ANALYSIS AND COMPARISON BETWEEN THE GENOME OF CRISPR BACTERIOPHAGES

Submitted By:

Mumtasin Muntaha

ID: 14336004

B.Sc. in Biotechnology

Department of Mathematics and Natural Sciences

BRAC University

December 2019

ANALYSIS AND COMPARISON BETWEEN THE GENOME OF CRISPR BACTERIOPHAGES

By

Mumtasin Muntaha

14336004

A dissertation submitted to BRAC University in partial fulfillment of the requirements for the degree of Bachelor of Science in Biotechnology

Department of Mathematics and Natural Sciences

BRAC University

December 2019

Declaration

It is hereby declared that

1. The thesis submitted is my/our 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/We have acknowledged all main sources of help.

Muntasin Muntaha
14336004

Approval

The thesis titled "Analysis and Comparison between the Genome of CRISPR bacteriophages" submitted by Mumtasin Muntaha (14336004) of Summer, 2014 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Biotechnology on 5th December 2019.

Examining Committee:

Supervisor:
(Member)

_____

Tokee Mohammad Tareq
Lecturer, Department of Mathematics & Natural Sciences
BRAC University

Program Coordinator:
(Member)

_____

Iftekhar Bin Naser
Assistant Professor, Department of Mathematics & Natural Sciences
BRAC University

Departmental Head:
(Chair)

_____

A F M Yusuf Haider
Professor, Department of Mathematics & Natural Sciences
BRAC University

Ethics Statement

This thesis has been composed solely by me and it has not been submitted, in whole or in part, in any previous institution for a degree or diploma. All explanations that have been adopted literally or analogously are marked as such.

Abstract

Viruses that infect bacteria are called bacteriophages or shortened as phages. Like all viruses, Phages are absolute parasites, although carry all the information to direct their reproduction in an appropriate host, they have no machinery for generating energy and no ribosomes for making proteins. They are defined as an obligate intracellular bacterial parasite that lacks an independent metabolism. Therefore the CRISPR Bacteriophages has got some unique features for which their genome is a very interesting topic in the field of genetic engineering. Comparison between their genome and investigating their specificity and special characteristics, phage genome comparison shows they are highly specific in their characteristics by their own unique features.

Keywords: Bacteria, Bacteriophage, Genome.

Dedication

Dedicated to my family and friends for their unconditional love and support.

Acknowledgement

First and foremost, I would like to express my thanks to Almighty Allah because He has given me the opportunity and strength to finish this research. I'm thankful for His blessings to my daily life, good health and healthy mind. I acknowledge my esteem to Professor **A F M Yusuf Haider** and Chairperson of MNS Department and Assistant Professor **Iftekhar Bin Naser** for allowing me and encouraging me to complete my undergraduate thesis.

My regards, gratitude, indebtedness and appreciation go to my respected supervisor **Tokee Mohammad Tareq**, Lecturer, Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University for his constant supervision, constructive criticism, expert guidance, enthusiastic encouragement to pursue new ideas and never ending inspiration throughout the entire period of my research work. I would like to thank and express my deepest gratitude for guiding me in my report writing and providing time to time suggestions regarding setting of experimental designs, interpretation of results and subsequent directions for the whole work without being a bit of impatient. It would have been impossible to submit my report without his cordial help. I would thank Tushar Ahmed Shishir, Teaching Assistant, Department of Mathematics and Natural Sciences, BRAC University for his constant guidance and active support throughout my work.

Table of Contents

CHAPTER 1

INTRODUCTION

Viruses that infect bacteria are called bacteriophages or shortened as phages. Like all viruses, Phages are absolute parasites, although carry all the information to direct their reproduction in an appropriate host, they have no machinery for generating energy and no ribosomes for making proteins. They are defined as an obligate intracellular bacterial parasite that lacks an independent metabolism. A century of phage research has revealed that, these viruses are extremely different and ubiquitously present in the biosphere, preying on Eubacteria and Archaea in a broad range of biological niches. Accordingly, the genome sizes of phages vary enormously, from a few thousand base pairs up to 498 kilobase in phage G, the largest phage sequenced to date. Although the size of this genome resembles that of an average bacterium, even phage G lacks genes for essential bacterial machinery like ribosomes, emphasizing the purely parasitic nature of these organisms (Ceyssens&Lavigne, 2010). The High level of specificity, durability, long-term tolerance and the inherent potentials to reproduce rapidly in an appropriate host contribute to their maintenance of dynamic balance among the great diversity of bacterial species in any natural ecosystem (Guttman et al., 2005)

Based on their size and shapes, bacteriophages can be categorized into three. Icosahedron Bacteriophages, Filamentous Bacteriophages, and Complex Bacteriophages. 1. Icosahedron bacteriophages: an almost spherical shape, with twenty triangular facets, the smallest is icosahedron phages are about 25nm in diameter. 2. Filamentous bacteriophages: long tubes formed by capsid protein assembled into the helical structure; they can be up to about 900nm in diameter. 3. Complex bacteriophages: icosahedral heads attached to helical tails; may also poses base plates and tail fibers.

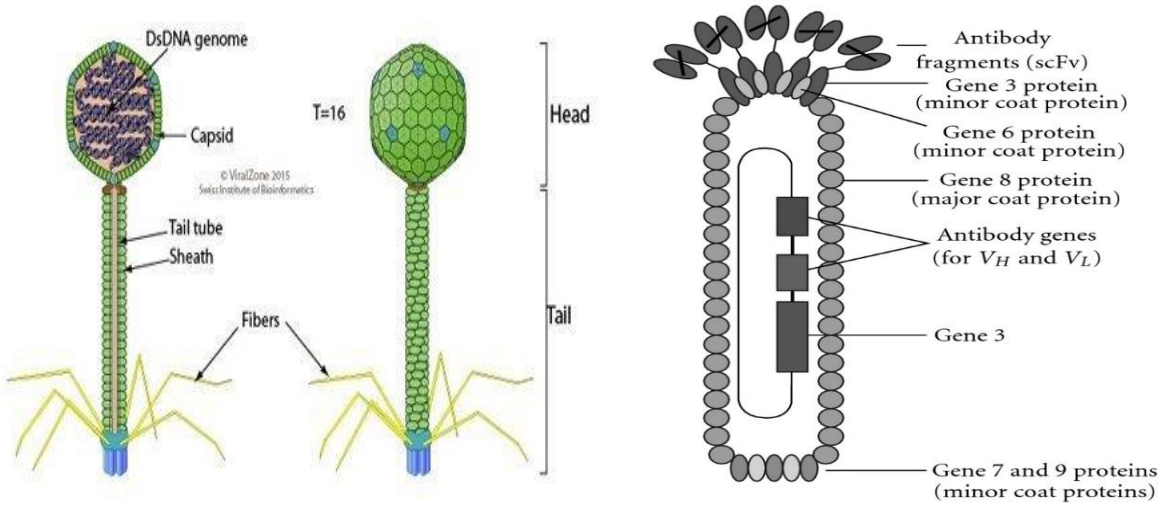
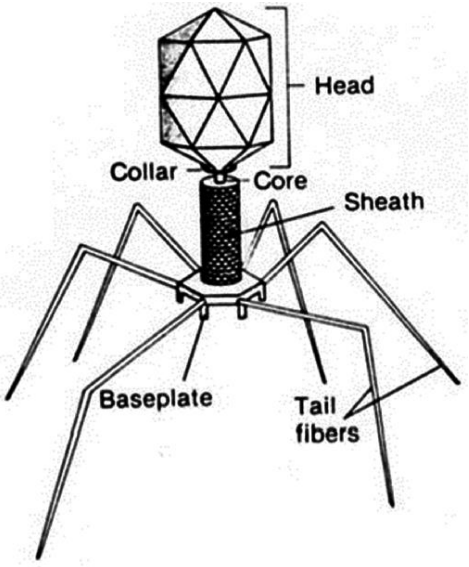


Fig 1: Shapes of bacteriophages Icosahedron(left), Filamentous(right), Complex



Bacteriophages were first discovered in 1915 by William Twort, and in 1917 by Felix d'Herelle realized that they had the potential to kill bacteria. After a pre-antibiotic era heyday they were then essentially disregarded as significant therapeutic agents in the West, primarily due to the comparative ease by which antibiotics could be administered. Research and the practice of using bacteriophages did continue in some countries such as Georgia (as part of the former USSR), where they were, and continue to be routinely isolated and used to treat a large number of diseases (Chanishvili&Sharp, 2009). Bacteriophage research then focused on a number of model phages which primarily infected *E. coli*. These studies provided the back-bone of modern molecular biology, for example phages were used to identify the basis of genetic material, and that 3 nucleotides code for an amino acid (Clokie & Kropinski, 2009). They also allowed the identification of restriction enzymes. For several decades, only a handful of phages were studied in great detail. The recent renaissance seen in phage biology has been triggered due to a growing awareness of the number of phages in all bacterial dominated environments (as revealed by epiflourescent and electron microscopy, and from molecular studies), and indeed in the genomes of bacteria following whole genome sequencing projects. This checkered history has resulted in a patchy knowledge of phage biology but with enough observations for scientists to realize that phages are dictating many aspects of Bacterial/Archaeal biology.

In 1896, a British bacteriologist Ernest Hanbury Hankin, working as the Chemical Examiner and Bacteriologist to the Government of the United Provinces and of the Central Provinces of India, demonstrated that the waters from the Indian rivers Ganga and Yamuna contained a biological principle that destroyed cultures of cholera-inducing bacteria. This substance could pass through milli - pore filters, known to be able to retain larger microorganisms such as bacteria. He published his work in the Annals of the Pasteur Institute (Hankin, 1896). In 1915, while he was studying the growth of vaccinia virus on cell-free agar media, Frederick Twort, a British microbiologist, noted that “pure” cultures of bacteria may be associated with a filter- passing transparent material which may entirely break down bacteria of a culture into granules (Twort, 1915). This transparent material, which was found to be unable to grow in the absence of bacteria,

was described by Twort as a ferment secreted by the microorganism for some purpose not clear at that time.

Two years after this report, Félix d'Herelle independently described a similar experimental finding, while studying patients suffering or recovering from bacillary dysentery. He isolated from stools of recovering shigellosis patients a so-called "anti-Shiga microbe" by filtering stools that were incubated for 18 h. This active filtrate, when added either to a culture or an emulsion of the Shiga bacilli, was able to cause arrest of the culture, death and finally lysis of the bacilli (Ttwort, 1917). D'Herelle described his discovery as a microbe that was a "veritable" microbe of immunity and an obligate bacteriophage. He also demonstrated the activity of this anti-Shiga microbe by inoculating laboratory animals as a treatment for shigellosis, seeming to confirm the clinical significance of his finding by satisfying at least some of Koch's postulates. He even introduced treatment with intravenous phage for invasive infections, and he summarized all these findings and observations in 1931 (Twort, 1931). More than a half a century later, in 1969, Max Delbrück, Alfred Hershey, and Salvador Luria were awarded the Nobel Prize in Physiology or Medicine for their discoveries of the replication of viruses and their genetic structure.

The ability of phages to infect bacteria led d'Hérelle to examine their therapeutic potential against bacterial infection. Even in his first paper, he noted that the presence of phages correlated with disease clearance in patients with dysentery, and he carried out a rabbit study, in which phages provided protection from infection with *Shigella*. Most early phage research conducted in the 1920s and 1930s focused on the development of phage therapy for the treatment of bacterial infections, and companies began to market phage preparations. However, in the late 1930s, the Council on Pharmacy and Chemistry of the American Medical Association concluded that the efficacy of phage therapy was ambiguous and that further research was required (Summers et al., 2001). These concerns, and the success of emerging antibacterials, led to a decline in interest in phage therapy, although research continued in the former Soviet Union and other Eastern European countries. During this period, insights into fundamental phage biology were limited, and up until the 1940s the viral nature of phages was still disputed. Visualization of phages by electron microscopy in the early 1940s proved their particulate nature (Ruska et al., 1940).

In order to appreciate the roles of phages in nature, an understanding of their possible interactions with their hosts is necessary. Phages have various possible life cycles which, along with interaction with their physical environment, dictate their role in bacterial/archaeal biology. The lytic life cycle is where phages infect and rapidly kill their infected host cells, thereby shaping

bacterial population dynamics and occasionally assisting in their long term evolution via generalized transduction (Weinbauer&Rassoulzadegan, 2009). The lysogenic life cycle in contrast, is where phages instead of directly killing their hosts, integrate into their host genome, or exist as plasmids within their host cell (Little, 2005). This lysogenic life cycle can be stable for thousands of generations and the bacteriophage may alter the phenotype of the bacterium by expressing genes that are not expressed in the usual course of infection in a process known as lysogenic conversion. A well-known example of this is the gene associated with Vibrio chlolerae which encodes the toxins that cause cholera symptoms (Los et al., 2010). Phages may also have a psuedolysogenic component to their life cycle. This is a controversial concept, and has many different definitions within phage biology (Abedon, 2008). We define it here as the situation that occurs when a phage has entered a bacterial cell and doesn't integrate in a stable fashion, but will stay in this 'mode' until conditions occur which trigger them to enter into the lytic or lysogenic life cycle (Wilson et al., 1996). Finally there is the chronic infection lifestyle found in some archaeal viruses, in filamentous phages (rod shaped single stranded DNA phages), and in plasmaviruses which infect Mycoplasma. In this life cycle phages are slowly shed from the cell over a long time period without obvious cell death.

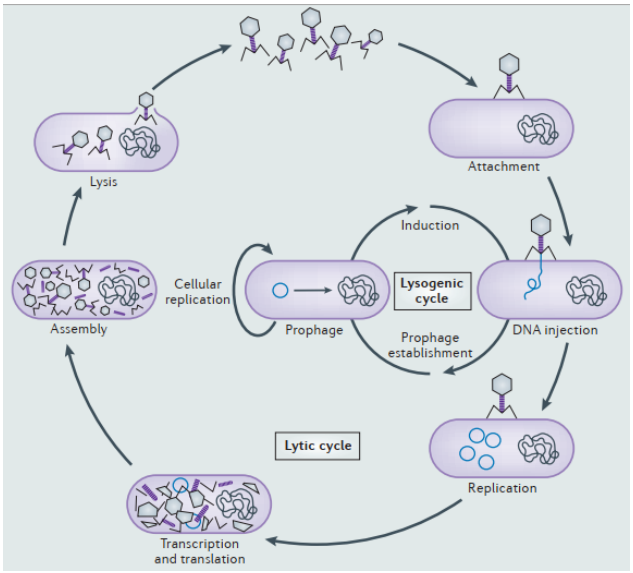


Fig 2: The stages of lytic and lysogenic life cycle of bacteriophage

Since their discovery, phages have had an immense and unforeseen impact on our understanding of the wider biological world. Their 'simplicity' enabled our understanding of core biological processes that are relevant to all biology. Phages provided tractable model systems that gave rise to molecular biology and provided many useful reagents, including restriction enzymes, *en route*. In addition, their influence on nutrient cycles, pathogenicity and bacterial evolution further underlines their central role in global ecology and evolution. Furthermore, the inexorable rise of antibiotic resistance has provided added impetus for 'back to the future' phage-based solutions to bacterial infection. We are also currently witnessing incredible advances in the biotechnological exploitation of CRISPR–Cas phage defence systems, which are revolutionizing both prokaryotic and eukaryotic molecular biology research.

The success of fundamental and translational phage research had high impact. Many researchers began to focus on bacteria other than *E. coli* and more complex eukaryotic organisms that became genetically tractable, owing to the new phage-inspired molecular biology 'toolkit'. However, research from the late 1980s to the 2000s also reinvigorated our understanding of basic phage biology. Before 1989, phage numbers in aquatic environments were thought to be low, until up to 2.5 × 108 viruses per ml were detected in natural waters (the typical range is ~106–107), suggesting that phages are important in the turnover of microorganisms and in gene transfer in the environment (Bergh, 1989). Subsequently, it became clear that viruses were abundant and active partners in food webs and in carbon and nitrogen cycling in the oceans (Suttle, 1990). Furthermore, some marine phage genomes carry auxiliary metabolic genes, such as those encoding proteins involved in photosynthesis, which are thought to assist infection by 'complementing' rate-limiting steps in host metabolism (Clokie, 2011). Metagenomics of ocean samples then revealed the true extent of phage abundance and diversity (Breitbart, 2012). The abundance and diversity of phages in almost all natural environments, and in association with plants and animals, are now widely accepted (Clokie, 2011).

Improvements in sequencing technologies generated a surge in phage genome data, revealing evolutionary relatedness and genomic mosaicism (Hendrix, 1999) with concomitant implications for phage classification. However, despite the deluge of phage genomic data, particularly since the emergence of next-generation sequencing methods, bacterial viruses often encode viral proteins with no known homologues33, and presumably these 'unknowns' define new biological processes. Therefore, there are still enormous gaps in our knowledge of phages and their life cycles.

To understand the ecological and evolutionary role of phages, their co-evolution with their bacterial hosts must be considered. During his research in 1917, d'Hérelle observed increased phage numbers in stool samples taken from patients before their recovery from dysentery, perhaps an early ecological insight into the dynamics of the interaction between phages and bacteria2. Co-evolutionary experiments in natural and laboratory settings have begun to show that co-evolution promotes the rate of phage and bacterial evolution, sustains both genetic and phenotypic variation and can alter microbial community structures. Indeed, it is perhaps unsurprising that phages and bacteria constitute the greatest genetic diversity on the planet. The ease and speed with which large bacterial and phage populations can be manipulated in laboratory microcosms has favored their application towards addressing more general questions in evolutionary theory, an approach termed experimental evolution (Buckling et al., 2009). The power of rapid next

generation sequencing to elucidate genetic changes underpinning phenotypic and population shifts will further increase the utility of phage–bacterium systems as ecological and experimental evolution models. Finally, the ability to analyse CRISPR–Cas systems to link host bacteria with phages will undoubtedly improve our understanding of their interactions in complex ecosystems and in an evolutionary context.

Bacteriophages play significant roles in a large number of biological and environmental processes, it is estimated that phages can kill and lyse between 15% and 40% of the ocean's bacteria every day which influences the ratio of particulate to dissolve carbon, rates of phytoplankton productivity and oxygen production, perhaps even global climate and weather patterns (Danovaro et al., 2011) and they also play as significant drivers in the evolution of bacteria, especially temperate phages which are prominent agents of horizontal gene transfer. It is believed worldwide that phages mediate gene transfer events between bacteria through transduction process up to 20 million times per second (Chibani&Chennoufi, 2004).

The impact of phages on the evolution of bacteria is underscored by the estimation that, globally, ~2 × 1016 phage-mediated gene transfer events occur every second (Brussow et al., 2004). Furthermore, in 1996, the discovery that the *Vibrio cholerae*toxin, a key virulence factor, is encoded in the genome of the transferrable filamentous phage CTXΦ highlighted the importance of phages in the evolution of bacterial pathogenicity (Waldor et al., 1996). This process, termed lysogenic conversion (or phage conversion), was first observed in the 1950s (Freeman, 1951) and it describes a situation in which a prophage provides additional genes that benefit the lysogen. The importance of phages for pathogenicity in bacteria was further demonstrated in the 1990s, when genome sequencing revealed the abundance of prophages and established that they account for the main genetic variability between closely related bacterial strains (for example, pathogens versus nonpathogens) (Brussow et al., 2004). For example, in *Streptococcus pyogenes*, ~10% of the genome consists of prophages, which encode multiple virulence factors, and in *E. coli* O157:H7 str. Sakai, 18 prophages constitute 16% of its genome (Brussow et al., 2004). In some cases, the prophages define core aspects of bacterial pathogenesis. For example, in Shiga toxin-producing *E. coli*, prophage induction upregulates the toxin genes, and cell lysis is important for toxin release (Neely & Friedman, 1998). Finally, prophages that exhibit regions of homology can drive evolutionary changes through inversions or deletions and other chromosomal rearrangements (Brussow et al., 2004) . For example, an *S. pyogenes*M3 strain isolated in Japan differs from an isolate from the United States owing to a chromosomal inversion between two different prophages, which encourages reshuffling of the prophage virulence genes (Brussow et

al., 2004). Inversions and deletions can also modulate fitness through selection events that drive rapid evolutionary changes. Phages are also responsible for horizontal gene transfer (HGT) between bacteria. For example, by inducing bacterial lysis, phages promote release of bacterial DNA, which can then be acquired by neighboring competent cells. Furthermore, much of the phage-derived HGT occurs by generalized transduction, where bacterial DNA is accidentally packaged during phage replication and then delivered into neighboring cells. This phenomenon was discovered in 1952 by Norton Zinder and Joshua Lederberg, which contributed to Lederberg's shared 1958 Nobel Prize in Physiology or Medicine "for his discoveries concerning genetic recombination and the organization of the genetic material of bacteria". Specialized transduction is the transfer of DNA located adjacent to the integrated prophage after imprecise excision. Transduction facilitates the mobilization of antibiotic resistance and virulence genes, and antibiotic exposure can promote these processes (Modi et al., 2013). A form of 'constitutive generalized transduction' is promoted by gene transfer agents (GTAs), which have a significant role in HGT in bacteria (McDaniel et al., 2010). GTAs are prophage-like elements encoded in bacterial genomes that package random host DNA but cannot package enough to enable the transmission of their own genes (Lang et al., 2012). GTAs might have evolved from mutant prophages that became defective and subsequently decayed. Phage-inducible chromosomal islands (PICIs) can hijack phages to assist in their transfer, giving rise to high-efficiency transduction, where the islands are transferred to neighboring bacteria (Novick et al., 2010). For example, the *Staphylococcus aureus*pathogenicity islands (SaPIs) encode super antigens and 'parasitize' phages for high frequency transduction. As identification of PICIs and GTAs is challenging, their general contribution to gene transfer is probably underestimated. In addition to providing virulence genes to bacteria, phages themselves may have been co-opted by bacteria during the evolution of R-type pyocins and type VI secretion systems (Novick et al., 2010). Type VI secretion systems use a phage tail-like cell puncturing mechanism to deliver effector proteins into both eukaryotic and prokaryotic cells. R-type pyocins are phage tail-like structures encoded in bacterial genomes that are released during cell lysis and bind to and kill other bacteria. Notably, some marine bacteria release arrays of tail-like structures, which can induce metamorphosis in marine tubeworms, suggesting that phage-like structures can be used by bacteria to interact with eukaryotic organisms (Shikuma et al., 2014). Collectively, these studies demonstrate that there are many ways in which phages contribute to bacterial virulence and host interactions.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)

Bacteria dominate many natural habitats, including inhospitable environments, despite challenging conditions that include pred- atory viruses. To survive, bacteria have developed many ways to fend off invaders (Labrie et al., 2010), including the recently described CRISPR system. CRISPR is an acronym for clustered regularly interspaced short palindromic repeats. CRISPR loci contain short, partially palindromic DNA repeats that occur at regular intervals and form loci that alternate repeated elements (CRISPR repeats) and variable sequences (CRISPR spacers). These peculiar loci were first observed in 1987 (Ishino et al., 1987), but they received little attention until similar, idiosyncratic loci were described in microbial genome drafts (Jansen et al., 2002).

These loci are typically flanked by accompanying CRISPR-associated (cas) genes. Their biological role remained elusive until 2005, when three groups reported that the spacers were homologous to foreign genetic elements, including viruses and plasmids (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). These reports lead to the hypothesis that CRISPRs might function as an immune system (Makarova et al., 2006). Evidence for CRISPR-mediated immune function quickly followed (Barrangou et al., 2007), and subsequent studies established that CRISPR-mediated defense involves sequence-specific, RNA-mediated (Brouns et al., 2008) targeting of mostly DNA (Marraffini and Sontheimer, 2008), and occasionally RNA (Hale et al., 2009). Since then, many studies have delved in the molecular underpinnings of the CRISPR-Cas system genetics, mechanisms, and applications.

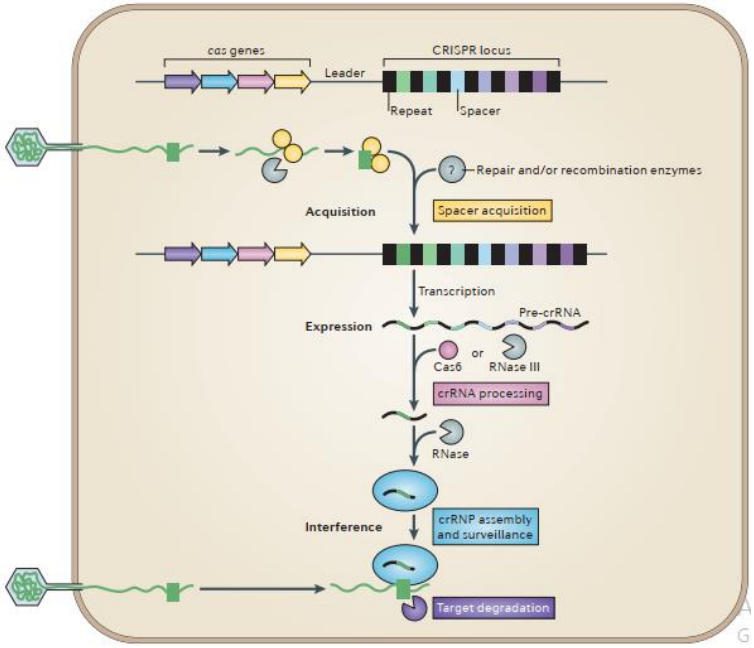


Fig 3: Overview of the CRISPR system.

Generally, CRISPR-Cas systems and their elements are hypervariable and differ broadly in terms of occurrence, genes, sequences, number, and size across genomes. Indeed, CRISPR repeats can vary widely (23–55 nt), though they are typically 28–37 nt, and their partially palindromic nature allows them toform hairpin structures. Likewise, CRISPR spacers can also vary in size (21–72 nt), though they are typically 32–38 nt. Quantitatively, arrays that contain up to 588 repeats have been reported (in Haliangiumochraceum), but they are less than 50 units in most cases. Likewise, although up to 19 distinct loci have been reported in Methanocalcococcus, and 25 putative CRISPR loci have been suggested inMethanotorrisigneus, organisms typically contain one to two CRISPR loci (Grissa et al., 2007). According to the database CRISPRdb, CRISPRs occur in nearly half (1,126/2,480, or ~45%) of bacterial genomes and the large majority (125/150, or ~83%) of archaea (Grissa et al., 2007).

CRISPR-Cas systems have been classified into three major types, namely type I, type II and type III, and 12 subtypes, given their genetic content and structural and functional differences (Makarova et al., 2011a, 2013). The core defining feature of CRISPR-Cas types and subtypesare the cas genes and the proteins they encode, which are highly genetically and functionally diverse, illustrating the many biochemical functions that they carry throughout the different steps of CRISPR-mediated immunity. Noteworthy, the RNA recognition motif is widespread in many Cas proteins, and most of the Cas families of proteins carry functional domains that interact with nucleic acids, such as DNA binding, RNA binding, helicase, and nuclease motifs (Makarova et al., 2002, 2006, 2011a, 2011b, and 2013). Genetically, cas1 and cas2 universally occur across types and subtypes, whereas cas3, cas9, and cas10 have been defined as the signature genes for type I, type II, and type III, respectively. Phylogenetically, type II systems have solely been identified in bacteria, thus far, and there is a bias for type I systems in bacteria and type III systems in archaea and hyperthermophiles.

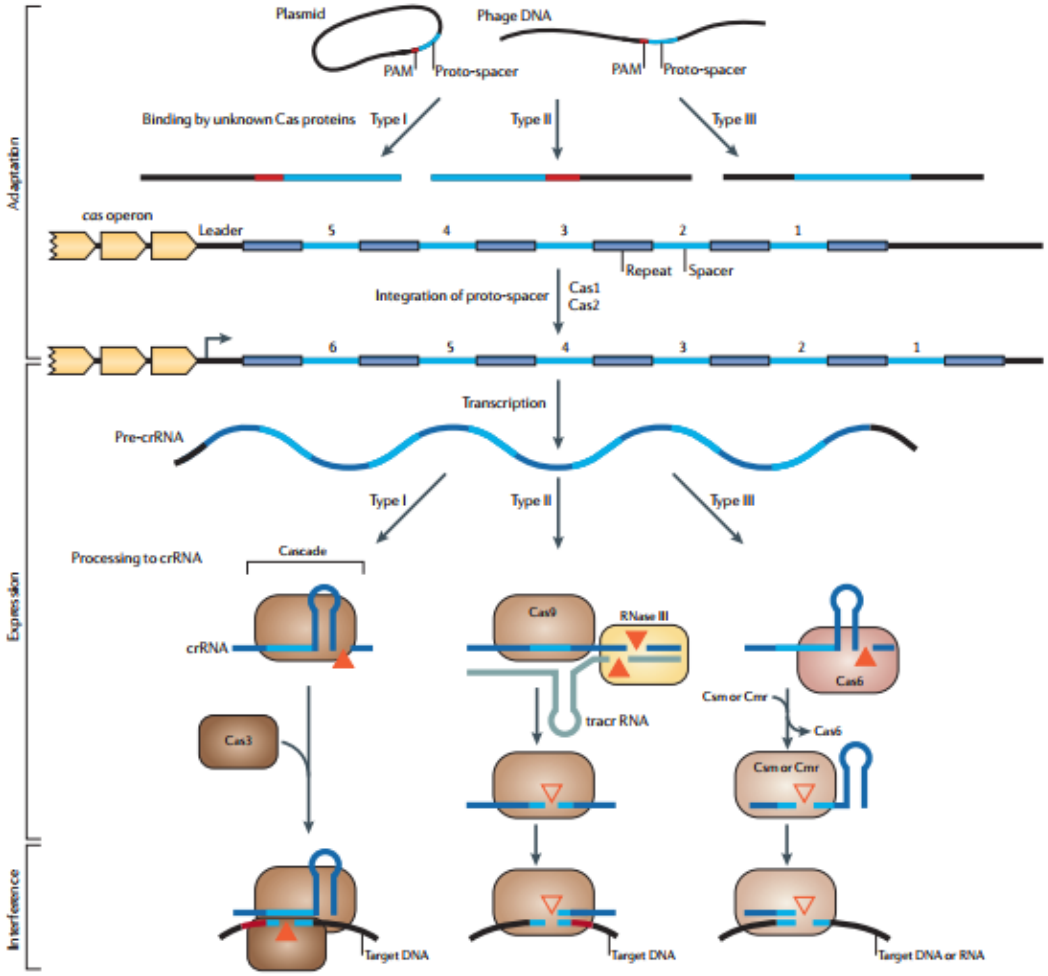


Fig 4: The three major CRISPR-Cas system

Overall, CRISPR-Cas immune systems function in three steps. The first step is adaptation, in which new spacers are acquired from exogenous nucleic acid into the CRISPR locus. The adaptation step is followed by CRISPR RNA (crRNA) biogenesis, in which CRISPR arrays are transcribed and processed into small interfering crRNAs. The final step is targeting, in which crRNAs guide Cas nucleases for specific cleavage of homologous sequences. These steps have been described in a number of reviews published since 2008, including several recent extensive or focused detailed reviews (Barrangou, 2013; Barrangou and Horvath, 2012; Fineranand Charpentier, 2012; Marraffini, 2013; Reeks et al., 2013; Sorek et al., 2013; Westra et al., 2012; Wiedenheft et al., 2012). Comparative analyses have unraveled potential commonancestry between CRISPR- Cas system components and core elements that define mobile genetic elements (notably transposases), as well as other defense systems such as toxin-anti toxin and restriction-modification systems (Makarova et al., 2013).

Although there are several innate-immunity-like systems in bacteria, such as abortive infection, receptor mutation, and restriction-modification, the recently characterized CRISPR-Cas system has been described as an adaptive immune system, which provides specific and acquired immunization against exogenic mobile genetic elements.

The first biological evidence that CRISPR-Cas systems have a role in adaptive immunity was reported in 2007 when *S.thermophilus* CRISPR loci were shown to acquire novel spacers derived from the invasive phage DNA (Barrangou et al., 2007). Theacquisition of phage DNA into the leader end of a CRISPR array led to sequence-specific, inheritable immunity against phages bearing homologous sequences. Only a small proportion of the population gained CRISPR encoded immunity, but this immuniza- tion, albeit infrequent, provided a high level of resistance (Barran- gou et al., 2007; Deveau et al., 2008; Levin et al., 2013). Several studies of the *S.thermophilus* system have shown that most areas of the viral genome are targeted, including both DNA strands, coding and noncoding sequences, and all transcription modules. Nevertheless, a recent study showed that sampling of the viral genome is biased (Paez-Espino et al., 2013), which might be due to DNA structural or composition features. There is a danger that CRISPR systems can target host sequences. Indeed, self-tar- geting is a very rare and lethal event (Paez-Espino et al., 2013). Given this danger, CRISPRs must have a way to distinguish self from nonself that is yet to be characterized. Several subsequent studies in *S.thermophilus* confirmed this adaptive immunity phenomenon and characterized genetic elements involved (Deveau et al., 2008; Horvath et al., 2008). A study in 2010 showed this CRISPR-Cas system can also vaccinate cells against plasmid uptake by adaptive spacer acquisition (Garneau et al., 2010). Noteworthy, spacers were acquired against antibiotic resistance genes, which could prevent the uptake of any plasmid bearing complementary sequences.

Several studies of model CRISPR-Cas systems have also shown adaptive spacer uptake in E. coli type I-E systems (Datsenko et al., 2012; Swarts et al., 2012; Yosef et al., 2012) from plasmid exposure, as well as in a type III system from Sulfolobussolfataricus (Erdmann and Garrett, 2012). Interestingly, in these studies, a phenomenon described as ‘‘priming’’ showed that the first immunization events influence subsequent acquisition events, again supporting the adaptive nature of these processes.

Taxonomic group	Genomes analyzed	Genomes containing *cas1*	Proportion of genomes containing *cas1*	Genomes containing a type I system (*cas7* and *cas3*)	Genomes containing a type II system (*cas9*)	Genomes containing a type III system (*cas10*)
Archaea						
Crenarchaeota	17	15	0.88	15	0	16
Euryarchaeota	47	37	0.79	33	0	23
All Archaea	67	54	0.81	50	0	40
Bacteria						
Actinobacteria	72	26	0.36	28	15	8
Aquificae	7	5	0.71	7	1	4
Bacteroidetes–Chlorobi group	32	16	0.50	14	2	6
Chlamydiae–Verrucomicrobia group	10	2	0.20	0	1	1
Chloroflexi	10	9	0.90	9	2	7
Cyanobacteria	14	7	0.50	7	1	7
Firmicutes	126	56	0.44	40	17	23
Proteobacteria	318	107	0.34	117	20	22
Spirochaetes	13	3	0.23	2	1	0
Thermotogae	11	10	0.91	10	0	9
All Bacteria	639	256	0.40	245	65	99

Fig 5: Taxonomic distribution of three CRISPR-Cas system types

The sequence in the exogenous nucleic acid element corresponding to a CRISPR spacer has been defined as a protospacer (Deveau et al., 2008), which is flanked by a system-specific, highly conserved CRISPR motif, subsequently renamed proto- spacer adjacent motif (PAM) (Mojica et al., 2009) (Figure 2). PAMs are typically 2–5 nt highly conserved sequence motifs immediately flanking one side of the protospacer (within 1–4 nt of one extremity). These motifs have been identified in several type I systems (where they primarily consist of 2–3 nt motifs occurring on the 50 end of the protospacer) (Figure 2A) and many type II systems (where they range between 2–5 nt and primarily occur on the 30 end of the protospacer) (Figure 2B), while they are yet to be characterized in type III systems (Mojica et al., 2009; Sorek et al., 2013). PAMs have been implicated in both immunization (sampling of the exogenous DNA for spacer uptake) (Paez-Espino et al., 2013) and targeting (because PAM mutations preclude target cleavage) (Gasiunas et al., 2012; Jiang et al., 2013a; Jinek et al., 2012; Sapranauskas et al., 2011; Sashital et al., 2012). This is consistent with the preferential mutation of the PAM by viruses to escape CRISPR immunity (Semenovaet al., 2011; Sun et al., 2013).

During co-evolution with phages, bacteria have acquired many antiviral strategies, and new types are still being uncovered (Dyet al., 2014). Fundamental studies of these resistance systems have produced serendipitous findings that enabled powerful innovation and translational applications. R–M systems are the example of these unexpected and unpredictable rewards (Pingoud, 2014). Another resistance class providing significant commercial benefit are the abortive infection

systems, which give population-level protection through the ‘altruistic cell suicide’ of phage-infected bacteria (Dy et al., 2014). More recently, the discovery and characterization of CRISPR–Cas adaptive immune systems has again demonstrated that basic phage research often yields sophisticated tools with wide-ranging biotechnological utility. The first demonstration that CRISPR–Cas provides immunity against phages was reported in 2007 (Barrangou, 2007), and after only 5 years the possibility of exploiting these systems for genome editing was realized (Jinek, 2012)

Since 2012, there has been an explosion in the number of applications based on the CRISPR–Cas system (Jinek, 2012). These advances are centred on the type II CRISPR–Cas system, owing to the ability of one protein (Cas9) and an engineered single-guide RNA (sgRNA) to direct double-strand breaks in complementary DNA sequences. By either non-homologous end joining, or homology directed repair, spontaneous or specific mutations can be generated in diverse organisms, including phages, bacteria, fungi, plants and animals. In addition, a nuclease deficient Cas9 yields an RNA-guided protein that binds to specific DNA regions to repress gene expression or, when fused to activator domains, enhances transcription. Further developments include the generation of large sgRNA libraries, which enable genome-wide screening for gene function to assist drug target identification (Doudna&Charpentier, 2014). CRISPR–Cas9 is readily applicable to both model and non-model organisms, and is driving widespread studies of gene function and the development of modified crops and animals in the agricultural sector and for diverse biomedical applications (Doudna&Charpentier, 2014). For example, cells were engineered to target HIV, and this disrupted both latent viral genomes and protected cells from new viral infection85. There is also interest in using CRISPR–Cas9 to modify ecosystems to control pest species or eliminate disease vectors, such as mosquitoes. Indeed, a CRISPR–Cas9 gene drive was able to rapidly spread a mutant allele through *Drosophila* in laboratory experiments (Gantz et al., 2015). In bacteria, CRISPR–Cas has also been used to generate mutations in several bacterial genomes with potential use synthetic biology and metabolic pathway engineering (Selle et al., 2015) and has been explored as a novel antimicrobial strategy. For example, CRISPR–Cas can kill bacteria in a sequence specific manner to selectively eliminate particular strains, to select for less virulent survivors owing to the loss of pathogenicity islands, and to inhibit antibiotic-resistant bacteria by targeting resistance genes (Selle et al., 2015). In addition, CRISPR–Cas9 can be used to manipulate phage genomes to study phage biology with greater ease and precision (Martel &Moineau, 2014). Although promising, these developments also raise many ethical and regulatory questions (Liao et al., 2015).

In recent studies, a few stealthy phages have been discovered that can bypass this bacterial protection mechanism through mutation or deletion of CRISPR-targeted regions (Deveau, H. et al., 2008) or through acquisition of anti-CRISPR genes (Bondy-Denomy et al., 2013). Metagenomic studies have also identified CRISPR/Cas systems in viral genomes (Minot, S. et al., 2012) but their biological relevance has not been proposed. Enter Seed and colleagues, who elegantly demonstrate that phage genomes are not just ammunition and targets for CRISPR/Cas sequences, certain *V. cholera* phages have hijacked the entire system for their own defense and persistence.

*V. cholerae*is the cause of cholera, which affects hundreds of thousands of people each year (WHO, 2012). Phages are among the factors that may modulate the burden of cholera in endemic regions, so understanding the interactions between the bacteria and their infecting phages is of interest. Seed *et al*. analysed the genomes of 11 phages isolated from stool samples of patients with cholera, and found that five contained a CRISPR/ Cas system. When the authors examined the sequence of the spacers in the phages, they found that they matched regions in the genome of the host bacteria. Specifically, the spacer sequences matched an 18-kilobase 'genomic island' that is also present in several other strains of *V. cholerae*. This genomic element resembles phageinducible chromosomal islands (PICIs), which are found in some bacteria, including some *Staphylococcus aureus*strains. In *S. aureus*, these regions are known as SaPIs, and they represent pathogenicity islands that contain virulence-factor-encoding genes (Novick et al., 2010). When a SaPI-containing cell is infected by certain phages, the SaPI sequence excises from the bacterial chromosome, circularizes and replicates presumably to exit the infected bacterium. During this process, the bacterium also activates a largely uncharacterized defense system in an attempt to stop phage propagation, and thereby ensure its own persistence and the persistence of the surrounding phage-susceptible bacterial population (Ram G. R. et al., 2012). Seed *et al.* demonstrate that the 18-kb element in *V. cholerae*also circularizes after phage infection, and that it encodes an active anti-phage system. Consequently, the authors refer to it as a PICI-like element, or PLE. Further studies of one of the isolated phages, which carried a CRISPR/Cas system with two PLE-targeting spacers, showed that it could replicate and kill a PLE-harbouring*V. cholera* strain that had been isolated from the same stool sample. However, Seed and colleagues show that a mutant version of this phage that lacks the matching spacer cannot replicate in the PLE-harboring strain, but can replicate in a mutated *V. cholerae*strain lacking the PLE, further supporting the targeted action of the system. The authors also performed an elaborate set of experiments to confirm the hallmarks of an active CRISPR/Cas system. For example, they

show that crRNAs are transcribed and processed from the phages, and that derivative phages that have acquired new CRISPR spacers targeting the PLE can be isolated. Overall, these results demonstrate that phages can hijack a functional, adaptive immune-evasion system to benefit their own multiplication. And, as stated by the authors, because bacterial cell death and DNA damage are inherent in virulent phage infection, CRISPR-mediated DNA cleavage of the targeted bacterial genome does not have a negative impact on phage proliferation. Seed and colleagues' study illustrates another extraordinary turn of events in the evolution of phages and bacteria, in which the phages defeat the bacteria outright by using one of its own weapons against it. How frequently such an event occurs and whether a phage that contains a CRISPR/Cas system stays stable remains to be seen. Nevertheless, these findings will certainly fuel selected applications. For example, the discovery of other phages with a CRISPR/Cas system that targets host genes or more phages with anti CRISPR genes8 may provide additional leverage to design an efficient cocktail of natural or engineered phages to prevent or treat bacterial contamination or infection. On the other hand, this finding suggests that biotechnological industries that rely solely on CRISPR/Cas systems to protect key bacterial strains from phage infection should be ready to go back to the drawing board. Because, as always, phages will find a way.

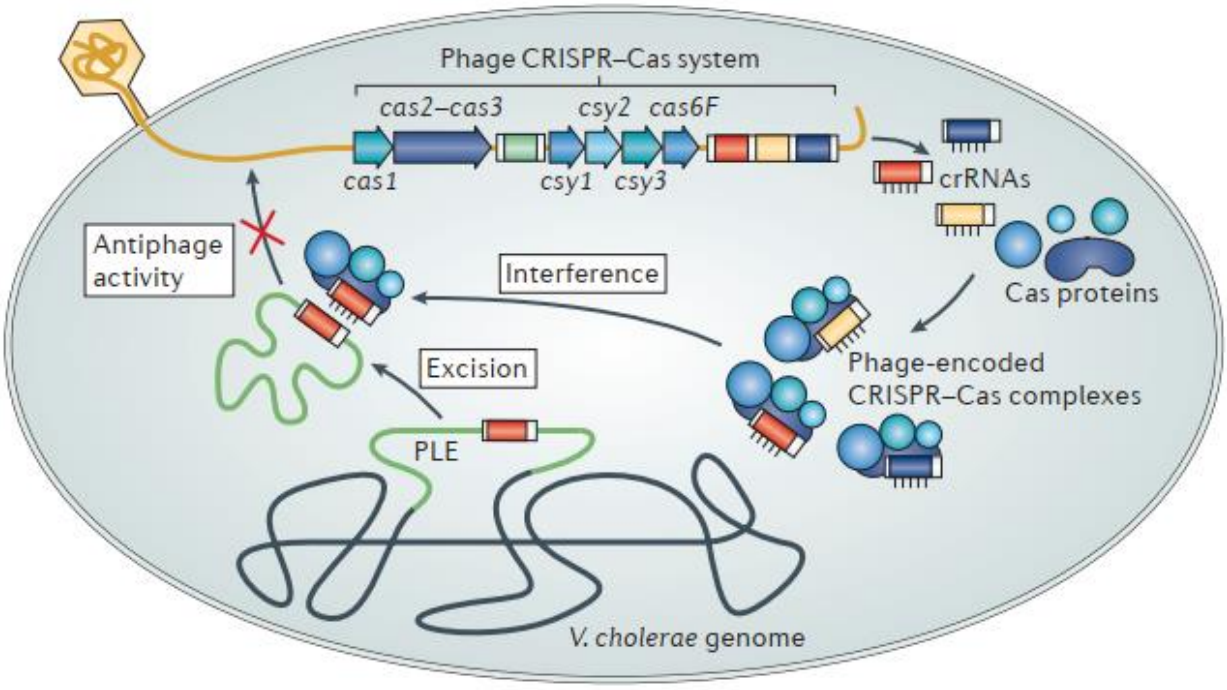


Fig 6: The CRISPR–Cas system of *Vibrio cholerae*phages

CHAPTER 2

MATERIALS AND METHODS

2. Methods and Materials:

Genomics is a multifaceted field of biology focusing on the function, structure, evolution, mapping and editing of genomes. A genome is an organism's complete set of DNA, which includes all of its genes. It targets at the collective characterization and qualification of all of an organism's genes, their interrelations and influences on the organism. Genomics also involves the sequencing and analysis of genomes by using of high throughput DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes (Culver, 2002).

In this project, correlation between bacterial genome and spacers have been investigated through some computational approaches such as genome data downloading, finding the spacers, annotating the whole genome bacterial sequences etc. by using NCBI database, Minced respectively.

2.1Bacterial genome data download from NCBI:

Bacterial genome sequencing which was started by an approach made on genome analysis through sequencing and assembly of unselected pieces of DNA to get the complete nucleotide sequence of the genome from the whole chromosome in the year of 1995 which led to a promising breakthrough in the area of microbiology and infectious disease research.

NCBI, The National Centre for Biotechnology Information advances science and health by providing access to biomedical and genomic information. NCBI has a multi-disciplinary research group that consists of computer scientists, molecular biologists, mathematicians, biochemists, research physicians concentrating on basic and applied research in computational molecular biology. NCBI assumed responsibility for the GenBank DNA sequence database in October 1992, the stuff of NCBI who have the advanced training in molecular biology build the database from various sequences submitted by different individual laboratories and by exchanging data with European Molecular Biology Laboratory (EMBL), the DNA Database of Japan (DDBJ) and the international nucleotide sequence databases. NCBI has the arrangements with the U.S. Patent and Trademark Office that enable the integration of patented sequence data.

It also supports and distributes a multifarious of databases for the medical and scientific communities, these databases include the Molecular Modeling Database (MMDB) of 3D protein structures, the Online Mendelian Inheritance in Man (OMIM), a Gene Map of the Human

Genome, the Taxonomy Browser, and the Cancer Genome Anatomy Project (CGAP), in collaboration with the National Cancer Institute.

From NCBI, a number of 12000 bacterial genome sequences have been downloaded.

2.2 Determining Spacers number using MinCED:

Clustered Regularly Interspaced Palindromic Repeats (CRISPRs) are a novel type of direct repeat found in a wide range of bacteria and archaea. CRISPRs work by defending their hosts against invading extrachromomal elements such as viruses. The CRISPR arrays are identified using minCED (mining CRISPRs in environmental data sets), a derivative of CRISPR Recognition Tool that is more conservative in repeat calling and allows more flexible user outputs (Bland, 2007). It is a program to find Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in full genomes or environmental datasets such as assembled contigs from metagenome. Custom code determines the orientation of the repeats, generates the consensus repeat sequences, and returns the number of repeats, indicating the size of the array. After the identification of CRISPR loci, the types and subtypes are assigned by using the presence or absence of genes, detect multiple systems in a genome, and by identifying the missing repeats and cas proteins it determines the completeness of the system.

2.3 BLAST:

Basic Local Alignment Search Tool, an algorithm for finding regions of similarity between biological sequences through comparing nucleotide or protein sequence databases and calculates the statistical significance, one of the most widely used bioinformatics programs for sequence searching. BLAST uses the heuristic algorithm which is much faster than other approaches for example: calculating an optimal alignment. It is more time-efficient than FASTA by searching only for the more significant patterns in the sequences, yet with comparative sensitivity. BLAST determines which bacterial species have a protein that is lineage to a certain protein with known amino-acid sequence. It also helps to find out genes which encode proteins that exhibit structures or motifs such as ones that have never been identified. BLAST takes FASTA or Genbank format sequences and weight matrix as input.

The output of BLAST comes in a variety of formats such as, HTML, plain text, XML formatting etc. For NCBI's web-page, the default format for output is HTML. When performing a BLAST on NCBI, the results are given in a graphical format showing the hits found, a table showing

sequence identifiers for the hits with scoring related data, as well as alignments for the sequence of interest and the hits received with corresponding BLAST scores for these

In Unix, the BLAST result contains various columns such as: query id, subject id, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, e-value, bit score, sequence.

2.4 FastANI:

FastANI is developed for fast alignment-free computation of whole-genome Average Nucleotide Identity (ANI). ANI is defined as mean nucleotide identity of orthologous gene pairs shared between two microbial genomes. FastANI supports pairwise comparison of both complete and draft genome assemblies. Its underlying procedure follows a similar workflow as described by Goris et al. 2007. However, it avoids expensive sequence alignments and uses Mashmap as its MinHash based sequence mapping engine to compute the orthologous mappings and alignment identity estimates. Based on experiments with complete and draft genomes, its accuracy is on par with BLAST-based ANI solver and it achieves two to three orders of magnitude speedup. Therefore, it is useful for pairwise ANI computation of large number of genome pairs.

It performs fast, alignment-free computation of whole-genome Average Nucleotide Identity (ANI). ANI is defined as mean nucleotide identity of orthologous gene pairs shared between two microbial genomes. A common use of computing ANI values for assemblies would be to compare a binned assembly from a metagenome against a reference isolate to check if it is the same species as the reference isolate (>=95% ANI). In KBase, FastANI supports pairwise comparison of both complete genomes and draft genome assemblies as input. The App generates a report with the ANI values, with a PDF containing a visualization of the pairwise comparison.

2.5 Synteny:

Synteny provides a framework in which conservation of homologous genes and gene order is identified between genomes of different species. The availability of human and mouse genomes paved the way for algorithm development in large-scale synteny mapping, which eventually became an integral part of comparative genomics. Synteny analysis is regularly performed on assembled sequences that are fragmented. Synteny analysis is a practical way to investigate the evolution of genome structure. Visualization of synteny linkages was made by R (v3.3.1)

CHAPTER 3

RESULTS AND DISCUSSION

3. Results and Discussion:

3.1 Genome Characteristics and Phylogeny:

The 26 CRISPR Bacteriophages genomes were analyzed from isolates that were collected from NCBI. The maximum likelihood phylogenetic analysis grouped both of the whole-genome and core-genome alignments into several general clusters. The vibrio ones were seen in one cluster. This cluster also contains ten of the twenty six total CRISPR-Bacteriophage in the dataset. The next cluster contains the rest sixteen CRISPR Bacteriophages, which are unique and individual amongst each other, so they grouped in another cluster.

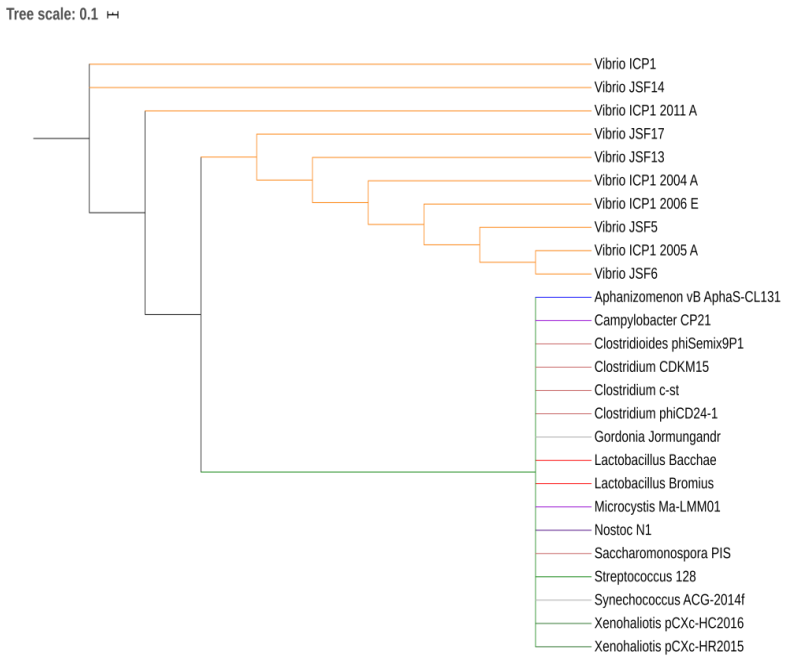


Fig 7: Phylogenetic tree of the analyzed sequences

Dendrogram representing phylogenomic relatedness based on nucleotide similarity of whole genomes of 26 CRISPR Bacteriophages. Maximum likelihood phylogenetic trees (unrooted) were constructed based on a multiple whole-genome alignment of all 26 random CRISPR Bacterio phage sequences. The blue lines connect identical leaves between trees to indicate relative phylogeny. The scale bar measures nucleotide substitutions per base pair. Distinct phylogenetic clusters are shaded grey. Vibrio bacteriophages 10 Overall, the topologies between the whole-genome and the core-genome trees were highly similar, with only minor exceptions. Both shared identical clustering and had almost no leaf-level differences within them.

3.2 Genome Alignment Visualization:

The consensus genomic sequence that was constructed from the whole-genome alignment was 125,128 bp long and contained all of the coding and non-coding regions from each genome. It was used as a reference to map the genomes and to visualize the overall multiple-genome alignment especially amongst the vibrio bacteriophages by using BRIG. Variable regions of insertions and deletions were visible as gaps in the circular alignment. Similar to the whole-genome phylogeny, there was not a clear progression of sequence divergence based on isolation chronology. No large regions of GC content difference were observed.

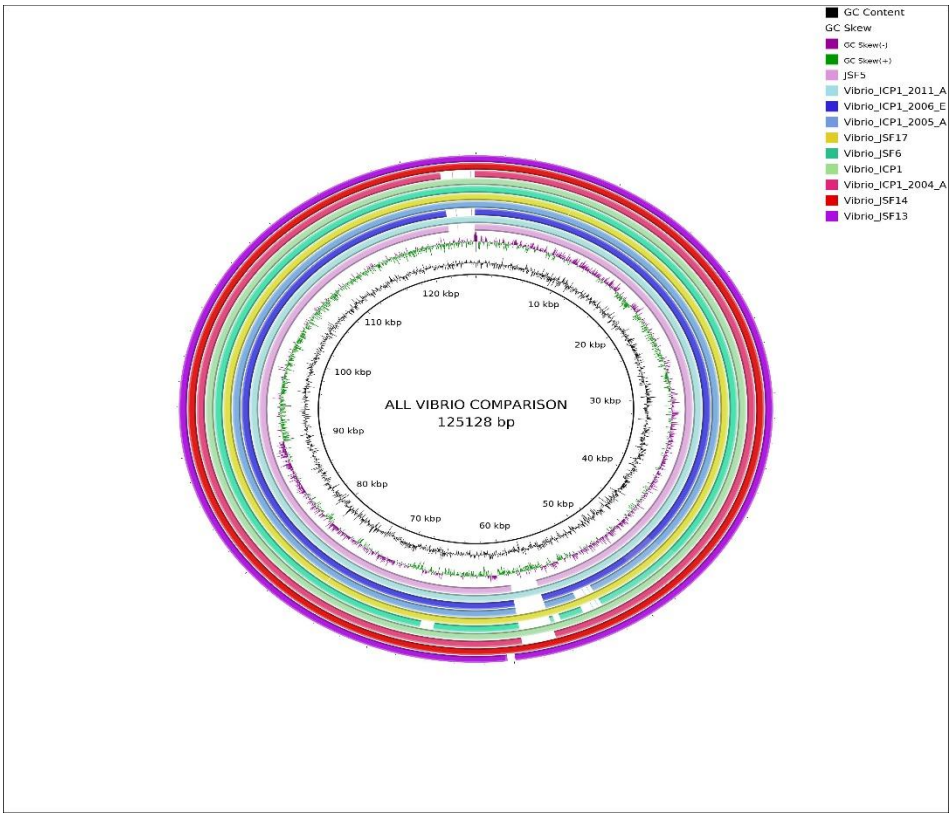


Figure 8: ICP1 pan-genome consensus alignment. A BLASTn-based whole genome alignment of all 10 vibrio bacterio phage genomes using the MAUVE alignment consensus sequence as a reference. The innermost ring is the consensus sequence. The next ring represents the GC content for that region. The following 10 rings display the alignment for each genome and are colored by phylogenetic.

3.3 FastANI:

Fast Alignmentfree computation of whole genome Average Nucleotide Identity (ANI) is done as mean identity of orthologous gene pairs shared between two bacteriophages genomes. Here FastANI was done to check average nucleotide identity amongst 26 bacteriophages randomly. Few of their results are shown in the table:

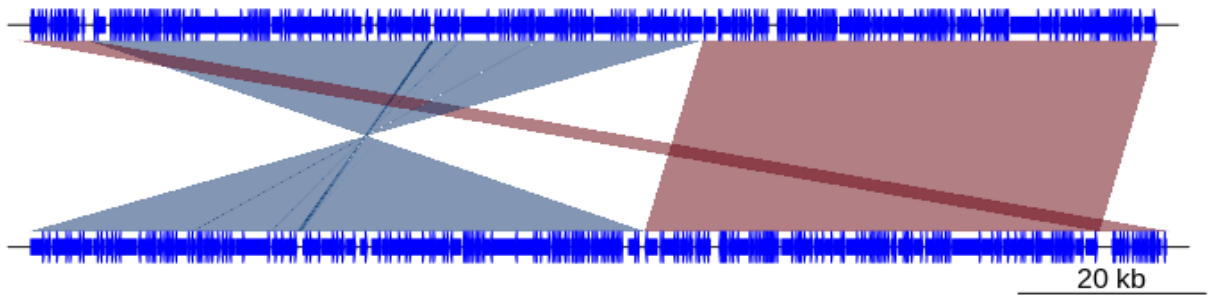
Query Sequence	Reference Sequence	Total	Number of Matches	ANI Estimate (in 100%)
Campylobacter_CP21	Clostridium_c-st	365	43	73.5732
Clostridioides_phiSemix9P1	Clostridium_c-st	113	22	73.5944
Clostridium_CDKM15	Clostridium_c-st	101	20	73.801
Lactobacillus_Bacchae	Lactobacillus_Bromius	282	248	94.5552
Vibrio_ICP1_2004	Vibrio_JSF17	256	241	99.6055
Vibrio_ICP1_2005	Vibrio_JSF5	258	252	99.6589
Vibrio_ICP1_2011_A	Vibrio_JSF17	253	248	99.7935
Vibrio_ICP1	Vibrio_JSF5	250	237	99.3413
Vibrio_JSF6	Vibrio_JSF17	267	241	98.7281
Xenohaliotis_pCXc-HR2015	Xenohaliotis_pCXc-HC2016	71	71	99.9777

Table 1: FastANI Output

3.4 Synteny:

It was done to see to see whether genes reside on the same chromosome on different bacteriopages. Most of the bacteriophages has shown their uniqueness hence less orthologous genes were visible. But in vibrio phages they shown synteny in their orthologus genes and it was identified in which area of regions they match particularly.

Vibrio ICP1



Vibrio JSF17

Fig 9: Synteny between Vibrio ICP1 and Vibrio JSF 17, The occurrence of synteny of orthologous genes in two vibrio bacteriophages are highly conservative. Block linkages, the

orthologous ones in the same orientation are labeled in red, while those in inverted orientation are labeled in blue.

3.5 Diversified Spacers:

Diversity of spacers in the CRISPR arrays carried by the bacterio phages, specially in the vibrio ones. However, there were diferences in the sequences, as well as in number of spacers in the CRISPR array carried by various phages analyzed in the present study. Each of the phages designated JSF5 and JSF6 carried a total of 8 spacers. Although the CRISPR arrays in JSF5 and JSF6 phages were identical, their entire CRISPR-Cas regions were not identical. 2 spacers of JSF14 were identical to spacers of ICP1 phages6 . JSF17 was found to carry a total of 11 spacers and 4 of these were identical to spacers of ICP1 phages. So, even if they are identical, their spacers are not identical in all the cases, which actually determines their way of working in a different way.

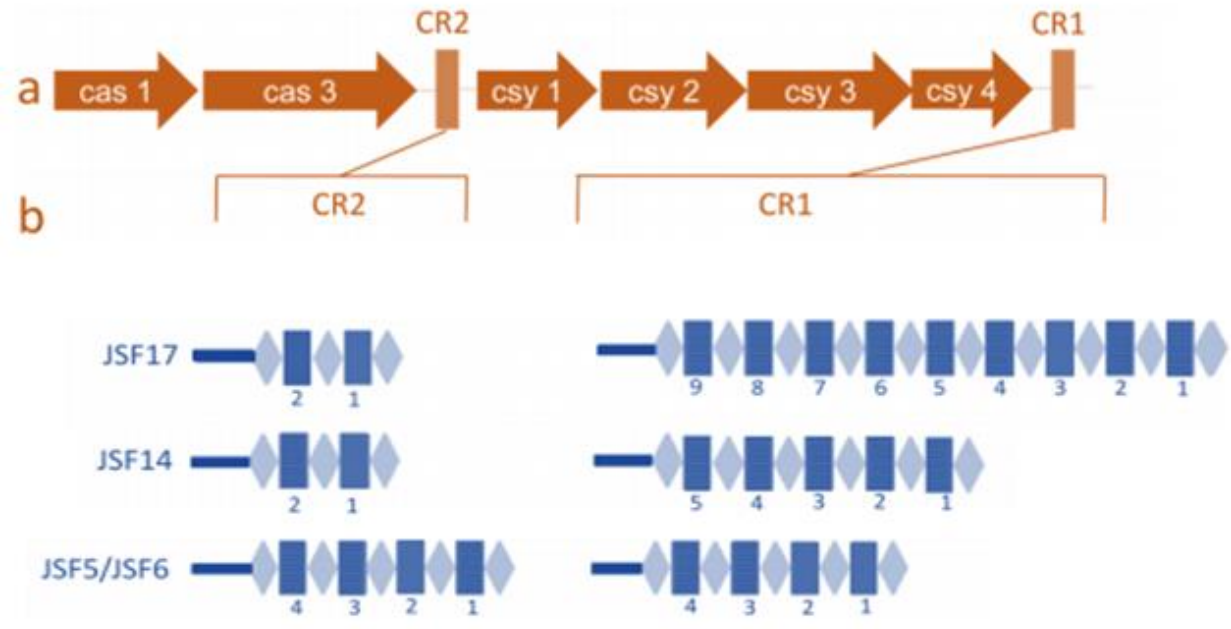


Figure 10: Structure of CRISPR-Cas system carried by phages JSF5, JSF6, JSF14, JSF17, and ICP1_2005_A showing the diversity and arrangement of the spacers. Spacers are marked with vertical, horizontal or no lines to show their 100% identity. The repeats are shown as the light blue diamonds and the spacers are shown as the deep blue diamonds respectively.

Discussion:

In the present study we conducted genomic analysis of representative 26 CRISPR bacterio phages and *V*ibrio collected from NCBI. (Hypothetical).The emerging diversity of the CRISPR-Cas system carried by the phages, in view of recurring and extensive phage-bacterial interactions. The CRISPR-Cas system has been described as a microbial adaptive immune system in that the system extends its range of targets by continually acquiring new spacers matching protospacer regions in the invading nucleic acids diversification of the CRISPR arrays in terms of number and variety of their spacers.

The results further showed that the number of spacers matching the neocleotide sequences of the spacers in a CRISPR array carried by the phages should be identical to a region. While most of these ORFs encode hypothetical proteins of unknown function, we propose that one or more of these gene products may influence the process that restricts phage replication. *V*ibrio specific phages encoding their own functional CRISPR-Cas system to neutralize a bacterial defense mechanism against phages has been discovered recently6, but the origin of the CRISPR-Cas system carried by the phages remains unknown. Mechanisms such as genome rearrangements, and genomic exchange with other viral or microbial genomes to acquire new traits allow phages to evolve rapidly, facilitated by their genomic plasticity.

However, monitoring of the CRISPR-Cas arrays in phages allows to understand the genetic variability and phage-bacterial co evolution. This knowledge may be useful in designing engineered phages targeting various regions of the bacterial anti-phage genomic determinants, in potential phage therapy or environmental interventions to control cholera. In summary, we have demonstrated the emerging diversity of the CRISPR-Cas system in phages by acquisition of new spacers to expand their ability.We performed a comprehensive phylogenetic analysis on all available, well-sequenced ICP1 isolates to elucidate their genetic divergence over time, and to provide a platform on which to develop future ICP1-related bioinformatic analyses.

We have found that the genomes of ICP1 were surprisingly well-conserved between all isolates of vibrio phages over the twelve-year period in which they were isolated. This is demonstrated in the whole-genome phylogeny which, while resolvable into distinct phylogenetic clusters, still only represents a maximum variation of approximately one nucleotide substitution per 100 base pairs between the most divergent isolates many of which are likely silent or non-coding. A high degree of genomic conservation is also indicated by the relatively large core-genome that is shared between all isolates. This conservation is not only surprising due to the amount of time that the core-genome remained stable but is also surprising.

However, what may be the most telling is that 80% of all them do not possess homology to any conserved domain, indicating that there is a great deal left to be learn about the interaction between them. However, we were conservative in our methods and are confident that a very large proportion of the calls are accurate. It will also be important to continue to track their phylogeny and genetic composition. This will help us to develop a better understanding of ICP1's co-evolution with the phages and will assist us in disentangling the complex molecular and ecological interactions that may play an important role in future genetic engineering sector one day.

CHAPTER 4

CONCLUSION

Future direction and Conclusion:

Through this study, we have attempted to highlight the importance of phage in understanding the coevolution between bacteria and phages. Even our comprehension of bacterial resistance mechanisms and phage infectivity is continually improving. The relatively recent discovery of CRISPRs as a defense against phage (Barrangou et al., 2007; Andersson&Banfield, 2008) has already been expanded to include anti-CRISPR counter-measures by phages (Bondy-Denomy et al., 2012) and to show that phages can also carry a CRISPR-cas system to target a chromosomal island of the bacterial host (Seed et al., 2013). Furthermore, the importance of bacterial suicide upon phage infection has recently been demonstrated in E. coli and was found to be a low cost strategy for reducing the population-wide impact of phage (Refardt et al., 2013) that is favored in spatially structured environments (Berngruber et al., 2013). However, the coevolutionary implications of these new mechanisms have not yet been explored, and this avenue is ripe for empirical testing using an experimental coevolution approach and for examination of natural patterns in the field.

Another key advance of the field will be incorporation of both theoretical and empirical examination of coevolution between bacteria and temperate, as well as filamentous phages. There are a number of reasons to expect the coevolutionary process to differ for these interactions relative to those with lytic phage. Primarily, many of these phages confer a strong fitness benefit to their hosts and thus will act more as mutualists than parasites. This can shift dynamics from parasite-mediated negative frequency-dependent selection (where hosts are constantly evolving to defend themselves against the common parasite) to positive frequency-dependent selection, where for example, carrying the lysogenic phage confers resistance to the same phage in the lytic form and therefore the benefit of being a lysogen increases with the frequency of other lysogens. Similarly, filamentous phages can increase the fitness of their hosts through toxin production and increased pathogenicity, as has been found for V. cholerae, the causative agent of cholera (Waldor&Mekalanos, 1996). Both filamentous and temperate phage systems have proven amenable to in vitro experimentation, but, to our knowledge, have not been used to test for coevolution. One-sided experimental evolution of the filamentous phage f1 demonstrated increased virulence (in terms of larger impact on population density) when horizontal transmission among hosts was increased relative to vertical transmission within a dividing bacterial lineage (Messenger et al., 1999), but it remains to be determined whether the bacterial population would respond by evolving increased resistance under these same conditions. One-sided experimental evolution of the lysogenic phage k was also used to select for altered

sensitivity and threshold for the switch from lysogenic to lytic phage life cycle (Refardt& Rainey, 2010).

Finally, further exploration of the similarities and differences between bacteriophages and other viruses will both help inform the utility of in vitro coevolution studies as a basis for building predictions for other virus–cell interactions, and uncover any unique adaptations of phages to their bacterial hosts. For example, examination of the archaeon, Sulfolobusislandicus, and its associated viral parasites isolated from hot springs suggests a clear biogeographic structure, such that viral genomes were found to be specifically associated with each local host population (Held & Whitaker, 2009). This system has led the way in uncovering the parallel role of CRISPR systems in archaea–virus interactions and reinforces evidence from bacteria–phage systems that demonstrate a role for viruses in maintaining host diversity (Held et al., 2010). In addition, an examination of temporal dynamics of archaea–virus interactions in a hypersaline lake suggests ample change over the course of both months and years, indicating similar timescales and mechanisms for these interactions as observed with bacteria–phage systems (Emerson et al., 2013). The other similarities between archaea–virus and bacteria–phage interaction have been reviewed elsewhere (Snyder & Young, 2011). Finally, it remains to be seen whether our increasing understanding of bacteria–phage coevolution will prove useful in studies of eukaryote-virus interactions, but at the least, each body of work could help shape the questions addressed in and techniques utilized by the other (Brockhurst et al., 2007a; Sharp & Simmonds,2011).

CHAPTER 5

REFERENCES

Abedon ST., (2008). Phages, Ecology, Evolution. In: Abedon ST, editor. Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses. Cambridge University Press; 2008. pp. 1–28

Ackermann HW, Prangishvili D., (2012). Prokaryote viruses studied by electron microscopy. Arch Virol. 2012 Oct;157(10):1843-9.

Ackermann HW., (1987). Bacteriophage taxonomy in 1987. PMID: 3153614.

Ackermann HW. (1996)., Frequency of morphological phage descriptions in 1995. ArchVirol. 1996;141(2):209-18. DOI:10.1007/bf01718394.

Anilionis, A., Ostapchuk, P., and Riley, M. (1980) Identification of a second cryptic lambdoid prophage locus in the E. coli K12 chromosome. Mol Gen Genet 180:479–481

Bacteriophages: characteristics and 1st stages of a classification. PatholBiol (Paris). 1969 Nov;17(21):1003-24.

Banks, D.J., Beres, S.B., and Musser, J.M. (2002) The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol 10: 515–521.

Barrangou R (2015). The roles of CRISPR-Cas systems in adaptive immunity and beyond. CurrOpinImmunol. 2015 Feb; 32:36-41. doi: 10.1016/j.coi.2014.12.008 PMID: 25574773.

Barrangou R1, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007 Mar 23;315(5819):1709-12. DOI: 10.1126/science.1138140 PMID: 17379808.

Bertani,E.,andSix,E.(1988)TheP2-like phages and their parasite P4.In The Bacteriophages, Vol. 2. Calendar, R. (ed.). New York: Plenum Press, pp.73–143.

Bland C, Ramsey TL, Sabree F, et al. (2007). CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics. 2007 Jun 18;8:209. DOI: 10.1186/1471-2105-8-209 PMCID: PMC1924867.

CampbellA, SchneiderS, SongB.(1992).Lambdoid phages as elements of bacterial genomes (integrase/phage21/Escherichia coli K-12/icd gene) Genetica.1992;86:259–267

Campbell A. (1963). Episomes. Adv. Genet. 1962;11:101–145

Campbell, A. (1994) Comparative molecular biology of lambdoid phages. Annu Rev Microbiol 48: 193–222

Casjens, S., and Hendrix, R. (2003) Bacteriophage roles in bacterial chromosome evolution. In The Bacterial Chromosome. Higgins, P. (ed.). Washington, DC: American Society for Microbiology Press, (inpress).

Casjens, S., Van Vugt, R., Tilly, K., Rosa, P.A., and Stevenson, B. (1997). Homology throughout the multiple 32- kilobase circular plasmids present in Lyme disease spirochetes. J Bacteriol 179:217–227.

Chang, C.C., Gilsdorf, J.R., DiRita, V.J., and Marrs, C.F. (2000) Identification and genetic characterization of Haemophilusinfluenzae genetic island 1. Infect Immun 68: 2630–2637

Culver KW, Labow MA (8 November 2002). Genomics. In Robinson R (ed.). Genetics. Macmillan Science Library. Macmillan Reference USA. ISBN 978-0-02-865606-9.

d'HerelleF.(1917).Surun microbe invisible antagonist edes bacillus dysentériques.CRAcad Sci Paris.1917;165:173–5.

d'HerelleF.(1931).Bacteriophage as a treatment in acute medical and surgical infections. Bull N Y Acad Med.1931;7:329–48

Deghorain M. and Melderen L. (2012). The Staphylococci Phages Family: An Overview. Viruses. 2012 Dec; 4(12): 3316–3335.

Faubladier, M., and Bouche, J.P. (1994) Division inhibition gene dicF of Escherichia coli reveals a widespread group of prophage sequences in bacterial genomes. J Bacterial 176: 1150–1156

FauquetCM1,PringleCR.(2000).Abbreviations for bacterial and fungal virus species names. Arch Virol. 2000;145(1):197-203. DOI:10.1007/s007050050017.

FoutsD.(2006).Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res.2006;34:5839–5851

Fuhrman JA, Noble RT. (1998). Viruses and protists cause similar bacterial mortality seawater. LimnolOceanogr. 1995;40:1236–1242

Gratia, J.P. (1989) Products of defective lysogeny in Serratiamarcescens SMG 38 and their activity against Escherichia coli and other Enterobacteria. J Gen Microbiol 135: 25–35

Greener, A., and Hill, C.W. (1980) Identification of a novel genetic element in Escherichia coli K- 12. J Bacteriol 144: 312–321

Hankin EH. (1896). L'action bactericide des eaux de la Jumna et du Gangesur le vibrion du cholera. Ann Inst Pasteur (Paris) 1896;10:511–23.

Holmes, R. K. and Jobling, M. G. (1996). Genome organization. In S. Barron (Ed.), Medical microbiology (4th ed.). Galveston, TX: University of Texas Medical Branch at Galveston.

Humphrey, S.B., Stanton, T.B., Jensen, N.S., and Zuerner, R.L. (1997) Purification and characterization of VSH- 1, a generalized transducing bacteriophage of Serpulinahyodysenteriae. J Bacteriol 179: 323–329.

JedA.Fuhrman(1999).Marinevirusesandtheirbiogeochemicalandecologicaleffects.Nature 399, 541-548(1999).

Kaiser, K., and Murray, N.E. (1979) Physical characterisation of the 'Racprophage' in E. coli K12. Mol Gen Genet 175:159–174.

Kaiser, K., and Murray, N.E. (1979) Physical characterisation of the 'Racprophage' in E. coli K12. Mol Gen Genet 175:159–174.

Krogh, S., O'Reilly, M., Nolan, N., and Devine, K.M. (1996) The phage- like element PBSX and part of the skin element, which are resident at different locations on the Bacillus subtilis chromosome, are highly homologous. Microbiology 142: 2031–2040.

Lackie, J. M., ed. (2007), The Dictionary of Cell & Molecular Biology (4th ed.), Burlington, MA: Academic Press, p. 394

Landy A, Ross W. (2007). Viral integration and excision: structure of the lambda att sites. Science. 1977;197:1147–1160.

Lang, A.S., and Beatty, J.T. (2001). The gene transfer agent of Rhodobactercapsulatus and 'constitutive transduction' in prokaryotes. Arch Microbiol 175: 241–249

Lang, A.S., Beatty, J.T., LeBlanc, H., Towers, G., Harris, J., Lang, G., et al. (2000) Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of Rhodobactercapsulatus. ProcNatlAcadSci USA 97: 859–864

Lindsey,D.F.,Mullin,D.A.,andWalker,J.R.(1989)Characterization of the cryptic lambdoid prophage DLP12 of Escherichia coli and overlap of the DLP12 integrase gene with the tRNA gene argU. J Bacteriol 171:6197–6205.

Little JW. (2005). Loysogeny, Prophage Induction, and Lysogenic Conversion. In: Waldor MK, Friedman DI, Adhya S, editors. Phages Their Role in Bacterial Pathogenesis and Biotechnology. Washington DC: ASM Press; 2005. pp. 37–54

Los M, Kuzio J, McConnell MR, Kropinski AM, Wegrzyn G, Christie GE. Lysogenic conversion in bacteria. In: Sabour PM, Griffiths MW, editors. Bacteriophages in the Control of Food- and Waterborne Pathogens. Washington, DC: ASM Press; 2010

Masui, S., Kamoda, S., Sasaki, T., and Ishikawa, H. (2000) Distribution and evolution of bacteriophage WO in Wolbachia, the endosymbiont causing sexual alterations in arthropods.JMolEvol 51:491–497

Mizuno, M., Masuda, S., Takemaru, K., Hosono, S., Sato, T., Takeuchi, M., et al. (1996) Systematic sequencing of the 283 kb 210 degrees- 232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes. Microbiology 142:3103–3111.

Nakayama, K., Takashima, K., Ishihara, H., Shinomiya, T., Kageyama, M., Kanaya, S., et al. (2000) The R- type pyocin of Pseudomonas aeruginosa is related to P2 phage, and the F- type is related to lambda phage. MolMicrobiol 38:213–231.

National Human Genome Research Institute (8 November 2010). "A Brief Guide to Genomics". Genome.gov. Retrieved 2011-12-03.

Nguyen, A.H., Tomita, T., Hirota, M., Sato, T., and Kamio, Y. (1999) A simple purification method and morphology and component analyses for carotovoricinEr, a phage- tail- like bacteriocin from the plant pathogen Erwinia carotovoraEr. Biosci Biotechnol Biochem 63: 1360–1369.

Ojaimi, C., Brooks, C., Casjens, S., Rosa, P., Elias, A., Barbour, A., et al. (2003) Profiling temperature- induced changes in Borreliaburgdorferi gene expression using whole genome arrays. Infect Immun 71: 1689–1705.

Okamato, K., Mudd, J., Mangon, J., Huang, W.M., and Marmur, J. (1968) Properties of the defective phage of Bacillus subtilis. J MolBiol 34: 413–428.

Rakonjac, J. (2012). Filamentous bacteriophages: Biology and applications. In eLS. Chichester, UK: John Wiley and Sons.

Rapp, B., and Wall, J. (1987) Genetic transfer in De sulfovibrio desulfuricans. ProcNatlAcad Sci USA 84: 9128–9130.

Rausch H, Lehmann M., (1991).Structural analysis of the action phage phi C31 attachment site. Nucleic Acids Res.1991;19:5187–5189

Richardson EJ, Watson M., (2013). The automatic annotation of bacterial genomes. Brief Bioinform. 2013 Jan;14(1):1-12. doi: 10.1093/bib/bbs007.

Rohwer F, Edwards R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol. 2002 Aug;184(16):4529-35. DOI: 10.1128/jb.184.16.4529-4535.2002.

Rohwer F., Edwards R. (2002)., The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol. 2002 Aug;184(16):4529-35.

Rohwer,F.(2003).Global phage diversity.Cell113,141.doi:10.1016/S0092-8674(03)00276- 9

Rohwer,F.,andEdwards,R. (2002).The phage proteomic tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535. doi:10.1128/JB.184.16.4529-4535.2002

Ruzin, A., Lindsay, J., and Novick, R.P., (2001) Molecular genetics of SaPI1 – a mobile pathogenicity island in Staphylococcus aureus. MolMicrobiol 41: 365–377

Sajia Akhter,1,* Ramy K. Aziz,2,3 and Robert A. Edwards1,2,4,* (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res. 2012 Sep; 40(16): e126. doi: 10.1093/nar/gks406.

Seemann T1 (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068-9. doi: 10.1093/bioinformatics/btu153.

Shimada K, Weisberg R, Gottesman M., (1972). Prophage lambda at unusual chromosomal locations. I. Location of the secondary attachment sites and the properties of the lysogens. J. Mol. Biol. 1972;63:483–503.

Starich, T., Cordes, P., and Zissler, J., (1985) Transposon tagging to detect a latent virus in Myxococcusxanthus. Science 230:541–543.

Suttle CA., (2005). Viruses in the sea. Nature. 2005;437:356–361. doi: 10.1038/nature04160.

Takemaru, K., Mizuno, M., Sato, T., Takeuchi, M., and Kobayashi, Y., (1995) Complete nucleotide sequence of a skin element excised by DNA rearrangement during sporulation in Bacillus subtilis. Microbiology 141:323–327.

Thaler, J.O., Baghdiguian, S., and Boemare, N. (1995) Purification and characterization of xenorhabdicin, a phage tail- like bacteriocin, from the lysogenic strain F1 of Xenorhabdusnematophilus. Appl Environ Microbiol 61: 2049–2052

Weinbauer MG, Rassoulzadegan F. (2004). Are viruses driving microbial diversification and diversity Environ Microbiol. 2004;6:1–11. doi: 10.1046/j.1462-2920.2003.00539.x.

Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: The unseen majority. ProcNatlAcadSci USA. 1998;95:6578–6583

Wilhelm SW, Suttle CA. (1999). Viruses and nutrient cycles in the sea: Viruses play critical roles in the structure and function of aquatic food webs. Bioscience. 1999;49:781–788

William S. Klug, Michael R. Cummings, Charlotte A. Spencer, Michael A. Palladino (2012). Concepts of genetics (10th ed.). San Francisco: Pearson Education. ISBN 978-0-321-72412-0

Yen,H.C.,Hu,N.T.,andMarrs, B.L. (1979)Characterization of the gene transfer agent made by an overproducer mutant of Rhodopseudomonas capsulata. J MolBiol 131:157–168.

Zink, R., Loessner, M.J., and Scherer, S. (1995) Characterization of cryptic prophages (monocins) in Listeria and sequence analysis of a holin/endolysin gene. Microbiology 141:2577–2584