interaction Data Integration and Analysis System (PHIDIAS) is a web-based database system. It allows researchers to search, compare, and analyze integrated genome sequences, conserved domains, and gene expression data related to pathogen-host interactions (PHIs) for pathogen species including bacteria and phages (Xiang, 2020). It also works as a source of several other databases as well.

2.2 Tools and Software Programs

In addition to that, three software tools such as Prokka (a command line software tool which can rapidly annotate a whole prokaryotic genome), BLAST (refers to a suite of programs to find regions of similarity between nucleotide or protein sequences and to calculate the statistical significance), PhiSpy (a computer program that can be used to identify the prophage regions in a prokaryotic genome) was used as well to conduct various purposes, and PSI-BLAST, a position-specific scoring matrix from multiple sequence alignment of sequences that are detected above a given score threshold via using protein–protein BLAST. PSI-BLAST provides a means of detecting distant relationships between proteins (Bhagwat, 2020).

2.3 Methodology

To begin the process, virulence bacterial genome sequence was collected from PATRIC database. Due to bacterial increased pathogenicity over the years that are observed everywhere, genome sequences of *Pseudomonas*, *Klebsiella*, and *Neisseria* were divided in two timeframes, starting from 18th century to 1990 and from 1990 to the present times which were then subdivided to three regions that are - Asia, USA and the rest of the world including Africa and Europe. Two different time period would give idea about the increased pathogenicity in the course of time while three various regions would uncover bacterial characteristics changes noticed worldwide. For every genus, 20 genome sequences were accumulated from each region and time period contributing to a total of 360 virulence bacterial genome sequences.

After that, those unannotated sequences were annotated using the software program Prokka. Genome annotation was done to describe the function of a predicted gene, including a nucleotide-

level annotation, to identify the physical location of DNA sequences which would determine where components such as genes, RNAs, and repetitive elements are located, a protein-level annotation, to discern the possible functions of genes that a given organism might or might not have, and a process-level annotation, to specify the pathways and processes in which different genes interact, assembling an efficient functional annotation (*Next-Generation Sequencing and Data Analysis: Strategies, Tools, Pipelines and Protocols*, 2018). Since Prokka delivers a variety of information, annotated protein data for each sequence was separated for further inquiry.

Next thing was the vast collection of virulence bacterial protein sequences from VFDB database which would go for BLAST in order to find regions of similarity between previously annotated protein sequences and virulence bacterial protein sequences. Sample sequences that were more than 90% similar to the protein sequences were accepted to go for the next process as it re-assured the highly virulence properties of the bacteria.

Moreover, those highly pathogenic protein sequences went through PhiSpy so that we could detect the prophage regions integrated into the bacterial protein sequences. The prophage regions were compiled together which also provided us with the virulence phage proteins embedded in those regions.

Furthermore, sequences of pathogenic phage proteins and pVOGs, gathered from PHIDIAS and pVOGs database respectively, went for BLAST to determine similarities with the compiled prophage regions with the intention of reassurance that these prophage regions contained virulence phage proteins.

After all of these, compiled protein ids created via PhiSpy allowed for the separation of pathogenic bacterial proteins from pathogenic phage proteins, and finally, total number of pathogenic proteins

were counted in each sequence along with the number of pathogenic bacterial and phage proteins.

To conclude the analysis, PSI-BLAST was used to map genome of targeted strains from two different timetables to illustrate the actual inclusion of phage genome into the more pathogenic bacteria.

Chapter 3

Result and Discussion

3.1 Significant differences found between bacteria of the two time periods

To investigate bacterial increased pathogenicity throughout many years, all the above steps were done and analyzed with precision, leading to an intriguing outcome of the hypothesis which is clearly illustrated by the following figures.

Figure 5 shows the comparison of the changes observed in several characteristics potentially related to pathogenicity in the three genera *Pseudomonas, Klebsiella, and Neisseria* concerning the previously divided time period (before 1990 and after 1990).

The time period was selected considering the fact that genome or protein sequence before 2000 was hard to get due to the technological barriers.

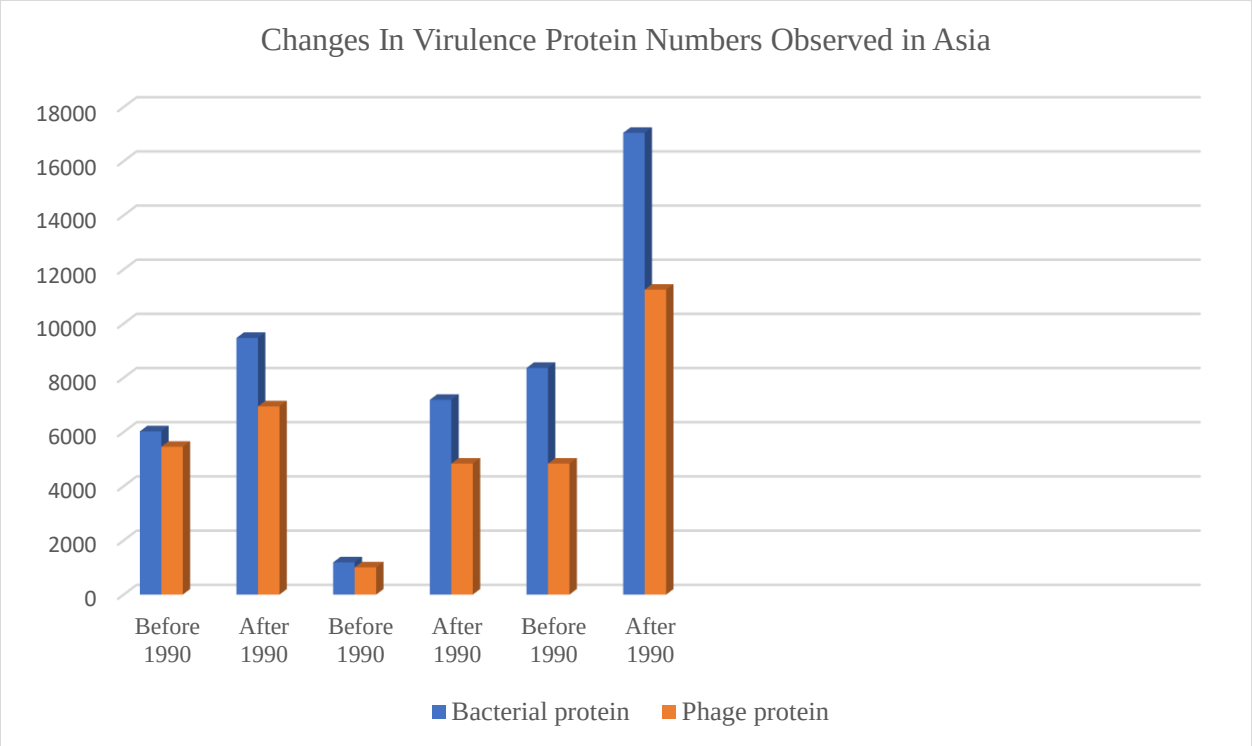
Besides, to analyze any data over the span of many years, a minimum of 25 years or more difference is crucial since a microorganism takes decades to change even a little. Thus, a time segregation of 30 years or more would give valid data regarding the changes observed in microorganisms. In the charts below, among six columns, the 1st two columns (from left to right) represent *Pseudomonas,* the 2nd two *Klebsiella,* and the rest signify *Neisseria.*

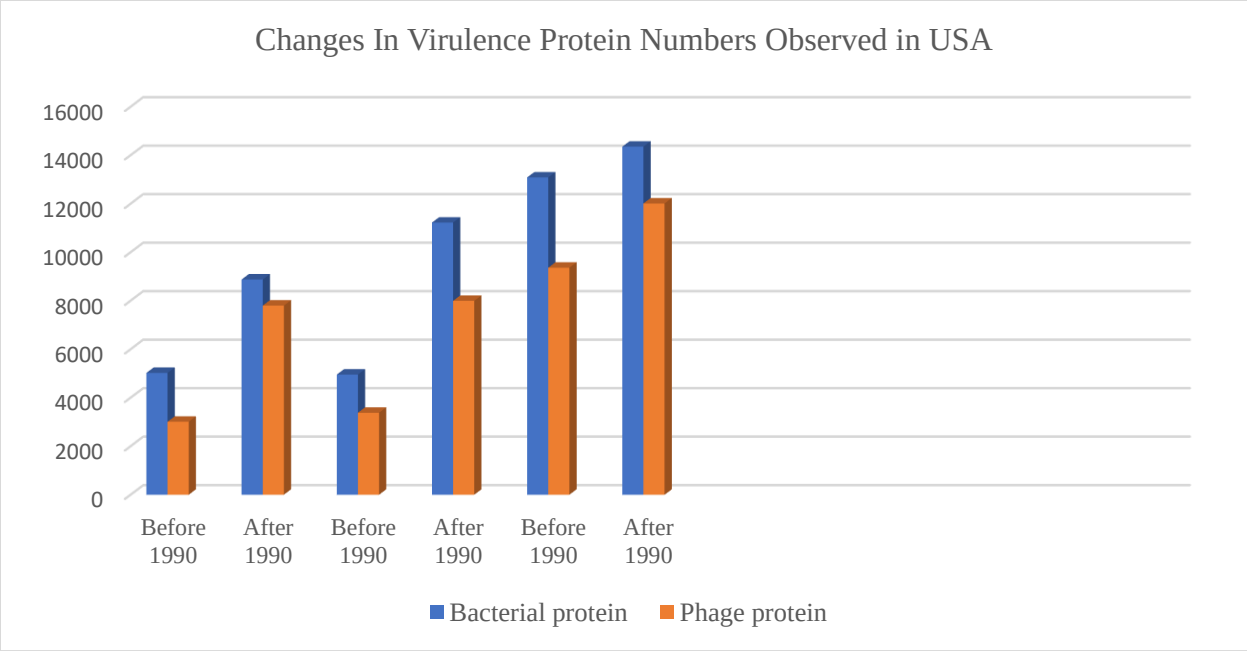
Pathogenic bacteria isolated after 1990 have significantly higher number of virulence proteins in their genome including both the bacterial and phage proteins than that of bacteria before 1990.

In Asia, among the three genera, *Neisseria* shows highest number of growths in respect to acquire more virulence proteins (Fig. 5A) while *Klebsiella* takes the lead spot in the USA (Fig. 5B).

However, bacteria experimented throughout all over the world, the genera that achieved highest growth in obtaining virulence proteins is again *Neisseria* itself (Fig. 5C).

Neisseria takes the lead at the highest expansion in pathogenic proteins numbers between the two time periods, it also consists the most numbers of pathogenic proteins in their genome, whereas *Klebsiella* contains the least numbers of pathogenic proteins in their genome among the three genera regardless of their geographic location.





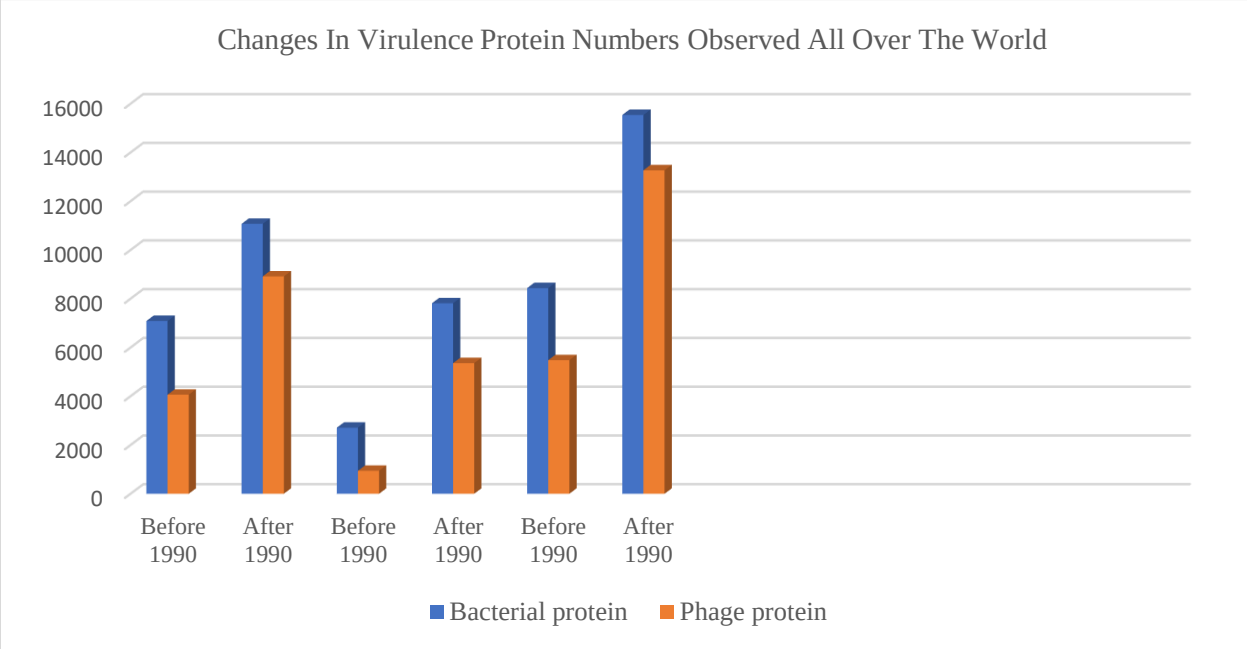
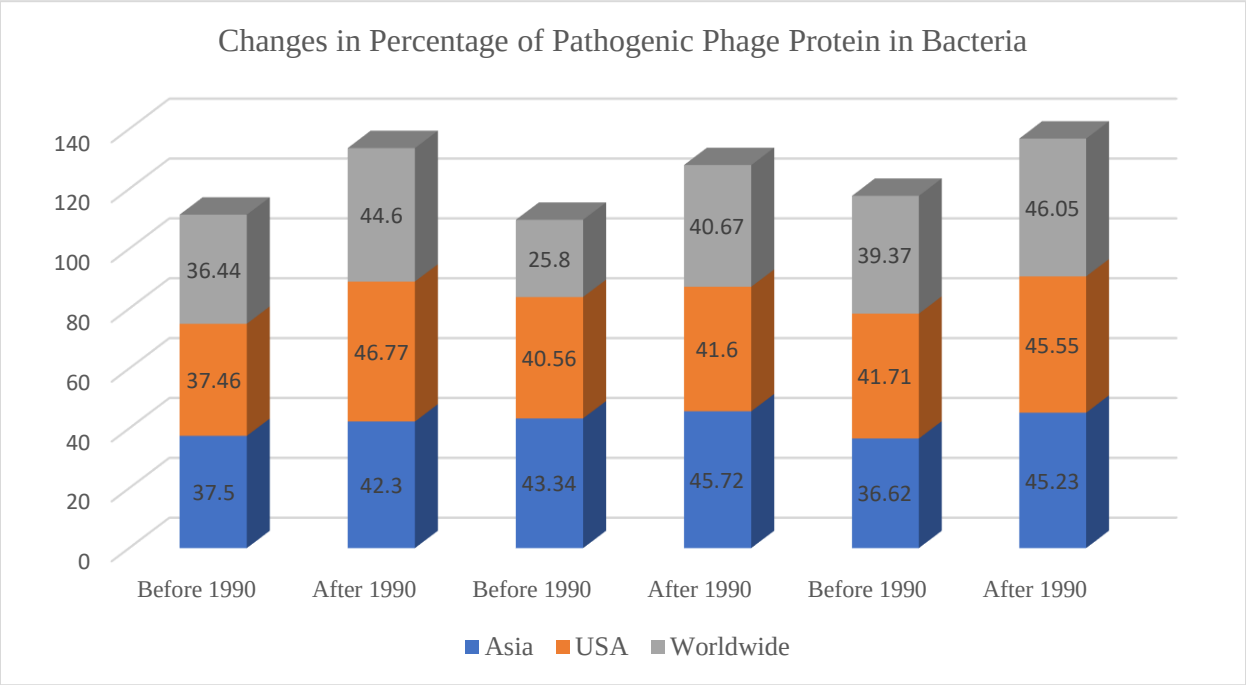


Figure 05: Significant differences between the pathogenic strains of the three genera from both the time period (before 1990 and after 1990). For all panels, each two sets of columns represent (left to right) *Psedomonas*, *Klebsiella* and *Neisseria.*

3.2 Notable increase in pathogenic phage proteins in latter bacteria

It is observed in the analysis that the pathogenic proteins are increasing in numbers from the period before 1990 and the period after. Pathogenic phage proteins acquisition has also increased in numbers among the genera irrespective of their isolated location in course of time. Substantial changes observed in percentage of pathogenic phage protein numbers within the three genera (Fig.



6A). Inclusion of more pathogenic phage protein has been observed in all the genera after 1990 in a notable amount.

Thus, the percentage of pathogenic phage protein prevalence seems to be more likely in the bacteria isolated after 1990 (Fig. 6B) where *Pseudomonas* was found to have the greatest acquisition in phage pathogenic protein with an increase of 8.7% while *Klebsiella* and *Neisseria* had risen by 7.08% and 7.1% respectively.

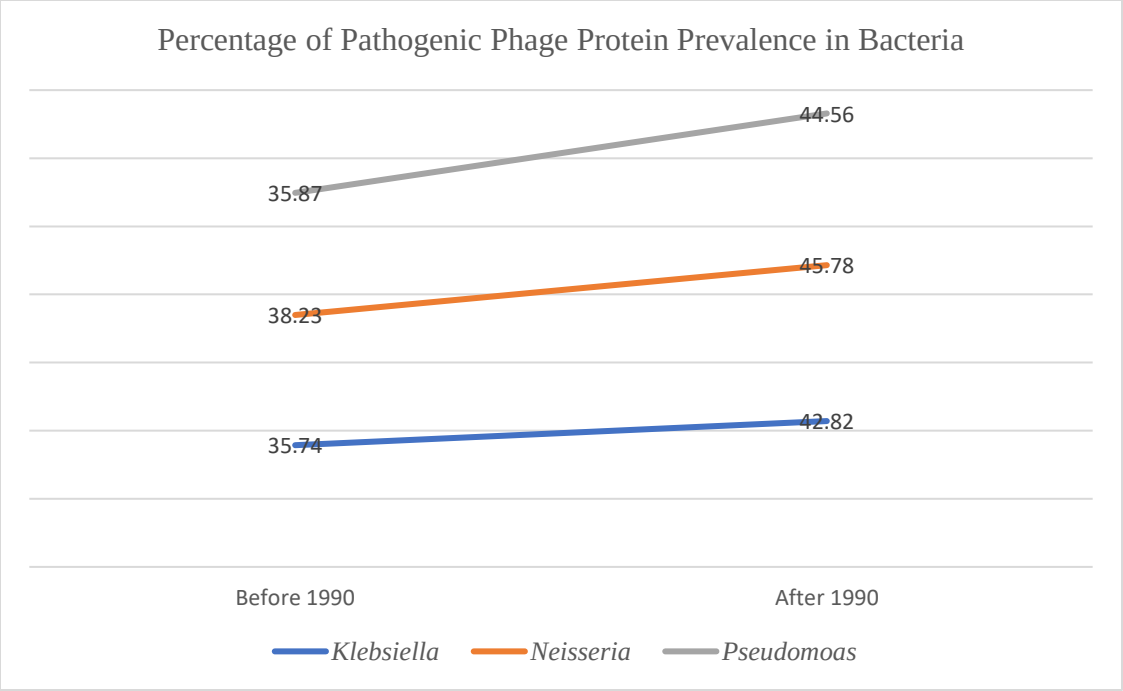


Figure 06: Significant differences in pathogenic phage proteins between the pathogenic strains of the three genera from both the time period (before 1990 and after 1990). A) Each two sets of columns represent (left to right) changes in percentage of pathogenic phage proteins in *Psedomonas, Klebsiella* and *Neisseria.* B) Prevalence of pathogenic phage proteins among all three genera in respect to time.

3.3 Phage proteins are the source of increased pathogenicity among bacteria

As time goes by, bacteria acquire more pathogenic genes consisting both bacterial and phage virulence genes in a significant amount making virulent bacteria deadlier than ever. Bacteria procure a hefty number of genes from other bacteria using conjugation process. For instance, the fertility plasmid of *E. coli* present in one or two copies per cell is very large which is about 100 kilobases (Sapkota, 2020). From the donor or F^+ cell, the F element once in every 10,000 F^+ cells becomes associated with the main bacterial chromosome where a copy of the chromosome instead is transferred through the conjugation tube to the recipient cell F^-, thus making this phenomenon

quite infrequent and being responsible for widespread transfer among bacterial populations of genes conferring several functions. Interplay of commensal and pathogenic bacteria may lead towards mutation that alters commensal bacterial genome by changing normal genes to the manufacturer of pathogenic proteins. Compositional changes in microbiota, hence head to decreased protective and increased aggressive species (Packey & Sartor, 2008, p. 599). Traditionally, mutations within individual genes is considered as the major source of phenotypic variation leading to adaptation. It generates diversity among species, contributing to the initial development of gene functions and fine-tuning (Ochman, 2001, p. 1098). Changes observed in genome repertoire, occurring through gene acquisition and deletion, are another major event underlying the emergence and evolution of bacterial pathogens. Henceforth, obtaining more bacterial pathogenic proteins is noticed all over the globe (Fig. 5A, 5B, 5C).

Increasing phage genes inclusion has definitely strengthen bacterial pathogenicity over the years. The most common way for the acquisition of phage genome is HGT, where transduction makes it happen. As biosphere is filled with virus and bacteria, so HGT is quite frequent at times to allow genetic reassortment. Pathogenic proteins both from bacteria and phages are increasing in numbers while pathogenicity is greatly achieved due to the increasing phage genes. Increasing phage genes allows for the production of more pathogenic phage proteins observed in the experiment (Fig. 6A and 6B). These phage genes are involved in the pathogenicity of bacteria through facilitating several important functions to invade the host. The proteins that secret toxins, bind with the regulatory regions of bacteria that are important enough to kill the host or destroy the host barrier completely. In this case, *P. aeruginosa* produces cytotoxin, necessary for the neutralization and invasion in the host, translated from gene ctx which is a phage encoded gene from phage ΦCTX, without this toxin *P. aeruginosa* would lose its pathogenicity. Similarly, a protein able to alter

antigenicity known as membrane protein of *N. meningitidis* is also encoded by Pnm1 phage (Brüssow et al., 2004, p. 562). Hundreds of phage genes inclusion ultimately produces many pathogenic proteins aiding bacteria in the process like mentioned above. For that reason, increased percentage of pathogenic phage protein prevalence noticed in bacteria after 1990 largely imply enhanced and developed pathogenicity of bacteria.

Phage mediated increased pathogenicity can be more evident with the match of phage provided virulence factors in the bacteria. That's why pVOGs come into the scene. Phage genes contains pVOGs, highly related to providing bacteria with pathogenicity, which was found to have unprecedented matches to the bacterial genome (Fig. 07). Bacteria samples after 1990 showed remarkable matches to pVOGs than the earlier ones. Though pVOGs do not have great affect on the growth of bacteria, yet these help bacteria by adding pathogenicity (Busby et al., 2012, p. 311). Due to their vital match, it highly indicates towards acquired pathogenicity via prophage.

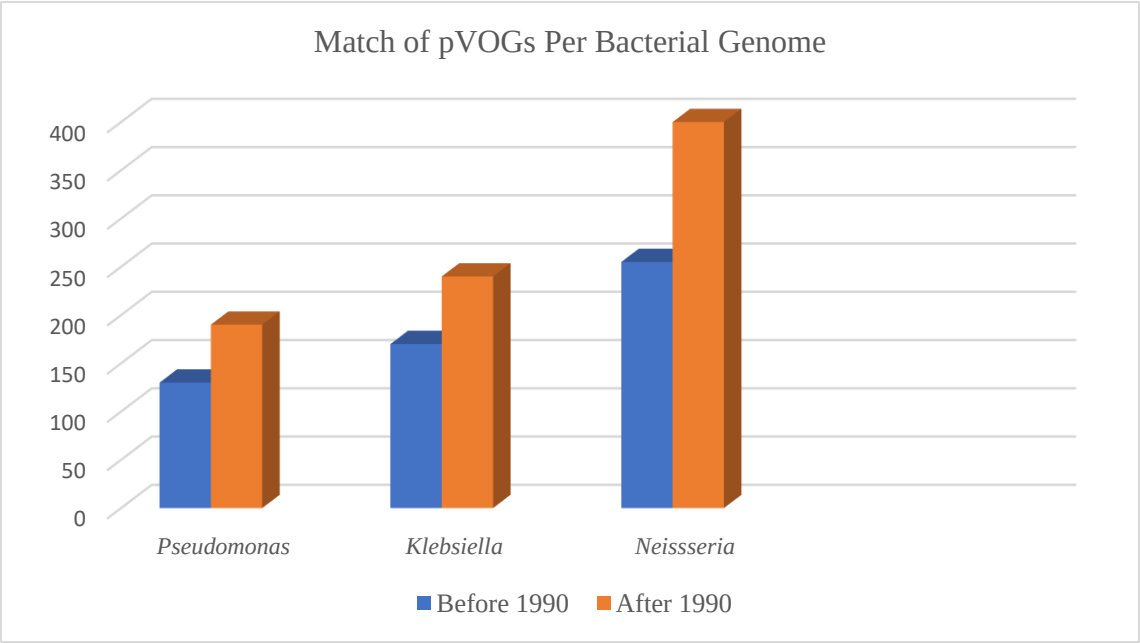
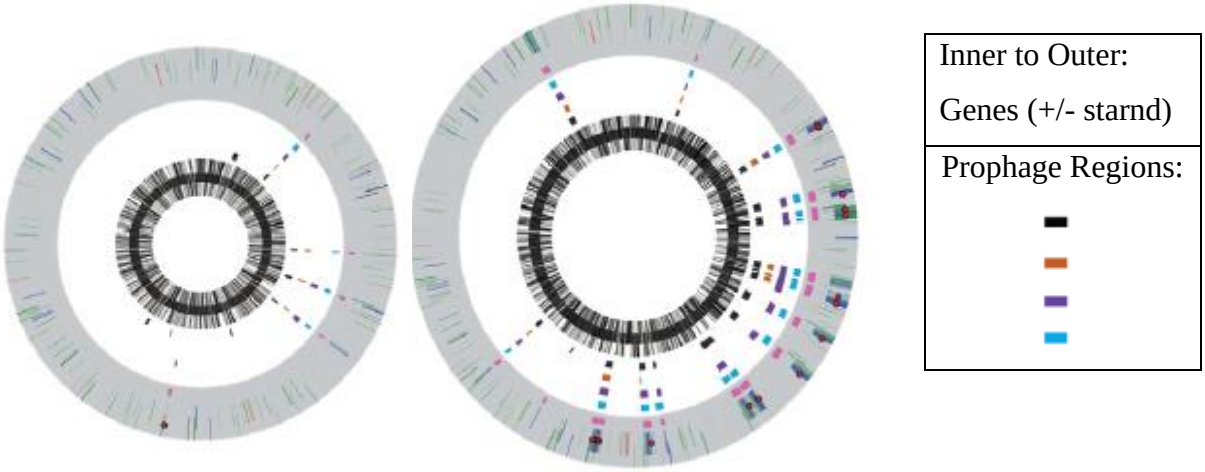


Figure 07: pVOGs matches in each bacterial genome from both the time periods.

The genomic mapping of two closely related pathogenic *Pseudomonas* strains, distinguished from one another on the basis of size, could also lead towards quite a revelation. *Pseudomonas aeruginosa* strain PA87 which is collected in 2018, differs vastly with *Pseudomonas aeruginosa* strain PA33, collected in 1989 in regard to genome (Fig. 08). The size of the strain PA87 expanded by 19.56% giving a whopping 698 phage genes with comparison to just 207 phage genes which also includes a gene named lom that gives the bacteria improved binding capability to mammalian cells. Hereby, addition of phage genes in a pathogen is truly equipping it to better exercise its virulence to another level.



Collected in 1989 *Pseudomonas aeruginosa* **(Strain PA33)**	Collected in 2018 *Pseudomonas aeruginosa* **(Strain PA87)**
4.6 Mbp 4143 Genes 207 Phage matches	5.5 Mbp 5228 Genes 698 Phage matches

Figure 08: Noteworthy differences between two *Pseudomonas aeruginosa* strains representing two different time periods.

The results of the analysis suggest that genome expansion by prophage integration is indeed a major route of evolution by which pathogenic bacteria convert from their familiar dangerous state to more destructive condition.

Chapter 4

Conclusion

Prophages can modify the lifestyle, fitness, virulence, and evolution of their bacterial host in numerous ways. A striking example outside of three genera analyzed here is a disease-causing *S. aureus* that upon chromosomal integration of a phage, ΦN315 converts to a more pathogenic form, secreting enterotoxin P to show minor symptoms to finally sepsis and kill a previously healthy individual within hours by toxic shock syndrome (Brüssow et al., 2004, p. 562).

We currently grasp only a small fraction of the total phage diversity, both in terms of genetic content, but also in terms of novel molecular mechanisms enabling modulation of bacterial phenotypes and lifestyles, including virulence (Fortier & Sekulovic, 2013, p. 357).

Genome sequencing would inexorably continue for the time to come, however genomic data sets, on the contrary are getting too large and complex to use due to brute force algorithms. Hence, simple and efficient approaches like the one applied here for phage gene detection are becoming especially useful to analyze large data sets.

The current research into prophages from bacterial pathogens are largely descriptive and static. We need more prophage genomic studies in large time-series of ecologically or epidemiologically well-defined strain collections using microarray technology.

How many of the emerging bacterial diseases are actually driven by the acquisition of new prophages or new combinations of bacteriophage-encoded virulence factors? We also need more

insight into the fascinating and dynamic interplay between bacteria and prophages both at the level of mRNA expression and at the level of genome architecture.

Therefore, this will certainly represent one of the major challenges over the next era of phage research.

References

1. Wagner, P. L., & Waldor, M. K. (2002). Bacteriophage Control of Bacterial Virulence. *Infection and Immunity*, *70*(8), 3985–3993. https://doi.org/10.1128/iai.70.8.3985-3993.2002

2. Fortier, L.-C., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354–365. https://doi.org/10.4161/viru.24498

3. Boyd, E. f., & Brüssow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in Microbiology*, *10*(11), 521–529. https://doi.org/10.1016/s0966-842x(02)02459-9

4. Boyd, E. F. (2012). Bacteriophage-Encoded Bacterial Virulence Factors and Phage–Pathogenicity Island Interactions. *Bacteriophages, Part A*, 91–118. https://doi.org/10.1016/b978-0-12-394621-8.00014-5

5. Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, *9*(8), 605–618. https://doi.org/10.1038/nrg2386

6. Cross, A. S. (2008). What is a virulence factor? *Critical Care*, *12*(6), 197. https://doi.org/10.1186/cc7127

7. *Definition of Scarlatina*. (2017, January 24). MedicineNet. https://www.medicinenet.com/script/main/art.asp?articlekey=17685

8. Tsao, Y.-F., Taylor, V. L., Kala, S., Bondy-Denomy, J., Khan, A. N., Bona, D., Cattoir, V., Lory, S., Davidson, A. R., & Maxwell, K. L. (2018). Phage Morons Play an Important Role

in Pseudomonas aeruginosa Phenotypes. *Journal of Bacteriology*, *200*(22), 1–15. https://doi.org/10.1128/jb.00189-18

9. Cumby, N., Davidson, A. R., & Maxwell, K. L. (2012). The moron comes of age. *Bacteriophage*, *2*(4), e23146. https://doi.org/10.4161/bact.23146

10. Hendrix, R. W., Lawrence, J. G., Hatfull, G. F., & Casjens, S. (2000). The origins and ongoing evolution of viruses. *Trends in Microbiology*, *8*(11), 504–508. https://doi.org/ 10.1016/s0966-842x(00)01863-1

11. Tobe, T., Beatson, S. A., Taniguchi, H., Abe, H., Bailey, C. M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., & Pallen, M. J. (2006). An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. *Proceedings of the National Academy of Sciences*, *103*(40), 14941–14946. https://doi.org/10.1073/pnas.0604891103

12. Boyd, E. F., Carpenter, M. R., & Chowdhury, N. (2012). Mobile effector proteins on phage genomes. *Bacteriophage*, *2*(3), 139–148. https://doi.org/10.4161/bact.21658

13. Rath, S., Rud, T., Karch, A., Pieper, D. H., & Vital, M. (2018). Pathogenic functions of host microbiota. *Microbiome*, *6*(1), 1–13. https://doi.org/10.1186/s40168-018-0542-0

14. Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, *3*(1), 3–31. https://doi.org/10.1186/s40168-015-0094-5

15. Ribet, D., & Cossart, P. (2015). How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes and Infection*, *17*(3), 173–183. https://doi.org/10.1016/j.micinf.2015.01.004

16. *From friend to foe: How benign bacteria evolve into virulent pathogens*. (2013, December 12). ScienceDaily. https://www.sciencedaily.com/releases/2013/12/131212185831.htm

17. Miskinyte, M., Sousa, A., Ramiro, R. S., de Sousa, J. A. M., Kotlinowski, J., Caramalho, I., Magalhães, S., Soares, M. P., & Gordo, I. (2013). The Genetic Basis of Escherichia coli Pathoadaptation to Macrophages. *PLoS Pathogens*, *9*(12), e1003802. https://doi.org/10.1371/journal.ppat.1003802

18. World Health Organization: WHO. (2017, February 27). *WHO publishes list of bacteria for which new antibiotics are urgently needed*. World Health Organization. https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed

19. EUZéBY, J. P. (1997). List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. *International Journal of Systematic and Evolutionary Microbiology*, *47*(2), 590–592. https://doi.org/10.1099/00207713-47-2-590

20. Campbell, D. (2017, December 2). *Death toll from third superbug soars*. The Guardian. https://www.theguardian.com/uk/2007/nov/18/health.society

21. BBC News. (2012, February 15). *Pseudomonas needs central monitor*. https://www.bbc.com/news/uk-northern-ireland-17040588

22. Campbell, D. (2017b, December 2). *Death toll from third superbug soars*. The Guardian. https://www.theguardian.com/uk/2007/nov/18/health.society

23. *Klebsiella infection | Genetic and Rare Diseases Information Center (GARD) – an NCATS Program*. (2018, June 2). National Center for Advancing Translational Sciences. https://rarediseases.info.nih.gov/diseases/10085/klebsiellainfection#:%7E:text=Klebsiella%20is%20a%20type%20of,responsible%20for%20most%20human%20illnesses.

24. *Klebsiella pneumoniae pathogenesis - microbewiki*. (2016, February 11). Microbe Wiki. https://microbewiki.kenyon.edu/index.php/Klebsiella_pneumoniae_pathogenesis#:%7E:text=Complete%20genome%20here,Description,fermenting%2C%20rod%2Dshape%20organism.&text=Although%20found%20in%20the%20microbiota,urinary%20tract%20infections%2C%20and%20meningitis.

25. *Pathogenicity - Gonorrhea - STD Information from CDC*. (2017, March 31). Centers for Disease Control and Prevention.https://www.cdc.gov/std/gonorrhea/lab/pathogenicity.htm

26. *Woman at NJ Conference Dies of Meningitis*. (2002, February 12). Infection Control Today. https://www.infectioncontroltoday.com/view/woman-nj-conference-dies-meningitis

27. Tawfeek, F. (2017, November 23). *Pupil dies after being diagnosed with meningitis*. Egypt Independent news. https://egyptindependent.com/student-dies-after-being-diagnosed-with meningitis/

28. Gillespie, J. J., Wattam, A. R., Cammer, S. A., Gabbard, J. L., Shukla, M. P., Dalay, O., Driscoll, T., Hix, D., Mane, S. P., Mao, C., Nordberg, E. K., Scott, M., Schulman, J. R.,

Snyder, E. E., Sullivan, D. E., Wang, C., Warren, A., Williams, K. P., Xue, T., … Sobral, B. W. (2011). PATRIC: The Comprehensive Bacterial Bioinformatics Resource with a Focus on Human Pathogenic Species. *Infection and Immunity*, *79*(11), 4286–4298. https://doi.org/10.1128/iai.00207-11

29. Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E. K., Olson, R., Overbeek, R., Pusch, G. D., Shukla, M., Schulman, J., Stevens, R. L., … Sobral, B. W. (2013). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*, *42*(D1), D581–D591. https://doi.org/10.1093/nar/gkt1099

30. Yang, J., Chen, L., Sun, L., Yu, J., & Jin, Q. (2007). VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. *Nucleic Acids Research*, *36*(Database), D539–D542. https://doi.org/10.1093/nar/gkm951

31. Xiang, Z. (2020, September 12). *PHIDIAS: a pathogen-host interaction data integration and analysis system*. PubMed. https://pubmed.ncbi.nlm.nih.gov/17663773/

32. *Next-Generation Sequencing and Data Analysis: Strategies, Tools, Pipelines and Protocols*. (2018, January 1). ScienceDirect. https://www.sciencedirect.com/science/article/pii/B9780128046593000117

33. Bhagwat, A. M. L. (2020, September 12). *PSI-BLAST Tutorial - Comparative Genomics - NCBI Bookshelf*. NCBI. https://www.ncbi.nlm.nih.gov/books/NBK2590/

34. Sapkota, A. (2020, May 3). *Bacterial Conjugation- Definition, Principle, Process, Examples*. Microbe Notes. https://microbenotes.com/bacterial-conjugation/#:~:text=In%20certain%20F%2B%20bacterial%20cells,from%20donor%20to%20recipient%20cell

35. Packey, C. D., & Sartor, R. B. (2008). Interplay of commensal and pathogenic bacteria, genetic mutations, and immunoregulatory defects in the pathogenesis of inflammatory bowel diseases. *Journal of Internal Medicine*, *263*(6), 597–606. https://doi.org/10.1111/j.1365-2796.2008.01962.x

36. Ochman, H. (2001). Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis. *Science*, *292*(5519), 1096–1099. https://doi.org/10.1126/science.1058543

37. Busby, B., Kristensen, D. M., & Koonin, E. V. (2012). Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. *Environmental Microbiology*, *15*(2), 307–312. https://doi.org/10.1111/j.1462-2920.2012.02886.x

38. Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews*, *68*(3), 560–602. https://doi.org/10.1128/mmbr.68.3.560-602.2004

39. Brüssow, H., & Hendrix, R. W. (2002). Phage Genomics. *Cell*, *108*(1), 13–16. https://doi.org/10.1016/s0092-8674(01)00637-7

40. Frobisher, M., and J. Brown. 1927. Transmissible toxicogenicity of streptococci. Bull. Johns Hopkins Hosp. 41:167–173.

41. Spanier, J. G., & Cleary, P. P. (1980c). Bacteriophage control of antiphagocytic determinants in group A streptococci. *The Journal of Experimental Medicine*, *152*(5), 1393–1406. https://doi.org/10.1084/jem.152.5.1393

42. Guan, S., Bastin, D. A., & Verma, N. K. (1999). Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX. *Microbiology*, *145*(5), 1263–1273. https://doi.org/10.1099/13500872-145-5-1263

43. Desiere, F., McShan, W. m., van Sinderen, D., Ferretti, J. J., & Brüssow, H. (2001). Comparative Genomics Reveals Close Genetic Relationships between Phages from Dairy Bacteria and Pathogenic Streptococci: Evolutionary Implications for Prophage-Host Interactions. *Virology*, *288*(2), 325–341. https://doi.org/10.1006/viro.2001.1085

44. Guttman B, Raya R, Kutter E, Kutter E, Sulakvelidze A. Basic phage biology. In: Kutter E, vidze A, eds. *Bacteriophages: biology and applications.* Boca Raton: CRC Press, 2005.

45. Wang, I.-N., Smith, D. L., & Young, R. (2000). Holins: The Protein Clocks of Bacteriophage Infections. *Annual Review of Microbiology*, *54*(1), 799–825. https://doi.org/10.1146/annurev.micro.54.1.799

46. Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004b). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews*, *68*(3), 560–602. https://doi.org/10.1128/mmbr.68.3.560-602.2004

47. Cumby, N., Edwards, A. M., Davidson, A. R., & Maxwell, K. L. (2012). The Bacteriophage HK97 gp15 Moron Element Encodes a Novel Superinfection Exclusion Protein. *Journal of Bacteriology*, *194*(18), 5012–5019. https://doi.org/10.1128/jb.00843-12

48. Vaca-Pacheco, S., Paniagua-Contreras, G. L., García-González, O., & de la Garza, M. (1999). The Clinically Isolated FIZ15 Bacteriophage Causes Lysogenic Conversion in Pseudomonas aeruginosa PAO1. *Current Microbiology*, *38*(4), 239–243. https://doi.org/10.1007/pl00006794

49. Stavrinides, J., Ma, W., & Guttman, D. S. (2006). Terminal Reassortment Drives the Quantum Evolution of Type III Effectors in Bacterial Pathogens. *PLoS Pathogens*, *2*(10), e104. https://doi.org/10.1371/journal.ppat.0020104

50. Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., Kodama, T., Abe, H., Nakayama, K., Kurokawa, K., Tobe, T., Hattori, M., & Hayashi, T. (2009). Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic Escherichia coli. *Proceedings of the National Academy of Sciences*, *106*(42), 17939–17944. https://doi.org/10.1073/pnas.0903585106

51. Hacker, J., & Kaper, J. B. (2000). Pathogenicity Islands and the Evolution of Microbes. *Annual Review of Microbiology*, *54*(1), 641–679. https://doi.org/10.1146/annurev.micro.54.1.641

52. Napolitano, M. G., Almagro-Moreno, S., & Boyd, E. F. (2011). Dichotomy in the evolution of pathogenicity island and bacteriophage encoded integrases from pathogenic Escherichia coli strains. *Infection, Genetics and Evolution*, *11*(2), 423–436. https://doi.org/10.1016/j.meegid.2010.12.003

53. Tamboli, C. P. (2004). Dysbiosis in inflammatory bowel disease. *Gut*, *53*(1), 1–4. https://doi.org/10.1136/gut.53.1.1

54. *Microbes Quotes (18 quotes)*. (n.d.). Goodreads. https://www.goodreads.com/quotes/tag/microbes