

**ANALYSIS OF THE STABILITY OF A TYPE III SECRETION SYSTEM
CONTAINING PATHOGENICITY ISLAND (VPI-3) IN THE HUMAN
PATHOGEN *VIBRIO CHOLERAE***

by

Molly C Peters

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the Honors Bachelor of Science Degree in Biological Sciences with a concentration in Cell and Molecular Biology & Genetics with Distinction

Spring 2015

© 2015 Molly Peters
All Rights Reserved

**ANALYSIS OF THE STABILITY OF A TYPE III SECRETION SYSTEM
CONTAINING PATHOGENICITY ISLAND (VPI-3) IN THE HUMAN
PATHOGEN *VIBRIO CHOLERAE***

by

Molly C. Peters

Approved: _____
E. Fidelma Boyd, PhD.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Ramona Neunuebel, PhD.
Committee member from the Department of Department Name

Approved: _____
Emily Day, PhD.
Committee member from the Board of Senior Thesis Readers

Approved: _____
Michael Arnold, Ph.D.
Directory, University Honors Program

ACKNOWLEDGMENTS

There are many people without whom this accomplishment would have been close to impossible. I would first like to thank Dr. Boyd for welcoming me into her lab and investing significant time and resources so that I could have this experience, which was truly one of the most educational experiences of my academic career thus far. Additionally, I would like to extend my gratitude to my committee members, Drs. Neunuebel and Day who gave me invaluable feedback throughout the course of this process. I would also like to especially thank Megan Carpenter, my mentor who tirelessly steered me in the right direction and showed me what it takes to be a scientist. Thank you to all past and present members of Boyd Lab; although this document technically belongs to me, each of you has significantly impacted and aided the work that I have accomplished. To the University of Delaware Biological Sciences Department and Undergraduate Research Program, thank you for building the framework and providing the funding that made this entire process possible. I learned a great deal outside of the classroom over the past three years, and few other universities offer the support and resources to do so. Finally, to my friends and family, thank you all for the unconditional love and support, whether it be listening to my poster presentation countless numbers of times or pretending to care as I ramble about bacteria at the dinner table. Although he was a businessman and not a scientist, I've learned that Henry Ford was correct in saying, "Whether you think you can, or you think you can't--you're right."

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT	ix
 1 INTRODUCTION	 1
1.1 Introduction to <i>Vibrios</i>	1
1.2 <i>Vibrio cholerae</i> : Background and history	3
1.3 <i>Vibrio cholerae</i> : A Public Health Threat	4
1.4 Non-O1/Non-O139 <i>Vibrio cholerae</i>	6
1.5 Type III Secretion System (T3SS) in <i>Vibrio</i> species	7
1.6 Pathogenicity Islands	10
1.7 Integrase Background	13
 2 MATERIALS AND METHODS	 15
2.1 Bacterial Strains, Plasmids, & Growth Conditions	15
2.2 Mutant Construction	16
2.2.1 PCR amplification and cloning	16
2.2.2 Bacterial Transformation	17
2.2.3 Conjugation Protocol	19
2.3 Excision Assays	21
2.4 Bioinformatics	23
2.4.1 Annotation of VPI-3 in ARTEMIS	23
2.4.2 ARTEMIS Comparison Tool (ACT) Alignments	24
2.4.3 Construction of Phylogenetic Trees	25
 3 RESULTS	 27
3.1 Structure of the Type III Secretion System Containing PAI in NRT36S	27
3.2 Excision of VPI-3	29
3.3 ORFs within the NRT36S VPI-3 Region	32
3.4 Presence of VPI-3 in other <i>V. cholerae</i> strains	35

3.5	Discovery of a repeated region of homology in T3SS ⁺ strains of <i>V. cholerae</i>	38
3.6	Phylogenetic Analysis of the T3SS ATPase of VPI-3	42
4	DISCUSSION.....	46
4.1	Excision behavior of T3SS containing PAI	46
4.2	Evidence for the mosaic evolution of <i>V. cholerae</i> PAIs	46
4.3	Evolution of new effector proteins due to multiple repeated sequences ..	47
4.4	Relevance to the scientific community	48
	REFERENCES	50
APPENDIX		
A	ALIGNMENTS OF REPEATED REGIONS IN AM-19226	59
B	EXAMINING PAI EXCISION IN RECA, HIMA, AND FIS MUTANTS IN <i>V. CHOLERA</i> E	60

LIST OF TABLES

Table 1 List of Plasmids and bacterial strains used in this study.	15
Table 2 Oligonucleotide primers used for PCR in this study. Primers were designed for the study and prepared by Integrated DNA Technologies.....	23
Table 3 VPI-3 Open Reading Frames, gene name, amino acid length, and % homology.....	33
Table 4 Background information of <i>V. cholerae</i> strains used in bioinformatic comparisons of the VPI-3 region.....	37
Table 5 Mutants examined, size of selected gene in wild-type, and size of selected gene	60

LIST OF FIGURES

Figure 1 Relative rates of laboratory-confirmed infections with <i>Campylobacter</i> , Shiga toxin-producing <i>Escherichia coli</i> (STEC) O157, <i>Listeria</i> , <i>Salmonella</i> , and <i>Vibrio</i> in the United States during 1996-2012, compared with 1996-1998 rates, by year.	2
Figure 2 Estimated percentage change in incidence of laboratory-confirmed bacterial and parasitic infections in the United States in 2012, compared with average annual incidence during 2006-2008, by pathogen.	3
Figure 3 Diagram of the type III secretion system (T3SS) needle complex and associated structures.	9
Figure 4 VPI-1 and VPI-2 of <i>V. cholerae</i> N16961.....	12
Figure 5 Conserved domains of phage integrase proteins.....	14
Figure 6 Confirmation of the single crossover event for construction of $\Delta intV2$	20
Figure 7 PCR amplification using AD and flanking primer sets confirmed the deletion of the <i>intV2</i> gene in the mutant strain.....	21
Figure 8 PCR amplification of the <i>attB</i> and <i>attP</i> sites of wild type <i>V. cholerae</i> N16961 and NRT36S strains.....	22
Figure 9 Gene maps of VPI-3 and VPI-2 regions of NRT36S and N16961.	28
Figure 10 Artemis comparison tool (ACT) analysis of differences between VPI-2 and VPI-3 in <i>V. cholerae</i> strains N16961 (top), NRT36S (middle), and AM-19226 (bottom).	29
Figure 11 The two-stage <i>attB</i> PCR amplification assay was used to detect excision of VPI-3 in $\Delta intV2$ and wild-type NRT36S strains.....	31
Figure 12 Conserved domains of the integrase protein present on VPI-3 of <i>V. cholerae</i> strain CP1115.	31
Figure 13 Artemis comparisons of the T3SS containing PAIs of 10 <i>V. cholerae</i> strains.....	37

Figure 14 Examination of the repeated region at 22.4 kb between <i>V. cholerae</i> strains NRT36S and AM-19226.....	40
Figure 15 Comparison of the repeated region of NRT36S to the corresponding region of 9 different <i>V. cholerae</i> strains.....	41
Figure 16 A. Phylogenetic tree of T3SS ATPases in <i>V. cholerae</i> strains with an associated integrase protein encoded on a PAI.	44
Figure 17 Phylogenetic tree of the T3SS ATPase protein in multiple species of gram-negative bacteria.	45
Figure 18 BLAST alignments of the 201 base pair repeated region found in <i>V. cholerae</i> AM-19226	59
Figure 19 Diagram of excision related proteins examined in this study.	61
Figure 20 Growth curve analysis of mutants included in this study.	61
Figure 21 VPI-1 <i>attB1</i> Excision Assay Results.....	62
Figure 22 VPI-2 <i>attB2</i> Excision Assay results.	62

ABSTRACT

Vibrio cholerae is an enteric pathogen that is the causative agent of the secretory diarrhea, cholera, that affects millions of people each year. While the *V. cholerae* O1 serogroup pathovar strains are well studied due to their propensity to cause epidemic and pandemic cholera, a second pathovar has been identified that causes inflammatory diarrhea. Strains of this pathovar encode a Type III Secretion system (T3SS) that is present on a Pathogenicity Island (PAI), a mobile genetic element integrated into the chromosome of some non-O1 serogroup strains. Our hypothesis is that this PAI was horizontally acquired and was an essential acquisition in the emergence of this pathovar. The aims of this study were to investigate the genetic structure of T3SS island regions, examine the excision behavior of the region in strain NRT36S and reconstruct the evolutionary history of the region. To accomplish this, we first performed a bioinformatics analysis among a group of strains that contain a T3SS. We constructed a genetic deletion of cognate T3SS island integrase, *intV2*, and determined the excision phenotype of the island using a two stage nested PCR assay. We showed that *intV2* is necessary for the excision of the region. Lastly, we performed a phylogenetic analysis of the conserved T3SS ATPase protein and associated integrases to show unique clustering patterns, indicating that this island is mosaic in structure. Additionally, phylogenetic analysis revealed the presence of two different variants of T3SSs within *V. cholerae*, T3SS $\square$ and T3SS $\square$. The grouping pattern on the tree showed a close relationship between the T3SS ATPase of *V. cholerae* NRT36S, *V. parahaemolyticus*, and *V. mimicus* suggesting that this virulence

system may have been passed horizontally among these different species of bacteria in the past.

Chapter 1

INTRODUCTION

1.1 Introduction to *Vibrios*

The genus *Vibrio* is a diverse group of aquatic bacteria in the Class Gamma-Proteobacteria. These bacteria thrive in marine, estuarine, or riverine environments and can be free swimming or survive due to commensal or pathogenic relationships with aquatic organisms from fish to phytoplankton (Romalde et al., 2013, Paillard et al., 2004). While most *Vibrio* species thrive in the mesophilic temperature range, some *Vibrio* species have also been found in extreme environments such as hydrothermal deep-sea vents (Colwell 2006). *Vibrios* have fermentative metabolic processes and it is thought that some environmental bacteria in this genus play an important role in the nutrient cycling and nitrogen fixation of the ecosystems in which they live (Colwell 2006). Although the importance of environmental *vibrios* is clear, the scientific community has typically focused on pathogenic species that pose a threat to human health. In 2012, incidence of infection for reported cases of *Vibrio*-related illness increased by 43% (Centers for Disease Control and Prevention, 2013). Comparatively, incidence rates for other foodborne pathogens such as *Listeria*, *Salmonella*, and *Shigella* did not increase during the same two-year time period, as shown in Figures 1 and 2 (Centers for Disease Control and Prevention, 2013). *Vibrio parahaemolyticus*, *Vibrio vulnificus*, and *Vibrio cholerae* have been extensively studied because they cause human illness.

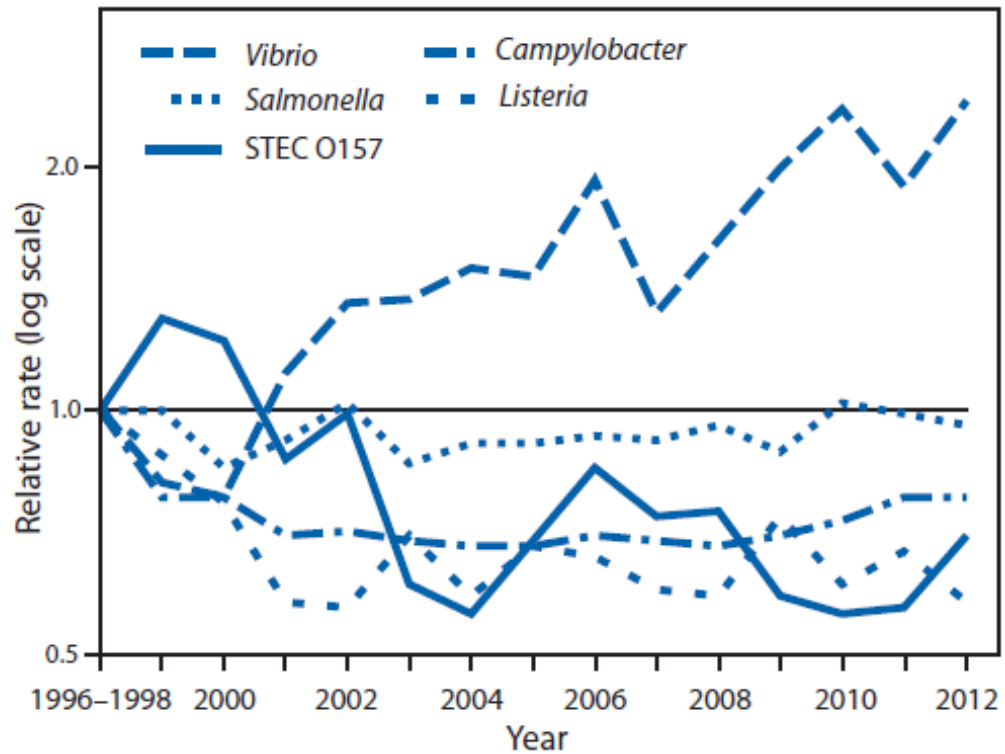


Figure 1 Relative rates of laboratory-confirmed infections with *Campylobacter*, Shiga toxin-producing *Escherichia coli* (STEC) O157, *Listeria*, *Salmonella*, and *Vibrio* in the United States during 1996-2012, compared with 1996-1998 rates, by year. In comparison with 1996-1998, incidence of infection was significantly lower for *Campylobacter*, *Listeria*, *Shigella*, STEC O157, and *Yersinia*, whereas incidence of *Vibrio* infection was higher (Centers for Disease Control and Prevention, 2013).

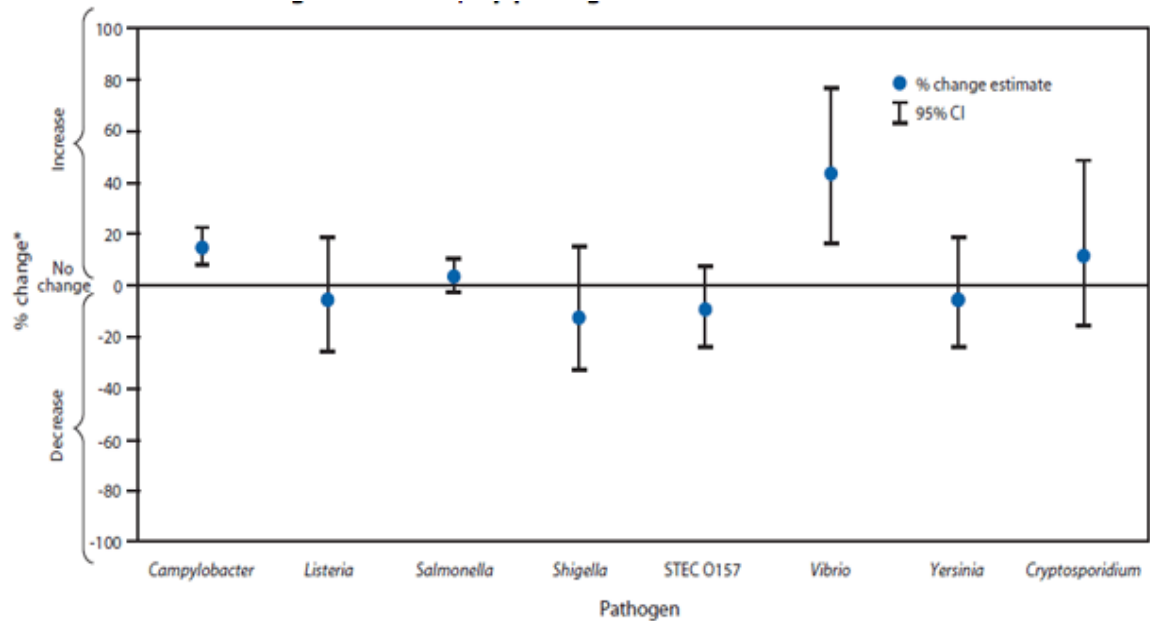


Figure 2 Estimated percentage change in incidence of laboratory-confirmed bacterial and parasitic infections in the United States in 2012, compared with average annual incidence during 2006-2008, by pathogen. The estimated incidences of infection were higher in 2012 compared with 2006-2008 for *Campylobacter* (14% increase; 95% confidence interval: 7%-21%) and *Vibrio* (43% increase; 95% confidence interval: 16%-76%) and unchanged for other pathogens (Centers for Disease Control and Prevention, 2013).

1.2 *Vibrio cholerae*: Background and history

Sanskrit records dating from around 500 BC describe one of the first recorded cases of cholera. The origin of the name of the disease is thought to come from the ancient Greek words “chole” meaning bile and “rein” meaning flow (Colwell 2006). To follow the historical epidemiology of cholera is to follow the course of human history. It has been speculated that cholera, or a disease like it, affected the outcome of historically important battles such as Gallipoli, El Alamein, and the Oman war (Colwell 1996). The second cholera pandemic, which began in Russia in 1829, spread across the globe to New York City. The arrival of this pandemic to London in 1849

sparked the birth of the public health field of epidemiology (Colwell 1996). Modern sanitation technologies have reduced the number of global instances of cholera, but the seventh pandemic, which began in 1961 in Indonesia and spread to South America in 1991, is still occurring today (Colwell 1996). Developing areas with poor sanitation, crowded living conditions, and low hygiene standards, or countries experiencing natural disasters and turmoil, are more likely to report outbreaks or epidemics of cholera. In today's global world, an epidemic can more easily become a pandemic. In 1996, Colwell et al. showed a correlation between the sea surface temperatures surrounding Bangladesh and the number of cases of cholera in the region. Thus, as the world's oceans continue to warm due to climate change, cholera may become an even greater public health threat.

1.3 *Vibrio cholerae*: A Public Health Threat

Over the past 20 years, the incidence of most food and water borne bacterial infections has either decreased or remained constant. This fact does not hold true for *Vibrio* infections, which have steadily increased in incidence since 2000 (WHO, 2014). In 2013, 47 countries across the globe reported 129,064 cases of cholera to the WHO. 2,102 of these cases were fatal, equaling a case fatality rate of 1.63% (WHO, 2014). *V. cholerae* is often misreported or underreported; therefore, the global burden of cholera is likely much greater than is represented by the data available. An estimated 1.4-4.3 million cases of cholera occur each year, leading to up to 142,000 un-reported deaths (WHO, 2014).

Researchers are frequently isolating newly evolved strains of *V. cholerae*. *V. cholerae* is naturally competent (able to take up DNA) in the presence of chitin, a substance that constitutes the shells of many sea-dwelling crustaceans and bivalves

(Meibom et al., 2005). Thus, *V. cholerae* in their natural environment are able to take up foreign DNA and potentially integrate this DNA into their chromosomes (Meibom et al., 2005). Newly evolved strains pose a greater public health threat, especially in areas where cholera is endemic, such as India and Bangladesh, because the acquired immune response does not protect against novel strains. Some strains isolated from these regions also demonstrate multi-drug antibiotic resistance (Faruque et al 2007). While antibiotics are usually used only in the most severe cases of cholera infection, resistance to drugs commonly employed in rural or impoverished communities could potentially increase the case fatality rate in these areas where cholera outbreaks are common (WHO, 2014).

As of 2013, there are two vaccinations used to mitigate the effects of a cholera outbreak. Dukoral and Shanchol are approved by the World Health Organization and are Oral Cholera Vaccines (OCVs). Unlike vaccines for the measles or chickenpox that confer lifetime immunity to the infection, these two OCVs only provide limited immunity. Dukoral protects against O1 serogroup strains of *V. cholerae*. 85% of recipients gain immunity to infection for 4-6 months and 60% gain immunity for up to 2 years. Additionally, Dukoral must be given in a minimum of two doses, a problem for public health professionals working in areas where patient follow-up is extremely difficult (“Weekly Epidemiological Record: Cholera 2013,” 2014). Shanchol is another OCV option that protects against both O1 and O139 serogroups of *V. cholerae*. 67% of patients administered 2 doses of Shanchol gain immunity for 2 years with no booster doses, and 65% are protected for 5 years (CDC, 2014). Due to the fact that recent outbreaks and the use of vaccines in areas with endemic cholera have depleted the global vaccine stockpile, the WHO notes that OCVs are primarily used as

a form of outbreak control rather than as a form of primary prevention (“Weekly Epidemiological Record: Cholera 2013,” 2014).

1.4 Non-O1/Non-O139 *Vibrio cholerae*

Each pandemic of cholera to date has been caused by O1 or O139 serogroup strains of *V. cholerae*. Due to the obvious health threat that they pose to the public, O1/O139 strains have also been the predominant strains focused on by researchers studying *V. cholerae*. However, non-O1/O139 toxigenic serogroups of *V. cholerae* have emerged that cause disease symptoms such as inflammatory diarrhea, gastroenteritis, wound infections, and septicemia (Ceccarelli et al., 2015). Non-O1/O139 strains have been associated with individual infections or localized outbreaks, especially in locations where cholera is endemic (Chatterjee et al., 2009). These genetically diverse strains do not carry the cholera toxin (CT) or toxin co-regulated pilus (TCP) genes, but can encode a variety of genes that contribute to alternative methods of virulence. Many non-O1/O139 strains encode genes for hemolysin, protease, mobility, Type VI Secretion systems (T6SS), Type III Secretion systems (T3SS), and biofilm formation proteins (Chatterjee et al., 2009). While very few non-O1/O139 isolates secrete the cholera toxin, some isolates encode genes for various other toxins including the MARTX toxin, a multifunctional autoprocessing repeats-in-toxin, *rtxA*, heat-stable enterotoxin (Nag-St), and the Cholix toxin (Schirmeister et al., 2014, Ceccarelli et al., 2015). A 2013 study conducted in the Chesapeake Bay area discovered that between 5-10% of all *Vibrio* infections are due to *V. cholerae* non-O1/O139 strains (Jones et al., 2013). Furthermore, it has been observed that non-O1/O139 *V. cholerae* isolates are more likely to be resistant to

penicillin and ampicillin, as well as sporadically resistant to other commonly used antibiotics such as erythromycin (Ceccarelli et al., 2015).

V. cholerae NRT36S is a serogroup O31 strain that is cholera toxin (CT) negative and heat-stable enterotoxin (Nag-ST) positive (Chen et al, 2007). This strain causes severe inflammatory diarrhea in patients, which is different from the secretory diarrheal symptomatic of infection with O1/O139 strains that encode CT, but can be just as severe (Dziejman et al., 2005, Chen et al, 2007). *V. cholerae* NRT36S, relies on a T3SS to cause disease in its host due to its lack of CT and Toxin coregulated pilus (TCP) genes (Shin et al., 2011).

1.5 Type III Secretion System (T3SS) in *Vibrio* species

Bacterial T3SSs are contact-dependent protein delivery organelles. A T3SS is a needle-like structure anchored in the bacterial cell membranes and spans the gap between the bacterial cell and its host (Galan et al., 2014). After making contact with a eukaryotic host cell, the T3SS can actively pump proteins, called effector proteins, from the bacterial cell cytosol through the host's plasma membrane into the host cytosol. Once in the host cytosol, effector proteins can disrupt multiple cellular processes such as actin remodeling or signal transduction, causing disease (Galan et al., 2014). The presence of a T3SS can also cause damage in the absence of effector proteins. It has been shown that the translocation of the needle tip of the *Pseudomonas aeruginosa* T3SS into the host cell membrane is enough to cause host lung cell damage (Audia et al., 2013). This machinery is thought to be evolutionarily derived from the structure of the bacterial flagella due to the fact that the architecture of the membrane components resembles those of the bacterial flagella (Figure 3) (Galan et al., 2014). Additionally, *Mycoccus xanthus* bacteria display an intermediate non-

functional flagella-like structure that is capable of secreting proteins, providing experimental evidence for this theory (Konovalova et al., 2010, Galan et al., 2014). Of the proteins included in the T3SS apparatus, it has been found that both the inner membrane export apparatus proteins and the cytosolic component proteins are highly conserved, even amongst T3SSs of different species and genera (Galan et al., 2014). Of the conserved cytosolic proteins, the T3SS ATPase protein is often used for evolutionary and phylogenetic analysis due to its highly conserved nature.

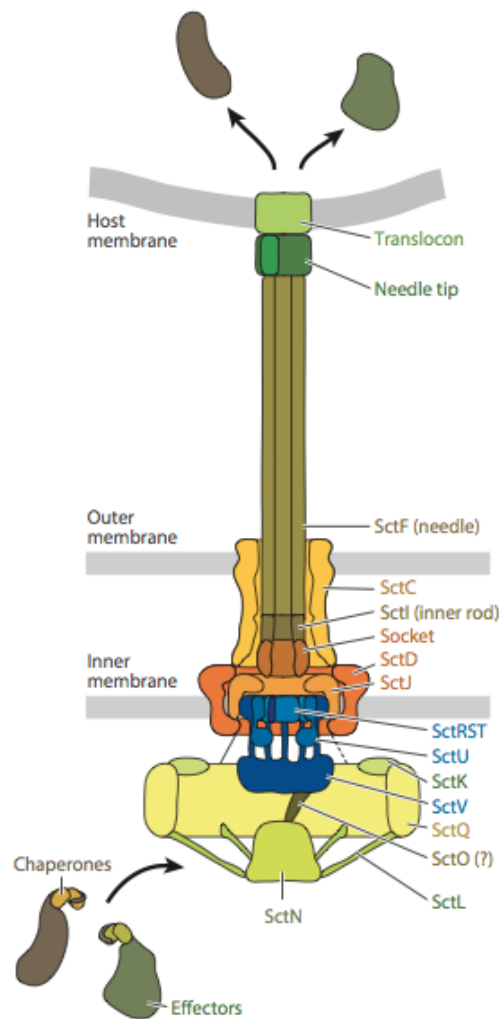


Figure 3 Diagram of the Type III Secretion System (T3SS) needle complex and associated structures. SctN indicates T3SS ATPase in this diagram (Galan et al., 2014).

T3SS play a crucial role in the pathogenesis of non-O1/O139 *V. cholerae* species. In the absence of the CT and TCP genes, these strains must utilize other means of virulence. T3SS genes have been found in clinical strains isolated from patients with diarrhea that lack CT and TCP virulence factors (Dziejman et al., 2005).

In addition to causing disease through the translocation of effector proteins, it has also been shown that the T3SS contributes to effective colonization in the absence of TCP genes in the infant mouse model (Dziejman et al., 2005). In fact, in *V. cholerae* AM-19226, a T3SS positive strain, both colonization and the secretory diarrheal response were found to be just as severe as infections with the cholera causing strain N16961 in the infant rabbit colonization model (Dziejman et al., 2005).

In this study we use *V. cholerae* NRT36, which encodes a T3SS that shares 99% sequence similarity to that of strain AM-19226 (Chen et al., 2007). A functional T3SS is required for the virulence of *V. cholerae* AM-19226 and presumably NRT36S (Shin et al., 2011). In NRT36S, the T3SS genes are encoded on a pathogenicity island that bears significant homology to VPI-2 of *V. cholerae* N16961 (Chen et al., 2007, Murphy and Boyd, 2008).

1.6 Pathogenicity Islands

Pathogenicity islands (PAIs) are large regions of DNA that contain virulence factors in the chromosome of pathogenic isolates and are absent from non-pathogenic isolates. (Hacker & Kaper, 2000). Pathogenicity islands have the following main characteristics that distinguish them from core chromosomal DNA. PAIs have a codon usage and a GC content that is different from that of the host core chromosome, and they integrate into the host chromosome via site specific integration at a tRNA gene. This process is mediated by an integrase and results in direct repeat sequences flanking the integrated PAI (Hacker & Kaper, 2000). These features indicate that PAIs have been acquired by horizontal gene transfer. PAIs can include genes that encode alternate metabolic proteins, confer antibiotic resistance, or increase bacterial virulence such as the T3SS. PAIs can also encode genes for self-regulation such as

toxT and *tcpPH* present in the *Vibrio* pathogenicity island-1 (VPI-1) (Hacker & Kaper, 2000). Although not self-mobilizable like a phage, PAIs are able to excise out of the core chromosome of the bacteria under certain conditions such as stress, ultra-violet light, or antibiotic exposure (Almagro-Moreno, Napolitano, & Boyd, 2010). Excision occurs when two direct repeat sites on the left and right sides of the island, *attL* and *attR* respectively, recombine and the island exists as a circular intermediate outside of the core genome of the bacteria (Rajanna et al., 2003). When these islands occur in a circular intermediate state, they can be transferred to other bacteria via conjugation or transduction (Hacker & Kaper, 2000). Throughout the course of evolution, PAIs can lose mobility genes such as integrase proteins and recombination directionality factors (RDFs) and become a permanently integrated part of the core chromosome of the bacteria, as can be seen in some *Salmonella* and *E. coli* spp (Marcus, Brumell, Pfeifer, & Finlay, 2000). This occurs when evolution favors the selection of bacteria containing the PAI due to the increased fitness it confers to the bacteria, either in terms of metabolism or other functions (Hacker & Kaper, 2000).

Pathogenicity Islands have been characterized in both *E. coli* and *Salmonella* species. In *Salmonella*, T3SS machinery has been found to be associated with PAIs (Galan, 1999). Additionally, PAIs present in uropathogenic and EPEC strains of *E. coli* were some of the first PAIs studied (Hacker et al., 1997). Consistent with what was later found in the PAIs of *V. cholerae*, *E. coli* PAIs also integrate site-specifically near tRNA synthase genes (Hacker et al., 1997). In EPEC strains of *E. coli*, the PAI present is responsible for the enterohaemorrhagic effects of the bacteria due to the presence of a T3SS on the island (Hacker et al., 1997, Marcus et al., 2000).

In choleraogenic *V. cholerae* serogroup O1 and O139 isolates, the *Vibrio* pathogenicity island-2 (VPI-2) region is integrated at a tRNA-ser gene and contains genes for an integrase, a type I restriction modification system (RM), a sialic acid catabolism (SAC) region, a phage-like region, and two RDFs named *vefA* and *vefB* (Figure 4) (Jermyn & Boyd, 2002). Although it was originally thought that this region was only present in choleraeagenic strains of *V. cholerae*, homologous integrase, RDF, and SAC proteins have been found in AM-19226 and NRT36S, among other strains (Chen et al., 2007, Dziejman et al., 2005, Murphy & Boyd 2008, Jermyn & Boyd 2005). It is theorized that the two versions of VPI-2, one encoding a RM system and a phage, and one encoding a T3SS, have a common ancestor and have diverged throughout the course of evolution.

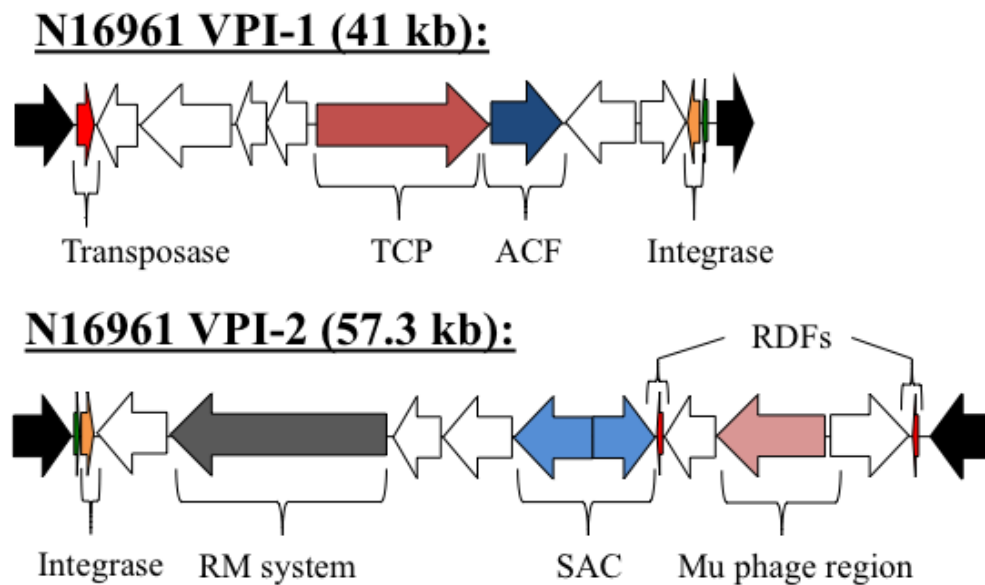


Figure 4 VPI-1 and VPI-2 of *V. cholerae* N16961. The genes shown in orange represent the cognate integrase of the islands, *intV1* and *intV2* respectively. Shown in green is the tRNA synthase locus of integration.

1.7 Integrase Background

The excision of PAIs is mediated by the function of a protein called an integrase. Integrases are site-specific recombinase proteins that are necessary for the integration and excision of the PAI from the bacterial chromosome (Rajanna et al., 2003, Murphy & Boyd, 2008, Almagro-Moreno, Napolitano, & Boyd, 2010). Phage integrase proteins have three conserved domains, a domain of unknown function (DUF) that is commonly found on the N-terminus of phage integrase proteins, a phage integrase sterile alpha-like motif (SAM) that is also related to the n-terminus of the protein, and the phage integrase family domain (Figure 5) (Marchler-Bauer et al., 2015). The phage integrase family domain is located at the C-terminus of the protein, and houses the active site where double stranded DNA is cut during integration and excision (Marchler-Bauer et al., 2015). PAI excision has been found to be dependent on integrase proteins in both *E. coli* and *Yersinia pseudotuberculosis* strains as well as in *V. cholerae* (Lesic et al., 2004, Hochhut et al., 2006). In *V. cholerae* VPI-1, Rajanna et al. found that the island could excise without the cognate integrase, *intV1*, present (Rajanna et al., 2003). This could be due to the presence of *vpiT*, a transposase gene encoded on the island, or possibly crosstalk between VPI-1 and *intV2*-- the integrase protein of the second PAI present in *V. cholerae*, VPI-2 (Rajanna et al., 2003, Carpenter and Boyd, unpublished data) (Figure 4). It was also found that while inactivation of *intV1* did not abolish excision of VPI-1, the complementation of the gene on a plasmid in the mutant increased the rate of excision of the island (Rajanna et al., 2003). This result was not consistent with what is found in VPI-2. When the cognate integrase of VPI-2, or *intV2*, is deleted, excision of the island no longer occurs (Murphy & Boyd, 2008, Almagro-Moreno, Napolitano, & Boyd, 2010). The lack of a transposase gene on VPI-2 could explain this key difference in excision behavior

between two PAIs in *V. cholerae*. Two of the aims of this study focus on the cognate integrase of VPI-3, the VPI-2 variant containing a T3SS in strain NRT36S. The first aim is to show that, like VPI-2, the integrase encoded on VPI-3 is necessary for excision. The second aim of this research is to study the evolution of T3SS containing PAIs using the sequence of the cognate integrase encoded on the island. Integrase sequence has been previously shown to be a reliable target of analysis in *E. coli* PAI evolution (Napolitano et al., 2011).

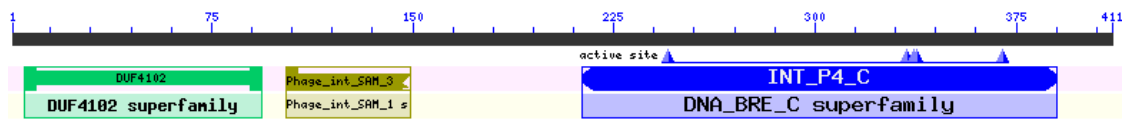


Figure 5 Conserved domains of phage integrase proteins. Figure adapted from the NCBI conserved domains database (Marchler-Bauer et al., 2015).

Chapter 2

MATERIALS AND METHODS

2.1 Bacterial Strains, Plasmids, & Growth Conditions

Bacterial strains were grown in 3mL lysogeny broth (LB) (Fisher Scientific) overnight (12 hours), at 37° C with shaking at 225 rpm in aerobic conditions unless otherwise noted. In the case of *E. coli* β 2155 Δ *pir*, the LB media was supplemented with Diaminopimelic acid (DAP) to a final concentration of 0.3 mM. When needed, the media was also supplemented with antibiotics in the following concentrations: 200 μ g/ml streptomycin (Sm) (Fisher Scientific), 25 μ g/ml chloramphenicol (Cm) (Fisher Scientific) and 100 μ g/ml ampicillin (Amp) (Fisher Scientific). LB agar was supplemented with 10% sucrose for double cross screening.

Table 1 List of Plasmids and bacterial strains used in this study.

Strain Name	Background	Source
<i>V. cholerae</i> Strains:		
NRT36S	O31, VPI-2, partial VSP-II, Sm ^R	Chen et al. 2007
NRT36S Δ <i>intV2</i>	NRT36S, Δ VCI758, Sm ^R	This study
N16961	O1, El Tor strain, VPI-1, VPI-2, VSP-I, VSP-II, Sm ^R	Heidelberg et al., 2000
<i>E. coli</i> strains:		
β 2155 Δ <i>pir</i>	Donor bacteria in conjugation experiment, DAP auxotroph, produces the π protein needed for plasmid replication outside of the bacterial chromosome	
DH5 α <i>pir</i>	Laboratory designed strain	

	of <i>E. coli</i> for cloning, high transformation frequency, produces the π protein needed for plasmid replication outside of the bacterial chromosome	
Plasmid strains:		
pJET1.2	cloning vector, Am ^R	Fermentas
pJET1.2 Δ intV2	Cloning vector, Am ^R , contains truncated Δ VC1758 product	This study
pDS132	Suicide Vector, SacB, Cm ^R	Philippe, Alcaraz et al. 2004
pDS132 Δ intV2	pDS132 containing truncated Δ VC1758 product, Cm ^R	This study

2.2 Mutant Construction

2.2.1 PCR amplification and cloning

The truncated gene was created using the Splicing by Overlap-Extension (SOE) method of PCR (Horton et al., 1989). The restriction enzyme cut sites used in this primer design were SacI (GAGCTC) and XbaI (TCTAGA). All primers were designed to create an in-frame gene deletion and were ordered from Integrated DNA Technologies. 2.0 μ l of 1x dNTPs, 5.0 μ l of 10x PCR Buffer, 2.0 μ l of each primer, and 2.38 μ l of Taq polymerase was used in each PCR reaction (Fisher), 1 μ l of 10 ng/ μ l template DNA was also added to all reactions besides the control. In the SOE method, the AB and CD products of the gene are amplified individually in a 50 μ L PCR reaction. The PCR conditions used to amplify the AB product were: 95°C for 5 mins, 29 cycles of 94° C for 30 sec, 48.3° C for 30 sec, and 72° C for 30 sec, 72° C for 7 min. Conditions used to amplify the CD product were: 95°C for 5 mins, 29 cycles of 94° C for 30 sec, 58.3° C for 30 sec, and 72° C for 41 sec, 72° C for 7 min. The AB

primers (VC1758A & NRTintv2B) amplified a region of 523 base pairs, and the CD primers (NRTintv2C & NRTintv2D) amplified a region of 678 base pairs. These products were used as DNA template in a third 50 μ L PCR reaction, where the reverse complement of primer B, which is included in the sequence of primer C, anneals the AB and CD products together. The conditions used for this third PCR were: 95°C for 5 mins, 29 cycles of 94° C for 30 sec, 55.4° C for 30 sec, and 72° C for 1:15, 72° C for 7 min resulting in formation of a truncated intV2 gene product. This truncated 1,161 base pair product is then cloned into the cloning vector pJET1.2 using the blunt end protocol (ThermoScientific). The blunt end protocol was used so that the DNA fragment can ligate directly into pJET1.2. A 1:3 vector to insert reaction was set up using 10 μ L reaction buffer, 1 μ L of T4 DNA ligase enzyme, and 19 μ L of water. This reaction was kept at 16°C for 14 hrs to ensure optimal ligation of the truncated Δ intV2 gene into pJET1.2

pJET1.2 is a plasmid designed to facilitate the cloning of DNA fragments and the positive selection of bacteria containing the truncated gene product. The plasmid contains the *rep* gene that is responsible for the replication of the plasmid, an origin of replication, the *bla* gene that confers ampicillin resistance to bacteria containing the plasmid, and the *eco47IR* gene that is lethal to the plasmid unless disrupted. The multiple cloning site of the plasmid is located within the *eco47IR* gene; therefore, when the truncated Δ intV2 gene is ligated into pJET1.2, this lethal gene is disrupted and the plasmid survives.

2.2.2 Bacterial Transformation

Once Δ intV2 is ligated into pJET1.2, it is then necessary to transform the plasmid pJET1.2 Δ intV2 into *E. coli* DH5 α , an *E. coli* strain with a high

transformation frequency. pJET1.2 $\Delta intV2$ is transformed into *E. coli* DH5 α using a CaCl₂ and heat shock reaction. The pDNA enters the *E. coli* cell when the plasma membrane is disrupted at a high temperature (42°) in the presence of CaCl₂. Consequently, all *E. coli* bacteria that obtain the pJET1.2 $\Delta intV2$ plasmid through transformation also acquire ampicillin resistance from the plasmid. This resistance allows the *E. coli* containing the plasmid to be isolated on selective media. Once enough copies of the plasmid containing the truncated gene have been obtained through bacterial replication, pJET1.2 is purified from *E. coli* DH5 α (QIAGEN Plasmid Mini Kit (2005)), and digested at specific cut sites with the SacI and XbaI restriction enzymes (ThermoScientific). The digested plasmid is run out on a 0.8% agarose gel. When viewed under a UV light, there are two bands present on the gel. The larger band is the linear plasmid, and the smaller band is the truncated $\Delta intV2$ product that has been cut out by restriction enzymes. This product is purified from the agarose using gel cut purification. The purified product is then ligated into another plasmid, pDS132 overnight at 16° C (Philippe, Alcaraz et al. 2004).

pDS132 is a suicide vector that cannot extrachromosomally replicate in strains of bacteria that do not encode the *pir* gene. In the case that pDS132 is transformed into a strain of bacteria that does not encode the *pir* gene, such as *V. cholerae* NRT36S, the plasmid can only replicate if incorporated within the chromosome during DNA replication. Additionally, pDS132 contains the *cat* gene that confers Cm resistance. *Cat* encodes the enzyme chloramphenicol acetyltransferase, which detoxifies Cm to the bacteria by attaching an acetyl group to the antibiotic. This prevents the antibiotic from attaching to and disabling the bacterial ribosomes. Once the truncated gene product is ligated into pDS132, pDS132 $\Delta intV2$ is transformed into *E. coli* β 2155 using

the same previously mentioned CaCl_2 transformation protocol. *E. coli* $\beta 2155$ serves as the donor strain in bacterial conjugation with *V. cholerae* NRT36S.

2.2.3 Conjugation Protocol

Conjugation was used in order to transfer pDS132 $\Delta intV2$ from *E. coli* $\beta 2155$ to *V. cholerae* NRT36S. Following overnight growth, *E. coli* $\beta 2155$ was plated on LB agar with 0.3mM of DAP and incubated at 37° C for five hours. *E. coli* $\beta 2155$ is a DAP auxotroph, therefore adding DAP to the media was necessary for growth. However, no antibiotics were spread onto this plate because unlike *E. coli* $\beta 2155$ + pDS132, *V. cholerae* NRT36S is not resistant to Cm. Following this incubation, 150 μL of exponential phase *V. cholerae* NRT36S was plated onto the *E. coli* $\beta 2155$ plate and incubated at 37° C for 12 hours. The lawn of bacterial overnight growth was then resuspended in 3 mLs of LB, and 150 μL was spread on an LB agar plate supplemented with Sm and Cm. The absence of DAP and the presence of Sm on this plate ensured that NRT36S grew, but not *E. coli* $\beta 2155$. The presence of Cm did not allow any *V. cholerae* NRT36S to grow that had not received the pDS132 plasmid. Thus, any bacteria grown on this plate should be *V. cholerae* NRT36S + pDS132 $\Delta intV2$.

Several conjugated colonies were selected and passaged on LB agar + Cm + Sm plates until robust, single colonies were obtained and confirmed by PCR (Figure 6). The double cross was then screened for. The selective media used to identify the double crossover event is LB agar + 10% sucrose + Str. pDS132 contains a sucrose sensitivity gene, *sacB*, which encodes the enzyme levansucrase. This enzyme catalyzes the synthesis of levans, which are fructose polymers toxic to *V. cholerae*. If the plasmid is incorporated into the genome of NRT36S, as it is after the single cross

has occurred, the bacterial colonies will appear lysed and mucoid in the presence of sucrose due to the presence of a functional *sacB* gene. In the event of the double crossover, the plasmid is removed from the core genome of the bacteria leaving only the truncated AD product. If the plasmid is no longer integrated, the bacteria are no longer sensitive to sucrose and their growth should appear phenotypically normal on the selective LB+ 10% sucrose media. Both PCR analysis and genomic sequencing (Genewiz) confirmed the expected deletion of the 1,108 base pairs between primer B and primer C in the IntV2 region of NRT36S (Figure 7).

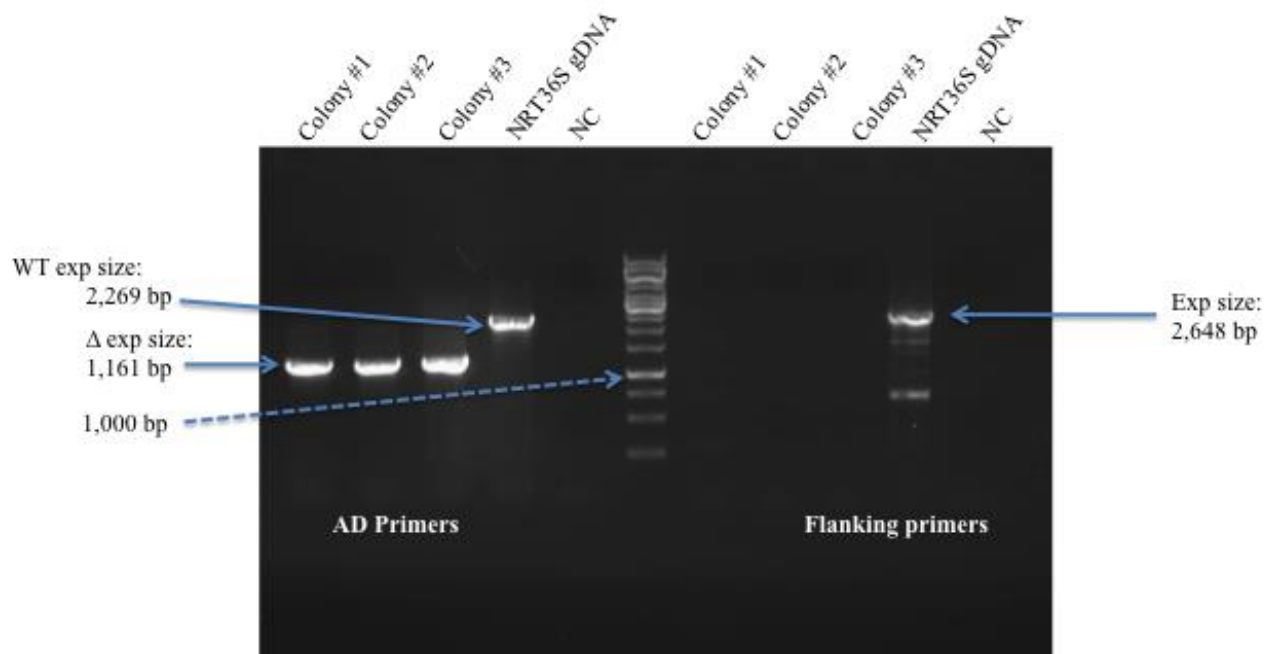


Figure 6 Agarose gel confirming the occurrence of the single crossover event. Because the plasmid becomes incorporated into the core chromosome of NRT36S after single crossover occurs, the PCR product using the flanking primers is too large to amplify using PCR.

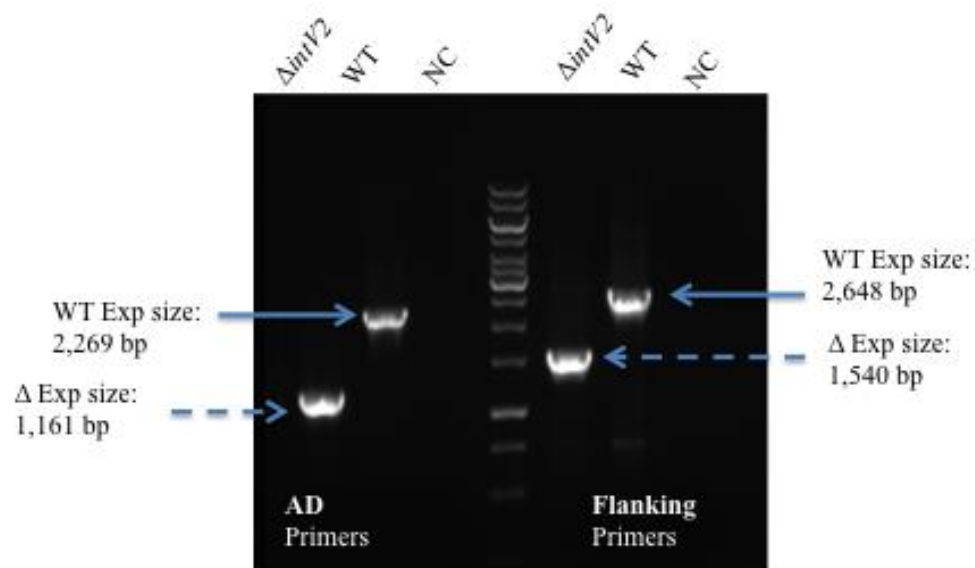


Figure 7 PCR amplification using AD and Flanking primer sets confirmed the deletion of the *intV2* gene in the mutant strain. Amplification of a correctly sized product by the flanking primers indicates that the double crossover event has occurred and the plasmid is no longer incorporated into the bacterial chromosome.

2.3 Excision Assays

In order to detect whether the excision of VPI-2 has occurred in the absence of *IntV2*, a two-stage nested PCR was done to amplify the *attB2* site. After one round of PCR, no *attB2* product can be seen on the gel; however, after two rounds a strong band can be visualized if excision has occurred (Figure 8). The two stages of this assay are necessary due to the fact that the excision of the PAI is rare, and a more sensitive assay is needed to detect the genomic *attB2* site that is left behind after the island excises. The first round PCR conditions to detect *attB2* are 95° C for 5 min, then 30 circuits of 94° C for 30 sec, 60° C for 30 sec, 72° C for 1 min 30 sec, concluded by 72° for 5 min. The primers used in this round were VPI-2fk1 and VPI-2fk2 (Table 2). A 1:10 dilution of the genomic DNA used in this assay was made and heated at

60°C for 20 min. After heating, the concentration of the 1:10 diluted gDNA was taken with the Nanodrop 2000. A subsequent dilution was made after the concentration was taken in order to bring the final concentration of the gDNA to 10 ng/μl. 1 μl of gDNA is used as template in the PCR reaction. The PCR conditions for the second round are as follows: 95° C for 5 min, then 30 circuits of 94° C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec, concluded by 72° for 5 min. The primers used in this second round were VPI-2attF and VPI-2attR (Table 2). 1 μl of PCR product from the first round of PCR was used as template for the second round of the excision assay. The sequences for all primers used in these excision assays can be found in Table 2.

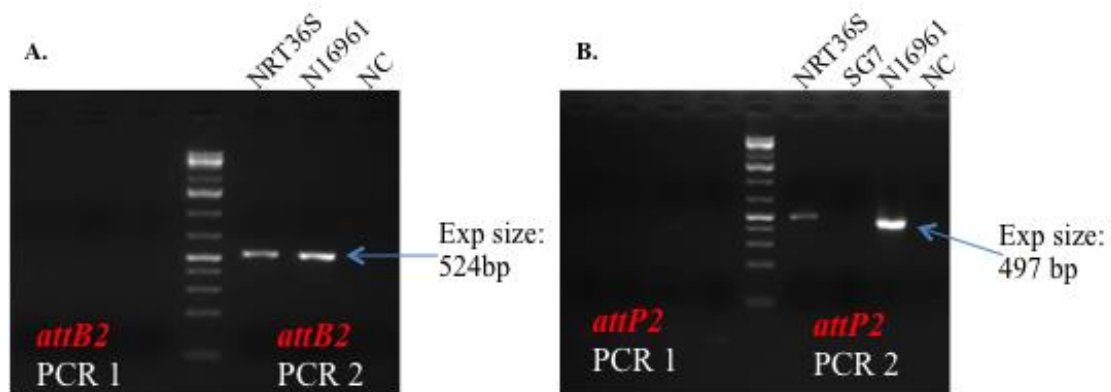


Figure 8 PCR amplification of the **A.** *attB* and **B.** *attP* sites of wild type *V. cholerae* N16961 and NRT36S strains. Due to the low excision rate of this island, no product from PCR1 was detected; therefore, product from PCR 1 was used as template for PCR 2 to amplify *attB2* and *attP2*. In **B.**, *V. cholerae* SG7 was included as a negative control, as SG7 does not contain VPI-2. These results demonstrated that VPI-3 of NRT36S can excise from the chromosome and can be detected through our *attB2* and *attP2* assays

Table 2 Oligonucleotide primers used for PCR in this study. Primers were designed for the study and prepared by Integrated DNA Technologies

Primer Name	Sequence	T _m (°)	GC(%)
VC1758FF	AAGCAAACGCACTCAATGCG	58	50
NRTintV2FR	GGATTTCTGCCTACTACCGT	54	50
VC1785A*	TCTAGAGATTCGGTGAGTTGTCCGAG	56	55
NRTintV2B	CATGAGCGAGAATTACTTGG	52	45
NRTintV2C**	<i>CCAAGTAATTCTCGCTCATGAAGCTACAGTGTCGCTGGTG</i>	58	55
NRTintV2D***	GAGCTCGCCTGTGGCGAGATAGAGAC	58	60
VPI-2fk1	CTGGCTATGAGCTGATTTG	51	47
VPI-2fk2	AGGGATTGGCTTTGAGG	52	53
VPI-2attF	AGAGTGAAAGTCGCCAAAG	53	47
VPI-2attR	GGTGCAATTTTCGCATGTTGC	59	52

* Bases in bold represent XbaI restriction enzyme cut sites

** Bases in italics represent the reverse complement of primer NRTintV2B

*** Bases in bold represent SacI restriction enzyme cut sites

2.4 Bioinformatics

2.4.1 Annotation of VPI-3 in ARTEMIS

The Artemis program is a bioinformatics tool developed by the Sanger Institute to analyze genome sequence data. Genbank files can be imported from outside databases into Artemis, which allows the user to edit these sequences or perform other functions such as analysis of RNA sequencing data, calculation of GC percentage, and also calculation of certain statistics such as the percent of pseudogenes, gene density, and protein coding bases (Rutherford et al., 2000).

In this study Artemis was used to annotate VPI-3 in *V. cholerae* NRT36S from raw whole genome sequence data. The FASTA sequence of the island was assembled from unordered contig level data obtained from Colin Stine's lab using the homologous region of *V. cholerae* AM-19226 as a guideline to determine gene order. This was particularly useful when genes were located sequentially on AM-19226, but on separate contigs in NRT36S. This FASTA file was then imported into Artemis, which detected the open reading frames (ORFs) in the sequence. Each ORF was then run through NCBI's protein BLAST algorithm and annotated in one of three ways. If the ORF was located in the T3SS region, it was annotated using the locus tag of the most homologous gene in *V. cholerae* AM-19226. If the ORF was located in the Sialic Acid Catabolism region of the island, it was annotated using the locus tag of the most homologous gene in *V. cholerae* N16961. If the ORF did not blast to a protein from either of these strains, the locus tag of the protein from the BLAST result with the highest homology and query cover was chosen for annotation. Recombination directionality factors *vefA* and *vefB* were identified by searching the sequence manually due to the fact that the ORFs were too small for the default settings of Artemis to find in the initial ORF search. The annotation was then saved in Genbank file format for further use.

2.4.2 ARTEMIS Comparison Tool (ACT) Alignments

Artemis is designed primarily for the viewing, analysis, and annotation of a single sequence, in comparison Artemis Comparison Tool is used for comparing and viewing homology between multiple large genomic sequences. Generating an ACT comparison, such as the one shown in Fig. 3, requires 2 different software programs. ACT allows the user to view a comparison and an external program is used to generate

a blastout or crunch file that can then be imported into ACT for viewing. Sanger offers a web based platform called WebACT that accomplishes this task and allows the user to download the blastout files to save for use with ACT (www.webact.org). On this web-based platform, FASTA files for up to 5 sequences can be uploaded for analysis and the program will generate 4 sequential blastout files. This means that a specific sequence is only compared to those sequences directly before and after it. When the sequences and the comparison files are loaded into ACT, the program generates the visual comparison. In ACT a red band connecting two sequences indicates a region of homology between the two sequences. The “intensity” of the band will vary depending on the percent of homology between the two regions (Carver et al. 2005). A blue band between the two sequences indicates a reverse match. Each ORF in NRT36S was given a number from 1 (*intV2*) to 70 (*vefB*), which corresponds to the numbers labeling the ORFs in Table 3. Annotations for *V. cholerae* N16961 and AM-19226 were taken in Genbank file format from the NCBI database and loaded into ACT. Nine additional strains with T3SS containing PAIs were selected for analysis with ACT (Table 4). The FASTA sequence of VPI-3 for each of these strains was downloaded from either the NCBI or IMG database. Background information for each of these strains including serogroup, location of isolation, year of isolation, and accession number for the genome can be found in Table 4.

2.4.3 Construction of Phylogenetic Trees

All Phylogenetic trees were constructed using the software program MEGA. Protein amino acid sequences were aligned by the ClustalW algorithm using the default parameters. A tree was then constructed from the aligned sequences using the neighbor-joining statistical method. Evolutionary distances were computed using the

Poisson correction method (Saitou & Nei, 1987, Zuckerkandl & Pauling, 1965, Tamura et al., 2011). Positions lacking data or containing gaps were removed from the analysis. All *V. cholerae* amino acid sequences were collected from NCBI and IMG genome databases. Analysis of T3SS ATPase proteins located near PAI integrase proteins was crucial, as we only wanted to examine T3SSs present on PAIs. Cognate integrase sequences on PAIs lacking a T3SS were also excluded from the analysis. The “Gene Neighborhood” tool in the IMG database was used to find associated integrase and T3SS ATPase proteins. 11 *V. cholerae* strains with T3SS containing PAIs, in addition to strain NRT36s, were selected for phylogenetic analysis due to the availability and assembly status of their genomes (Table 4).

The phylogenetic tree shown in Figure 17 was generated from MEGA using the same method and parameters as previously described (Saitou & Nei, 1987, Zuckerkandl & Pauling, 1965, Tamura et al., 2011). However, T3SS ATPases located outside of PAIs were not excluded from this analysis. A total of 27 T3SS ATPase sequences were chosen for the final analysis. The main criterion when selecting sequences was genus and species level diversity in the final data set.

Chapter 3

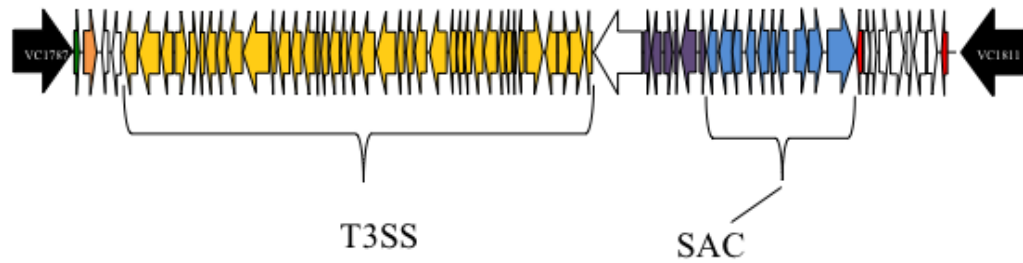
RESULTS

3.1 Structure of the Type III Secretion System Containing PAI in NRT36S

Construction of NRT36S VPI-3 from raw sequence data was the crucial first step in understanding the excision behavior and evolutionary history of the island. The presence of a VPI-2 like island in pathogenic *V. cholerae* non-O1/O139 was previously shown in a number of studies (Jermyn and Boyd, 2005, Chen et al., 2007, Murphy and Boyd, 2008). In this study we show that the island in *V. cholerae* NRT36S is made up of 70 Open Reading Frames (ORFs), which were sequentially numbered and will be referred to as ORF1-ORF70 (Table 3). The island shows extremely high homology with a VPI-2-like region of *V. cholerae* strain AM-19226 (Figure 10). We have named this island region in NRT36S as VPI-3. In NRT36S, VPI-3 ORFs 1-43 encode the various components of a Type III secretion system, including effector proteins. Homologous T3SS proteins are present in *V. cholerae* strains AM-19226, HE-25, TMA-21, VC12129, EM1676-A, VC35, CP1115, P-18785 and NHCC-008D (Table 4). ORFs 54-65 encode genes necessary for the metabolism of sialic acids. This sialic acid metabolism region is not only homologous in all of the strains listed above; it is also homologous to a similar region found in VPI-2 of O1 strains such as *V. cholerae* N16961. Gene maps of VPI-3 of NRT36S and VPI-2 of N16961 can be found in Figure 9. Figure 10 shows an ACT alignment of VPI-2 of N16961, VPI-3 of NRT36S, and VPI-3 of AM-19226. This figure clearly depicts

regions of significant genetic difference in the structures of VPI-2 in an O1/O139 strain of *V. cholerae* and VPI-3 of two pathogenic non-O1/O139 strains.

NRT36S VPI-3 (67.9 kb):



N16961 VPI-2 (57.3 kb):

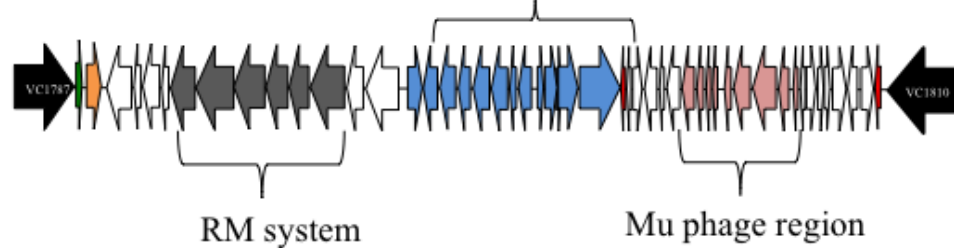


Figure 9 Gene maps of VPI-3 and VPI-2 regions of NRT36S and N16961 respectively. The conserved integrase of the islands, IntV2, is shown in orange. Green ORFs represent tRNA-ser, the locus of island integration. The T3SS is depicted as yellow ORFs. The purple region shown in VPI-3 is a variable region that can be divergent among multiple species containing this island. The SAC region (blue) contains the amino sugar sialic acid catabolism region. RM (grey) contains a restriction modification system. Red ORFs represent recombination directionality factors *vefA* and *vefB*.

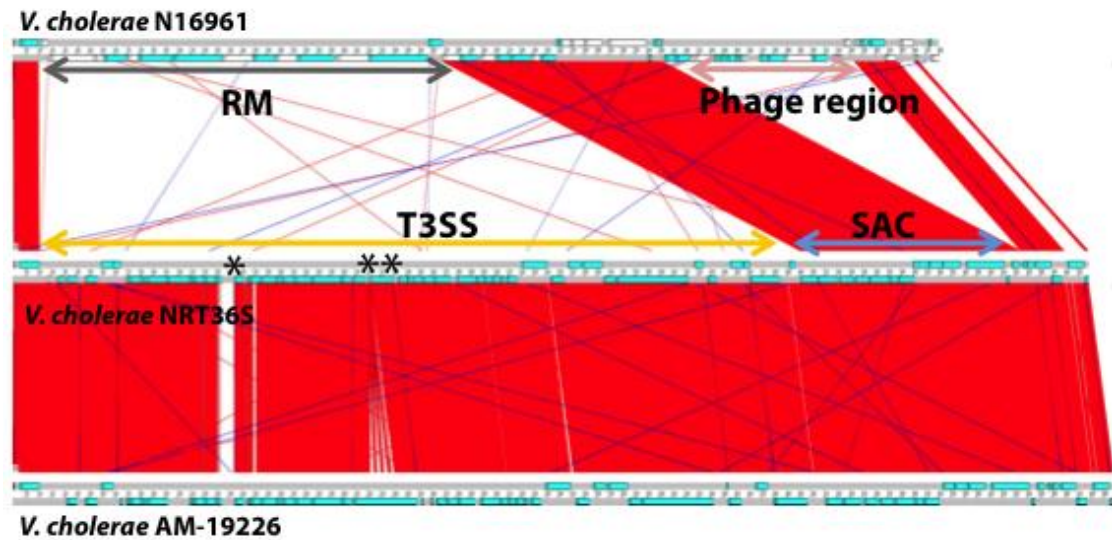


Figure 10 The Artemis comparison tool (ACT) was used to examine sequence differences between VPI-2 and VPI-3 in *V. cholerae* strains N16961 (top), NRT36S (middle), and AM-19226 (bottom). Red blocks represent forward regions of homology between the three island regions and blue lines indicate reverse matches (Carver et al. 2005). In NRT36S, the T3SS is found inserted between the conserved integrase and SAC region rather than the RM system found in N16961. NRT36S also lacks the Mu phage region present in N16961. Two major differences are presented between the otherwise homologous T3SS region of NRT36S and AM-19226: 1) a 1,051 bp region at the 13 kb position (*) which comprises ORF17 and the intergenic region between ORF17 and ORF18 of NRT36S and 2) repeated sequence region (**) indicative of several 201 bp duplications between ORF 26 and ORF 27. All of the default settings were used in ACT, and all comparison files were generated using the webACT tool available at www.webact.org.

3.2 Excision of VPI-3

It has been previously shown that VPI-2 of *V. cholerae* El Tor biovar N16961 has the capability to excise out of the core chromosome of the bacteria, and that this excision is dependent upon the presence of the cognate integrase encoded on the island (Murphy and Boyd, 2007, Almagro-Moreno, Napolitano, & Boyd, 2010). The ability of VPI-3 to excise out of the core chromosome of NRT36S was tested using a two-stage nested PCR reaction. Two sets of primers (VPI2 fk1, VPI2 fk2 & VPI-2 attF,

VPI-2 attR) were designed in NRT36S to amplify a 524 bp area, including the 22 bp *attB* site, which could only be amplified by PCR if the island was not present in the core genome of the bacteria. When wild-type strains were tested (Figure 8a) the expected band was observed for both NRT36S and N16961, a positive control strain. This indicates that the rare excision event is occurring, allowing the primers to bind and the PCR to amplify the 524 bp site that includes the *attB* site. The same result is found when the wild type strains are tested for amplification of a region containing the 497 bp *attP* site (Figure 8b).

In order to test whether the cognate integrase of NRT36S VPI-3 (*intV2*) is also necessary for the excision of VPI-3, the same assay described above was performed on NRT36SΔ*intV2* (Figure 10). No expected band was observed following our two round PCR reaction for NRT36SΔ*intV2* (Figure 11). The primers were not able to amplify the 524 bp site due to the fact that the island was unable to excise; therefore, the area between the primers was too large for amplification via PCR. The absence of the cognate integrase from the island causes this result, suggesting that this protein is crucial for the excision of VPI-3 as well as in the strains where it has been previously shown in *V. cholerae* (Rajanna et al., 2003; Murphy and Boyd, 2008, Almagro-Moreno, Napolitano, & Boyd, 2010). All PCR reactions described above were repeated a minimum of three times. When examining the amino acid sequence of *V. cholerae* strain CP1115, a premature stop codon was found that prevented the translation of the first 60 amino acids of the cognate integrase protein of VPI-3. As a result of this, the DUF domain of this integrase was truncated (Figure 12).

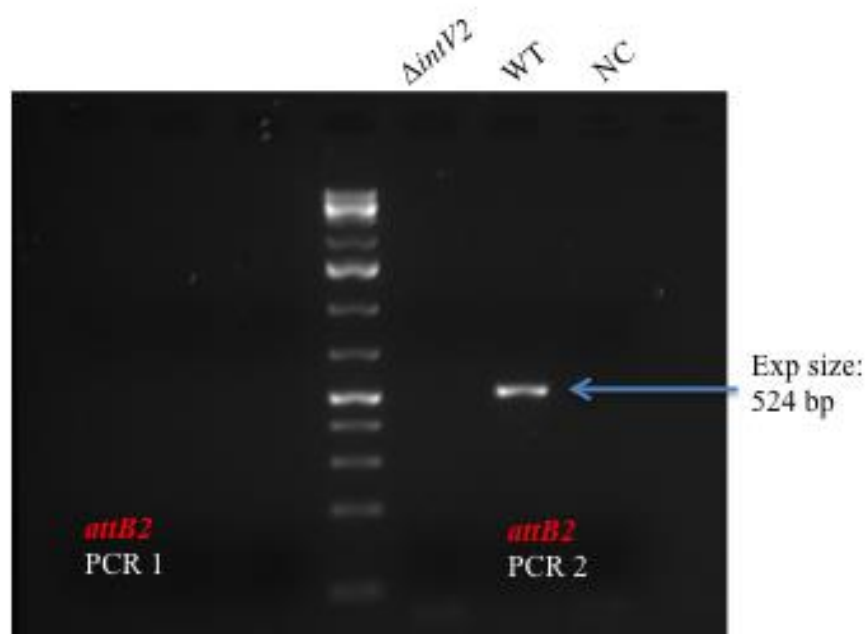


Figure 11 The two-stage *attB* PCR amplification assay was used to detect excision of VPI-3 in $\Delta intV2$ and wild-type NRT36S strains. As expected, no PCR product was detected in the first round. A wild-type *attB* excision product of 524 bp was detected in the second round, however no product was detected for $\Delta intV2$ indicating excision had been abolished in this strain.

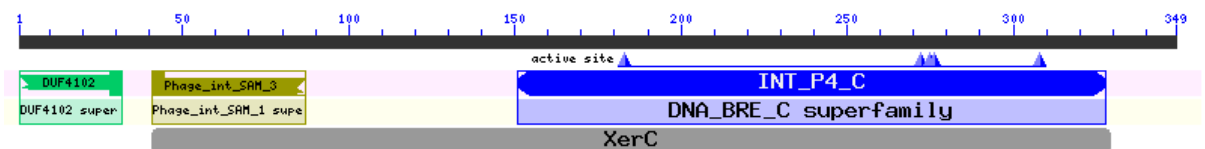


Figure 12 Conserved domains of the integrase protein present on VPI-3 of *V. cholerae* strain CP1115. The truncated DUF domain at the N-terminus of the protein due to the disruption of translation is shown in green. Figure adapted from the NCBI conserved domain database (Marchler-Bauer et al., 2015).

3.3 ORFs within the NRT36S VPI-3 Region

The 62.9 kb region that makes up the VPI-3 region of NRT36S is made up of 70 Open Reading Frames (ORFs) (Table 3). Most of these ORFs fall within the two distinct regions of the island, the Type III Secretion system region (ORFs 3-45) and the Sialic Acid catabolism region (ORFs 54-65). ORFs 1-14 are 99% homologous to A33_1660 to A33_1671 in AM-19226. ORF16 has 97% identity with its homolog in AM-19926 (A33_1673), and lies in a region that is 96% identical to homologous sequence in AM-19226. ORF 17 has 78% identity with its homolog in AM-19226, and lies in a region of sequence with significant non-homology to AM-19226 (13117-14005). ORF 18 has 97% homology to its homolog A33_1675, and lies within a region of sequence that is 97% homologous to corresponding sequence in AM-19226. ORF 19 is 99% homologous to its AM homolog, but lies within a region of sequence that is 97% identical to corresponding AM sequence. ORF 20 is 87% identical to its AM-19226 homolog, A33_1677 and lies within a region of non-homology between the two sequences (15185-15356). The sequence containing ORFs 21- 27 is 99% homologous to the corresponding sequence in AM-19226 (A33_1678-A33_1683). The sequence containing ORFs 29-40 is 99% homologous to the corresponding sequence in AM-19226 except for a 22 bp gap at 29.6 KB (29652-29674). ORFs 41-52 are encoded in a region of sequence that is 99% homologous to between NRT36S and AM-19226. The sequence containing ORFs 53-71 is 99% homologous to the sequence containing A33_1707-AM_1723. Although the nucleotide sequence is identical between the two strains in this region, no annotated protein in AM_19226 is homologous to ORF 53 found by ARTEMIS (hypothetical protein). ORFs 66-70 are

all 100% homologous to their AM-19926 homologs, except for ORF 72, which contains 2 gaps between in homology between it and A33_1724.

Table 3 VPI-3 Open Reading Frames, gene name, amino acid length, and % homology

OR F	Gene Name	Length (aa)	AM-19226 Homolog (% ID)	N16961 Homolog (% ID)	Other Homolog (% ID)
1	Integrase	414	A33_1660 (100)	VC_1758 (100)	
2	Transposase and inactivated derivatives	56	A33_1661 (100)		
3	Acf A	126			VCB_002800 (99) ¹
4	Hypothetical protein (VopE)	204	A33_1662 (100)		
5	Hypothetical Protein (VopX)	252	A33_1663 (100)		
6	ToxR (transcriptional regulator)	250	A33_1664 (99)		
7	Hypothetical protein	161			VCB_RS14480 (100)
8	Hypothetical protein	77	A33_1666 (100)		
9	Conserved hypothetical protein	247	A33_1667 (100)		
10	Hypothetical protein	95	A33_1668 (100)		
11	T3SS ATPase	420	A33_1669 (100)		
12	VcsC2	485	A33_1670 (99)		
13	NAD-specific glutamate dehydrogenase	71	A33_1671 (98)		
14	VcsT2	228	A33_1672 (98)		
15	FliP, flagellar biosynthesis pathway component	224	A33_1673 (97)		
16	Hypothetical protein	99	A33_1674 (78)		
17	ToxR2	183	A33_1675 (97)		
18	Conserved domain protein	71	A33_1676 (99)		
19	Hypothetical protein	216	A33_1677 (87)		
20	Hypothetical protein (VopH)	334	A33_1678 (98)		
21	Hypothetical protein	115	A33_1679 (98)		
22	Hypothetical protein (VopA)	255	A33_1680 (99)		
23	VscU2	350	A33_1681 (100)		
24	VcsV2	627	A33_1682 (99)		
25	Hypothetical protein	203	A33_1683 (100)		
26	VsaC lipoprotein	260	A33_1684 (100)		VCB_002822 (100)

¹ Locus tags VCB_XXXXXXX indicate proteins found in *V. cholerae* TMA-21

27	Hypothetical protein	493	A33_1684 (99)	
28	Putative Dimethyladenosine transferase	191		VCB_RS14595 (100)
29	Hypothetical protein	136		VCB_RS14600 (100)
30	ATP dependent exoDNase	193	A33_1687 (99)	
31	Hypothetical protein	242	A33_1688 (99)	
32	VspD	349	A33_1689 (99)	
33	Hypothetical protein (VopW)	413	A33_1690 (99)	
34	Hypothetical regulator protein	162	A33_1691 (99)	
35	Conserved hypothetical protein	152	A33_1692 (100)	
36	VcsJ2	177	A33_1693 (99)	
37	Conserved hypothetical protein	84	A33_1694 (100)	
38	Hypothetical protein	213	A33_1695 (99)	
39	Hypothetical protein (VopF)	530	A33_1696 (99)	
40	Hypothetical protein (VopG)	277	A33_1697 (100)	
41	Spermidine/ putrescine ABC transporter permease	270	A33_1698 (99)	
42	Hypothetical protein (VopK)	424	A33_1699 (100)	
43	Conserved hypothetical protein (VopY)	249	A33_1700 (92)	
44	AcfD, large exoprotein involved in heme utilization	1507	A33_1701 (99)	
45	Trh	180	A33_1702 (100)	
46	alpha beta hydrolase fold	189	A33_1703 (99) ²	
47	Hypothetical protein (VopZ)	260	A33_1704 (100)	
48	Hypothetical protein	146		VCB_RS14705 (100)
49	Bacterial Ig-like domain family	631	A33_1706 (99)	
50	Hypothetical protein	134		VCG_RS0118140 (99) ³
51	Mutatrotase, YjhT family	355	A33_1707 (100)	VC_1773(100)
52	Mutatrotase, n-acetylneuraminic acid	384	A33_1708 (100)	VC_1774 (100)
53	Hypothetical protein, pos transcriptional regulator	278	A33_1709 (100)	VC_1775 (99)
54	Lyase, n-acetylneuraminase	298	A33_1710 (100)	VC_1776 (100)
55	ABC transporter permease	427	A33_1711 (100)	VC_1777 (100)
56	C4-dicarboxylate ABC transporter permease	173	A33_1712 (100)	VC_1778 (99)
57	C4-dicarboxylate-binding periplasmic protein	321	A33_1713 (100)	VC_1779 (100)

² 85% Query Cover

³ Locus tags VCG_XXXXXXX indicate proteins found in *V. cholerae* VC12129

58	Hypothetical protein N-acetylmannosamine-6-phosphate	132		VC_1780 (97) ⁴
59	2-epimerase	240	A33_1714 (100)	VC_1781 (99)
60	N-acetylmannosamine kinase N-acetylglucosamine-6-phosphate	287	A33_1715 (100)	VC_1782 (100)
61	deacetylase	378	A33_1716 (100)	VC_1783 (100)
62	Neuraminidase	803	A33_1717 (99)	VC_1784 (99)
63	VefA	68	A33_1718 (100)	VC_1785 (100)
64	RadC	158	A33_1719 (100)	VC_1786 (100)
65	Hypothetical protein	77	A33_1720 (100)	
66	Hypothetical protein	148	A33_1721 (100)	
67	GTPase, hypothetical protein	328	A33_1722 (100)	
68	Conserved hypothetical protein CopG family transcriptional	213	A33_1723 (100)	
69	regulator	340	A33_1724 (99)	
70	VefB	76	A33_1725 (100)	VC_1809 (100)

3.4 Presence of VPI-3 in other *V. cholerae* strains

Using the ARTEMIS comparison tool, the T3SS containing island of NRT36S was compared to nine other T3SS containing strains of *V. cholerae* (Figure 13). While the main components of the T3SS and the sialic acid metabolism region were generally conserved amongst the ten strains, there were regions of non-homology present, suggesting divergence throughout the evolution of VPI-3 in *V. cholerae*. The first major gap in homology is at 13.2 kb in NRT36S. The gap extends through the region between ORF16 and ORF18 (ToxR2), including the absence of most of ORF17 (hypothetical protein). These ORFs lie within the T3SS of NRT36S. This gap in homology is present in each strain selected for comparison except for CP115. At 33 kb, the region of non-homology between *V. cholerae* AM-19226 and TMA-21 spans the region between A33_1695 (hypothetical protein) and A33_1697 (Zinc finger domain protein). In AM-19226 this region encodes effector protein VopF.

⁴ 73% Query Cover

Interestingly, when we examine the region at 33 kb in NRT36S an ORF with homology to VopF was not identified. The absence of an effector protein in two otherwise highly homologous strains of *V. cholerae* is noteworthy, especially if VopF is translocated into the host cell via the T3SS. Another significant gap in homology exists between NRT36S and *V. cholerae* strains TMA-21 and V51 at around 29 kb, a region that encodes a hypothetical protein. These gaps in homology could indicate hot spots for recombination events that allow virulence elements to enter and exit the island during horizontal gene transfer or the evolution of novel effector proteins. For example, in *V. cholerae* El Tor biovar N16961 the Mu-phage region is found inserted between *vefA* and *vefB*, whereas this region is absent in NRT36S. Gaps within the T3SS itself could lend insight into the evolution of this virulence system.

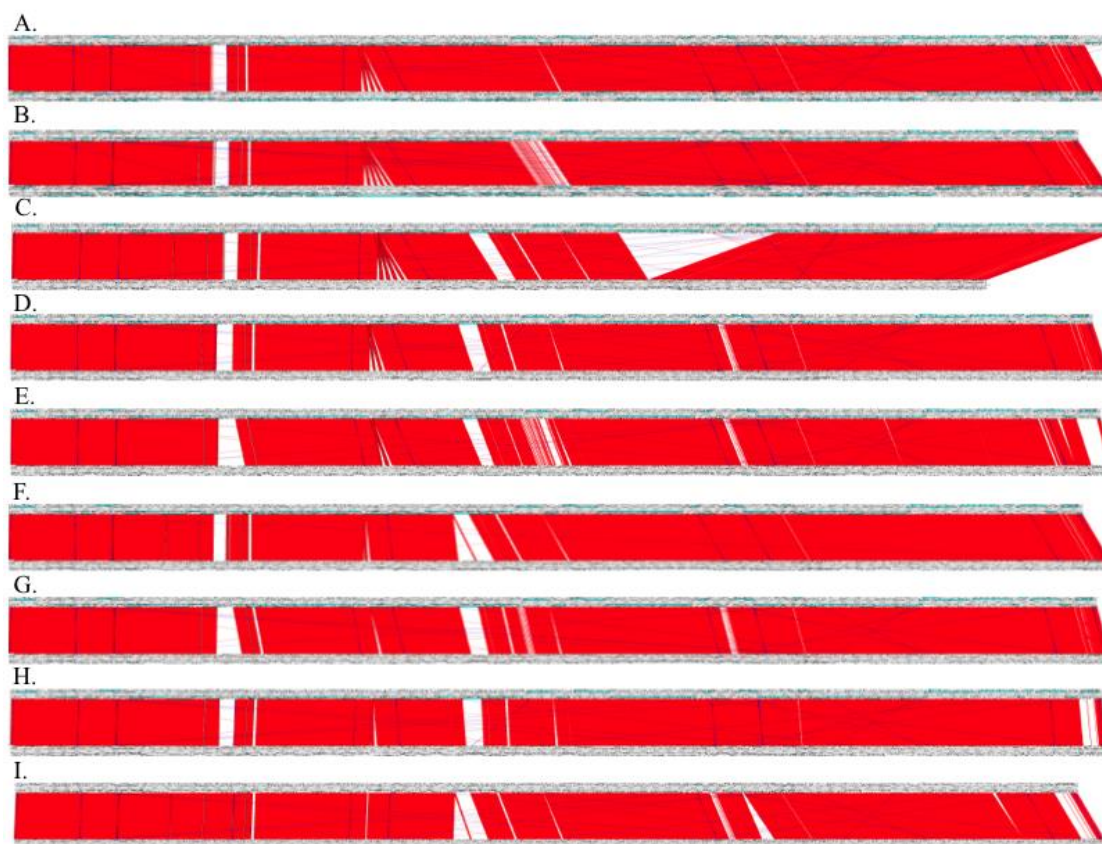


Figure 13 Artemis comparisons of the T3SS containing PAIs of 10 *V. cholerae* strains. In each comparison, *V. cholerae* NRT36S is sequence one. Strains compared to NRT36S in A-I are AM-19226, TMA-21, VC12129, EM-1676A, NHCC008D, VC35, HE-25, P-18785, and CP1115 respectively. Two prominent and conserved sites of non-homology are the gaps at the 13 kb and 29 kb positions.

Table 4 Background information of *V. cholerae* strains used in bioinformatic comparisons of the VPI-3 region

Table Strain Name	Sero group	IMG Taxon ID	NCBI Accession number	Location	Source of Isolation	Year
NRT36S	O31			Japan	Clinical Isolate	1990
AM-19226	O39	647533244	GCA_000153785.1	India	Bangladesh Clinical isolate	2010

TMA-21		645058807	GCA_000174295.1	Brazil	Seawater	1982
HE25		2521172724	GCA_000279265.1	Haiti	Wastewater	2010
VC12129	Inaba (O1)	645058798			Water	1985
EM1676-A	O1	2537561651	GCA_000348345.1	Bangladesh	Human Stool	2010
VC35	O1	2551306303	GCA_000299495.1	Kedah, Malaysia	Human Stool	2004
NHCC-008D	O1	2537561651	GCA_000348425.1	Bangladesh	Human Stool	2010
VC1587	O12	640963020	GCA_000168895.1	Peru	Clinical Isolate	2006
CP1115	O75	2554235147	GCA_000387605.1	Florida, USA	Clinical Isolate	2011
P18785	O1	2518285605	GCA_000338215.1	Rostov-on-Don, Russia	Human Stool	2014
V51	O141	2562617060	GCF_000152465.1	USA	Clinical Isolate	2007

3.5 Discovery of a repeated region of homology in T3SS⁺ strains of *V. cholerae*

During the ACT comparison of the highly homologous VPI-3 regions of NRT36S and AM-19226 a series of 3 repeated sequences was discovered beginning at approximately 22.4 kb. The three repeats are 201 base pairs long and are 95% identical between the sequence of NRT36S and AM-19226 (Figure 14a). In AM-19226 each of these repeats is separated by 234 base pairs; the regions separating the repeats are 96% identical, and the repeats themselves are between 95-100% identical. The repeats fall in an intergenic region between NRT36S ORFS 26, a VsaC lipoprotein, and ORF 27, a hypothetical protein. In contrast, two separate ORFs are not found in the annotation of AM-19226. In that region a 1,328 amino acid hypothetical protein, A33_1684, has been identified and annotated (Alam et al., 2011). In order to confirm that A33_1684 is the only ORF present in that region of AM-

19226, the sequence was entered into NCBI's ORFinder program (Figure 14b). This program confirmed the presence of a single large ORF in the region containing the repeated sequences. Further analysis revealed the presence of a 46 bp smaller repeated sequence directly downstream of the 201 bp repeated sequence in NRT36S. This sequence is also repeated three times in AM-19226 and is 82% identical to the sequence that originates in NRT36S. In AM-19226 two of the repeats are identical and one has a one bp T to C substitution at the 19 bp site, a G to A substitution at the 103 bp site, an A to G substitution at the 113 bp site, a T to C substitution at the 118 bp site, and a C to G substitution at the 145 bp site (Figure 18).

Variations of this repeated region were found in nine strains of *V. cholerae* containing a T3SS, including AM-19226. Background information on these strains can be found in Table 3. When the sequence of the island region of these strains was compared to that of NRT36S, it was found that each of these strains has a unique number of repeats, and the sequence that is repeated between the comparison strain and NRT36S can vary (Figure 15). In seven of the nine strains, the repeats disrupt the transcription of the ORFs in the region, leading to the transcription of a large ORF similar to what was discovered in AM-19226 (Figure 15).

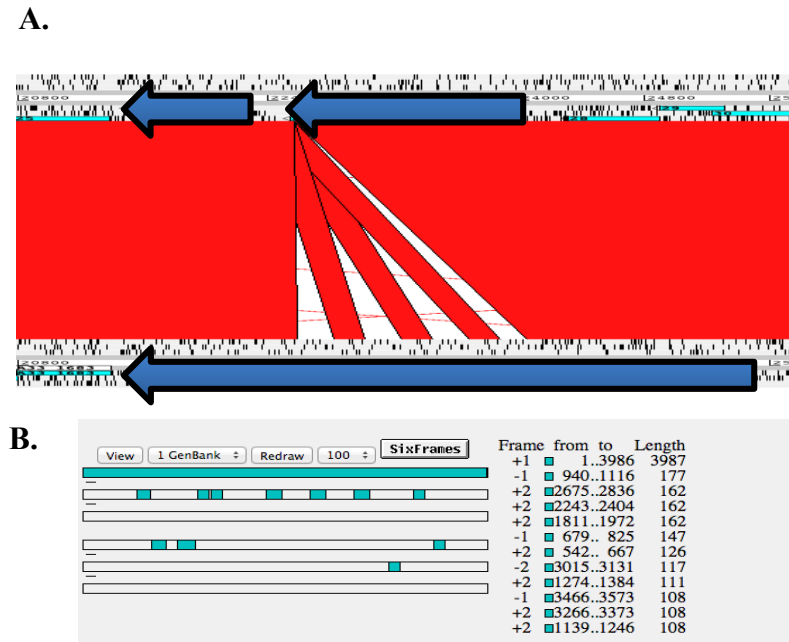


Figure 14 Further examination of the repeated region at 22.4 kb between *V. cholerae* strains NRT36S and AM-19226. **A.** Enlarged Artemis comparison of NRT36S and AM-19226 reveals the difference in annotation of the ORFs in this region. In NRT36S, 2 ORFs separated by a 265 bp intergenic region were detected. However, in AM-19226, one large ORF (A33_1684) encompassing 1328 amino acids is detected in the corresponding region. A33_1648 encodes a probable effector protein (Alam et al. 2011). **B.** NCBI ORFinder display confirming that no alternative ORFs could be found in the region encoding A33_1684 in AM-19226.

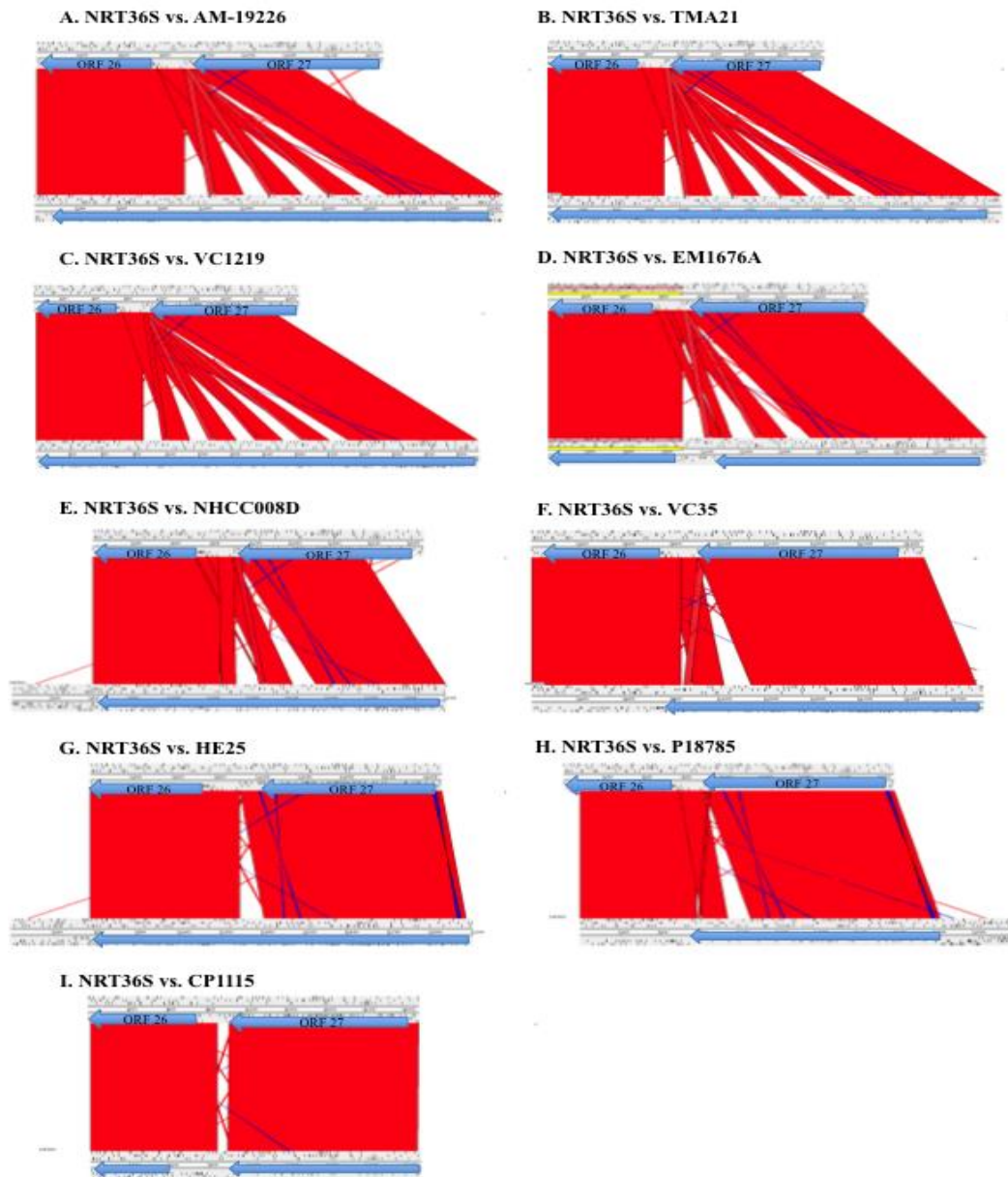


Figure 15 Comparison of the repeated region of NRT36S to the corresponding region of 9 different *V. cholerae* strains. The number of repeats varies in all cases; however, the repeat always originates in the intergenic region of NRT36S VPI-3 between ORF 26 and ORF 27.

3.6 Phylogenetic Analysis of the T3SS ATPase of VPI-3

The evolution of a PAI can be speculated by examining the evolution of its component biological function genes. The phylogenetic trees of integrase proteins and T3SS ATPase proteins of various *V. cholerae* PAIs were compared (Figure 16). The T3SS ATPase amino acid sequence was chosen for analysis as it is conserved not only across T3SSs of different species and genera of bacteria, but across different types of T3SS as well, such as T3SS2 α and T3SS2 β . Only T3SS containing islands with an associated integrase were included in the phylogenetic analysis. Figure 16a shows that the T3SS ATPase proteins found on VPI-3 of *V. cholerae* strains NRT36S, AM-19226, CP1110, EM-1676A, V51, TMA21, 12129, VC35, HE-25 and P-18785 are all highly homologous. These strains encode a T3SS2 α . *V. cholerae* strains 1587 and 623-39 cluster together separately because they encode the variant T3SS2- β . A phylogenetic analysis of the cognate integrase of VPI-3 was also performed (Figure 16b). Amino acid sequences of integrase proteins found on PAIs containing T3SS were included. The 11 *V. cholerae* strains that all encode highly homologous T3SS ATPase proteins display a different clustering pattern when analyzing the cognate integrase of VPI-3. It is important to note the difference in scale between the two phylogenetic trees. On the T3SS ATPase tree, one scale bar equals 0.01 amino acid substitutions per site. On the integrase tree this number is much lower; one scale bar equals 0.002 amino acid substitutions per site. Hence, the different clades *V. cholerae* integrase proteins are more highly related than the two distinct clades of T3SS ATPase proteins.

The T3SS ATPase from 27 different *Vibrionaceae* and *Salmonella* bacteria was also chosen for phylogenetic analysis. The tree shown in Figure 17 illustrates the relationships between the two varieties of T3SSs found in different *Vibrios* and

Salmonella. The *Salmonella* strains examined contain a T3SS1, whereas the *V. cholerae* strains encode a T3SS2. *V. parahaemolyticus* strains studied have either T3SS1 or T3SS2, and strains such as RIMD2210633 acquired both systems on separate chromosomes or other mobile genetic elements. There are two types of T3SS2, T3SS2 α and T3SS2 β . Various *Vibrio* species contain either the T3SS2 α or T3SS2 β . The clustering pattern of the phylogenetic tree shown in Figure 17 indicates that the T3SS ATPase of strain NRT36S is most closely related to strain AM-19226. The T3SS of NRT36S is also closely related to certain strains of *V. parahaemolyticus*, *V. mimicus* and *G. hollisae* (Figure 17). The clustering of T3SS containing strains of bacteria of different genera and species on the tree suggests that the T3SS was transferred amongst these species horizontally. The exact direction of this transfer is unknown.

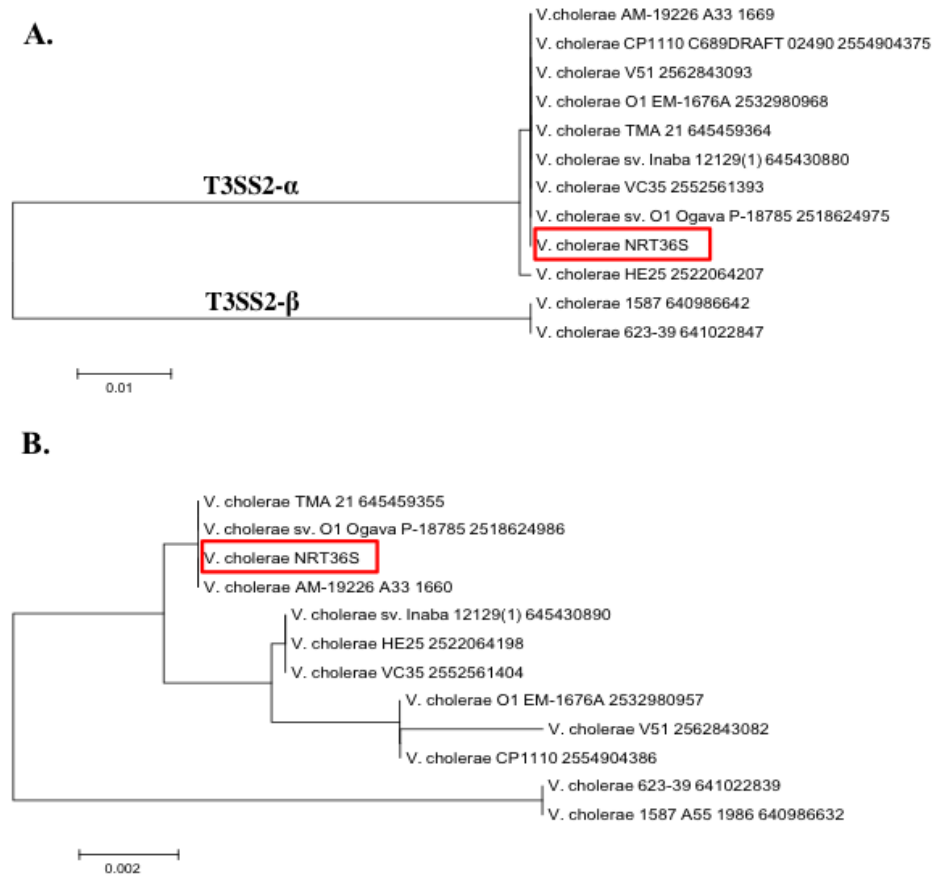


Figure 16 A. Phylogenetic tree of T3SS ATPases in *V. cholerae* strains with an associated integrase protein encoded on a PAI. B. Phylogenetic tree of integrase proteins encoded on PAIs containing T3SS. Sequences were obtained from either NCBI or IMG databases and uploaded into MEGA where they were aligned using the ClustalW algorithm (Tamura et al., 2007). After alignment, MEGA was used to generate a neighbor-joining tree, using the bootstrap method (Saitou & Nei, 1987; Felsenstein, 1985). 1,000 bootstraps were used. In all phylogenetic trees the entry after the strain name indicates the locus tag of the protein on the tree, and the second entry is the IMG database gene ID number. Note the difference in the scales of the two trees.

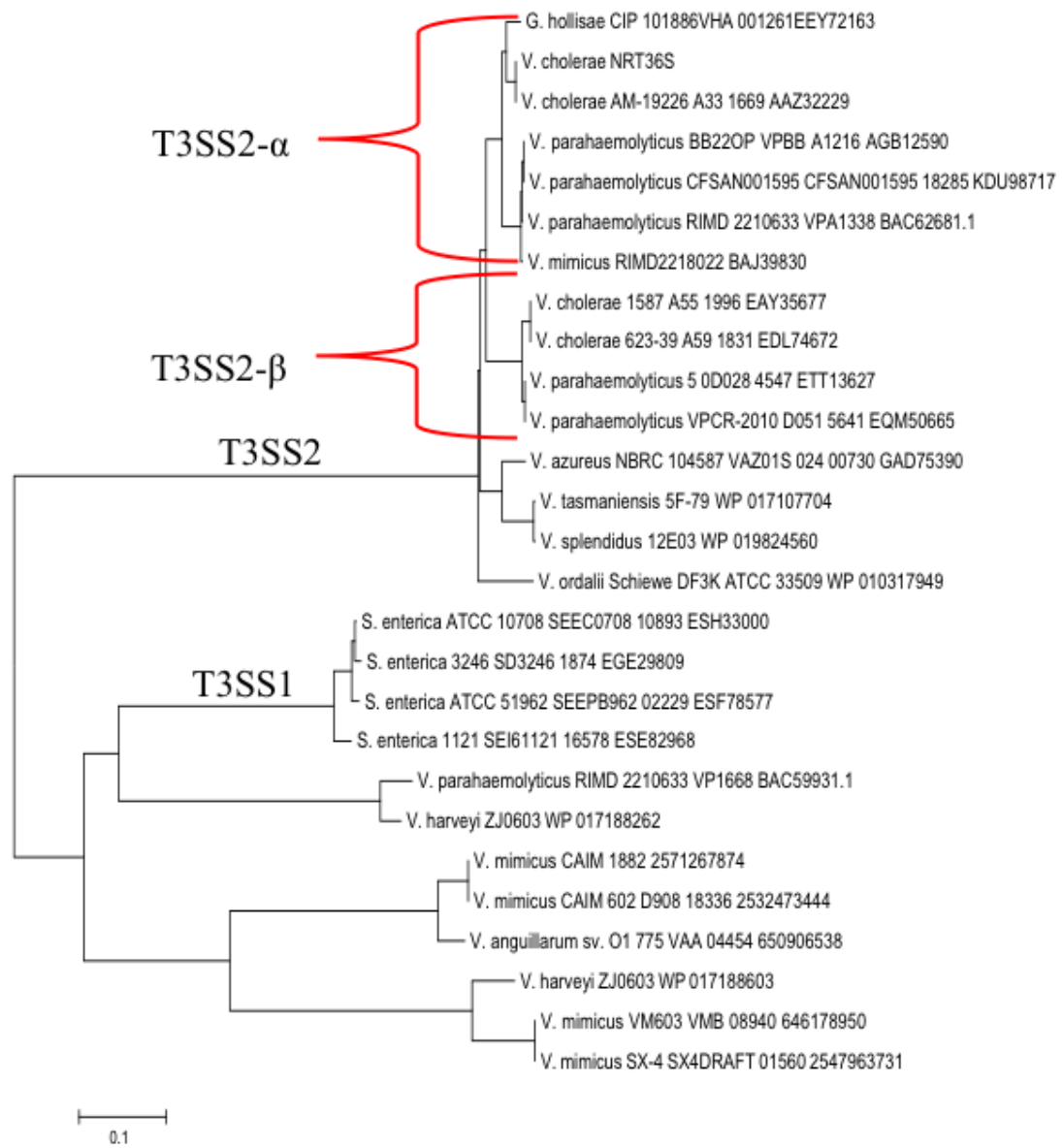


Figure 17. Phylogenetic tree of the T3SS ATPase protein in multiple species of Gram-negative bacteria. Tree was constructed using the neighbor-joining method described in Figure 13.

Chapter 4

DISCUSSION

4.1 Excision behavior of T3SS containing PAI

It has been demonstrated that the cognate integrase of VPI-2 is necessary for the excision of VPI-2 in *V. cholerae* (Rajanna et al., 2003, Murphy & Boyd, 2008, Almagro-Moreno, Napolitano, & Boyd, 2010). However, excision in a variant of VPI-2 containing a T3SS, which we named VPI-3, was not characterized before this study. It was confirmed that the cognate integrase of VPI-3 is necessary for the excision of the island. This observation is relevant not only because the biological function genes on the islands are drastically different, but also because VPI-3 is approximately 10 kb larger than VPI-2 in *V. cholerae* N16961. It is possible that the truncated DUF domain could affect the folding or functioning of VPI-3's cognate integrase in strain CP1115; however, this result is unlikely due to the fact that the truncated domain is not present in the phage integrase domain nor is it near the active site of the protein. Further investigation into the effects of this mutation is required.

4.2 Evidence for the mosaic evolution of *V. cholerae* PAIs

It is known that PAIs are involved in the evolution of novel and virulent strains of bacteria; however, less is understood about the evolution of the PAI regions themselves. It is possible that PAIs evolve as one defined unit over time, or more fluidly as a mosaic-like structure where biological function genes could evolve within the island. In this model, the evolution of the island does not follow a sequential path where mutations occur in one region while the rest of the island stays consistent. Instead, the island evolves in a manner where the components of the island are not a

discrete unit and have evolved independently over the course of time. If the PAI followed the sequential path of evolution, the biological function genes such as the T3SS proteins, would have evolved in a similar manner to the structural components of the island, such as the integrase protein. In strains of *V. cholerae* containing the T3SS2, the ATPase proteins are separated into two clades representing T3SS proteins present in T3SS2 α and T3SS2 β , but are otherwise identical (Figure 15a). However, the integrase protein is not genetically identical in the majority of these strains, as shown in the phylogenetic tree above (Figure 15b). The incongruent structure of the two trees suggests that the integrase and T3SS proteins did not co-evolve as would be expected if the island was a discrete entity. There are two evolutionary distinct clades present on the T3SS ATPase tree representing the ATPase present in the T3SS2 α or T3SS2 β . In contrast, the integrase proteins cluster into four separate clades that are very closely related, even between strains with T3SS2 α and T3SS2 β present on the island. If these two proteins evolved together over time, the structures of the trees presented in Figure 15 would be the same and the same strains would consistently cluster together. This divergence implies that the T3SS was acquired at a different point in the course of the evolution of VPI-3 for each of these strains, providing evidence to support the theory that PAIs can evolve in a mosaic-like manner.

4.3 Evolution of new effector proteins due to multiple repeated sequences

This study identified a repeated region in multiple strains of *V. cholerae* encoding T3SS PAIs (Figure 15). The ORFs shown by blue arrows in the comparison strains were found by NCBI's ORFinder program (Figure 15). Although the exact reason and mechanism for these duplications is unknown, we hypothesize that they do affect the transcription of nearby genes by causing frameshift mutations, disrupting

start and stop codons, and potentially contributing to the formation of novel effector proteins. Figure 14 shows one such disruption when comparing the DNA sequence between NRT36S and AM-19226. It appears as if the duplications that occur in the sequence of AM-19226 disrupt the stop codon of ORF 27, leading to the transcription of one large ORF in that region as opposed to the two smaller ORFs seen in NRT36S. Alam et al. found that this ORF in AM-19226 (A33_1684) was translocated into the host cell, providing experimental evidence that this protein may be an effector (Alam et al., 2011). This fusion has occurred in all strains examined except for *V. cholerae* EMA-1676A and CP1115. It has been previously shown that the fusion of ORFs is a contributing factor to the evolution of effector proteins (Alam et al., 2011). This region of repeats may be evidence of this evolutionary phenomenon.

4.4 Relevance to the scientific community

Understanding the evolution of non-O1/O139 strains of *V. cholerae* is an extremely important aspect of *V. cholerae* research. Because these strains do not present classical choleraeagenic symptoms, they can often go misdiagnosed as “acute watery diarrhea” in treatment centers (WHO, 2014). Thus, while the mortality rate from *V. cholerae* strains lacking CT and TCP may be low, it is estimated that the morbidity levels from non-O1/O139 strains causing localized and patient level outbreaks is much higher than reported. Furthermore, as discussed previously, these non-O1/O139 strains evolve rapidly and may have acquired antibiotic resistance. They are not recognized by the immune response in areas where cholera is endemic and native populations have acquired immune responses to O1/O139 strains. No vaccines are available to provide protection against non-O1/O139 strains of *V. cholerae* if a major outbreak were to occur, and in strains that encode a T3SS, disease symptoms

can be just as severe as infection with classical or El Tor strains. Expanding the existing knowledge base to include the evolution of the molecular mechanisms of pathogenesis in these non-O1/O139 strains is crucial if the medical and public health fields are to effectively prepare for future outbreaks and possible epidemics.

REFERENCES

- Alfano, J. R., Charkowski, A. O., Deng, W.-L., Badel, J. L., Petnicki-Ocwieja, T., Dijk, K. van, & Collmer, A. (2000). The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proceedings of the National Academy of Sciences*, 97(9), 4856–4861. <http://doi.org/10.1073/pnas.97.9.4856>
- Almagro-Moreno, S., Napolitano, M. G., & Boyd, E. F. (2010). Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA. *BMC Microbiology*, 10, 306. <http://doi.org/10.1186/1471-2180-10-306>
- Baker-Austin, C., Trinanes, J. A., Taylor, N. G. H., Hartnell, R., Siitonen, A., & Martinez-Urtaza, J. (2013). Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Nature Climate Change*, 3(1), 73–77. <http://doi.org/10.1038/nclimate1628>
- Bhowmik, S. K., Pazhani, G. P., & Ramamurthy, T. (2014). Phylogenetic and in silico functional analyses of thermostable-direct hemolysin and tdh-related encoding genes in *Vibrio parahaemolyticus* and other Gram-negative bacteria. *BioMed Research International*, 2014, 576528. <http://doi.org/10.1155/2014/576528>
- Boyd, E. F., Almagro-Moreno, S., & Parent, M. A. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends in Microbiology*, 17(2), 47–53. <http://doi.org/10.1016/j.tim.2008.11.003>
- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G., & Parkhill, J. (2005). ACT: the Artemis comparison tool. *Bioinformatics*, 21(16), 3422–3423. <http://doi.org/10.1093/bioinformatics/bti553>

- Ceccarelli, D., Chen, A., Hasan, N. A., Rashed, S. M., Huq, A., & Colwell, R. R. (2015). *Vibrio cholerae* non-O1/non-O139 carrying multiple virulence factors and *V. cholerae* O1 in the Chesapeake Bay, Maryland. *Applied and Environmental Microbiology*, AEM.03540–14.
<http://doi.org/10.1128/AEM.03540-14>
- Centers for Disease Control and Prevention. [Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 1996–2012]. *MMWR* 2013;62(15):[283-287]
- Chatterjee, S., Ghosh, K., Raychoudhuri, A., Chowdhury, G., Bhattacharya, M. K., Mukhopadhyay, A. K., ... Nandy, R. K. (2009). Incidence, Virulence Factors, and Clonality among Clinical Strains of Non-O1, Non-O139 *Vibrio cholerae* Isolates from Hospitalized Diarrheal Patients in Kolkata, India. *Journal of Clinical Microbiology*, 47(4), 1087–1095. <http://doi.org/10.1128/JCM.02026-08>
- Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J. A., ... Stine, O. C. (2007). The capsule polysaccharide structure and biogenesis for non-O1 *Vibrio cholerae* NRT36S: genes are embedded in the LPS region. *BMC Microbiology*, 7, 20. <http://doi.org/10.1186/1471-2180-7-20>
- Chen, Y., Johnson, J. A., Pusch, G. D., Morris, J. G., & Stine, O. C. (2007). The genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from those of O1 *Vibrio cholerae*. *Infection and Immunity*, 75(5), 2645–2647. <http://doi.org/10.1128/IAI.01317-06>
- Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., ... Colwell, R. R. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, 106(36), 15442–15447.
<http://doi.org/10.1073/pnas.0907787106>
- Colwell, R. R. (1996). Global Climate and Infectious Disease: The Cholera Paradigm*. *Science*, 274(5295), 2025–2031. <http://doi.org/10.1126/science.274.5295.2025>
- Colwell, R. R. (2006). A global and historical perspective of the genus *Vibrio*. In *The biology of Vibrios* (pp. 3–11). Washington, D.C: ASM Press.

- Dalsgaard, A., Forslund, A., Bodhidatta, L., Serichantalergs, O., Pitarangsi, C., Pang, L., ... Echeverria, P. (1999). A high proportion of *Vibrio cholerae* strains isolated from children with diarrhoea in Bangkok, Thailand are multiple antibiotic resistant and belong to heterogenous non-O1, non-O139 O-serotypes. *Epidemiology and Infection*, 122(2), 217–226.
- Dryselius, R., Kurokawa, K., & Iida, T. (2007a). Vibrionaceae, a versatile bacterial family with evolutionarily conserved variability. *Research in Microbiology*, 158(6), 479–486.
<http://doi.org/10.1016/j.resmic.2007.04.007>
- Dryselius, R., Kurokawa, K., & Iida, T. (2007b). Vibrionaceae, a versatile bacterial family with evolutionarily conserved variability. *Research in Microbiology*, 158(6), 479–486.
<http://doi.org/10.1016/j.resmic.2007.04.007>
- Dutilh, B. E., Thompson, C. C., Vicente, A. C., Marin, M. A., Lee, C., Silva, G. G., ... Edwards, R. A. (2014). Comparative genomics of 274 *Vibrio cholerae* genomes reveals mobile functions structuring three niche dimensions. *BMC Genomics*, 15(1), 654. <http://doi.org/10.1186/1471-2164-15-654>
- Dziejman, M., Serruto, D., Tam, V. C., Sturtevant, D., Diraphat, P., Faruque, S. M., ... Mekalanos, J. J. (2005). Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9), 3465–3470. <http://doi.org/10.1073/pnas.0409918102>
- Esposito, D., & Scoeca, J. J. (1997). Purification and Characterization of HP1 Cox and Definition of Its Role in Controlling the Direction of Site-specific Recombination. *Journal of Biological Chemistry*, 272(13), 8660–8670. <http://doi.org/10.1074/jbc.272.13.8660>
- Faruque, S. M., & Mekalanos, J. J. (2003). Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends in Microbiology*, 11(11), 505–510. <http://doi.org/10.1016/j.tim.2003.09.003>
- Galán, J. E. (1999). Interaction of *Salmonella* with host cells through the centisome 63 type III secretion system. *Current Opinion in Microbiology*, 2(1), 46–50.
- Galan, J. E. (2009). Common themes in the design and function of bacterial effectors. *Cell Host & Microbe*, 5(6), 571–579. <http://doi.org/10.1016/j.chom.2009.04.008>

- Galán, J. E., Lara-Tejero, M., Marlovits, T. C., & Wagner, S. (2014). Bacterial Type III Secretion Systems: Specialized Nanomachines for Protein Delivery into Target Cells. *Annual Review of Microbiology*, 68(1), null. <http://doi.org/10.1146/annurev-micro-092412-155725>
- Hacker, J., Blum-Oehler, G., Mühldorfer, I., & Tschäpe, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular Microbiology*, 23(6), 1089–1097. <http://doi.org/10.1046/j.1365-2958.1997.3101672.x>
- Hacker, J., & Kaper, J. B. (2000a). Pathogenicity Islands and the Evolution of Microbes. *Annual Review of Microbiology*, 54(1), 641–679. <http://doi.org/10.1146/annurev.micro.54.1.641>
- Hacker, J., & Kaper, J. B. (2000b). Pathogenicity Islands and the Evolution of Microbes. *Annual Review of Microbiology*, 54(1), 641–679. <http://doi.org/10.1146/annurev.micro.54.1.641>
- Haley, B. J., Choi, S. Y., Grim, C. J., Onifade, T. J., Cinar, H. N., Tall, B. D., ... Colwell, R. R. (2014). Genomic and phenotypic characterization of *Vibrio cholerae* non-O1 isolates from a US Gulf Coast cholera outbreak. *PloS One*, 9(4), e86264. <http://doi.org/10.1371/journal.pone.0086264>
- Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., ... Fraser, C. M. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*, 406(6795), 477–483. <http://doi.org/10.1038/35020000>
- Hensel, M., Nikolaus, T., & Egelseer, C. (1999). Molecular and functional analysis indicates a mosaic structure of *Salmonella* pathogenicity island 2. *Molecular Microbiology*, 31(2), 489–498. <http://doi.org/10.1046/j.1365-2958.1999.01190.x>
- Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., ... Hacker, J. (2006a). Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536. *Molecular Microbiology*, 61(3), 584–595. <http://doi.org/10.1111/j.1365-2958.2006.05255.x>
- Hochhut, B., Wilde, C., Balling, G., Middendorf, B., Dobrindt, U., Brzuszkiewicz, E., ... Hacker, J. (2006b). Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic

- Escherichia coli strain 536. *Molecular Microbiology*, 61(3), 584–595. <http://doi.org/10.1111/j.1365-2958.2006.05255.x>
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, 77(1), 61–68. [http://doi.org/10.1016/0378-1119\(89\)90359-4](http://doi.org/10.1016/0378-1119(89)90359-4)
- Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 1996–2012. (n.d.). Retrieved January 30, 2015, from <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6215a2.htm>
- Jermyn, W. S., & Boyd, E. F. (2002). Characterization of a novel *Vibrio* pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae* isolates. *Microbiology (Reading, England)*, 148(Pt 11), 3681–3693.
- Jermyn, W. S., & Boyd, E. F. (2005). Molecular evolution of *Vibrio* pathogenicity island-2 (VPI-2): mosaic structure among *Vibrio cholerae* and *Vibrio mimicus* natural isolates. *Microbiology*, 151(1), 311–322. <http://doi.org/10.1099/mic.0.27621-0>
- Jones, E. H., Feldman, K. A., Palmer, A., Butler, E., Blythe, D., & Mitchell, C. S. (2013). *Vibrio* Infections and Surveillance in Maryland, 2002–2008. *Public Health Reports*, 128(6), 537–545.
- Kaper, J. B., Morris, J. G., & Levine, M. M. (1995). Cholera. *Clinical Microbiology Reviews*, 8(1), 48–86.
- Karaolis, D. K., Johnson, J. A., Bailey, C. C., Boedeker, E. C., Kaper, J. B., & Reeves, P. R. (1998). A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), 3134–3139.
- Lan, N. P. H., Nga, T. V. T., Yen, N. T. T., Dung, L. T., Tuyen, H. T., Campbell, J. I., ... Baker, S. (2014). Two Cases of Bacteremia Caused by Nontoxigenic, Non-O1, Non-O139 *Vibrio cholerae* Isolates in Ho Chi Minh City, Vietnam. *Journal of Clinical Microbiology*, 52(10), 3819–3821. <http://doi.org/10.1128/JCM.01915-14>

- Li, F., Du, P., Li, B., Ke, C., Chen, A., Chen, J., ... Wang, D. (2014). Distribution of Virulence-associated Genes and Genetic Relationships in Non-O1/O139 *Vibrio cholerae* Aquatic Isolates from China. *Applied and Environmental Microbiology*. <http://doi.org/10.1128/AEM.01021-14>
- Marcus, S. L., Brumell, J. H., Pfeifer, C. G., & Finlay, B. B. (2000). Salmonella pathogenicity islands: big virulence in small packages. *Microbes and Infection*, 2(2), 145–156. [http://doi.org/10.1016/S1286-4579\(00\)00273-2](http://doi.org/10.1016/S1286-4579(00)00273-2)
- Marchler-Bauer A et al. (2015), "CDD: NCBI's conserved domain database.", **Nucleic Acids Res.**43(D)222
- Marchler-Bauer A et al. (2011), "CDD: a Conserved Domain Database for the functional annotation of proteins.", **Nucleic Acids Res.**39(D)225-9.
- Marchler-Bauer A et al. (2009), "CDD: specific functional annotation with the Conserved Domain Database.", **Nucleic Acids Res.**37(D)205-10.
- Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", **Nucleic Acids Res.**32(W)327-331.
- Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C.-Y., & Schoolnik, G. K. (2005). Chitin Induces Natural Competence in *Vibrio cholerae*. *Science*, 310(5755), 1824–1827. <http://doi.org/10.1126/science.1120096>
- Miold, S., Ehrbar, K., Weissmüller, A., Prager, R., Tschäpe, H., Rüssmann, H., & Hardt, W.-D. (2001). Salmonella Host Cell Invasion Emerged by Acquisition of a Mosaic of Separate Genetic Elements, Including Salmonella Pathogenicity Island 1 (SPI1), SPI5, and sopE2. *Journal of Bacteriology*, 183(7), 2348–2358. <http://doi.org/10.1128/JB.183.7.2348-2358.2001>
- Morita, M., Yamamoto, S., Hiyoshi, H., Kodama, T., Okura, M., Arakawa, E., ... Watanabe, H. (2013). Horizontal gene transfer of a genetic island encoding a type III secretion system distributed in *Vibrio cholerae*. *Microbiology and Immunology*, 57(5), 334–339. <http://doi.org/10.1111/1348-0421.12039>
- Morris, J. G. (2011). Cholera—Modern Pandemic Disease of Ancient Lineage. *Emerging Infectious Diseases*, 17(11). <http://doi.org/10.3201/eid1711.111109>

- Murphy, R. A., & Boyd, E. F. (2008). Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. *Journal of Bacteriology*, *190*(2), 636–647.
<http://doi.org/10.1128/JB.00562-07>
- Napolitano, M. G., Almagro-Moreno, S., & Boyd, E. F. (2011). Dichotomy in the evolution of pathogenicity island and bacteriophage encoded integrases from pathogenic *Escherichia coli* strains. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *11*(2), 423–436. <http://doi.org/10.1016/j.meegid.2010.12.003>
- Nelson, E. J., Harris, J. B., Morris, J. G., Calderwood, S. B., & Camilli, A. (2009). Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nature Reviews. Microbiology*, *7*(10).
<http://doi.org/10.1038/nrmicro2204>
- Octavia, S., Salim, A., Kurniawan, J., Lam, C., Leung, Q., Ahsan, S., ... Lan, R. (2013). Population Structure and Evolution of Non-O1/Non-O139 *Vibrio cholerae* by Multilocus Sequence Typing. *PLoS ONE*, *8*(6), e65342. <http://doi.org/10.1371/journal.pone.0065342>
- Okada, N., Matsuda, S., Matsuyama, J., Park, K.-S., de los Reyes, C., Kogure, K., ... Iida, T. (2010). Presence of genes for type III secretion system 2 in *Vibrio mimicus* strains. *BMC Microbiology*, *10*, 302.
<http://doi.org/10.1186/1471-2180-10-302>
- O'Shea, Y. A., Finnan, S., Reen, F. J., Morrissey, J. P., O'Gara, F., & Boyd, E. F. (2004). The *Vibrio* seventh pandemic island-II is a 26·9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43·4 kb genomic island in *V. vulnificus*. *Microbiology*, *150*(12), 4053–4063. <http://doi.org/10.1099/mic.0.27172-0>
- Paillard, C., Le Roux, F., & Borrego, J. J. (2004). Bacterial disease in marine bivalves, a review of recent studies: Trends and evolution. *Aquatic Living Resources*, *17*(04), 477–498.
<http://doi.org/10.1051/alr:2004054>
- Park, K.-S., Ono, T., Rokuda, M., Jang, M.-H., Okada, K., Iida, T., & Honda, T. (2004). Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. *Infection and Immunity*, *72*(11), 6659–6665. <http://doi.org/10.1128/IAI.72.11.6659-6665.2004>

- Rajanna, C., Wang, J., Zhang, D., Xu, Z., Ali, A., Hou, Y.-M., & Karaolis, D. K. R. (2003). The vibrio pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products. *Journal of Bacteriology*, *185*(23), 6893–6901.
- Romalde, J. L., Dieguez, A. L., Lasa, A., & Balboa, S. (2014). New *Vibrio* species associated to molluscan microbiota: a review. *Frontiers in Microbiology*, *4*. <http://doi.org/10.3389/fmicb.2013.00413>
- Rutherford, K., Parkhill, J., Horsnell, T., Rice, P., Rajandream, M.-A., & Barrell, B. G. (2000). Artemis: sequence visualization and annotation. *Bioinformatics*, *16*(10), 944–5.
- Schirmeister, F., Dieckmann, R., Bechlars, S., Bier, N., Faruque, S. M., & Strauch, E. (2014). Genetic and phenotypic analysis of *Vibrio cholerae* non-O1, non-O139 isolated from German and Austrian patients. *European Journal of Clinical Microbiology & Infectious Diseases*, *33*(5), 767–778. <http://doi.org/10.1007/s10096-013-2011-9>
- Sharma, C., Thungapathra, M., Ghosh, A., Mukhopadhyay, A. K., Basu, A., Mitra, R., ... Nair, G. B. (1998). Molecular Analysis of Non-O1, Non-O139 *Vibrio cholerae* Associated with an Unusual Upsurge in the Incidence of Cholera-Like Disease in Calcutta, India. *Journal of Clinical Microbiology*, *36*(3), 756–763.
- Shin, O. S., Tam, V. C., Suzuki, M., Ritchie, J. M., Bronson, R. T., Waldor, M. K., & Mekalanos, J. J. (2011). Type III Secretion Is Essential for the Rapidly Fatal Diarrheal Disease Caused by Non-O1, Non-O139 *Vibrio cholerae*. *mBio*, *2*(3), e00106–11. <http://doi.org/10.1128/mBio.00106-11>
- Siddique, A. K., Nair, G. B., Alam, M., Sack, D. A., Huq, A., Nizam, A., ... Sack, R. B. (2010). El Tor cholera with severe disease: a new threat to Asia and beyond. *Epidemiology & Infection*, *138*(03), 347–352. <http://doi.org/10.1017/S0950268809990550>
- Suzuki, Y., Kubota, H., Sato'o, Y., Ono, H. K., Kato, R., Sadamasu, K., ... Kamata, Y. (2015). Identification and characterization of novel *Staphylococcus aureus* pathogenicity islands encoding staphylococcal enterotoxins originating from staphylococcal food poisoning isolates. *Journal of Applied Microbiology*. <http://doi.org/10.1111/jam.12786>

- Takemura, A. F., Chien, D. M., & Polz, M. F. (2014). Associations and dynamics of Vibrionaceae in the environment, from the genus to the population level. *Frontiers in Microbiology*, 5. <http://doi.org/10.3389/fmicb.2014.00038>
- Taviani, E., Spagnoletti, M., Ceccarelli, D., Haley, B. J., Hasan, N. A., Chen, A., ... Colwell, R. R. (2012). Genomic analysis of ICEVchBan8: An atypical genetic element in *Vibrio cholerae*. *FEBS Letters*, 586(11), 1617–1621. <http://doi.org/10.1016/j.febslet.2012.03.064>
- Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology*, 3(9), 711–721. <http://doi.org/10.1038/nrmicro1234>
- Vezzulli, L., Guzmán, C. A., Colwell, R. R., & Pruzzo, C. (2008). Dual role colonization factors connecting *Vibrio cholerae*'s lifestyles in human and aquatic environments open new perspectives for combating infectious diseases. *Current Opinion in Biotechnology*, 19(3), 254–259. <http://doi.org/10.1016/j.copbio.2008.04.002>
- Wachsmuth, K., Blake, P. A., & Olsvik, Ø. (1994). *Vibrio cholerae and cholera: molecular to global perspectives*. Washington, D.C.: ASM Press.
- Weekly Epidemiological Record: Cholera 2013. (2014). *WHO*, 89(31), 345–356.
- Zarei, M., Eskandari, M. H., & Keshtkaran, S. (2014). Survival of Normal and Chlorine-Stressed Pathogenic and Non-Pathogenic *Vibrio parahaemolyticus* Under Adverse Conditions. *Jundishapur Journal of Microbiology*, 7(3), e9313. <http://doi.org/10.5812/jjm.9313>

Appendix A

ALIGNMENTS OF REPEATED REGIONS IN AM-19226

A.

Score	Expect	Identities	Gaps	Strand
366 bits(198)	1e-106	198/198(100%)	0/198(0%)	Plus/Plus
Query 1	TTGTGCATGCTCAACTTGC GCCCCTCTGTCACCCGCTCTGCGGCAACACCTTGTAAAGGT	60		
Sbjct 1	TTGTGCATGCTCAACTTGC GCCCCTCTGTCACCCGCTCTGCGGCAACACCTTGTAAAGGT	60		
Query 61	ATGGTGCCCTTTCACCTGGTGAAGTTCCGCTTCCGTTTGAGTGCCATTCTCGACAAATGT	120		
Sbjct 61	ATGGTGCCCTTTCACCTGGTGAAGTTCCGCTTCCGTTTGAGTGCCATTCTCGACAAATGT	120		
Query 121	TTGAGTCGTAACAGAGTTTTGCTCCACTGACGTCTGTGTTGAGCTGGACTCAACACTGGA	180		
Sbjct 121	TTGAGTCGTAACAGAGTTTTGCTCCACTGACGTCTGTGTTGAGCTGGACTCAACACTGGA	180		
Query 181	TTGAGTCGATCTGTTTTC	198		
Sbjct 181	TTGAGTCGATCTGTTTTC	198		

B.

Score	Expect	Identities	Gaps	Strand
324 bits(175)	9e-94	189/198(95%)	0/198(0%)	Plus/Plus
Query 1	TTGTGCATGCTCAACTTGC GCCCCTCTGTCACCCGCTCTGCGGCAACACCTTGTAAAGGT	60		
Sbjct 1	TTGTGCATGCTCAACTTGTGCCCCCTCTGTCACCCGCTCTGCGGCAACACCTTGTAAAGGT	60		
Query 61	ATGGTGCCCTTTCACCTGGTGAAGTTCCGCTTCCGTTTGAGTGCCATTCTCGACAAATGT	120		
Sbjct 61	ATGGTGCCCKTTCACRGTGAAGTTCCGCTTCCGTTTGAGTACCATTCTCGGCAAACGT	120		
Query 121	TTGAGTCGTAACAGAGTTTTGCTCCACTGACGTCTGTGTTGAGCTGGACTCAACACTGGA	180		
Sbjct 121	TTGAGTCGTAACAGAGTTTTGCTCGMCTGACGTCTGTGTTGAGCTGGACTCWACACTGGA	180		
Query 181	TTGAGTCGATCTGTTTTC	198		
Sbjct 181	TTGAGTCGATCTGTTTTC	198		

Figure 18 BLAST alignments of the 201 base pair repeated region found in *V. cholerae* AM-19226. A. The subject of the alignment is the first repeat in the sequence was and the query is the third repeat. B. The subject of the alignment is the first repeat in the sequence was and the query is the second repeat. Note that repeats one and three are 100% identical, while repeats one and two are only 95% identical.

Appendix B

EXAMINING PAI EXCISION IN RECA, HIMA, AND FIS MUTANTS IN *V. CHOLERA*

The genes *himA* and *fis* have been found to affect excision rates of the λ bacteriophage and SXT, which are mobile genetic elements that behave similarly to PAIs. Another gene, *recA* is known to influence homologous recombination, whereas site-specific recombination, the mechanism by which PAIs excise, can occur independently of *recA*. We used three mutant strains of *V. cholerae*, Bah-3 ($\Delta recA$), Sm463 ($\Delta himA$), and Sm465 (Δfis), to determine if the absence of these genes, which are located outside of the PAI, would affect excision of VPI-1 and VPI-2 in *V. cholerae*. We found that *himA*, *fis*, and *recA* do not play a significant role in VPI-1 and VPI-2 excision.

Table 5 Mutants examined, size of selected gene in wild-type, and size of selected gene

	Size of Mutant (bp)	Size of WT (bp)
Sm465 (Δfis)	192	399
Sm463 ($\Delta himA$)	75	297
Bah-3 ($\Delta recA$)	~540	1656

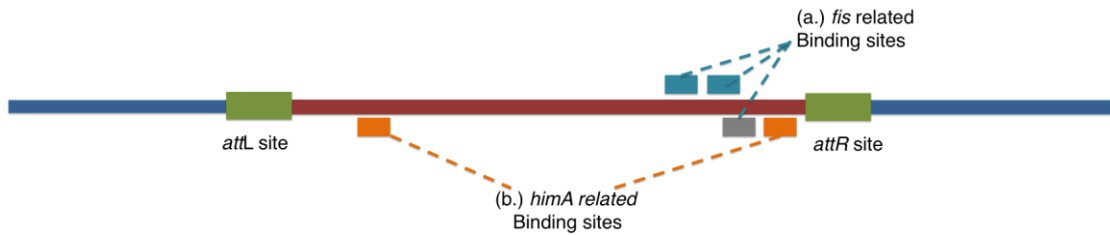


Figure 19 Diagram of excision related proteins examined in this study. A. Excision is found to be most efficient in the λ bacteriophage when the phage encoded excisionase (Xis) protein is bound to one of two binding sites (shown in teal) and the Fis protein is bound to a separate site (shown in grey). B. HimA is the alpha subunit of the protein Integration Host Factor (IHF). The binding of IHF to specific sites on the MGE chromosome (shown in orange) is required in order for the *attR* and *attL* sites on the bacterial chromosome to recombine and reform the *attB* and *attP* sites on the chromosome and the PAI (i.e for the excision event to occur).

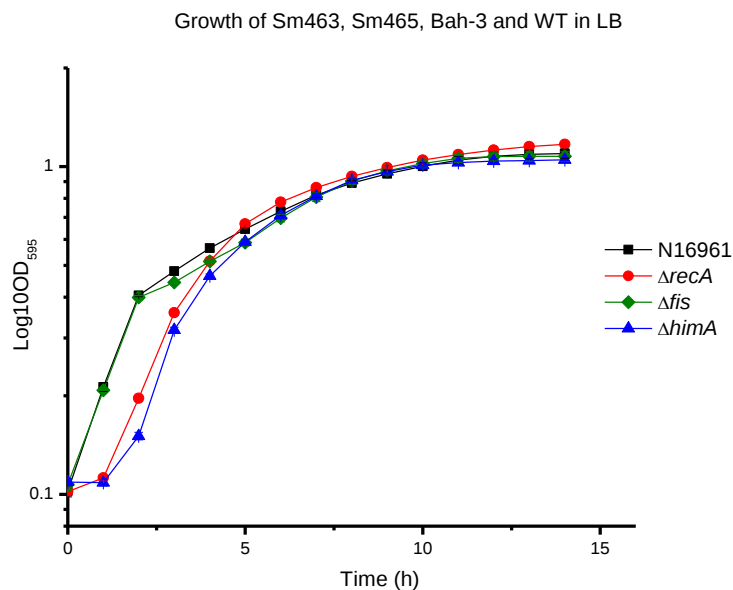


Figure 20 Growth curve analysis of mutants included in this study. No significant growth defects resulted from the deletion of the *recA*, *himA*, or *fis* genes.

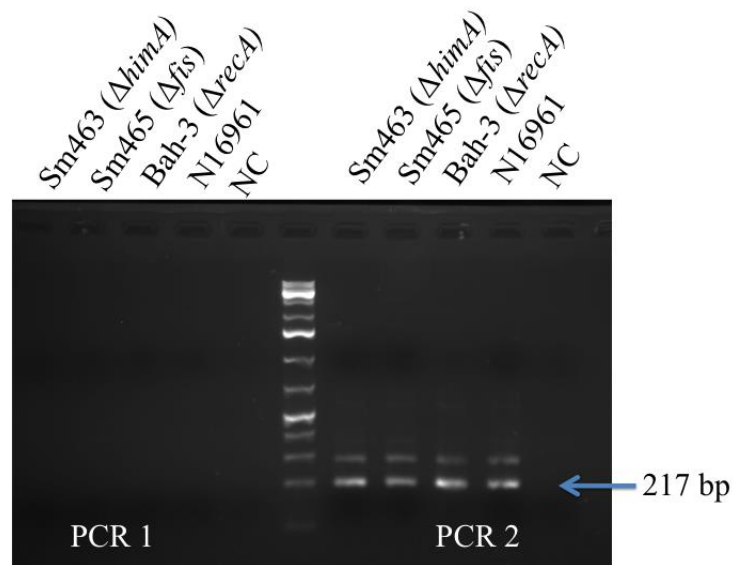


Figure 21 VPI-1 *attB1* Excision Assay Results. Excision occurs in all deletion mutants, indicating that the presence of the examined genes is not necessary for the excision of *V. cholerae* pathogenicity island VPI-1

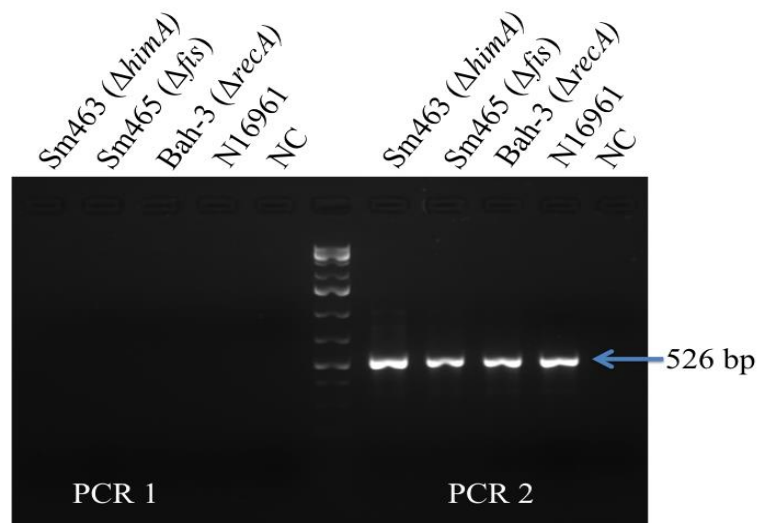


Figure 22 VPI-2 *attB2* Excision Assay results. Excision occurs in all deletion mutants, indicating that the presence of the examined genes is also not necessary for the excision of *V. cholerae* PAI VPI-2.