Exploring the possible function and acquisition pattern of self-targeting Spacers

In five common infectious bacteria

Submitted by:

Mutebi John Kenneth

Student ID: 15236014

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

Bachelor of Science in Biotechnology

Department of Mathematics and Natural Sciences

BRAC University

December 2019

Dedicated to my grandfather and to all my teachers at all levels in my Education

Declaration of Authenticity

I hereby declare that the research work embodying the results reported in this thesis entitled **"Exploring the possible function and acquisition pattern of self-targeting spacers in five common infectious bacteria"** has been written and submitted by me, Mutebi John Kenneth under the supervision of Md. Tokee Tareq, Senior Lecturer, Biotechnology Program, Department of Mathematics and Natural Sciences, BRAC University, Dhaka.

I further declare that this thesis presented here 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 except where stated otherwise by reference or acknowledgement.

Student's Full Name & Signature:

_____

Mutebi John Kenneth

ID: 15236014

Approval

This thesis titled —Exploring the possible function and acquisition pattern of self-targeting spacers in five common infectious bacteria" submitted by Mutebi John Kenneth (ID: 15236014) of Summer, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Biotechnology on 5th of December, 2019.

Examining Committee:

Supervisor:

 (Member)   **Md. Tokee Tareq**

Lecturer Biotechnology Program

Department of MNS

BRAC University

Departmental Head:

(Chairperson)   **A F M Yusuf Haider, PhD**

Professor and Chairperson, Department of MNS

BRAC University

Acknowledgement

First of all, I am thankful to Omnipotent God for my good health and wellbeing that were necessary for me not only to sail through my entire undergraduate studies but also this work at large to be done.

I am grateful to my family for always having faith in me, loving and supporting me in every aspect of my life and above all is their daily prayers. They have been a constant source of inspiration.

My sincere gratitude goes to Professor **A F M Yusuf Haider**, Chairperson of MNS Department of BRAC University for giving me the opportunity and the relevant support, without which I would not complete my undergraduate thesis.

I extend my gratitude to **Md. Tokee Tareq** Lecturer, Biotechnology Program, Department of Mathematics and Natural Sciences for not only introducing but also making me love this exciting field of Computational Biology.

My warmest gratitude goes to my close friends **Rifaquat Ahmed** and **Samiha Ashreen** for being my support and comfort during my bad times. I also appreciate my thesis mates Peya apu and Shishir vaia for their support and cooperation throughout my work.

Mutebi John Kenneth

List of Abbreviations
The following abbreviations have been used throughout the text

CRISPR   Clustered Regularly Interspaced Short Palindromic Repeats

Cas   CRISPR –associated

Bp   Base pairs

PAM   Proto-spacer Adjacent Motif

MGEs   Mobile Genetic Elements

NCBI   National Center for Biotechnology Information

MinCED   Mining CRISPR in Environmental Databases

BLAST   Basic Local Alignment Search Tool

Table of figures

Abstract

The adaptive immunity in prokaryotes known as **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (**CRISPR**) and its **C**RISPR **as**sociated (Cas) genes provide the encoding organisms with the ability to resist invasive genetic elements such as viruses, plasmids, and transposons. CRISPR-Cas system confers this immunity through incorporating nucleic acid derived spacers into CRISPR arrays which provide genetic memories that are a pre-requisite for subsequent neutralization of invading nucleic acids. The study analyzed the pattern by which self-targeting spacers are acquired among the 5 common infectious bacteria and the function of such spacers as autoimmunity. About 5% of the analyzed genomes contained self-targeting spacers which makes it a common phenomenon among CRISPR encoding bacteria. However, much of these self-targeting spacers were found matching on prophage sequences especially those bacteria that are commonly found in the environment. This not only proved the functionality of adaptive immunity provided by the CRISPR-Cas system but also showed how exposed such bacteria are to lysogenic viruses for being nearly in every environment. The acquisition of such spacers was found to be more in the proximal position to the leader sequence in the array which indicated that their detrimental nature makes then live shortly in bacterial genomes and are not inherited to the next progeny. Since 41% of self-targeting spacers were found to target bacterial proteins, bacteria must be having escape mechanisms from this autoimmunity. Of the 4 proposed mechanisms, this study focused on lack of Cas genes and it was found to be a reliable escape mechanism only in *S. aureus* strains. Autoimmunity was found to be occurring in all the 5 bacteria, therefore, redirecting it through genetic engineering to be against drug resistance genes can be a great strategy to fight the worrying increase in antibiotic resistance among these common infectious bacteria in our community

Table of Contents

Chapter – 1: Introduction

CHAPTER - 1

INTRODUCTION

Although bacteria and archaea constitute the most abundant life on earth, viruses have always been a threatening predator which would have put this fact about prokaryotes down. As a result, bacteria have developed a wide range of resistance mechanisms for viral infection. The recently discovered adaptive immune system in prokaryotes called CRISPR – Cas provides them with a sequence-specific immunity against viral attack, hence affecting entirely our understanding of virus-host bacteria interaction.

1.1 The Adaptive immunity of prokaryotes

The **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (**CRISPR**) and the **C**RISPR **as**sociated (Cas) system allows the encoding organisms to acquire and integrate short viral genome sequences of the encountered viruses and other mobile genetic elements into the cell's CRISPR locus. Such short sequences are known as spacers. They act as a form of immunization to the encoding cell and as a result, the acquiring cell develops a unique form of evolution in which it will not only resist its viral predators but also pass this resistance mechanism to its future progeny (Rollie et al., February 2018).

The mode of function of this adaptive immunity is that the acquired spacers are expressed into small RNAs called CRISPR – RNAs (crRNAs). These RNAs for a complex with Cas proteins which acts as a guide for recognizing viral genomic material through complementarity and then targeting it for degradation. The CRISPR –Cas system contains machinery of different nucleases that mediate the integration of 20 – 50 base pairs (bp) unique viral DNA or RNA fragments called spacers (or ―proto-spacers"). The invading viral proto-spacer will not be cleaved and degraded unless there is an adjacent Proto-spacer Adjacent Motif (PAM) sequence. PAM sequences are very essential in this adaptive immunity because they distinguish bacterial self from non-self DNA hence preventing the CRISPR locus from being targeted and destroyed by nucleases.

1.1.1 CRISPR Spacer acquisition

The Cas genes flanking CRISPR arrays encode proteins which play a very crucial role in the functioning of this adaptive immunity. Although these genes are so divergent and do not occur in every CRISPR locus, Cas1 and Cas2 genes are notably so much conserved across the major types of CRISPR systems. These two products are the only required Cas proteins during the acquisition of new spacers into the host cell's CRISPR- loci (Nunez et al., 2014). Mutation studies showed that removal of Cas1 or Cas2 stopped the process of acquiring new spacers despite having no effects on the functioning of the CRISPR immune system (Dugar et al., 2013).

When a bacterium is invaded by a virus (bacteriophage), the first response of its immunity is to capture unique fragments this phage's DNA (or RNA) and inserts them into the CRISPR locus to

form spacers. This process is mediated by two CRISPR associated proteins Cas1 and cas2 (Nunez et al., 2015). In the I-E system of *E. coli*, Cas1 and Cas2 form a complex in which Cas2 dimer bridges two Cas1 dimers. Cas2 performs as a non-enzymatic scaffold, binding the double-stranded fragments DNA (Nunez et al., 2015). On the other hand, Cas1 binds the single-stranded flanks of the invading DNA thus catalyzing their integration into the CRISPR array (Wang et al., 2015). The integration of new spacers occurs predominantly in the leader end of the CRISPR array (Swarts et al., 2012). The Cas1–Cas2 complex is capable of recognizing the leader repeat sequence, a property that none of these two proteins can offer alone (Nunez et al., 2014). This indicates that CRISPR loci can be relatable to molecular fossil records of past infections in that the recent memories are located at the leader end and the most ancestral spacers are located at the distal end concerning the leader sequence (Yosef et al., 2012). Therefore if the spacers are ordered chronologically, those from the recent invaders will be located in the leader end and will provide the most robust immunity as compared to those spacers in the downstream positions (McGinn et al., 2018).

More to that, the acquired spaces in CRISPR system types I and II are not uniformly sampled from viral genomes but they are rather chosen for excision purposes by Cas proteins from the fragments of short regions (3 -5bp) called Protospacer Adjacent Motifs (PAMs) that are adjacent to them. The PAM sequences are so short that it is more likely that they will be found in other parts of the host's DNA. This can lead to the acquisition of self-targeting spacers and when PAM sequences are present in CRISPR arrays, the CRISPR-Cas loci target themselves as a result (Sorek et al., 2013).
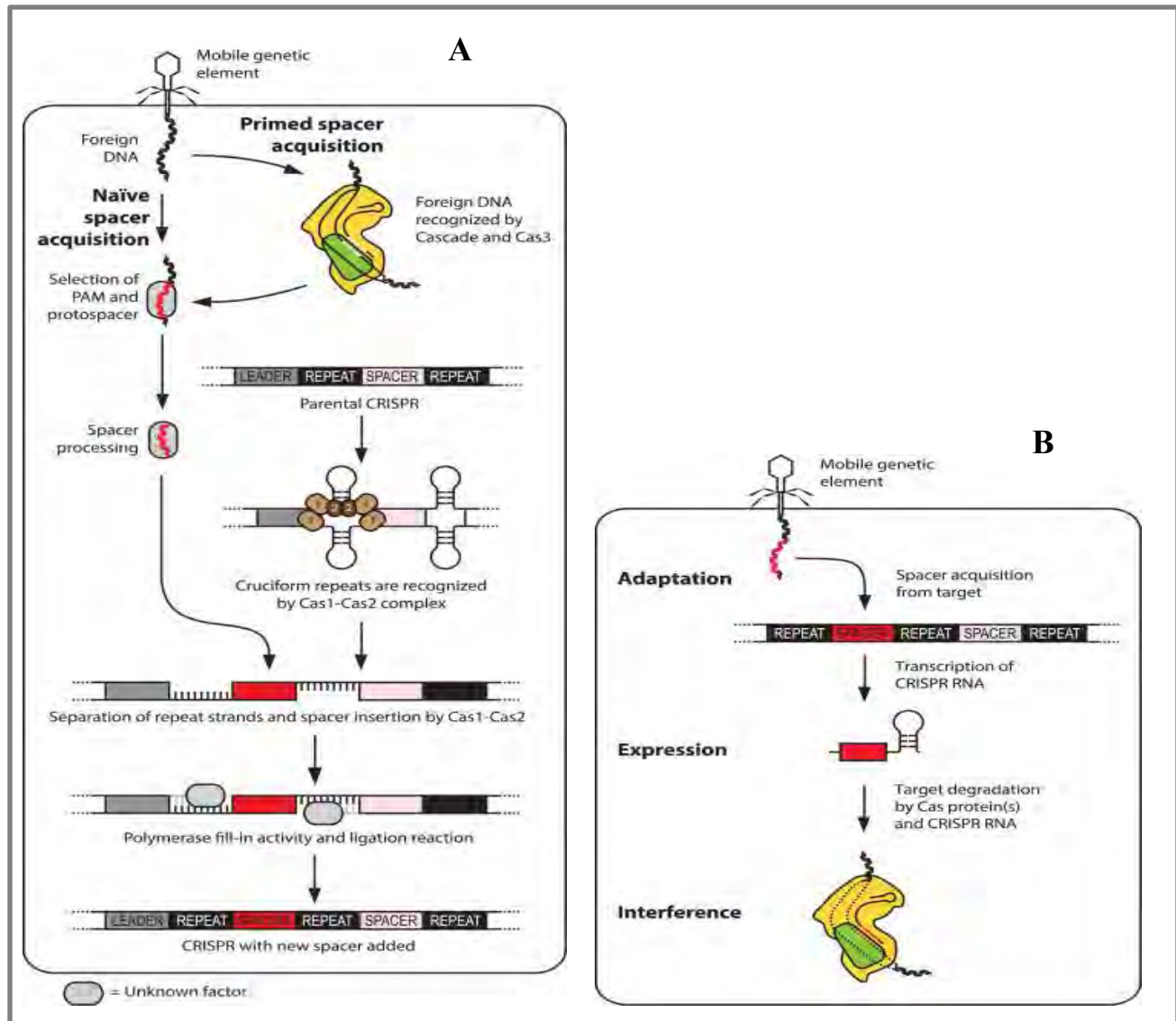
Figure 1.1 Spacer acquisition (A) and Mobile genetic element inhibition by (B) CRISPR –Cas System (Rath et al., 2015)

The adaptive immunity in bacteria has got two types spacers; self-targeting spacers which are of bacterial chromosomal origin that kill the bacteria by autoimmunity as well as the phage targeting spacers, which were acquired from the invading exogenous genetic material. In addition to matching mobile genetic elements like phages, transposons, and plasmids, spacers have been found to target sites on the bacterial or archaeal genome. An example can be drawn from *Yersinia spp* in which the majority of the analyzed 36 spacers were of bacteriophage origin and a total of 8 spacers matched sequences of the *Yersinia* chromosome. A similar pattern has been observed in a broader analysis that involved 4500 spacers from both bacteria and archaea. The analysis found out that 35% of the spacers matched sequences of the NCBI database which were derived from chromosomal DNA and are not related to phages or other MGEs whatsoever. Such unfortunate spacers are termed as self-targeting CRISPR spacers (Heussler et al, 2016).

Incorporating a self-targeting spacer into a CRISPR array can cause the death of the host cell (Shen et al., 2017) and it has been found that inserting the self-targeting spacer by a host cell is not a rare event (Heussler et al., 2016). Basing on the recent studies, 59 of the 330 (18%) CRISPR encoding organisms had at least one array which contained at least one self-targeting spacer and this was found to be in a ratio of 100: 23550 (0.4%). Moreover, more than half of the self-targeting spacer (53 of 100) were found targeting non-mobile genes, bring it to a conclusion that spacers targeting a host's genes is not a rare phenomenon (Stern et al., 2010; Shen et al., 2017)

Self-targeting spacers frequently target Mobile Genetic Elements (MGEs). This can be evidenced from a recent study where about half of all the self-targets were found to reside within elements of exogenous origin. These included provirus sequences, transposon sequences and established native plasmids (Stern et al., 2010). This exact much to the provirus sequences indicates that such an organism in question once encountered these viruses but they manage to escape the host's CRISPR degradation (Stern et al., 2010).

Acquiring a self-targeting space by a host can be as a result of an accidental CRISPR insertion mechanism (Stern et al., 2010), hence leading a deleterious effect to the host cell. A recently proposed model (Stern et al., 2010) indicates that CRISPR self-targeting spacers are as a result of leaky incorporation of self-nucleic acids into CRISPR arrays.
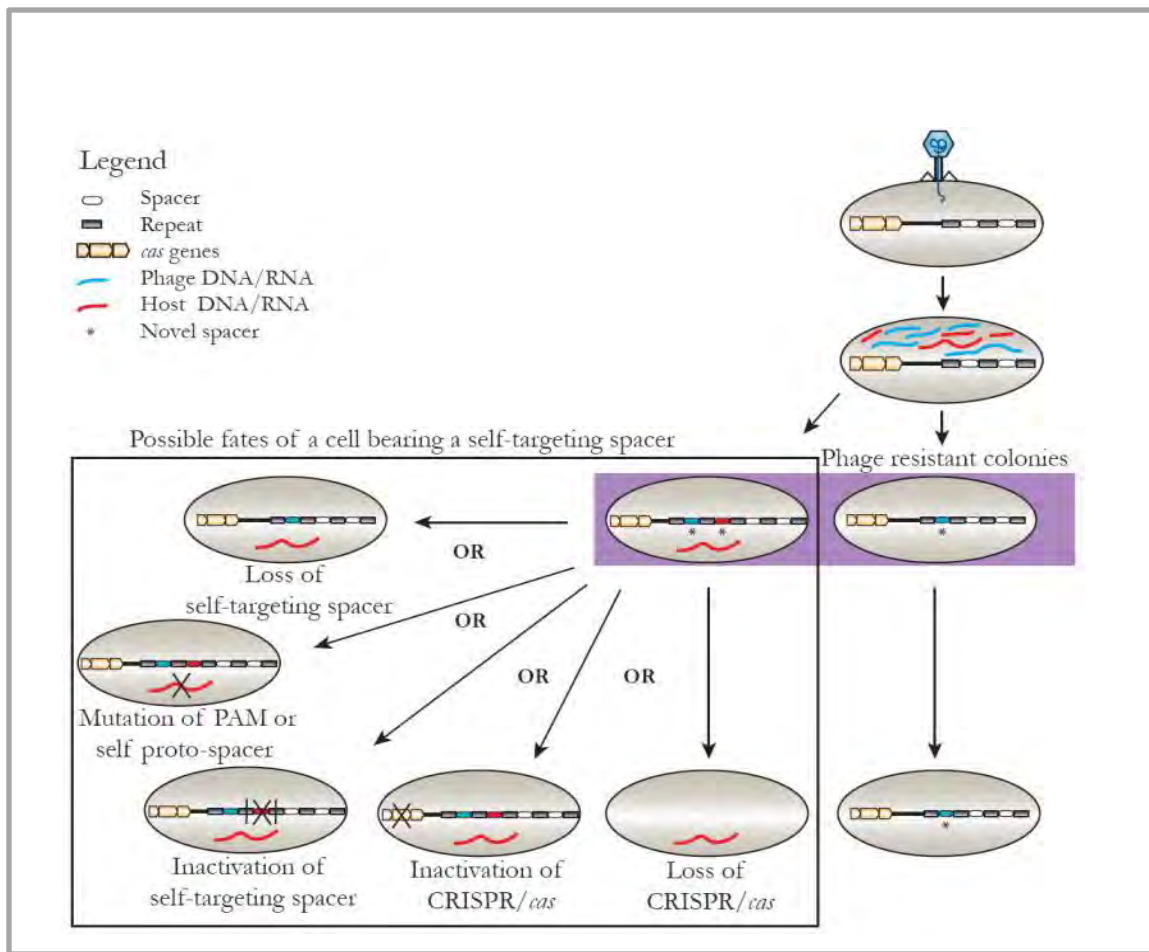
Figure 1.2 CRISPR autoimmunity and the possible outcomes to the host cell (Stern et al., 2010)

1.1.2 CRISPR Spacer auto-immunity

One of the major costs associated with the CRISPR-Cas system among encoding organisms is the self-targeting spacer related mortality (Bradde et al., 2019). Self –targeting is a form of autoimmunity whose adverse consequences lead to cell death. However, CRISPR systems have got various mechanisms by which to bias spacer acquisition to only foreign genetic elements (McGinn et al., 2018).

A single mutation on either CRISPR Cas system (which includes Cas genes, repeats, spacers and leader sequence) or on targeting sequences (which includes PAMs and Proto-spacers) can be an escape strategy to auto-immunity to a host cell. To assert whether this can be a reliable escape strategy, full matching spacers of *K. pneumoniae* were detected for their adjacent repeats, PAMs and leader sequences. It was found that mutation only occurred in the leader sequences (Shen et al., 2017). Other protective strategies against auto-immunity by CRISPR spacer include lack or inactivation of Cas genes, extended base pairing with the upstream flanking repeat and absence of PAM for some repeat clusters (Stern et al., 2010).

1.2 Campylobacter jejuni

Campylobacter jejuni is one of the most commonly identified bacterial causes of acute gastroenteritis in the world today. This gram-negative bacterium is a renowned cause of diarrheal disease as a result of food poisoning in Europeans countries and the United States of America. In the developing world, *Campylobacter* species are important causes of childhood morbidity by diarrheal illnesses. The most common route of infection by this pathogen in the developed world is the consumption and handling of chicken. For the developing, on the other hand, the route remains poor hygiene and sanitation as well as proximity to domesticated animals (Platts-Mills et al., 2015). Travelers from developed nations always suffer diarrheal infections as a result of *Campylobacter* species.

C. jejuni like most campylobacter species is resistant to cephalothin, therefore, its isolation from the stool of a suspected patient is by use of a culture medium that contains cephalothin. The post-infection complication of this pathogen is Guillain Barre Syndrome (GBS) which causes acute demyelination of the peripheral nervous system. Despite being a common trigger of GBS, the risk of developing GBS in the post-infection of *C. jejuni* is relatively smaller in the developed world (less than 1 case in 1000 C. jejuni after infections).

1.2.1 CRISPR-Cas system in C. Jejuni

C. jejuni encodes a minimal CRISPR- Cas system subtype II-C which is characterized by the presence of genes encoding different configurations of Cas1, Cas2, Cas4 and Cas9 proteins. In this subtype, these proteins are always accompanied by a trans-activated CRISPR –RNA (tracrRNA). This minimal CRISPR subtype requires the use of the host's RNase III during the expression of crRNA gene products (Sampson et al., 2013). Some strains of *C.jejuni* for example NCTC 11168 and PT14 have been reported to have subtype II –C CRISPR system which contains Cas proteins 1, 2, 9 as well as tracrRNA with no homology of Cas4 protein (Dugar et al., 2013). The CRISPR subtype II –C in *C. jejuni* lacks Cas4 proteins yet it is known that bacteriophages of campylobacter species harbor CRISPR Cas4 –like proteins (Hooton et al., 2013). The absence of Cas4 proteins hinders spacer acquisition in *C. jejuni* hence affecting the adaptive immunity against bacteriophages. To recouple a Cas4 component results in the

incorporation of self-derived genetic materials which makes it possible to modify the host's gene expression and evolution (Brathwaite et al., 2013).

The acquisition of these self-derived genetic materials increases the risk of CRISPR-mediated autoimmunity in *C. jejuni* (Stern et al., 2010) for which the majority of the cases lead to cell death (Yosef et al., 2012). It is believed that bacteriophages mobilize the acquisition of host-spacers as a decoy for preventing the phage-DNA acquisition; meaning that the expression of Cas4 proteins is an anti-CRISPR measure. Therefore, autoimmunity as a result of self-targeting spacers may be profoundly directing the shape of evolution in *C. jejuni* (Brathwaite et al., 2013).

1.3 Klebsiella pneumoniae

Klebsiella pneumonia is a gram-negative bacterium that normally lives in the human gut system and feces. It will always be harmless while kept in the gut system but on spreading to other parts of the body (especially when an individual is ill); *K. pneumoniae* can cause severe infections. The symptoms of the infections by this pathogen always depend on the location of the infection that is; lungs, bladder, liver, eyes, among others; and so does the treatment (Gerhard &Nunez, 2019). Although people with weak immunity and those on long spells of antibiotics use are at a higher risk of contracting the infections of this pathogen, individuals who work in health care centers like clinics and hospitals, given their strong immunity, are also likely to contract hospital-acquired *K. pneumoniae* infections (Gerhard & Nunez, 2019).

Despite its hyper virulence, the pathogen has higher rates of antibiotic resistance and this can be attributed to the presence of a high molecular weight plasmid in this pathogen which harbors multidrug resistance traits (Mathers et al., 2015; Lee et al., 2017).

1.3.2 CRISPR-Cas system in *K. pneumoniae*

CRISPR –Cas systems have been associated with antibiotic susceptibility in various bacterial species for example *Streptococcus pyogenes* and *E.coli* (Li et al., 2018). System-F, in particular, has the potential to interfere with the acquisition of such antibiotic-resistant plasmids, hence being a dominant system in antibiotic susceptible E.coli (Aydin et al., 2017). However, this association is unclear among *K. pneumoniae* CRISPR –Cas encoding strains.

Nevertheless, the characterized CRISPR –Cas systems in the published genome, as well as clinical isolates of *K. pneumoniae* obtained from urinary tract infection or bloodstream infection patients in Taiwan, were found to be of subtype I –E*. These strains were found to be more susceptible to certain antibiotics, for example, Gentamicin (87.2%), Cefazolin (79.5%), Ampicillin-Sulbactam (76.9%), among others (Cheng-Yen, Jing-Joi, Ming-Cheng, et al., 2018). This can be backed up by the fact that there were no subtype I –E* CRISPR –Cas system targeting spacer sequences that were found to be complementary to any of known antibiotic resistance genes. Since CRISPR subtype I –E* in *K. pneumoniae* interferes with foreign DNA acquisition to maintain antibiotic susceptibility, incorporation of chromosomal derived spacers

into CRISPR arrays is likely to be higher and so is the possibility of autoimmunity among *K. pneumoniae* strains (Diancourt et al., 2005). Since a high ratio of self-targeting spacers was found in *K. pneumoniae,* it means that there exist known or unknown mechanisms by which the strains of this pathogen compromise the function of self-targeting CRISPR spacers (Shen et al., 2017).

Studies point out a possibility of CRISPR –Cas system horizontal transfer among *K. pneumoniae* isolates. Therefore, determining the efficacy of CRISPR typing in epidemiology and outbreak investigation of *K. pneumoniae* will be interesting research in the future (Li et al., 2018).

1.4 Clostridium Difficile

Clostridium difficile, also known as *Peptoclostridium difficile* is a strictly anaerobic spore-forming gram-positive bacterium which causes diarrhea and other serious intestinal conditions such as inflammation of the colon. The pathogen contributes about half of the illnesses in the United States of America per year (Mayo Clinic, 2019). *C. difficile* bacteria are common in the environment and health care centers but they will only be infectious to individuals who are taking antibiotics. This is so because antibiotics not only wipe away bad germs but also good ones which would protect the body against infection. The effects of antibiotics may last for several months in the body and if an individual comes into contact with *C. difficile* pathogens within this time, s/he will be infected (Mayo Clinic, 2019).

1.4.2 CRISPR –Cas system in *C. difficile*

Recent studies showed that *C. difficile* strains encode an active subtype I-B CRISPR –Cas system. This subtype is characterized by an unusually high number of CRISPR arrays (Hargreaves et al., 2014).

CRISPR –Cas system in *C. difficile* can be hijacked for genome editing in this bacterium (Maikova et al., 2018). The study also demonstrates a great potential for using the CRISPR –Cas system in developing new therapeutic strategies against *C. difficile* infections through redirecting this system towards autoimmunity. For example, the hfq gene of *C.difficile* strains R20291 and 630Δerm –a chromosomal gene was chosen to be targeted by the CRISPR –Cas system to evaluate the possibility of self-targeting of this system in this pathogen. This gene plays a major role in RNA metabolism and so its depletion suggested a pleiotropic role of its product (Kuehne et al., 2011). The experiment was a success; despite earlier attempts to inactivate the same gene but using other gene knockout systems. Therefore, self-targeting CRISPR spacers may prove to be advantageous in the development of control strategies against *C. difficile* infections.

1.5 Staphylococcus Aureus

Staphylococcus Aureus are gram-positive bacteria that cause a wide range of clinical infections. Among such infections include bacteremia, endocarditis, osteomyelitis, skin infections and several hospital-acquired infections like pneumonia (Tong et al., 2015). About 30% of the population is colonized by *S.aureus* bacteria however much of this population is of carriers who do not show any symptoms of infection by the pathogen. However, novel therapeutic measures are needed for *S.aureus* infection due to increased cases of drug resistance (Rice, 2007).

1.5.2 CRISPR –Cas system in *S.Aureus*

Most strains of *S.aureus* contain only one CRISPR locus (Zhao, et al., 2018) but it is tempting to make a speculation that CRISPR may be closely related to high environmental adaptability rates in this human pathogen (Schroder, et al., 2013). According to a recent study, the number of spacer sequences in each CRISPR locus is between 1 -15, the length being between 25 -31bp whereas direct repeats being between 2 and 16. The study continues to show that less than 10 % strains of *S. aureus* contained the two most essential Cas genes; that is –Cas1 and Cas2 genes for the normal active CRISPR –Cas system (Zhao et al., 2013).

CRISPR /Cas9 –mediated genome editing system is a novel development in the CRISPR –Cas system of *S. aureus*. It accelerates the physiological studies, for example, drug resistance and drug –target exploration in these bacteria (Chen et al., 2018). A CRISPR locus targeting a drug resistance gene on the chromosome of a recipient bacterium was constructed and it was found that a recipient bacterium with a drug resistance gene died as a result (Bakird, et al., 2012). Targeting of *S. aureus* chromosome by it CRISPR type III-A can modulate deletion and alteration of resistance or virulence genes in the bacterium such as Staphylococcal Cassette Chromosome Mec (SSCmec) gene (Guan et al., 2017).

1.6 Vibrio Cholerae

Cholera is an acute watery diarrheal infection which is caused by ingesting water and food substances that were previously contaminated with a bacterium *Vibrio Cholerae*. The patient is highly dehydrated as a result, leading to his /her death if not given immediate treatment. As an indicator of social inequality and underdevelopment, cholera is commonly found in places with poor sanitation, overcrowding, wars and famine (World Health Organization).

1.6.2 CRISPR –Cas system in *V. Cholerae*

As a causative agent of a serious epidemic, *V. Cholerae* interacts with numerous phages in the environment as well as Cholera patients. The role of CRISPR –Cas system in *V. Cholerae* (with experimental evidence) as an immune system against bacteriophage invasion is yet to be documented despite the *in silico* identification of intact CRISPR –Cas system . Spacer sequences with identifiable targets tend to match sequences of *V. Cholerae* phages.

V. Cholerae does not rely entirely on the CRISPR –Cas system for phage resistance. The strains have evolved to use Phage Inducible Chromosome Island (PICI) –like Elements (PLEs) to resist the prevalent lytic phage ICP1. PLEs are mobile genetic elements that reside in the small chromosomes of *V. Cholerae* but the mechanisms by which they inhibit the phage replication is unknown (McKitterick, et al., 2019). This inhibition of phage replication promotes cell lysis to prevent the production of new phage progenies (Naser et al., 2012).

However, *V. Cholerae* specific ICP1 phages have been found to carry a CRISPR –Cas system. Most of the spacer sequences of these phages' CRISPR arrays were found to be fully matching on different regions of the 18kb PLE which is carried by *V. Cholerae* biotype O1 E1 Tor strains (Seed, et al., 2012). Therefore, the function of PLEs to interfere with the phage replication in such strains can be inactivated by the phage-encoded CRISPR –Cas system.

1.7 Objectives

With the recent advancements in CRISPR technology, studies have shown that CRISPR –Cas system can be a useful typing tool in epidemiology and outbreak investigation especially when the technique is optimized into an advanced technique called CRISPR High –Resolution Melt Analysis (Louwen et al., 2014).

This study focuses on exploring the pattern by which Self-targeting Spacers are acquired in each of the 5 bacteria. The acquisition pattern can be useful in determining how often incorporating chromosomal derived spacers occurs and its application in designing strategies against such bacteria. Exploring the function of Self-targeting Spacers can as well be a strategy against these infectious bacteria by developing means of beating the bacterial mechanisms to escape autoimmunity.

Since all the 5 bacteria are highly infectious in our population /country, autoimmunity can be a useful strategy to save the population of the scary increasing rates of drug resistance among common infectious bacteria.

Chapter 2: Materials & Methods

CHAPTER -2

MATERIALS AND METHODS

2.1 Gathering genome assemblies

A total of 22613 genomes for 5 bacteria were retrieved from NCBI's Assembly database as of 25th August 2019 in compressed form. Of these, 7558 were for *K. pneumoniae*, 10409 for *S. aureus,* 1557 for *C. jejuni*, 2146 for *C. difficile*, and 943 genomes for *V. Cholerae*. The assembly level for all the retrieved genome assemblies was ―Complete genome", with full genome representation of a given strain.

2.2 Identification of CRISPR spacers

CRISPR spacers were identified by using a command-line based CRISPR identification program called MinCED. Mining CRISPRs in Environmental Database (MinCED) is a program for finding CRISPRs in full genomes or environmental datasets for example assembled contigs from metagenomes. The program runs depending on different options (Skennerton, 2012) that is;

> **minced [option] file.fna [Output.file]**

CRISPR spacers were identified from encoding genomes by using a minced option called ―Spacers", that is;

> **minced – Spacers file.fna Out.fna.spacers**

2.3 Identification of Self –targeting spacers

Self –targeting spacers were identified by comparing the obtained spacers with their genomes of origin, using a command-based **BLAST.** BLAST (Basic Local Alignment Search Tool) is a bioinformatics tool, whose algorithm is used to compare primary biological sequence information such as amino-acid or nucleotide sequences. With this tool, a researcher compares a given sequence (query sequence) with a database of sequences to identify sequences that resemble the researcher's query sequence, depending on a certain threshold value.

Genomes sequences with spacers for each of the 5 bacteria were used to make a blast database by calling ―makeblastdb" option of BLAST, that is;

> **makeblastdb – in file.fna – dbtype nucl**

BLAST was called for search with spacer sequences being the query sequences, that is;

> **blastn – query out.fna.spacers – db (db_name) – outfmt 6**

From the output results, self-targeting spacers were identified as those which produced more than one ―hit" to its genome. This means that such spacers matched not only their original location but also other locations within the same genome.

2.4 Genome annotation

Gene location and function for each of the spacer contained genomes were identified by using a prokaryotic genome annotation pipeline called **prokka.** Prokka is a rapid prokaryotic genome annotation software tool which produces standards-compliant output files, that is;

```
prokka file.fna - - outdir My_dir - - locustag My_tag
```

2.5 Identification of Phage Sequences

Phage sequences from the spacer contained genomes were identified by a command-line based phage identification tool called **PhiSpy.** PhiSpy is a computer program initially written by Sajia Akhter, a Ph.D student at Edwards Bioinformatics Lab; to identify prophages in complete bacterial genome sequences. From genome annotation results, GenBank files (.gbk file) were converted to SEED using the following command;

```
python genbank_to_seed.py file.gbk Out.seed
```

PhiSpy was then run; using the SEED converted files;

```
./PhiSpy – i Out.seed – o My_Output – t 25
```

The results were compared and visualized graphically by using a statistical software called **R/R-Studio**

Chapter -3: Results

CHAPTER -3

RESULTS

3.1 Encoding CRISPR/ Cas system

Although it has been proved to be an adaptive immunity against invading mobile elements, the CRISPR-Cas system is not encoded by all bacteria, despite the phage and other mobile genetic elements threats. The results show that the system is encoded in only 34.7 % genomes in this study (Figure 3.1). Of all the 5 bacteria, *C. difficile* strains encode the system more than in any other bacteria with 99.6 % genomes found to contain CRISPR spacers. Unlike in other 4 bacteria, encoding CRISPR-Cas system among *V. Cholerae* strains was found to be lower with only 134 of 943 (14.2%) genomes found to contain CRISPR spacers.
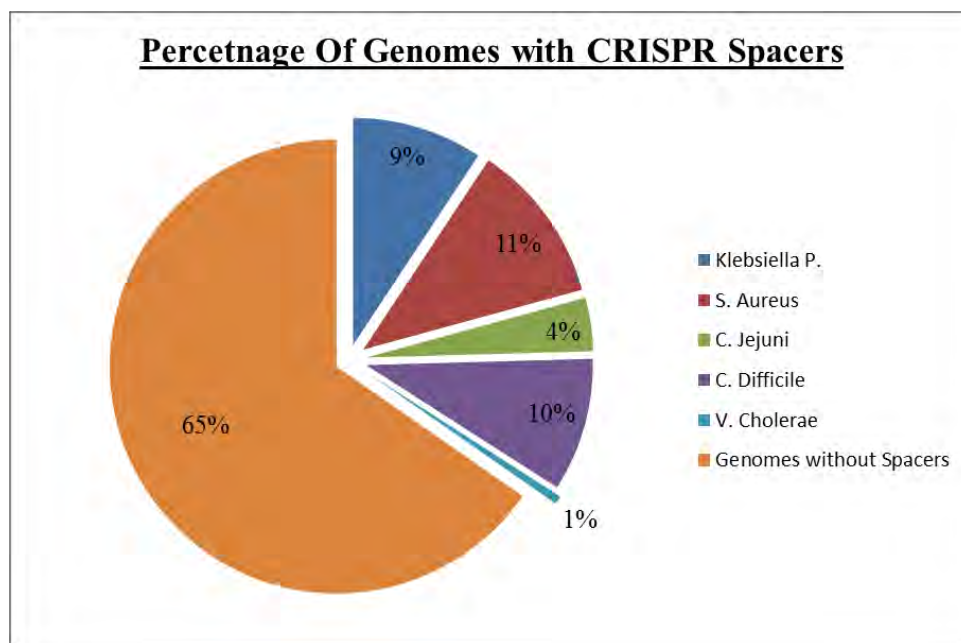
Figure 3.1 Presence of CRISPR Spacers in the Analyzed genomes

3.2 CRISPR spacers and Self-targeting spacers

A total of 277741 spacers were obtained from the 5 bacterial. Because the CRISPR-Cas system is encoded more than it is in other 4 bacteria, *C. difficile* strains contributed the highest number of spacers with close to 70% of all spacers coming from *C. difficile* genomes (Figure 3.2).

The study showed that less than 5% of the obtained spacers were self-targeting (Figure 3.3). All spacers were scanned for full chromosomal matching by blasting them against their respective

genomes. A spacer that had more than one hit against on its genome indicated that there was a possibility to find its sequence not only in the CRISPR array but also in other locations on the chromosome and so this spacer was regarded as a self-targeting spacer.
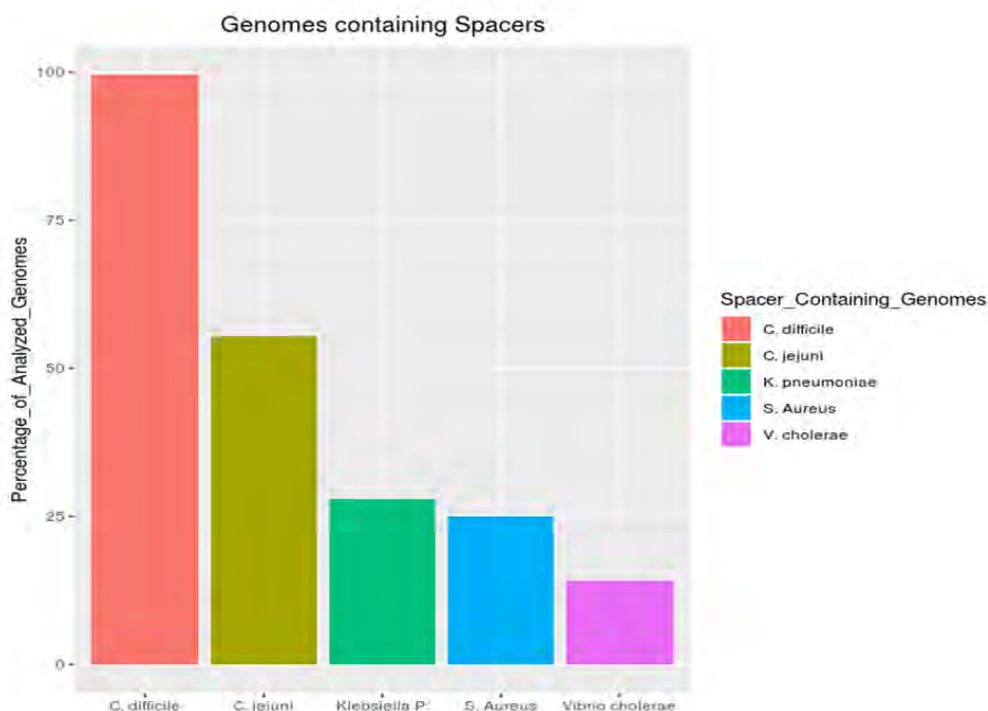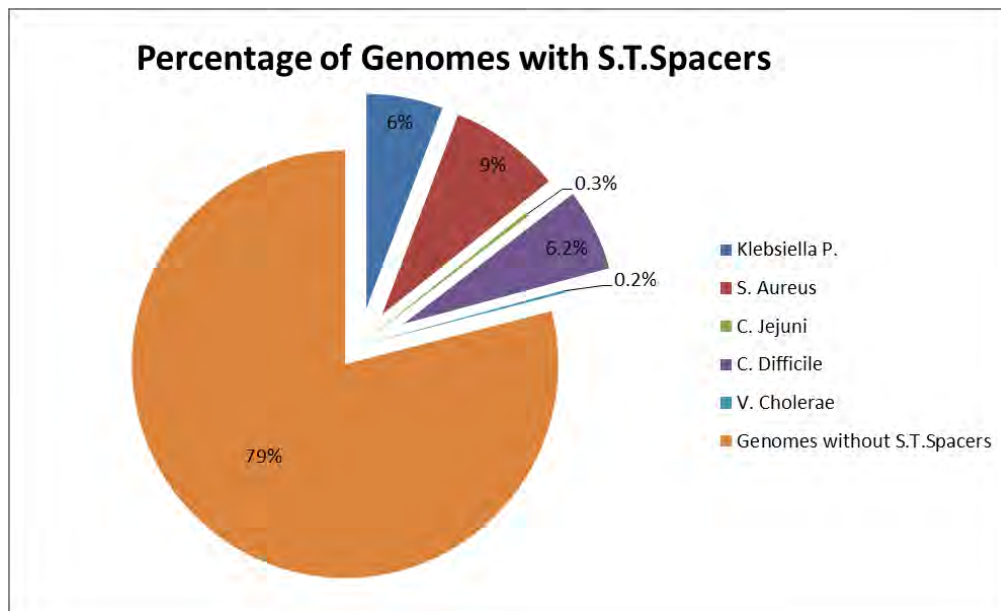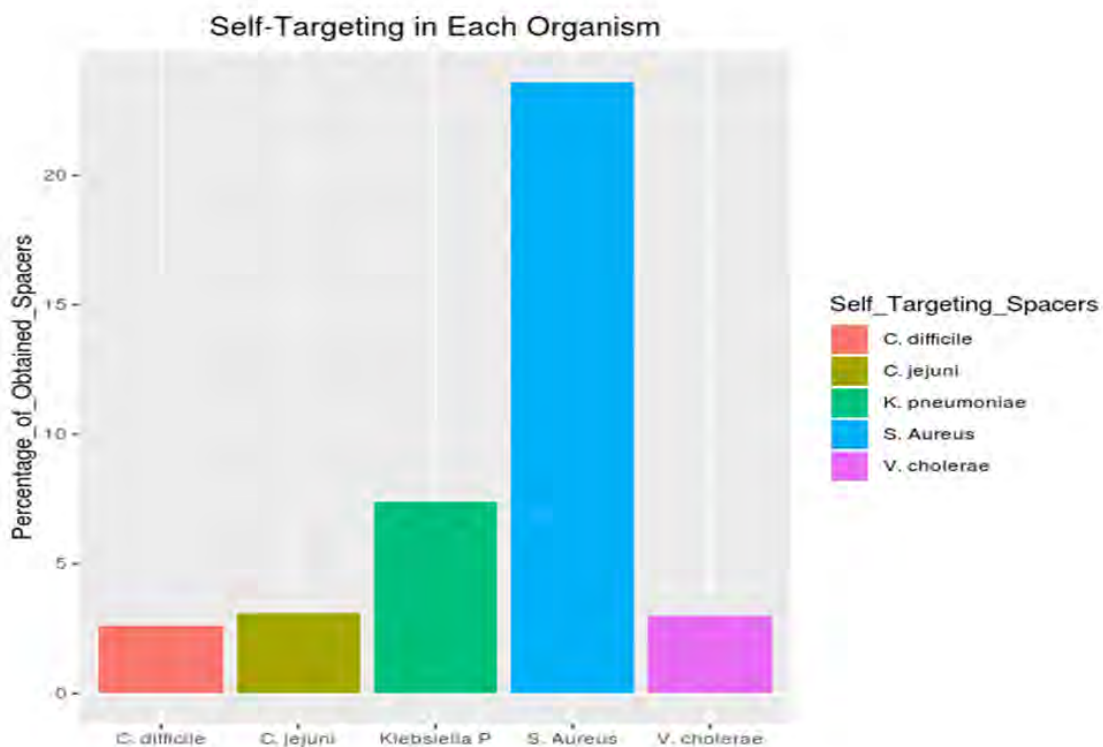

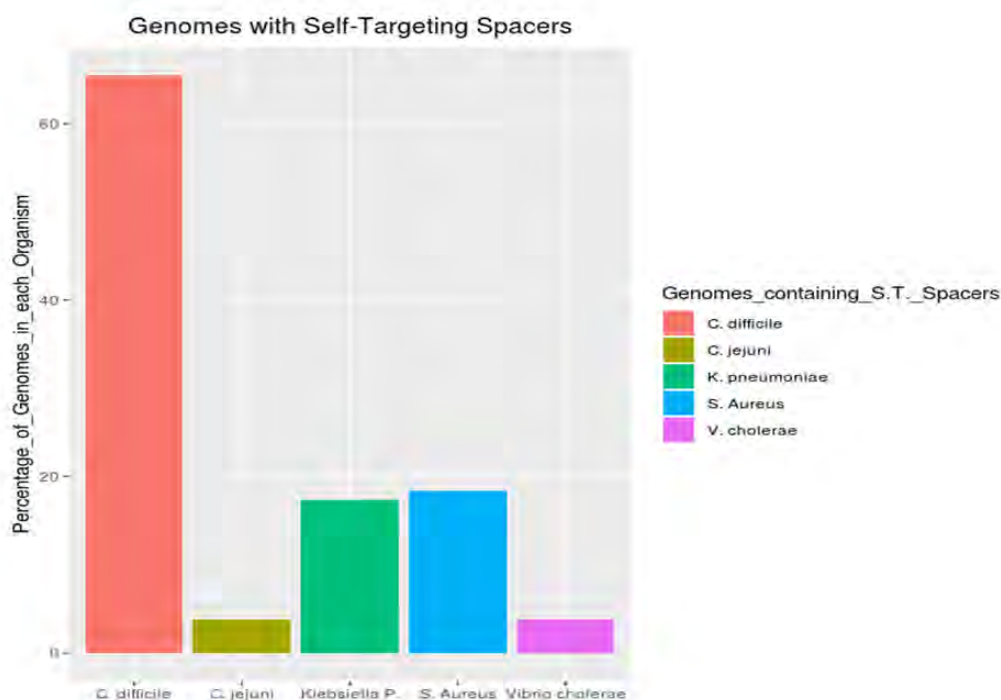**Figure 3.2 Spacer containing Genomes in each bacterium**

The self-targeting nature of spacers was found more in *S. aureus* genomes compared to other bacterial genomes. The study showed that 23.6% of spacers obtained from *S. aureus* genomes were self-targeting (Figure 3.4). For every spacer containing genome in *S. aureus* strains, the likelihood such genomes to contain self-targeting spacers was found to be 73.6%. This indicates it is more likely to incorporate chromosomal derived spacers in CRISPR encoding *S. aureus* strains. The probability of finding a random *C. difficile* genome with a self-targeting spacer was found to be higher than in the other 4 bacteria. The study showed that 65% of all *C. difficile* genomes in NCBI's database at the time of the study contained at least a self-targeting spacer (Figure 3.5). This indicates that incorporating chromosomal derived spacers in such strains is higher and so and the strains have got a strong escape mechanism for the like fitness cost of carrying spacers that target their chromosome. The study also found that a single genome could contain more than one self-targeting spacer, indicating that CRISPR/ Cas system can launch multiple attacks to its encoding cell if there are no mechanisms put in place to escape autoimmunity.

Figure 3.3 Presence of Self-targeting spacers in different bacterial Genomes

Figure 3.4 Self-targeting spacers as a percentage of all obtained spacers from each bacterium

Figure 3.5 Self-targeting containing genomes as a percentage of all genomes of each bacterium

3.3 Self-targeting spacer distribution and acquisition

A higher number of self-targeting spacers were found located in the proximal end to the leader sequence than the distal end. This means that self-targeting spacers are mostly of a recent acquisition event due to the chronological acquisition pattern of spacers into the CRISPR array. Self-targeting spacers in the proximal end were found to be with low targeting properties (those targeting the chromosomal DNA once or twice) while those with higher targeting properties are were found in the distal end. Although they are more concentrated proximal to the leader sequence, self-targeting spacers are normally distributed throughout the CRISPR array in these 5 bacteria under this study. The maximum number of times a spacer can target its genome was found to be 8 (Figure 3.6).
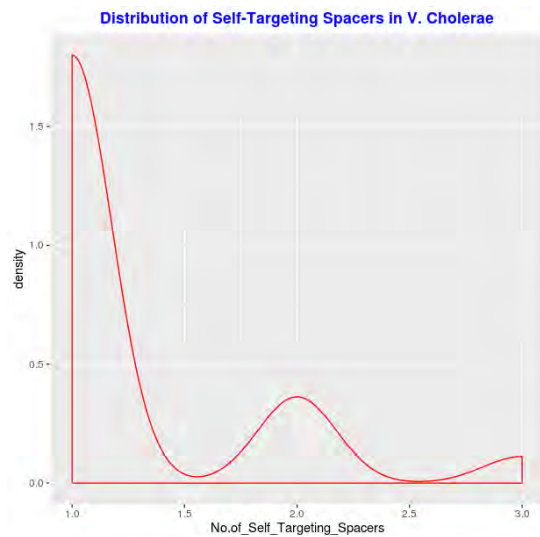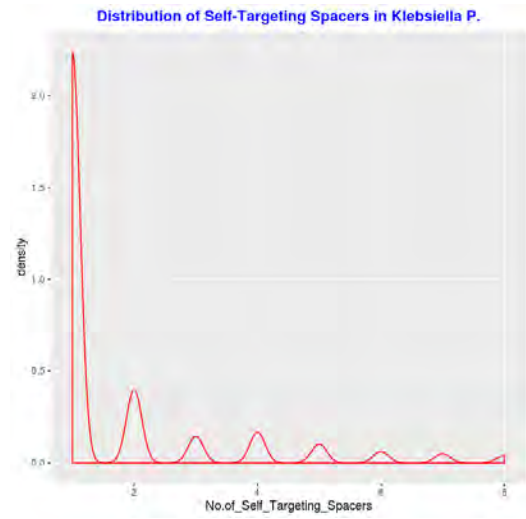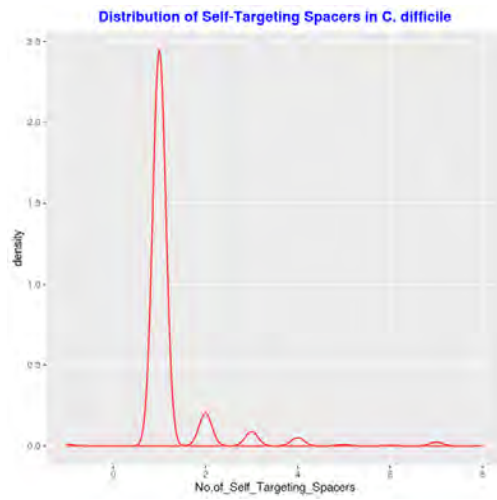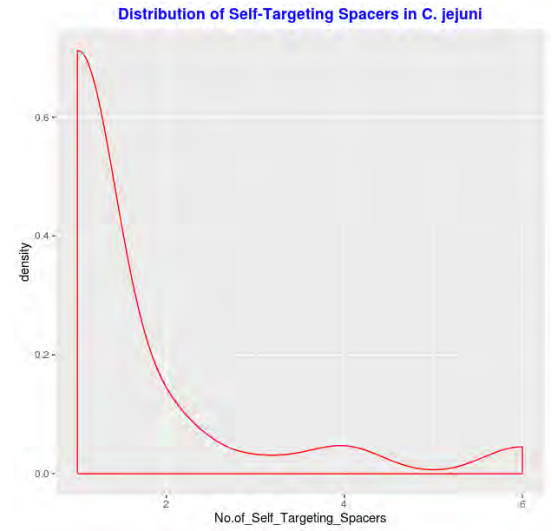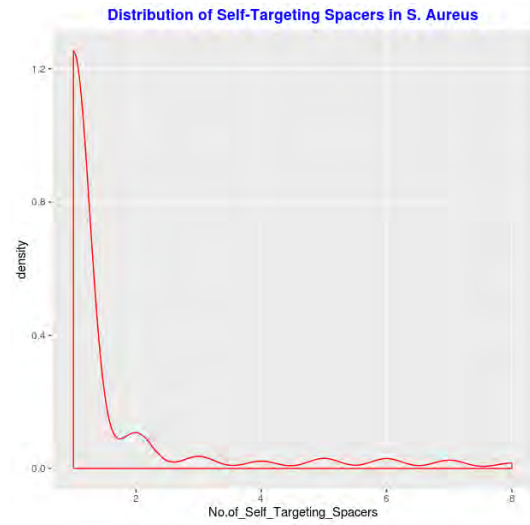
Figure 3.6 Distribution of Self-targeting spacers in different bacteria

3.4 The possible function of self-targeting spacers

A total of 12530 self-targeting spacers were found among genomes of the 5 bacteria under this study, most of which were found matching on phage sequences or bacterial proteins. 57.5% of these self-targeting spacers were found to be targeting prophages, 41% targeting bacterial proteins and 1.6% of them were unaccountable. In terms of percentages, *V. cholerae* contained the highest unaccounted for self-targeting spacers with 36.8% while in terms of numbers, *K. pneumoniae* contained the highest with 89 of 4602 (1.9%) unaccounted for self-targeting spacers.

The adaptive immunity of the CRISPR/Cas system was found functional among these 5 bacteria. 7218 of 21530 (57.6%) self-targeting spacers were found to be targeting invading mobile elements (mainly phages and prophages). The immunity was found to be highly functional in *K. pneumoniae* strains with 2793 of 4602 (60.7%) self-targeting spacers in such strains found to be targeting phages. The study also showed that much of the self-targeting spacers in *C. difficile* (59%) targeted phages than it was for other bacteria (Figure 3.7).

Figure 3.7 Self-targeting spacers targets in each bacterium

5135 of 12530 (41%) of self-targeting spacers were found to be targeting bacterial proteins. This showed that fitness costs as a result of self-targeting spacers is likely to be more in *C. jejuni* strains in which 85 of 120 (70.8%) of self-targeting spacers were found to be targeting bacterial proteins. *C. difficile* genomes contained the highest number of targeted proteins by self-targeting spacers with 2056 of 5098 (40.5%) spacers matching on bacterial gene sequences.

Self-targeting spacers were found to be targeting either known or hypothetical proteins although the study showed that hypothetical proteins were more targeted than known proteins (Figure 3.8). *S. aureus* genomes contain the highest number of targeted hypothetical proteins with 91.9% of targeted proteins in such genomes being hypothetical. This can be attributed to fewer genomic studies in *S. aureus* to reveal much of the gene products in this bacterial species. On the other hand, *C. difficile* genomes had the highest number of known targeted proteins with 27.6% of targeted proteins in such genomes being known. It was also found that a particular protein; ***Serine-aspartate repeat-containing protein*** was targeted in several genomes than other proteins and this trend did not occur in any of the other 4 bacteria.
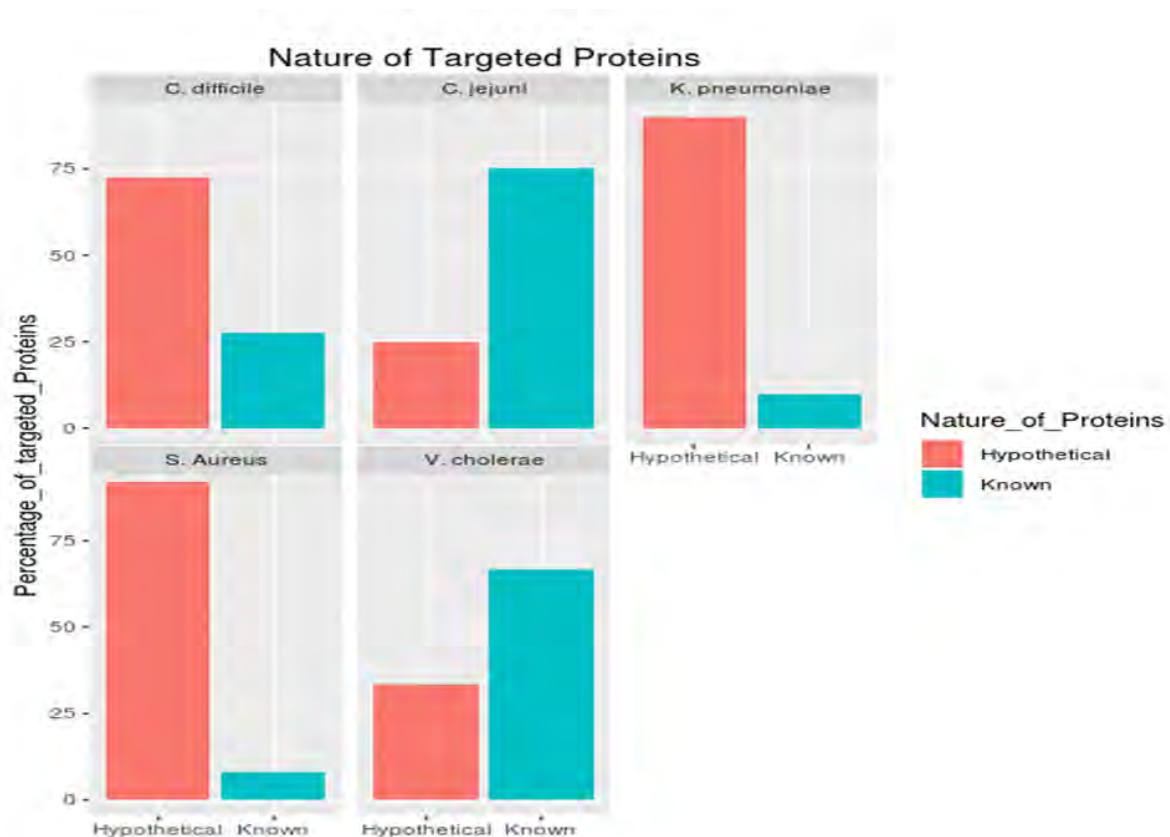
Figure 3.8 Nature of targeted proteins by Self-targeting spacers in each bacterium

As it was the case for self-targeting spacers, genomes could contain more than one targeted protein too. Apart from *V. cholerae* genomes, it was found in the other 4 bacterial genomes that a single protein could be targeted by more than one self-targeting spacer. This could be attributed to the alternative splicing nature of gene expression for which an incorporated chromosomal derived spacer could be encoding for multiple proteins as a result of differential splicing. In *C. jejuni* genomes, each self-targeting spacer was against one known protein (21 spacers against 21 proteins) but this was not the case for hypothetical proteins in the same genomes. The results also showed that targeted proteins among *S. aureus* genomes were the same. This indicated a possibility of a common chromosomal region among all *S. aureus* species which is being incorporated into their CRISPR arrays.

3.4.1 Escape mechanism from autoimmunity – lack of Cas genes

Of the four proposed mechanisms of how prokaryotes may escape from autoimmunity by self-targeting spacers (Stern et al., 2010), this study mainly focused on the lack of Cas genes in a self-targeting spacer containing genome. Although they are believed to be detrimental (Vercoe et al., 2013), the occurrence of self-targeting spacers among these 5 bacteria was not a rare phenomenon with 5% of the obtained spacers being self-targeting, therefore a need for escape mechanisms against the likely risks. Most of the proteins which were found to be targeted are very important for bacterial survival (Tables 3A and B) therefore, there must be a mechanism to save such bacterial strains at risk.

Table 3 A: Targeted proteins by Self-targeting spacers in *K. pneumoniae, S. aureus* and *V. cholerae* genomes

Name of Organism and Targeted proteins	
Klebsiella P.	**Targeted gene protein**
	16S_ribosomal_RNA IS66_family_transposase_ISEc22 IS110_family_transposase_ISKpn2 16S_ribosomal_RNA_(partial) putative_HTH-type_transcriptional_regulator Tyrosine_recombinase_XerD Lipoprotein_YlpA Plasmid_segregation_protein_ParM Plasmid-derived_single-stranded_DNA-binding_protein Pilin Fertility_inhibition_protein Nucleoid_occlusion_protein Coupling_protein_TraD Protein_PsiB Antitoxin_HigA1 Histidine-binding_periplasmic_protein Colicin_V_secretion_protein_CvaA 3-oxoacyl-[acyl-carrier-protein]_reductase_FabG Prophage_integrase_IntA putative_licABCH_operon_regulator Lysozyme_RrrD Beta-galactosidase CRISPR-associated_endonuclease/helicase_Cas3 Glutathione_import_ATP-binding_protein_GsiA 2%2C3-bisphosphoglycerate-independent_phosphoglycerate_mutase
S. Aureus	
	Serine-aspartate_repeat-containing_protein_E Serine-aspartate_repeat-containing_protein_D Serine-aspartate_repeat-containing_protein_C Clumping_factor_B Bone_sialoprotein-binding_protein Clumping_factor_A
V. Cholerae	
	S-adenosylmethionine_synthase

Table 3 B: Targeted proteins by Self-targeting spacers in *C. difficile* and *C. jejuni* genomes

C. difficile	Targeted gene protein
	D-alanine--D-alanine_ligase
	Tyrosine_recombinase_XerC
	Phosphoadenosine_phosphosulfate_reductase
	Inositol_2-dehydrogenase/D-chiro-inositol_3-dehydrogenase
	Demethyllactenocin_mycarosyltransferase
	DNA_polymerase_I
	Gamma-D-glutamyl-L-lysine_dipeptidyl-peptidase
	Error-prone_DNA_polymerase
	HTH-type_transcriptional_regulator_CysL
	HTH-type_transcriptional_regulator_ImmR
	Ribosome_hibernation_promotion_factor
	Alcohol_dehydrogenase_2
	Hemin_transport_system_permease_protein_HmuU
	PTS_system_fructose-specific_EIIABC_component
	D-aminoacylase
	Single-stranded_DNA-binding_protein_A
	Transcription_antiterminator_LicT
	Nicotinate_phosphoribosyltransferase
	putative_FMN/FAD_exporter_YeeO
	Flagellar_secretion_chaperone_FliS
	Tyrosine_recombinase_XerD
	16S_ribosomal_RNA

	Chromosome_partition_protein_Smc
	Methicillin_resistance_mecR1_protein
	50S_ribosomal_protein_L1
	putative_protein_YqbN
	ICEBs1_excisionase
	Chromosome_partition_protein_Smc
	5'-nucleotidase
	DNA_mismatch_repair_protein_MutS
	HTH-type_transcriptional_activator_RhaR
	Putative_ATP-dependent_DNA_helicase_YjcD
	DNA_polymerase_III_subunit_alpha
	Lipoate-protein_ligase_LplJ
	NADP-reducing_hydrogenase_subunit_HndD
	Glutamate_2%2C3-aminomutase
	1-deoxy-D-xylulose_5-phosphate_reductoisomerase
	Acetyl-coenzyme_A_carboxylase_carboxyl_transferase_subunit_beta
	Replicative_DNA_helicase
	Glutamate_2%2C3-aminomutase
	Putative_SPBc2_prophage-derived_single-strand_DNA-specific_exonuclease_YorK
	DNA_topoisomerase_3
	IS1595_family_transposase_ISCco3
	16S_ribosomal_RNA

C. jejuni	Targeted gene protein
	23S_ribosomal_RNA

Type_IIS_restriction_enzyme_Eco57I

Type_IV_secretion_system_protein_virB4
Type_IV_secretion_system_protein_PtlH
Motility_protein_B
Exoenzyme_S_synthesis_regulatory_protein_ExsA
GTP-binding_protein_TypA/BipA
Type_IV_secretion_system_protein_PtlH
tRNA_uridine_5-carboxymethylaminomethyl_modification_enzyme_MnmG
Type_IV_secretion_system_protein_virB4
GTP-binding_protein_TypA/BipA
Prophage_integrase_IntS
Alanine--tRNA_ligase
Glutamate_racemase
Carboxynorspermidine/carboxyspermidine_decarboxylase
L-asparaginase_2
Putative_DNA_repair_helicase_RadD |

To identify whether lack of Cas genes is an escape mechanism among these 5 bacteria, all targeted proteins genomes were annotated for Cas proteins. The results showed that the percentage of targeted proteins genomes with Cas proteins was higher than those without Cas proteins (86% and 14% respectively). The percentage of known targeted protein genomes containing Cas proteins was found to be higher (73%) than the case for the same genomes without Cas proteins. This indicated that such bacterial strains depend on other escape mechanisms rather than lack of Cas gene clusters. This could be evidenced from the study in *K. pneumoniae* and *C. difficile* for which only 2.6% and 20% respectively of their targeted protein genomes contained Cas proteins (Figure 3.9).
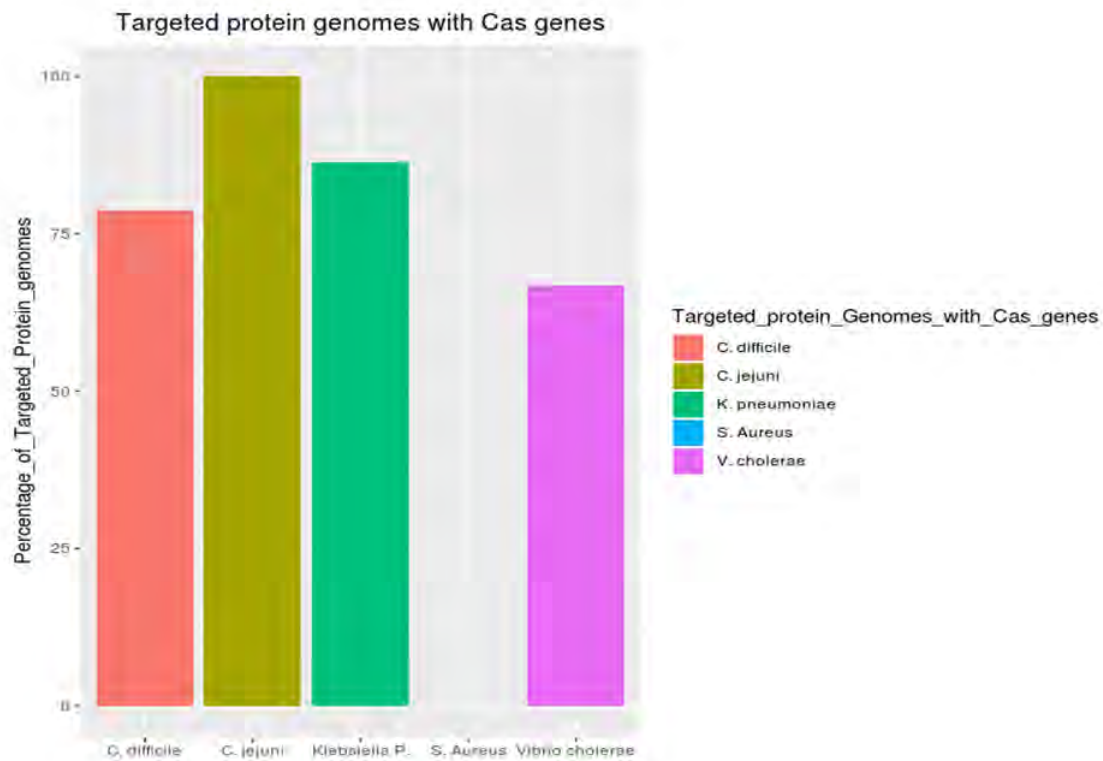
Figure 3.9 Cas proteins in targeted protein genomes in each bacterium

Unlike in these 2 bacteria, all targeted protein genomes in *S. aureus* were found not to contain any Cas proteins. It indicated that lack of Cas genes in self-targeting spacer containing genomes can be a reliable mechanism in *S. aureus* strains to escape autoimmunity. However, all targeted protein genomes in *C. jejuni* were found to contain Cas proteins, indicating that autoimmunity in self-targeting spacer containing strains of *C. jejuni* is escaped by other mechanisms rather than lack of Cas gene clusters. Unlike in *C. jejuni*, the presence of Cas proteins in targeted protein genomes of other bacteria was found not to be 100%. This meant that in such bacterial strains, lack of Cas genes and/or other mechanisms could help in escaping the risks of autoimmunity.

Chapter -4: Discussion

CHAPTER -4

DISCUSSION

The study aimed at giving insights about the pattern by which the CRISPR-Cas system acquires self-targeting spacers among 5 bacteria and the function of such spacers as autoimmunity. Although it is believed to be an acquired anti-viral system (Stern, et al., 2010), only 34.7% of the bacterial genomes in this study were found to contain CRISPR spacers. Some bacterial genomes were found to contain more spacers than others something that was observed to be a result of the environment such bacteria are commonly found. Genomes whose bacteria species are very common in the environment were found to contain more spacers. This indicates that such bacteria are at the risk of viral attack, hence highlighting the importance of the CRISPR-Cas system as immunity against viral infection in bacteria.

The study showed that CRISPR spacers are mainly derived from invading mobile elements like viruses and their addition in CRISPR arrays provides a genetic memory which is a pre-requisite for subsequent neutralization of invading nucleic acids (Boudry et al., 2015). Spacers were found to be acquired in a chronological order which revealed phylogenetic relations between the strains of different bacteria and also being a good technique for genotyping (Louwen, et al., 2014; Andersen et al., 2016). Some CRISPR types and subtypes are known to contain an unusually high number of arrays (Boudry, et al., 2015). In this study, the high of spacers in the genomes of *C. difficile* can be attributed to the unique CRISPR subtype I –B which is characterized by a high number of arrays.

However, there was a high ratio of self-targeting spacers among these 5 bacteria that were found to match fully to self-chromosomes. Although integrating a self-targeting spacer is regarded as an accidental leaky incorporation event of self-nucleic acid (Stern et al., 2010), the study showed that the event is somewhat frequent in some bacteria; making it a less accidental event especially in genomes with a higher number of self-targeting spacers.

A significant number of self-targeting spacers in each of these bacteria under this study was found to be targeting mobile elements especially prophages. These prophages were lysogenic viruses that attacked the bacteria and escaped the bacterial adaptive immunity hence got their genetic material lysogenized into the bacterial chromosome. This explains why the highest number of self-targeting spacers was found to match on prophage sequences than bacterial proteins, especially in those bacteria which are commonly found in the environment. However, in some bacteria, for example, *S. aureus*, the presence of spacers against chromosomal DNA indicates horizontal gene transfer events of virulent mobile genetic elements (Guan et al., 2017). Such elements, for example, Staphylococcal Cassette Chromosome Mec (SCCmec) can be altered or deleted from the bacterial genome by spacers that target them, hence attributing some of the obtained self-targeting spacers I such bacteria for this function. In *K. pneumoniae,* acquiring foreign genetic material is interfered with by CRISPR to maintain

antibiotic susceptibility in the bacteria. This means that there will be higher chances of incorporating chromosomal derived spacers hence explaining the higher number of self-targeting spacers in *K. pneumoniae* genomes.

On top of being acquired, spacers can also be lost, especially those that pose a fitness cost to the bacteria. The chronological acquisition indicated that such spacers in the proximal end to the leader sequence were not only recently acquired but also posed the highest threat to the cell. This explains why most self-targeting spacers were found to be in the proximal end than in the distal end to the leader sequence. All the bacteria except *S. aureus* contained the essential Cas proteins needed for spacer acquisition in their genomes; that is Cas1 and Cas2. However, studies have shown that lack of these 2 Cas proteins did not affect the function of the CRISPR-Cas system (Shen et al., 2017) hence concluding that the system functions normally in all the bacteria under this study.

Self-targeting spacers were found to function as autoimmunity among all the bacteria in this study. Although much of the targeted proteins were found to be hypothetical, a significant number of essential proteins were found to be targeted. This means indicated the existence of known or unknown escape mechanisms to compromise the detrimental nature of self-targeting spacers. Among the 4 proposed putative protection mechanisms against autoimmunity, this study focused on lack of Cas genes and how it fairs as one of the escape mechanisms. This study provided some lines of evidence supporting the role played by lack of Cas genes in protecting strains against autoimmunity.

Studies have shown that following the integration of self-targeting spacers, the CRISPR-Cas system is inactivated if the bacteria are to survive. Some organisms were found to contain highly degraded CRISPR system and so containing Cas pseudogenes (Stern et al., 2010). *S. aureus* targeted protein genomes were found to be lacking Cas proteins. This means that it is a reliable escape mechanism that cuts across all *S. aureus* strains and so other mechanisms are nearly not needed.

All targeted protein genomes in *C. jejuni* contained Cas proteins. This indicated that acquiring new spacers normally occurs in such strains but they depend on other strategies for autoimmunity protection rather than lack of Cas genes. In *K. pneumoniae*, the study showed that only 2.7% of the targeted protein genomes lacked Cas proteins. Previous studies had shown that mutations in leader sequences, adjacent repeats and PAMs in most of the full matching spacers is not an escape mechanism in *K. pneumoniae* strains (Shen et al., 2017). Given that it could be tolerated, it indicated that *K. pneumoniae* strains rely on the other 2 escape mechanism from autoimmunity or depend on different mechanisms which are yet to be known.

Other bacteria in the study were found to have some of their targeted protein genomes lacking Cas proteins. This means that in such bacteria, lack of Cas genes can be an escape mechanism but does not cut across all strains, therefore creating a need to rely on a combination of 2 or more

mechanisms in a given bacterial species. In S. aureus genomes, particular proteins were found to be targeted in different genomes. These were cell surface associated proteins namely *Serine-aspartate repeat-containing* protein C, E and D were found to be the most targeted proteins.

Apart from being detrimental, autoimmunity can be directed against drug resistance genes among these 5 infectious bacteria. The study showed the possibility of autoimmunity occurring among all the 5 bacteria yet there is a worrying increase in antibiotic resistance among them. Studies have shown that redirecting CRISPR-Cas system towards autoimmunity can be an effective genome editing tool (Maikova et al., 2019). This can be a great strategy to fight the increasing drug resistance in common infectious bacteria through genetic engineering to direct autoimmunity against drug resistance genes. Genome editing of this nature has already proved successful in knocking out certain genes in *C. difficile* (Maikova et al., 2019).

Basing on their acquisition pattern, self-targeting spacers represent a recent acquisition event implying that such spacers only survive for a short time among these 5 bacteria. On top of that, such spacers are not inherited to the next progeny. As it has been found out that autoimmunity occurs in all the 5 bacteria, genetic engineering can be used to redirect it against drug resistance genes which may be a good strategy for fighting antibiotic resistance among common infectious bacteria in our community.

Chapter – 5: References

CHAPTER -6

REFERENCES

Akhter, S., Aziz, R. K., & Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic acids research*, 40(16), e126. doi:10.1093/nar/gks406

Aliyari R, Ding SW (January 2009). RNA-based viral immunity initiated by the Dicer family of host immune receptors. *Immunological Reviews*. 227 (1): 176–188. doi:10.1111/j.1600-065X.2008.00722.x

American Academy of Dermatology (February 2013), "Five Things Physicians and Patients Should Question," Choosing Wisely: an initiative of the ABIM Foundation, *American Academy of Dermatology,* retrieved 5 November 2019

Andersen, J. M., Shoup, M., Robinson, C., Britton, R., Olsen, K. E., & Barrangou, R. (2016). CRISPR diversity and microevolution in Clostridium difficile. *Genome biology and evolution*, *8*(9), 2841-2855.

Andersson AF, Banfield JF (May 2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*. 320 (5879): 1047–1050. doi:10.1126/science.1157358

Angelina A. Kislichkina, Anastasia I. Lev, Ekaterina V. Komisarova, Nadezhda K. Fursova, Vera P. Myakinina, Tatiana N. Mukhina, Alexander A. Bogun, Nikolay V. Volozhantsev, Genome sequencing and comparative analysis of three hypermucoviscous Klebsiella pneumoniae strains isolated in Russia, *Pathogens and Disease*, Volume 75, Issue 4, June 2017, ftx024, https://doi.org/10.1093/femspd/ftx024

Aris-Edda Stachler, Israela Turgeman-Grott, Ella Shtifman-Segal, Thorsten Allers, Anita Marchfelder, Uri Gophna, High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon, *Nucleic Acids Research*, Volume 45, Issue 9, 19 May 2017, Pages 5208–5216, https://doi.org/10.1093/nar/gkx150

Aydin, S., Personne, Y., Newire, E., Laverick, R., Russell, O., Roberts, A. P., & Enne, V. I. (2017). Presence of Type IF CRISPR/Cas systems is associated with antimicrobial susceptibility in Escherichia coli. *Journal of Antimicrobial Chemotherapy*, *72*(8), 2213-2218.

Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B, et al. (January 2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair". *Molecular Microbiology*. 79 (2): 484–502. doi:10.1111/j.1365-2958.2010.07465.x

Bae J, Oh E, Jeon B (2014). Enhanced transmission of antibiotic resistance in Campylobacter jejuni biofilms by natural transformation. *Antimicrob. Agents Chemother*. 58 (12): 7573–5. doi:10.1128/AAC.04066-14

Barrangou R (2015). The roles of CRISPR-Cas systems in adaptive immunity and beyond.*Current Opinion in Immunology*. 32: 36-41. https://doi.org/10.1016/j.coi.2014.12.008

Barrangou R, van der Oost J (2013). CRISPR-Cas Systems : RNA-mediated Adaptive Immunity in Bacteria and Archaea. *Heidelberg: Springer*. p.6. ISBN 978-3-642-34656-9.

Bikard, D., Hatoum-Aslan, A., Mucida, D., & Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell host & microbe*, *12*(2), 177-186.

Birnie AJ, Bath-Hextall FJ, Ravenscroft JC, Williams HC (July 2008). Birnie AJ (ed.). Interventions to reduce Staphylococcus aureus in the management of atopic eczema. *The Cochrane Database of Systematic Reviews* (3): CD003871. doi:10.1002/14651858.CD003871.pub2

BLAST® Command Line Applications User Manual. https://www.ncbi.nlm.nih.gov/ Retrieved 2019-11-12.

Blot SI, Vandewoude KH, Hoste EA, Colardyn FA (October 2002). Outcome and attributable mortality in critically Ill patients with bacteremia involving methicillin-susceptible and methicillin-resistant Staphylococcus aureus. *Archives of Internal Medicine*. 162 (19): 2229–35. doi:10.1001/archinte.162.19.2229

Bogovazova, GG; Voroshilova, NN; Bondarenko, VM (April 1991). The efficacy of Klebsiella pneumoniae bacteriophage in the therapy of experimental Klebsiella infection. Zhurnal Mikrobiologii, Epidemiologii, *I Immunobiologii* (in Russian) (4): 5–8. ISSN 0372-9311

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (August 2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 151 (Pt 8): 2551–2561. doi:10.1099/mic.0.28048-0

Boudry, P., Semenova, E., Monot, M., Datsenko, K. A., Lopatina, A., Sekulovic, O., … Soutourina, O. (2015). Function of the CRISPR-Cas System of the Human Pathogen Clostridium difficile. *mBio*, 6(5), e01112–e1115. doi:10.1128/mBio.01112-15

Box AM, McGuffie MJ, O'Hara BJ, Seed KD. 2016. Functional analysis of bacteriophage immunity through a type I-E CRISPR-Cas system in *Vibrio cholerae* and its application in bacteriophage genome engineering. *J Bacteriol* 198:578–590. doi:10.1128/JB.00747-15.

Bradde, S., Mora, T., & Walczak, A. M. (2019). Cost and benefits of clustered regularly interspaced short palindromic repeats spacer acquisition. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences*, 374(1772), 20180095. doi:10.1098/rstb.2018.0095

Brathwaite, K. J., Siringan, P., Moreton, J., Wilson, R., & Connerton, I. F. (2013). Complete genome sequence of universal bacteriophage host strain Campylobacter jejuni subsp. jejuni PT14. *Genome Announc.*, *1*(6), e00969-13.

C Pourcel, G Salvignol, Gilles Vergnaud. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.. *Microbiology*, Microbiology Society, 2005, 151 (Pt 3), pp.653-63. ⟨10.1099/mic.0.27437-0 ⟩. ⟨hal-01158317⟩

Cassandra Cañez, Kurt Selle, Yong Jun Goh, Rodolphe Barrangou, Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*, *FEMS Microbiology Letters*, Volume 366, Issue 9, May 2019, fnz105, https://doi.org/10.1093/femsle/fnz105

Chang S, Sievert DM, Hageman JC, Boulton ML, Tenover FC, Downes FP, Shah S, Rudrik JT, Pupp GR, Brown WJ, Cardo D, Fridkin SK (April 2003). Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. *The New England Journal of Medicine*. 348 (14): 1342–7. doi:10.1056/NEJMoa025025

Chen, W., Zhang, Y., Yeo, W. S., Bae, T., & Ji, Q. (2017). Rapid and efficient genome editing in Staphylococcus aureus by using an engineered CRISPR/Cas9 system. *Journal of the American Chemical Society*, *139*(10), 3790-3795.

Clare Rollie, Shirley Graham, Christophe Rouillon, Malcolm F White, Prespacer processing and specific integration in a Type I-A CRISPR system, *Nucleic Acids Research*, Volume 46, Issue 3, 16 February 2018, Pages 1007–1020, https://doi.org/10.1093/nar/gkx1232

Clostridium difficile Infection Information for Patients | HAI | CDC. www.cdc.gov. Retrieved 22 October 2019.

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E (July 2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications*. 3: 945. doi:10.1038/ncomms1937

David Acheson, Ban Mishu Allos, Campylobacter jejuni Infections: *Update on Emerging Issues and Trends, Clinical Infectious Diseases*, Volume 32, Issue 8, 15 April 2001, Pages 1201–1206, https://doi.org/10.1086/319760

Davis, B; Waldor, M. K. (February 2003). Filamentous phages linked to virulence of Vibrio cholerae. *Current Opinion in Microbiology*. 6 (1): 35–42. doi:10.1016/S1369-5274(02)00005-X

Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (February 2008). Phage response to CRISPR-encoded resistance inStreptococcus thermophilus. *Journal of Bacteriology*. 190 (4): 1390–1400. doi:10.1128/JB.01412-07

Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A., & Brisse, S. (2005). Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. *Journal of clinical microbiology*, *43*(8), 4178-4182.

Díez-Villaseñor C, Guzmán NM, Almendros C, García-Martínez J, Mojica FJ (May 2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA Biology*. 10 (5): 792–802. doi:10.4161/rna.24023

Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, Sharma CM (May 2013). "High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates". *PLoS Genetics*. 9 (5): e1003495. doi:10.1371/journal.pgen.1003495

Dugar, G., Leenay, R. T., Eisenbart, S. K., Bischler, T., Aul, B. U., Beisel, C. L., & Sharma, C. M. (2018). CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the Campylobacter jejuni Cas9. *Molecular cell*, *69*(5), 893-905.

Erdmann S, Garrett RA (September 2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Molecular Microbiology*. 85 (6): 1044–1056. doi:10.1111/j.1365-2958.2012.08171.x

F M Colles, N D McCarthy, J C Howe, C L Devereux, A G Gosler, and M C J Maiden Dynamics of Campylobacter colonization of a natural host, Sturnus vulgaris (European Starling) Environ Microbiol. 2009 January; 11(1): 258–267. doi:10.1111/j.1462-2920.2008.01773.x

Food Safety: Campylobacter. U.S. Centers for Disease Control and Prevention. Retrieved 2019-11-12

Genome engineering of Clostridium difficile using the CRISPR-Cas9 system. Clinical microbiology and infection : the official publication of the *European Society of Clinical Microbiology and Infectious Diseases,* 24 10, 1095-1099 .

Gu, T., Zhao, S., Pi, Y., Chen, W., Chen, C., Liu, Q., … Ji, Q. (2018). Highly efficient base editing in *Staphylococcus aureus* using an engineered CRISPR RNA-guided cytidine deaminase. *Chemical science*, *9*(12), 3248–3253. doi:10.1039/c8sc00637g

Guan, J., Wang, W., & Sun, B. (2017). Chromosomal Targeting by the Type III-A CRISPR-Cas System Can Reshape Genomes in *Staphylococcus aureus*. *mSphere*, *2*(6), e00403-17. doi:10.1128/mSphere.00403-17

Han D, Lehmann K, Krauss G (June 2009). SSO1450—a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA. *FEBS Letters*. 583 (12): 1928–1932. doi:10.1016/j.febslet.2009.04.047

Hargreaves, K. R., Flores, C. O., Lawley, T. D., & Clokie, M. R. (2014). Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in Clostridium difficile strains and prophages target multiple phage types within this pathogen. *MBio*, *5*(5), e01045-13.

Hatoum-Aslan A, Maniv I, Marraffini LA (December 2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (52): 21218–21222. doi:10.1073/pnas.1112832108

He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., … Zhang, L. (2018). Cas1 and Cas2 From the Type II-C CRISPR-Cas System of *Riemerella anatipestifer* Are Required for Spacer Acquisition. *Frontiers in cellular and infection microbiology*, *8*, 195. doi:10.3389/fcimb.2018.00195

Heussler, Gary E. and O'Toole, George A., Friendly Fire: Biological Functions and Consequences of Chromosomal Targeting by CRISPR-Cas Systems (2016). *Open Dartmouth: Faculty Open Access Scholarship*. 1027. https://digitalcommons.dartmouth.edu/facoa/1027

Hille, F., & Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *371*(1707), 20150496. doi:10.1098/rstb.2015.0496

Hooton SPT and Connerton IF (2015) Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Front. Microbiol*. 5:744. doi: 10.3389/fmicb.2014.00744

Horvath P, Barrangou R (January 2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327 (5962): 167 -170. doi:10.1126/science.1179555

Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. (February 2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of Bacteriology*. 190 (4): 1401–1412. doi:10.1128/JB.01415-07

Illness & Symptoms Cholera CDC. www.cdc.gov. 2018-12-13. Retrieved 2019-11-12.

Klebsiella species – GOV.UK. www.gov.uk. Retrieved 13 January 2017.

Kuehne, S. A., Heap, J. T., Cooksley, C. M., Cartman, S. T., & Minton, N. P. (2011). ClosTron-mediated engineering of Clostridium. In *Strain Engineering* (pp. 389-407). Humana Press.

Lee, C. R., Lee, J. H., Park, K. S., Jeon, J. H., Kim, Y. B., Cha, C. J., ... & Lee, S. H. (2017). Antimicrobial resistance of hypervirulent Klebsiella pneumoniae: epidemiology, hypervirulence-associated determinants, and resistance mechanisms. *Frontiers in cellular and infection microbiology*, 7, 483.

Li B, Zhao Y, Liu C, Chen Z, Zhou D (2014). Molecular pathogenesis of Klebsiella pneumoniae. *Future Microbiology*. 9 (9): 1071–81. doi:10.2217/fmb.14.48

Li, H. Y., Kao, C. Y., Lin, W. H., Zheng, P. X., Yan, J. J., Wang, M. C., ... & Wu, J. J. (2018). Characterization of CRISPR-Cas systems in clinical Klebsiella pneumoniae isolates uncovers its potential association with antibiotic susceptibility. *Frontiers in Microbiology*, 9, 1595.

Louwen, R.P.L, Staals, L.M, Endtz, H.P, van Baarlen, P, & van der Oost, J. (2014). *The role of CRISPR-cas systems in virulence of pathogenic bacteria. Microbiology and Molecular Biology Reviews* (Vol. 78, pp. 74–88). doi:10.1128/MMBR.00039-13

Maikova A, Severinov K and Soutourina O (2018) New Insights Into Functions and Possible Applications of Clostridium difficileCRISPR-Cas System. *Front. Microbiol*. 9:1740. doi: 10.3389/fmicb.2018.01740

Maikova, A., Severinov, K., & Soutourina, O. (2018). New Insights Into Functions and Possible Applications of Clostridium difficile CRISPR-Cas System. *Frontiers in microbiology*, 9, 1740. doi:10.3389/fmicb.2018.01740

Marraffini LA, Sontheimer EJ (March 2010). CRISPR interference:RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics*. 11 (3): 181–190. doi:10.1038/nrg2749

Mathers, A. J., Stoesser, N., Sheppard, A. E., Pankhurst, L., Giess, A., Yeh, A. J., … Sifri, C. D. (2015). Klebsiella pneumoniae carbapenemase (KPC)-producing K. pneumoniae at a single institution: insights into endemicity from whole-genome

sequencing. *Antimicrobial agents and chemotherapy*, *59*(3), 1656–1663. doi:10.1128/AAC.04292-14

Mayo Clinic. C. difficile infection. https://www.mayoclinic.org/diseases-conditions/c-difficile/symptoms-causes/syc-20351691 Retrieved on 20 November, 2019

McAllister, K.N., Bouillaut, L., Kahn, J.N. et al. Using CRISPR-Cas9-mediated genome editing to generate C. difficile mutants defective in selenoproteins synthesis. *Sci Rep* 7, 14672 (2017) doi:10.1038/s41598-017-15236-5

McGinn, J., & Marraffini, L. A. (2019). Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nature Reviews Microbiology*, *17*(1), 7-12.

McKitterick, A. C., LeGault, K. N., Angermeyer, A., Alam, M., & Seed, K. D. (2019). Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: insights from a natural arms race. *Philosophical Transactions of the Royal Society B*, *374*(1772), 20180089.

Mick, E., Stern, A., & Sorek, R. (2013). Holding a grudge: persisting anti-phage CRISPR immunity in multiple human gut microbiomes. *RNA biology*, 10(5), 900–906. doi:10.4161/rna.23929

Naser, I.B., Hoque, M.M., Nahid, M.A. *et al.* Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of *Vibrio cholerae* in Bangladesh. *Sci Rep* **7,** 14880 (2017) doi:10.1038/s41598-017-14839-2

Nuñez JK, Harrington LB, Kranzusch PJ, Engelman AN, Doudna JA (November 2015). Foreign DNA capture during CRISPR-Cas adaptive immunity. Nature. 527 (7579): 535–538. doi:10.1038/nature15760

Nuñez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA (June 2014).Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*. 21 (6): 528–534. doi:10.1038/nsmb.2820

Nuñez, J. K., Lee, A. S., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, *519*(7542), 193.

Platts-Mills, James A, and Margaret Kosek. ̶Update on the burden of Campylobacter in developing countries." *Current opinion in infectious diseases* vol. 27,5 (2014): 444-50. doi:10.1097/QCO.0000000000000091

Pourcel C, Salvignol G, Vergnaud G (March 2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*. 151 (Pt 3): 653–663. doi:10.1099/mic.0.27437-0. PMID 15758212

Rath, D., Amlinger, L., Rath, A., & Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie, 117*, 119-128. doi: 10.1016/j.biochi.2015.03.025

Rice, K. C., Mann, E. E., Endres, J. L., Weiss, E. C., Cassat, J. E., Smeltzer, M. S., & Bayles, K. W. (2007). The cidA murein hydrolase regulator contributes to DNA release and biofilm development in Staphylococcus aureus. *Proceedings of the National Academy of Sciences*, *104*(19), 8113-8118.

Robins, WP; Mekalanos, JJ (2014). Genomic science in understanding cholera outbreaks and evolution of Vibrio cholerae as a human pathogen. *Current Topics in Microbiology and Immunology*. 379. pp. 211–29. doi:10.1007/82_2014_366

RStudio Team (2015). RStudio: Integrated Development for R. *RStudio, Inc*., Boston, MA URL http://www.rstudio.com/.

Saeed S. Banawas (21 February 2018). Clostridium difficile Infections: A Global Overview of Drug Sensitivity and Resistance Mechanisms. *Hindawi. BioMed Research International*. Retrieved 4 November 2019.

Sampson, T. R., & Weiss, D. S. (2013). Alternative roles for CRISPR/Cas systems in bacterial pathogenesis. *PLoS pathogens*, *9*(10), e1003621.

Schröder, W., Goerke, C., & Wolz, C. (2012). Opposing effects of aminocoumarins and fluoroquinolones on the SOS response and adaptability in Staphylococcus aureus. *Journal of Antimicrobial Chemotherapy*, *68*(3), 529-538.

Schwaber, Mitchell J.; Klarfeld-Lidji, Shiri; Navon-Venezia, Shiri; Schwartz, David; Leavitt, Azita; Carmeli, Yehuda (2008-03-01). Predictors of carbapenem-resistant Klebsiella pneumoniae acquisition among hospitalized adults and effect of acquisition on mortality. *Antimicrobial Agents and Chemotherapy*. 52 (3): 1028–1033. doi:10.1128/AAC.01020-07

Seed, K. D., Faruque, S. M., Mekalanos, J. J., Calderwood, S. B., Qadri, F., & Camilli, A. (2012). Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. *PLoS pathogens*, *8*(9), e1002917.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014 Jul 15;30(14):2068-9. PMID:24642063

Shabbir MAB, Tang Y, Xu Z, Lin M, Cheng G, Dai M, Wang X, Liu Z, Yuan Z and Hao H (2018) The Involvement of the Cas9 Gene in Virulence of Campylobacter jejuni. *Front. Cell. Infect. Microbiol.* 8:285. doi: 10.3389/fcimb.2018.00285

Shah SA, Erdmann S, Mojica FJ, Garrett RA (May 2013). Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biology*. 10 (5): 891–899. doi:10.4161/rna.23764

Shen, J., Lv, L., Wang, X., Xiu, Z., & Chen, G. (2017). Comparative analysis of CRISPR-Cas systems in Klebsiella genomes. *Journal of basic microbiology*, 57, 325-336. doi: 10.1002/jobm.201600589

Siddique, A.K.; Baqui, A.H.; Eusof, A.; Haider, K.; Hossain, M.A.; Bashir, I.; Zaman, K. (1991). Survival of classic cholera in Bangladesh. *The Lancet.* 337 (8750): 1125–1127. doi:10.1016/0140-6736(91)92789-5

Skennerton C (2013 -2019). MinCED - Mining CRISPRs in Environmental Datasets. https://github.com/ctskennerton/Crass

Sorek R, Lawrence CM, Wiedenheft B (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Review of Biochemistry*. 82 (1): 237–266. doi:10.1146/annurev-biochem-072911-172315

Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity?. *Trends in genetics : TIG*, 26(8), 335–340. doi:10.1016/j.tig.2010.05.008

Swarts DC, Mosterd C, van Passel MW, Brouns SJ (2012). CRISPR interference directs strand specific spacer acquisition. *PLOS ONE*. 7 (4): e35888. doi:10.1371/journal.pone.0035888

Swarts, D. C., Mosterd, C., van Passel, M. W., & Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PloS one*, 7(4), e35888. doi:10.1371/journal.pone.0035888

T A Tatusova, I Karsch-Mizrachi, J A Ostell, Complete genomes in WWW Entrez: data representation and analysis., *Bioinformatics,* Volume 15, Issue 7, July 1999, Pages 536–543, https://doi.org/10.1093/bioinformatics/15.7.536

Tong SY, Davis JS, Eichenberger E, Holland TL, Fowler VG (July 2015). Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management. *Clinical Microbiology Reviews*. 28 (3): 603–61. doi:10.1128/CMR.00134-14.

Tyson GW, Banfield JF (January 2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology*. 10 (1): 200–207. doi:10.1111/j.1462-2920.2007.01444.x

Vercoe, R. B., Chang, J. T., Dy, R. L., Taylor, C., Gristwood, T., Clulow, J. S., ... & Fineran, P. C. (2013). Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape

bacterial genomes and expel or remodel pathogenicity islands. *PLoS genetics*, *9*(4), e1003454.

W. Gerhard, K. Nunez. (2013).What you need to know about Klebsiella pneumoniae infection https://www.healthline.com/health/klebsiella-pneumonia Retrieved on 20 November, 2019

Wallis, M. R. (1994-03-01). The pathogenesis of Campylobacter jejuni. *British Journal of Biomedical Science.* 51 (1): 57–64. ISSN 0967-4845

Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, Wang Y (November 2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell.* 163 (4): 840–853. doi:10.1016/j.cell.2015.10.008

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., … Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science (New York, N.Y.)*, *350*(6264), 1096–1101. doi:10.1126/science.aac7041

Wiedenheft B, Sternberg SH, Doudna JA (February 2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 482 (7385): 331–338. doi:10.1038/nature10886

World Health Organization. Cholera https://www.who.int/news-room/fact-sheets/detail/cholera Retrieved on 20 November, 2019

Yosef I, Goren MG, Qimron U (July 2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Research*. 40 (12): 5569–5576. doi:10.1093/nar/gks216

Zhang, F., Zhao, S., Ren, C. *et al.* CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun Biol* **1,** 180 (2018) doi:10.1038/s42003-018-0184-6

Zhao, X., Yu, Z., & Xu, Z. (2018). Study the features of 57 confirmed CRISPR loci in 38 strains of Staphylococcus aureus. *Frontiers in microbiology*, *9*.

Zhu, Duolong; Sorg, Joseph A.; Sun, Xingmin (2018). Clostridioides difficile Biology: Sporulation, Germination, and Corresponding Therapies for C. difficile Infection. *Frontiers in Cellular and Infection Microbiology*. 8: 29. doi:10.3389/fcimb.2018.00029