

**EXCISION MECHANISMS OF  
PATHOGENICITY ISLANDS AND PHAGES AMONG VIBRIO PATHOGENS**

by

Megan R. Carpenter

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Sciences

Summer 2016

© 2016 Megan Carpenter  
All Rights Reserved

ProQuest Number: 10190438

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10190438

Published by ProQuest LLC (2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

**EXCISION MECHANISMS OF  
PATHOGENICITY ISLANDS AND PHAGES AMONG VIBRIO PATHOGENS**

by

Megan R. Carpenter

Approved: \_\_\_\_\_  
Robin W. Morgan, Ph.D.  
Chair of the Department of Biological Sciences

Approved: \_\_\_\_\_  
George H. Watson, Ph.D.  
Dean of the College of Arts and Sciences

Approved: \_\_\_\_\_  
Ann L. Ardis, Ph.D.  
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

E. Fidelma Boyd, Ph.D.  
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Thomas E. Hanson, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Julie A. Maresca, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

John H. McDonald, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

---

Salil A. Lachke, Ph.D.  
Member of dissertation committee

## **ACKNOWLEDGMENTS**

I would like to thank my advisor, Prof. E. Fidelma Boyd, for giving me the opportunity to pursue a graduate education in her laboratory. If not for her ability to see potential in non-traditional students, I would not be here today. Under her guidance I was able to master core molecular and microbiology skills, learn to critically evaluate literature, design and execute publishable projects, prepare manuscripts and much more than I can list here. She gave me the opportunity to manage other students in the lab, allowing me to create my own mentoring experience and ultimately teaching me ways to improve my management skills. Her mentoring style required great preparation before presenting new ideas and allowed me to identify potential pitfalls before starting an experiment, ultimately giving me the skills to become an independent scientist. In addition to the research competencies I have gained in her laboratory, she has taught me to work hard, treasure personal time and most importantly to believe in myself. She will remain a life-long mentor and friend. Her superb mentoring has shaped me into the independent scientist I am today and has thoroughly prepared me for the scientific endeavors that lie ahead.

Over the past five years my committee members have challenged me with tough questions, requiring examination of my research from different angles, provided valuable feedback and helped me realize when it was time to move on from an experiment. Their doors were always open, even outside of committee meetings. I am grateful for their time and contributions to my graduate career.

I would like to thank all my fellow lab members, past and present, for providing a great support network throughout graduate school. We collectively took time to ensure each student is thoroughly prepared for prelims, presentations, qualifying exams and committee meetings, all which created meaningful discussion and lasting bonds. I could not have asked for better coworkers and friends. I would particularly like to express gratitude for a former post-doc, Nitto Chowdhury, who trained me when I first entered the lab. He was instrumental in aiding my understanding of new molecular techniques which set me up to have a successful graduate career. Sai Kalburge, whom I entered the program with, has been a great friend and one of my biggest advocates. I have talked over experiments at nausea with Sai and Nate McDonald. Thank you for listening and providing significant feedback.

I would also like to thank Dr. Sharon Rozovsky and her group for generously allowing me to perform protein work and making me feel welcome in their lab. Sharon has been an additional mentor of mine the last few years of graduate school, as well as a friend. In her lab I learned the basics of protein purification and successfully purified three proteins in relatively short time. Thank you, Jun Li, Frank Zhang and Maggie Chen for their assistance during my time there.

There are additional department faculty I wish to thank. Dr. Melinda Duncan and Dr. Erica Selva for always being willing to discuss my research and offer advice on experimental design and Dr. Duncan for lending equipment. Betty Cowgill is an incredible individual that has been a lifesaver throughout my graduate degree by helping me solve almost any administrative issue – she goes above and beyond and is always thinking ahead.

My partner Karley has been my rock throughout grad school. She's constantly making sure I am maintaining my health -- exercising to relieve stress, eating well and drinking enough water -- all essential for optimal focus and energy as I completed my graduate degree. She was patient and listened to endless practice research talks.

My parents Neal and Peggy Carpenter have played a large part in contributing to the person I am today. They have worked their whole life to ensure I have a better one than theirs. Their dedication allowed me to attend college and receive a degree, something neither of them had the opportunity to do. They continued to support me through my graduate career, both emotionally and financially, when they could, to ease the financial burdens that accompany grad school. Their help allowed me to focus on my coursework and research. My sister, Devon has always encouraged me to accomplish this goal and has been nothing but supportive. My grandparents, Paul and Yoshi Carpenter, have consistently accepted and supported me throughout college and grad school. They funded a course abroad and traveled 8 hours through a blizzard with my parents to see me give my very first research talk. Their support and enthusiasm has pushed me to become the best scientist I could be.

I have many others to thank: my long-time friend Marcia who was always willing to proofread anything for me, from funding applications to research abstracts and always made time to listen; my closest grad school friends, Carrie Barnum and Katie Robbins, were great study partners and provided ease during stressful times, often in the form of coffee and cupcakes; and the Shorzulaks for providing weekend fun and trips, allowing me some time away from the lab.

Without the support of all of these people I am sure my graduate career would not have been as successful as it was.

## TABLE OF CONTENTS

LIST OF TABLES .....	xi
LIST OF FIGURES .....	xii
ABSTRACT .....	xiv

### Chapter

1	INTRODUCTION .....	1
	Mobile Integrative Genetic Elements (MIGEs) and Horizontal Gene Transfer (HGT).....	1
	Pathogenicity Islands (PAIs) .....	2
	<i>Vibrio cholerae</i> and the Diarrheal Disease Cholera .....	5
	Vibrio Pathogenicity Islands (VPIs).....	7
	PAI Mobility.....	15
	Dissertation Work.....	15
2	PATHOGENICITY ISLAND CROSS-TALK MEDIATED BY RECOMBINATION DIRECTIONALITY FACTORS FACILITATES EXCISION FROM THE CHROMOSOME.....	19
	Introduction .....	19
	Materials and Methods .....	23
	Bacterial strains, plasmids and growth conditions. ....	23
	Mutant strain construction. ....	24
	Excision assays. ....	25
	Mutant strain complementation. ....	26
	Quantitative real-time PCR. ....	27
	Protein purification of VefA and VefB. ....	27
	Electrophoretic Mobility Shift Assays. ....	29
	Results .....	30
	All three RDFs VefA, VefB and VefC can induce VPI-2 excision. ....	30
	<i>intV2</i> can mediate VPI-1 excision in the absence of <i>intV1</i> . ....	33
	RDFs are not essential for VPI-1 excision but can promote excision. ....	37
	RDFs are negative regulators of <i>intV1</i> and <i>intV2</i> . ....	39

	Over-expression of <i>intV2</i> can compensate for the absence of RDFs in VPI-2 excision .....	40
	Protein purification of VefA and VefB .....	41
	VefA and VefB can bind the <i>att</i> sites of VPI-1 and VPI-2. ....	44
	Distribution of RDFs among <i>Vibrio</i> .....	50
	Discussion.....	53
3	CONSERVED RECOMBINATION MODULES CARRIED ON EXCISABLE PATHOGENICITY ISLANDS WITH NOVEL CARGO GENES IN <i>VIBRIO CHOLERAE</i> .....	62
	Introduction .....	62
	Materials and Methods .....	65
	Bacterial Strains, Plasmids, & Growth Conditions. ....	65
	Mutant Construction. ....	66
	Excision Assays.....	67
	Comparative Genomic and phylogenetic analysis. ....	67
	Results .....	68
	Recombination modules and cargo genes of VPI-3 and VPI-6.....	68
	VPI-3 can excise from the <i>V. cholerae</i> NRT36S bacterial genome. ....	73
	VPI-6 can excise from the RC385 <i>V. cholerae</i> genome.....	74
	The cognate integrase of VPI-3 is necessary for island excision. ....	76
	Comparative and phylogenetic analyses of VPI-3 and VPI-6.....	77
	Discussion.....	84
4	QUORUM SENSING REGULATION OF EXCISION OF THE FILAMENTOUS PHAGE F237 IN <i>VIBRIO PARAHAEMOLYTICUS</i> .....	93
	Introduction .....	93
	Materials and Methods .....	97
	Bacterial strains used and growth conditions. ....	97
	Characterization of the f237 genome and its attachment <i>att</i> sites .....	98
	Excision assays for f237 (Detection of replicative form of f237) .....	99
	RNA isolation and cDNA synthesis .....	100
	Real-time quantitative PCR.....	100
	Protein purification of OpaR. ....	101
	Electrophoretic Mobility Shift Assays .....	103

Results .....	103
Genome organization of phage f237 .....	103
Detection of the replicative form of f237 .....	106
Higher levels of f237 circular intermediates detected in a <i>luxO</i> mutant ..	107
Higher levels of f237 <i>attB</i> products detected in a <i>luxO</i> mutant .....	108
Expression of f237 genes is increased in a <i>luxO</i> mutant .....	108
Protein purification of OpaR .....	111
OpaR can bind to predicted promoter regions in f237 phage.....	112
Discussion.....	113
5 PERSPECTIVES AND FUTURE DIRECTIONS.....	119
Cross-talk Mediated Pathogenicity Island Excision.....	119
Dual Functionality of VPI Excisionases.....	122
REFERENCES .....	126
Appendix	
A PERMISSION LETTER .....	140
B THE ROLE OF BACTERIAL HOST-ENCODED FACTORS AND ENVIRONMENTAL STIMULI ON VPI EXCISION .....	141
A.1 Host Encoded Factors .....	141
A.2 Intestinal Mucus and RDF Expression .....	144

## LIST OF TABLES

Table 1	Bacterial strains and plasmids used in this study. ....	58
Table 2	Primers used in this study. ....	60
Table 3	Bacterial strains and plasmids used in this study ....	87
Table 4	Primers used in this study. ....	88
Table 5	ORFs in <i>V. cholerae</i> NRT36S VPI-3 ....	89
Table 6	ORFs in <i>V. cholerae</i> RC385 VPI-6 and predicted gene products based on BLAST homology search. ....	92
Table 7	Strains used in this study. ....	116
Table 8	Primers used in this study. ....	117
Table 9	f237 ORF descriptions and predicted function. ....	118

## LIST OF FIGURES

Figure 1	Vibrio pathogenicity island 1(VPI-1).....	8
Figure 2	Vibrio pathogenicity island 2 (VPI-2).....	10
Figure 3	Vibrio seventh pandemic islands (VSP-I and VSP- II). ....	11
Figure 4	Vibrio pathogenicity island excision and integration. ....	13
Figure 5	Vibrio excision factors (vefs). ....	14
Figure 6	<i>Vibrio cholerae</i> wild-type excision. ....	31
Figure 7	RDFs promote VPI-2 excision. ....	33
Figure 8	Role of integrases and RDFs in excision of VPI-1.....	36
Figure 9	Conserved domains of IntV1 and IntV2.....	37
Figure 10	Predicted secondary structure of N-terminus VpiT and TorI. ....	38
Figure 11	RDFs act as transcriptional repressors of the integrases <i>intV1</i> and <i>intV2</i> . ....	40
Figure 12	SDS-PAGE of MBP-VefA purification: Expression test, Amylose column and TEV cleavage.....	42
Figure 13	SDS-PAGE of MBP-VefA purification: TEV cleavage and IMAC. ....	43
Figure 14	SDS-PAGE of MBP-VefB purification: Expression Test, TEV cleavage, IMAC and Heparin Cleanup.....	44
Figure 15	Mobile integrative genetic element (MIGE) recombination modules.....	45
Figure 16	VefA and VefB bind the att sites of VPI-1 and VPI-2.....	48
Figure 17	VefA and VefB bind the 1st half <i>attL2</i> site at a higher affinity. ....	49
Figure 18	Distribution of Vefs in Bacteria. ....	50
Figure 19	Vibrio island regions inserted at tmRNA genes containing RDFs.....	53
Figure 20	Gene organization and recombination modules of VPI-3 and VPI-6.....	70
Figure 21	VPI-3 can excise from chromosome 1 of <i>V. cholerae</i> NRT36S. ....	74

Figure 22	VPI-6 can excise from chromosome 1 of <i>V. cholerae</i> RC385. ....	76
Figure 23	The integrase ( <i>intV2</i> ) is essential for VPI-3 excision. ....	77
Figure 24	Evolutionary relationship between T3SS-containing VPIs. ....	79
Figure 25	T3SS ATPase phylogeny. ....	82
Figure 26	Comparison of VPI-6 like islands in <i>V. cholerae</i> strains. ....	84
Figure 27	f237 and other well-known filamentous phages. ....	99
Figure 28	f237 phage excision confirmed using attP and attB detection PRC assays. ....	107
Figure 29	The filamentous phage f237 and phage gene expression. ....	109
Figure 30	Bioinformatics identification of putative OpaR binding sites. ....	110
Figure 31	SDS-PAGE of OpaR protein purification. ....	111
Figure 32	OpaR can bind promoter regions in f237 phage. ....	112
Figure 33	OpaR EMSA negative control. ....	113
Figure 34	Vibrio pathogenicity island (VPI) crosstalk. ....	121
Figure 35	Model for VPI Excision. ....	123
Figure 36	Schematic of putative IHF binding sites on VPI-1 and VPI-2. ....	142
Figure 37	VPI-1 <i>attB1</i> Excision Assay Results. ....	143
Figure 38	VPI-2 <i>attB2</i> Excision Assay results. ....	143
Figure 39	Vef Expression in M9 Glucose and M9 Mucus. ....	145
Figure 40	<i>Vibrionaceae</i> PAI Integrase phylogeny. ....	146
Figure 41	<i>Vibrionaceae</i> PAI RDF phylogeny. ....	147
Figure 42	<i>Vibrionaceae</i> RNA polymerase B phylogeny. ....	148
Figure 43	Phylogeny of Vibrio MIGE Integrases (Condensed). ....	149

## ABSTRACT

Pathogenicity islands (PAIs) are mobile integrated genetic elements (MIGEs) that contain a diverse range of virulence factors and are essential in the evolution of pathogenic bacteria. PAIs are widespread among bacteria and integrate into the host genome, commonly at a tRNA locus, via integrase mediated site-specific recombination. The excision of PAIs is the first step in the horizontal transfer of these elements and is not well understood. The work in this dissertation examined the role of recombination directionality factors (RDFs) and their relationship with integrases in the excision of two PAIs essential for *Vibrio cholerae* host colonization: Vibrio pathogenicity island-1 (VPI-1) and VPI-2. VPI-1 does not contain an RDF, which allowed us to answer the question of whether RDFs are an absolute requirement for excision. We found that an RDF was required for efficient excision of VPI-2 but not VPI-1, and that RDFs can induce excision of both islands. Expression data revealed that the RDFs act as transcriptional repressors to both VPI-1 and VPI-2 encoded integrases. We demonstrated that the RDFs Vibrio excision factor (Vef) A and VefB bind at the attachment sites (overlapping the int promoter region) of VPI-1 and VPI-2, thus supporting this mode of integrase repression. In addition, *V. cholerae* RDFs are promiscuous due to their dual functions of promoting excision of both VPI-1 and VPI-2 and acting as negative transcriptional regulators of the integrases. This is the first demonstration of cross-talk between PAIs mediated via RDFs which reveals the complex interactions that occur between separately acquired MIGEs.

In chapter 3 we identify several islands with novel cargo genes and variant combinations of the VPI recombination modules described in chapter 2. An island we named VPI-3 in *V. cholerae* NRT36S contains a type three secretion system (T3SS) and an island we named VPI-6 in *V. cholerae* RC385 contains genes encoding for a CRISPR-Cas system and type VI secretion system (T6SS) were examined in this study. We showed that both VPI-3 and VPI-6 can excise from the bacterial chromosome. Evolutionary analysis of these island regions reveals a modular structure indicating that parts of these islands were likely acquired separately. These data demonstrate that identical recombination modules that catalyze integration and excision from the chromosome can acquire diverse cargo genes.

In chapter 4 we examine the role of host encoded factors on excision of a filamentous phage named f237, found in the human pathogen *Vibrio parahaemolyticus*. Here we demonstrate that the quorum sensing regulator LuxO is involved in regulating f237 excision and transcription of f237 encoded genes. Specifically, we present evidence for the direct regulation of phage f237 by the high cell density quorum sensing regulator OpaR. In a *luxO* mutant, cells are locked in a high cell density state and OpaR is constitutively expressed; in this mutant we showed that several f237 genes are more highly expressed, ranging from 3-fold to over 60-fold, relative to wild-type. We found that this increase in expression also correlated with a greater than 20-fold increase in production of f237 phage circular intermediates. Additionally, we found an increase in the amount of the *attB* excision product in a *luxO* mutant relative to wild-type. The function of this excision remains unknown and its effects on the physiology of the cell, if any. Given that increased f237 expression and circular intermediate (CI) production were observed in a *luxO* mutant, we

determined whether OpaR may be responsible for these phenotypes. We confirmed that OpaR could bind to these DNA regions using EMSA assays. All these data support the model that OpaR is directly and positively regulating the transcription of *f237* genes as well as production of *f237* circular intermediates.

## Chapter 1

### INTRODUCTION

#### **Mobile Integrative Genetic Elements (MIGEs) and Horizontal Gene Transfer (HGT)**

Mobile integrative genetic elements (MIGEs) are the vectors of horizontal gene transfer (HGT) among bacteria via the mechanisms of transformation, transduction or conjugation (Frost, Leplae et al. 2005). Members of MIGEs include bacteriophages, plasmids, integrative conjugative elements (ICEs), pathogenicity islands (PAIs), integrons and transposons and these MIGEs have been instrumental in shaping bacterial evolution (Boyd 2012). Commensal bacteria can be transformed into deadly pathogens through HGT of MIGEs that contain virulence factors, or a pathogen may become more virulent with the acquisition of additional virulence factors carried on a MIGE. The rapid evolution of a bacterium by HGT is exemplified by the emergence of several different pathovars of the important gut commensal *Escherichia coli* (Kaper, Nataro et al. 2004; Dobrindt, Chowdary et al. 2010). For example, acquisition of a bacteriophage carrying shiga-toxin contributed to the emergence of enterohaemorrhagic *E. coli* (EHEC) diarrhea, a leading cause of death in children under 5 in the developing world (Kaper, Nataro et al. 2004). The large 2011 outbreak of diarrhea causing haemolytic uremic syndrome in Germany was caused by an enteroaggregative *E. coli* (EAEC) O104:H4 strain that acquired the shiga toxin genes on a bacteriophage (Rasko, Webster et al. 2011; Grad, Lipsitch et al. 2012; Grad, Godfrey et al. 2013). *Vibrio cholerae*, the cause of the secretory diarrhea disease,

cholera, contains a filamentous bacteriophage that encodes the cholera toxin genes as well as two PAIs essential for virulence, which are all absent from non-pathogenic isolates (Taylor, Miller et al. 1987; Karaolis, Somara et al. 1999; Jermyn and Boyd 2002). PAIs are the least studied MIGEs.

### **Pathogenicity Islands (PAIs)**

The origin of PAIs is unknown. It was originally proposed that PAIs may have arisen from bacteriophages by deletion events that resulted in the loss of several phage-associated genes and mobility genes (Hacker and Kaper 2000). However, phylogenetic analysis of tyrosine recombinase (TR) integrases from PAIs and various MIGEs, including bacteriophage, ICEs and integrons show separate clustering patterns, distinct from PAI TR integrases. This data strongly suggests that PAIs are an evolutionarily distinct class of mobile elements, unrelated to phages (Boyd, Almagro-Moreno et al. 2009; Napolitano, Almagro-Moreno et al. 2010).

PAIs were first described in Uropathogenic *E. coli* (UPEC) and have since been described in most pathogenic Gamma-Proteobacteria including all pathovars of *E. coli*, *Salmonella enterica*, *Vibrio* species, *Yersinia pestis* and *Shigella* species, to name but a few (Carniel, Guilvout et al. 1996; Karaolis, Johnson et al. 1998; Hacker and Kaper 2000; Al-Hasani, Rajakumar et al. 2001; Jermyn and Boyd 2002; Hensel 2004; O'Shea, Finnan et al. 2004; Jermyn and Boyd 2005; Hurley, Quirke et al. 2006; Quirke, Reen et al. 2006; Murphy and Boyd 2008). PAIs range in size from 10-200 kb and have a GC content that differs from that of the core bacterial chromosome (Hacker and Kaper 2000). These island-encoded virulence factors are found integrated in the chromosome of pathogenic bacteria, but are absent from non-pathogenic strains or closely related species (Hacker and Kaper 2000). PAIs are found integrated in the

chromosome at the 3' end of tRNA loci (Hacker and Kaper 2000). In the chromosome they are found flanked by small direct repeats (DR) generated after integration. These DR sites may also be used for excision from the chromosome (Hacker and Kaper 2000). It is important to note that PAIs are not capable of autonomous replication or transfer (Boyd, Almagro-Moreno et al. 2009). Another defining characteristic of PAIs are the presence of an integrase.

The tyrosine recombinase (TR) family of integrases catalyze recombination between DNA substrates in the integration and excision of MIGEs including bacteriophage, ICEs and PAIs (Esposito and Scocca 1997; Rajeev, Malanowska et al. 2009). In the “branch migration” model, TR integrases catalyze the recombination event by using a tyrosine nucleophile to attack a scissile phosphate (Weisberg, Enquist et al. 1983). An intasome, which consists of four integrase monomers, binds at either ends of the DNA substrates – two monomers are active and two are inactive. The two active monomers cleave DNA on each strand, leading to strand exchange and formation of a Holiday Junction (Nunes-Duby, Matsumoto et al. 1987; Kitts and Nash 1988). Following this event, a conformation change occurs and the other two integrase monomers become active and perform the second set of DNA cleavage leading to strand exchange and formation of the recombinant product (Kitts and Nash 1988). Unlike homologous recombination, TR integrases require a short region of DNA homology as small as 6 -8 bp called the crossover region which is critical to the reaction and is flanked by recombinase binding sites (Rajeev, Malanowska et al. 2009). TR integrases are necessary for both integration and excision of several MIGEs (Ramsay, Sullivan et al. 2006; Lee, Auchtung et al. 2007; Flanigan and Gardner 2008; Fogg, Rigden et al. 2011). In prophage it was demonstrated that the

excision event also requires an additional protein, an excisionase or recombination directionality factor (RDF). (Sam, Papagiannis et al. 2002; Warren, Sam et al. 2003; Panis, Mejean et al. 2007; Coddeville and Ritzenthaler 2010; Singh, Plaks et al. 2013). In fact, it is believed that all TR integrases require an RDF to perform excision (Lewis and Hatfull 2001; Fogg, Rigden et al. 2011). TR integrases and RDFs have not been studied extensively in bacteriophage biology but little is known of their role in PAIs physiology.

RDFs are small DNA-binding accessory proteins that aid TR integrases in controlling the directionality of their site-specific recombination reactions (Lewis and Hatfull 2001). These proteins control directionality by favoring one reaction direction and interfering with the other. RDFs have been found to play an architectural role and in addition, sometimes a transcriptional role (Lewis and Hatfull 2001). RDFs are generally less than 100 amino acids and contain few residues that are highly conserved, making their identification through sequence comparison difficult (Lewis and Hatfull 2001). The majority of characterized RDFs are basic in charge, however several discovered are acidic (Lewis and Hatfull 2001). Due to the vast diversity of these proteins, in sequence and sometimes additional functions (transcriptional roles), it can be speculated that these proteins have more than one origin (Lewis and Hatfull 2001). Phylogenetic analysis of RDFs and their corresponding integrases from PAIs suggests that they co-evolve together (Napolitano, Almagro-Moreno et al. 2010). This co-evolution was also observed in the grouping patterns of bacteriophage and transposon integrases with RDFs (Lewis and Hatfull 2001).

## ***Vibrio cholerae* and the Diarrheal Disease Cholera**

*Vibrio cholerae* is a Gram-negative bacterium belonging to the *Vibrionaceae* family that is a member of the Class Gamma-Proteobacteria. *V. cholerae* inhabits brackish and estuarine waters, often in association with aquatic flora and fauna such as zooplankton, but has also been found as free swimming cells. *Vibrio cholerae* is the causative agent of the diarrheal disease cholera and can be serologically differentiated based on the O antigen present in the lipopolysaccharide. There are currently over 200 serovars of *V. cholerae* and only the O1 serogroup is associated with cholera pandemics. The O1 and O139 serogroups are associated with epidemic cholera (Chatterjee and Chaudhuri 2003). Infection occurs via ingestion of *V. cholerae* contaminated food or water and is prevalent in areas which lack the infrastructure to provide access to clean water.

The World Health Organization estimates that there are 3 – 5 million cases of cholera and 100,000 – 120,000 deaths world-wide each year, however these numbers are thought to be highly underestimated due to under reporting by countries in fear of travel and shipping restrictions. Since October of 2010, the start of the Haitian cholera outbreak, there have been over 470,000 reported cases and 6,631 deaths ([http://www.cdc.gov/haiticholera/haiti\\_cholera.htm](http://www.cdc.gov/haiticholera/haiti_cholera.htm)). It has now been confirmed that the strain responsible for the Haitian cholera outbreak is almost identical to the El Tor O1 strains prevalent in South Asia (Chin, Sorenson et al. 2011), and most likely brought over by United Nations peacekeepers following the January 2010 earthquake (Chin, Sorenson et al. 2011; Frerichs, Keim et al. 2012; Lantagne, Balakrish Nair et al. 2014; Bachmann, Kostiuk et al. 2015).

The genome of *V. cholerae* and many other *Vibrio* species have been shaped significantly by HGT (Boyd, Almagro-Moreno et al. 2009; Hazen, Pan et al. 2010;

Almagro-Moreno, Murphy et al. 2011; Faruque and Mekalanos 2012; Robins and Mekalanos 2014; Boyd, Carpenter et al. 2015). Several environmental factors contribute to signal HGT including low nutrient conditions as well as the ability of *Vibrio* to be naturally transformable in the presence of chitin (Meibom, Blokesch et al. 2005), which is found abundantly in their environmental habitat. *Vibrio cholerae* O1 El Tor, the cause of the seventh on going cholera pandemic, contains four PAIs. *Vibrio* pathogenicity island I (VPI-1) and VPI-2 are present in all cholera pandemic strains. The presence of two additional PAIs, named the *Vibrio* seventh pandemic island (VSP)-I and VSP-II, distinguishes biotype El Tor from the classical biotype *V. cholerae* strains that caused the first six pandemics (Karaolis, Johnson et al. 1998; Dziejman, Balon et al. 2002; Jermyn and Boyd 2002; O'Shea, Finnan et al. 2004).

The first genome sequence of a *V. cholerae* strain was N16961, a seventh pandemic isolate. The genome contains 4 million bp (Mbp), and characteristic of *Vibrio* species, it contains 2 circular chromosomes –2.9 Mbp and 1.1 Mbp in size. Chromosome 1, the larger of the two, contains essential genes for cell growth as well as many pathogenic genes, while chromosome 2 contains a large amount of hypothetical and unknown genes (Heidelberg, Eisen et al. 2000).

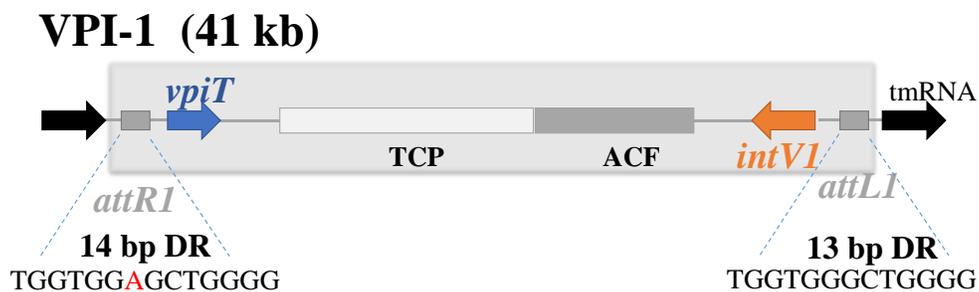
*V. cholerae* contains several factors that aid in virulence, including the cholera toxin (CT), toxin co-regulated pilus (TCP) and sialic acid scavenging, uptake, and catabolism genes. Each of these factors is found on a mobile genetic element and have been acquired by HGT (Waldor and Mekalanos 1996; Manning 1997; Karaolis, Johnson et al. 1998; Jermyn and Boyd 2002). CT is located in the temperate filamentous phage, CTX $\Phi$ , a 7 kb region which is integrated into chromosome I of *V. cholerae* biotype El Tor isolates (Waldor and Mekalanos 1996). The TCP, a type IV

bundle forming pilus, critical for intestinal colonization, is found within VPI-1, also on chromosome I (Taylor, Miller et al. 1987; Herrington, Hall et al. 1988; Manning 1997; Karaolis, Johnson et al. 1998). The TCP additionally serves as a receptor for the CTX $\Phi$  to enter the bacterial cell (Waldor and Mekalanos 1996). Therefore, only *V. cholerae* isolates that contain VPI-1 are capable of obtaining the CTX $\Phi$ , and thus the CT. A second PAI, VPI-2, also located on chromosome I and contains genes involved in sialidase production, sialic acid transport and catabolism (Jermyn and Boyd 2002). The sialidase acts to cleave higher order gangliosides on the intestinal cell wall. The result of this is twofold: the free sialic acid released from the intestinal mucosa may be used as a carbon source for *V. cholerae*, offering it a competitive advantage in the intestinal environment (Jermyn and Boyd 2002; Almagro-Moreno and Boyd 2009; Almagro-Moreno and Boyd 2009; McDonald, J.B. Lubin et al. 2016) and the cleavage exposes the GM1 receptor for the CT to bind enabling it to be taken up into the cell (Galen, Ketley et al. 1992).

### **Vibrio Pathogenicity Islands (VPIs)**

As mentioned previously, VPI-1 encodes the gene cluster required for the synthesis of the TCP, which is essential for host colonization of the small intestine as well as a gene cluster for the accessory colonization factor (ACF). Also VPI-1 contains genes that encode key regulators of virulence gene expression for both TCP and cholera toxin. VPI-1 is 41.2 kb in size with 30 open reading frames (ORFs) encompassing locus tags VC0817 to VC0847 and contains all the characteristics common to PAIs (Karaolis, Johnson et al. 1998) (**Figure 1**). VPI-1 has a TR integrase (*intVI*) located adjacent to a tRNA like gene, *ssrA* known as a transfer messenger (tm) RNA the site of insertion. VPI-1 is flanked by attachment (*att*) sites on either end

called direct repeats (DRs) and has a low GC content of 35% compared to that of the *V. cholerae* core chromosome, which is 48%. VPI-1 also encodes a transposase, VpiT (**Figure 1**). Karaolis and colleagues in a high profile paper proposed that VPI-1 was actually a second filamentous phage which they named VPIΦ (Karaolis, Somara et al. 1999). This was surprising since the 40 kb region did not contain any bona fide phage genes or a filamentous phage genome structure. In 2003, Mekalanos's group was unable to detect *tcpA* genes in phage preparations collected under identical conditions to that of Kaper's group. Additionally, they were unable to induce VPIΦ induction and several transduction experiments failed and they determined that VPI-1 was not a filamentous phage (Faruque, Zhu et al. 2003). Also in 2003, Karaolis' group published another paper doubting their original findings (Rajanna, Wang et al. 2003). It is now widely accepted that the VPI-1 region is not a phage but a typical PAI. Interestingly no one to date has re-examined the dynamics of VPI-1 excision in the *V. cholerae* chromosome.



**Figure 1 Vibrio pathogenicity island 1(VPI-1).** VPI-1 is a 41 kb region inserted at the transfer messenger (tm) RNA gene and contains genes encoding the toxin co-regulated pilus (TCP), accessory colonization factor (ACF) as well as an integrase (*intVI*) and a transposase-like gene (*vpiT*) and att sites *attL* and *attR*.

VPI-2 is 57.3 kb in size and consists of 52 open reading frames (ORFs) (VC1758 – V1809) (Jermyn and Boyd 2002) (**Figure 2**). In addition to the sialidase and sialic acid transport and catabolism genes, VPI-2 also contains a restriction modification system (Jermyn and Boyd 2002). Like VPI-1, this island contains a TR integrase (*intV2*), is flanked by DRs, has a GC content (42%), less than the core chromosome, has integrated into a tRNA Serine gene and contains no genes for autonomous replication or transfer (Jermyn and Boyd 2002; Murphy and Boyd 2008). Unlike VPI-1, two excisionases or RDFs, named *vefA* and *vefB*, were identified by bioinformatics analysis in VPI-2 of *V. cholerae* N16961 (**Figure 2**) (Boyd, Almagro-Moreno et al. 2009). A second pathovar of *V. cholerae* exists which is a prevalent cause of inflammatory diarrhea. This pathovar encodes a variant VPI-2 region named VPI-3 that along with the sialic acid metabolism genes contains a T3SS but an identical recombination module comprised of an integrase, RDFs and *att* sites (Dziejman, Serruto et al. 2005; Chen, Johnson et al. 2007; Murphy and Boyd 2008).

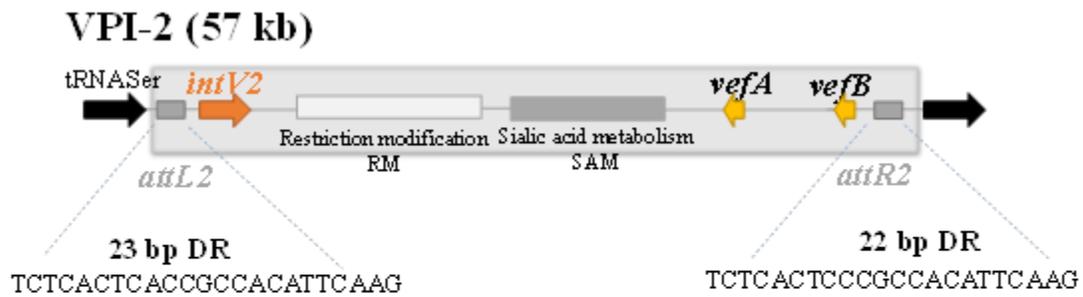
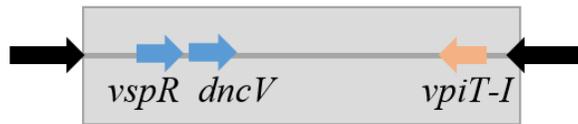


Figure 2 **Vibrio pathogenicity island 2 (VPI-2).** VPI-2 inserts into a tRNA-Serine locus and contains a restriction modification system (RMS), a sialic acid scavenging, transport and catabolism (SAM) region as well as an integrase (*intV2*), two recombination directionality factors (RDFs) named *vefA* and *vefB*, and two att sites *attL* and *attR*.

VSP-I is 16 kb in size, consists of 11 ORFs (VC0175 – VC0185) and has a GC content of 40% (Dziejman, Balon et al. 2002) (**Figure 3**). VSP-II is 27 kb in size, consists of 27 ORFs (VC0490 – VC0516) and has a GC content of 43% (Dziejman, Balon et al. 2002; O'Shea, Finnan et al. 2004). This island and its cognate TR integrase (*intV3*) is found adjacent to the tRNA Methionine gene and encodes an RDF named VefC (**Figure 3**) (O'Shea, Finnan et al. 2004). There is little known regarding the function of genes encoded within VSP-I and VSP-II. One study found that VC0179 of VSP-I encoded a dinucleotide cyclase (DncV) which produces a hybrid cyclic AMP-GMP molecule that acts to inhibit chemotaxis and promote colonization in the intestine (Davies, Bogard et al. 2012).

### VSP-I (16 kb)



### VSP-II (27 kb)

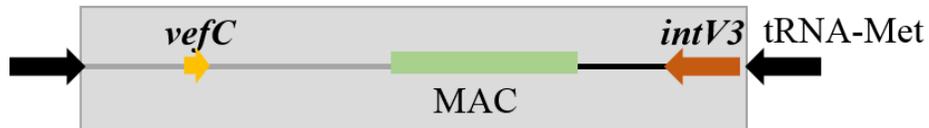


Figure 3 **Vibrio seventh pandemic islands (VSP-I and VSP- II).** VSP-I is 16 kb in size and consists of 11 ORFs (VC0175 – VC0185). VSP-I contains a novel di-nucleotide cyclase Vibrio (*dncV*) and transcriptional regulator *vspR* as well as a transposase like gene (*vpiT-I*). VSP-II is 27 kb in size, consists of 27 ORFs (VC0490 – VC0516) and is inserted at a tRNA-Methionine locus. VSP-II encodes genes for methyl-accepting chomotaxis (MAC) proteins as well as an integrase (*intV3*) and a RDF *vefC*.

The majority of work on PAI dynamics has been examined in the uropathogenic *E. coli* strain 536, which contains at least six PAIs (PAI I<sub>536</sub> – VI<sub>536</sub>). Five of these islands contain an intact TR integrase, which is flanked by DRs and has been shown to excise from the chromosome – PAI IV<sub>536</sub> is the exception (Hochhut, Wilde et al. 2006; Dobrindt, Chowdary et al. 2010). Additionally, Hochhut et al. found PAI II<sub>536</sub> integrase- mediated cross-talk via site-specific recombination at the *att* sites of PAI V<sub>536</sub>, meaning the PAI II<sub>536</sub> integrase could mediate both integration and excision of PAI V<sub>536</sub> (Hochhut, Wilde et al. 2006). Integration studies of PAI II<sub>536</sub> and PAI III<sub>536</sub> revealed that only 13-20 bp *att* sites were necessary for site specific recombination between *attP* and *attB* sites (integration), depending on the island.

Furthermore, the host-encoded DNA nucleoid structuring protein, integration host factor (IHF), enhanced excision of PAI III<sub>536</sub> and had a negative effect on integration, while recombination properties of PAI II<sub>536</sub> were unaffected (Wilde, Mazel et al. 2008). RDFs were not examined or identified in these studies, however subsequent bioinformatics identified a putative RDF within each PAI<sub>536</sub> (Napolitano, Almagro-Moreno et al. 2010).

PAI excision and the role of the island encoded integrase and RDF were examined in the High Pathogenicity Island (HPI) of *Yersinia pseudotuberculosis*. Excision of HPI requires the cognate island integrase and an RDF, named Hef. Hef was proposed to not act as a transcriptional activator of the integrase and was thought to play more of a structural role, although they did not demonstrate this (Lesic, Bach et al. 2004). Additionally, the she PAI of *Shigella flexneri* can excise from the chromosome (Rajakumar, Sasakawa et al. 1997; Al-Hasani, Rajakumar et al. 2001). It was found that Rox, the putative RDF, was required to do so, however its mechanism was not shown (Sakellaris, Luck et al. 2004).

In *V. cholerae*, PAI excision has been examined in VPI-1, VPI-2 and VSP-II (Rajanna, Wang et al. 2003; Murphy and Boyd 2008; Boyd, Almagro-Moreno et al. 2009) (**Figure 4**). Only one study has examined VPI-1 excision, and showed that the cognate TR integrase, *intVI*, and a transposases, *vpiT*, both found within the island, contributed to excision, and that one could compensate for the absence of the other (Rajanna, Wang et al. 2003). In the absence of both *intVI* and *vpiT*, excision of VPI-1 was abolished. This group also found that addition of mitomycin C did not result in increased excision of VPI-1 indicating the excision of this island is not SOS induced as is the case for many double stranded DNA prophages that are under the regulation

of LexA (Rajeev, Malanowska et al. 2009). The Rajanna study showed that VPI-1 is capable of excision, an important first step in the HGT of these elements, but it did not address the mechanism behind this excision and events thereafter, such as the genes involved in the excision event. No RDF is present on VPI-1 and this is believed to be an important requirement for TR integrase excision function. It is also puzzling that the transposase can compensate for the integrase in site specific recombination at the *att* sites of this island to aid in excision (Rajanna, Wang et al. 2003).

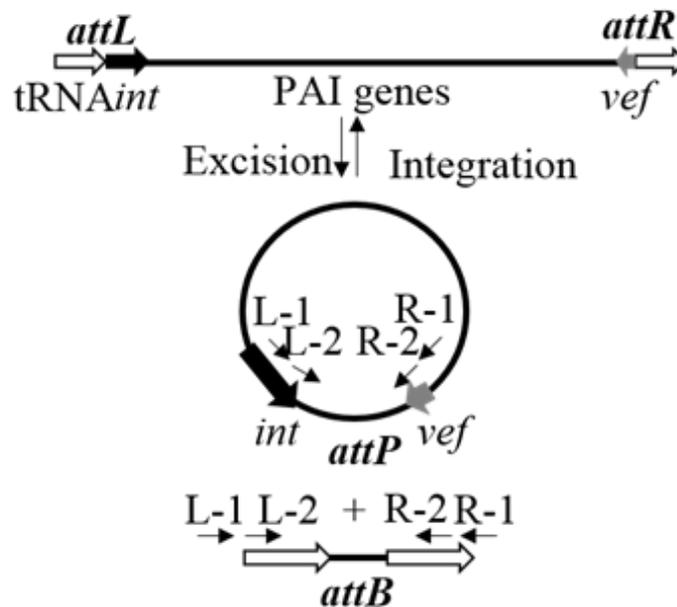
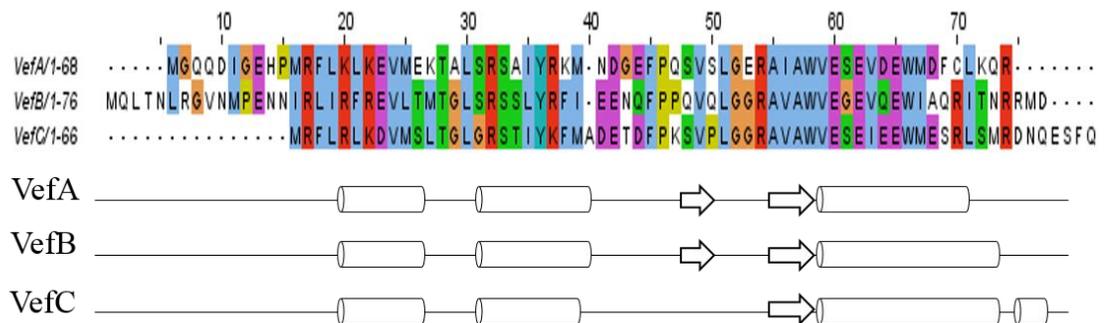


Figure 4 **Vibrio pathogenicity island excision and integration.** VPIs can excise from the chromosome and form a circular intermediate containing an *attP* site and left in the bacterial chromosome is a unique *attB* site. Primers used to detect the attachment (*att*) sites in two-stage nested PCR assays are represented by arrows on the left and right sides of the *attP* and *attB* sites.

Previous work in our laboratory examined the excision properties of VPI-2. This study revealed that (1) VPI-2 excision is not growth phase dependent, (2)

excision rates increased when exposed to low temperatures and sub-lethal UV-light irradiation and (3) excision decreased when cells were grown in minimal media supplemented with glucose. They additionally showed that transcript levels of *intV2* and *vefA* increased 10 fold and *vefB* increased 5 fold following exposure to UV light, indicating excision may be driven by the same mechanisms of that of the SOS pathway (Boyd, Almagro-Moreno et al. 2009). A total of three RDFs (VefA, VefB and VefC) are present in the genome of *V. cholerae* N16961 and each has a predicted protein structure that includes a helix-turn-helix, indicating DNA binding properties. Although they share this similar structure, their protein sequences are quite diverse with sequence identity of ~50% (**Figure 5**). The mechanisms of action of these three RDFs are unknown as is their requirement for PAI excision events.



**Figure 5 Vibrio excision factors (vefs).** Amino acid sequence alignment of VefA, VefB and VefC using ClustalW with default settings in JalView. Stars (\*) indicate conserved amino acids between all three Vefs. A 59% identity is shared between VefA and VefC, 46% between VefA and VefB and 49% between VefB and VefC. Predicted secondary structures of VefA, VefB and VefC were generated using JNet Secondary Structure Prediction (<http://www.jalview.org/help/html/webServices/jnet.html>). Arrows represent sheets and cylinders represent helices.

## **PAI Mobility**

While it is well understood that PAIs have the capability to integrate into and excise from bacterial chromosomes, little is known about how these elements mobilize between cells. Since PAIs do not contain their own mobility genes it is proposed that PAI can be transfer by transduction or taken up by transformation. It is possible that a conjugative plasmid might mobilize some PAIs. While many studies exist that claim to show mechanisms for PAI or genomic island transfer, the vast majority of these are not PAI regions but are instead defective phages or ICEs. Only one study has shed light on the mechanism of *Vibrio* PAI transfer and this study showed transduction was a viable mechanism *in vitro* (O'Shea and Boyd 2002). In this study, a generalized transducing phage was demonstrated to be capable of transferring VPI-1 from an O1 toxigenic strain to an O1 non toxigenic strain. The study demonstrated that the mobilized VPI-1 region integrated at the same tmRNA locus as in the toxigenic isolate (O'Shea and Boyd 2002). Another likely mechanism of transmission is transformation, particularly for *Vibrio* PAIs, given that the *V. cholerae* and other *Vibrio* species transformation machinery is induced in the presence of chitin, the second most common polymer found in the aquatic environment (exoskeleton of crabs, shrimps and molluscs) (Meibom, Blokesch et al. 2005; Sun, Bernardy et al. 2013).

## **Dissertation Work**

The research accomplished and presented in this dissertation set out to improve the overall understanding of mobile genetic elements, particularly the molecular mechanism and host control mechanisms of genetic elements excision from the bacterial chromosome.

Pathogenicity islands contribute significantly to the rapid evolution of a large number of bacterial pathogens. PAIs are also one of the least studied MIGE. It is important to assess the mechanisms of PAI transfer, which is a difficult question to address since PAIs do not contain genes for autonomous replication or transfer, however, excision from the bacterial chromosome is the first step in transfer of the island to a new host. The purpose of chapter two was to determine the mechanism of PAI excision. Two PAIs from *V. cholerae* N16961, VPI-1 and VPI-2, were chosen for examination not only because of the significant cargo genes they carry, as discussed in this introduction, but because of their unique recombination modules. Both contain cognate integrases however VPI-2 contains two RDFs while VPI-1 contains none. One of the first questions we sought to answer was how VPI-1 can excise from the chromosome without possessing an RDF, as all previously studied MIGEs that contain a TR integrase required an RDF for its excision function. We hypothesized that VPI-1 utilizes another RDF found elsewhere on the *V. cholerae* genome. While we found this to be correct, we unexpectedly found that excision of VPI-1 still occurs in the absence of all RDFs in the *V. cholerae* genome, while excision of VPI-2 was abolished. In an attempt to understand this observed phenotype we measured the levels of each island's integrase through qPCR, in the RDF-negative background. These findings revealed that there is nearly twice the amount of *intV1* than *intV2* in the RDF negative strains. Thus, we hypothesized that the higher levels of *intV1* overcome the necessity of an RDF for VPI-1 excision. To test this hypothesis we overexpressed *intV2* in the RDF-negative background and saw that the excision of VPI-2 was indeed restored. Thus, if enough integrase is provided, RDFs are no longer essential for excision.

In chapter 2 we also show that over-expression of each RDF resulted in super-excision of both VPI-1 and VPI-2, indicating that all RDFs can promote the excision of each island, regardless of not being located on the respective island. Not only did this data help answer how VPI-1 can excise despite not containing RDF, but it also revealed the extensive cross-talk that exists between these separately acquired non homologous elements. This was the first study to report PAI cross-talk mediated via RDFs. Furthermore, we went on to examine the function of these RDFs in excision. The above mentioned qPCR revealed that the RDFs were negative transcriptional regulators of the integrases. This may seem counterintuitive given that RDFs promote excision that is mediated by the integrase. However, upon examination of the recombination module, we speculated that since the promoter region of the integrases overlap with the *att* sites where RDFs bind to bend DNA and aid in excision, they were also inhibiting integrase transcription. To test this hypothesis, I purified RDF proteins VefA and VefB and demonstrated that they can bind the *att* sites of VPI-1 and VPI-2, is consistent with the proposed mechanism of repression. Shared sequence homology near the *att* sites of these two islands supports the binding and negative regulation of the integrase and also indicates a shared evolutionary history. Indeed, several PAIs which contained varying combinations of integrase (IntV1 and IntV2) homologs and RDF (VefA, VefB and VefC) homologs were found in the NCBI database, supporting the cross-talk seen in this study.

In Chapter 3 we explored the evolutionary history of these additional islands in more detail and showed the excision capability of PAIs from various *V. cholerae* strains with identical VPI-1 and VPI-2 recombination modules but with diverse cargo genes including those encoding for type III secretion systems (T3SS), type VI

secretion systems (T6SS) and CRISPR-Cas systems. We examined the strain *V. cholerae* NRT36S which contains a PAI, which we named VPI-3 that has an identical recombination module as that of VPI-2 but contains a T3SS in place of the restriction modification system. We showed that VPI-3 can excise from the chromosome and like VPI-2, its cognate integrase is necessary to do so. Several other *V. cholerae* strains which contained VPI-3 were identified. We performed phylogenetic analysis of various genomic regions within VPI-3 and found that VPI-3 has a modular structure whereby different functional regions were acquired by an identical recombination module. Additionally, an island we named VPI-6 was identified in *V. cholerae* RC385. This island contained an IntV1 integrase as well as VefA and VefB, a hybrid VPI-1/VPI-2 recombination module, as well as genes encoding for a T6SS and a CRISPR-Cas system. Phylogenetic analysis of this island also suggested a modular evolutionary history. Overall, work from this chapter suggests that similar recombination modules can associate with different cargo genes that can convert a strain into a pathogen with very different disease pathologies. Additionally, similar to the evolution of bacteriophage, PAIs have a modular structure where different functional regions are acquired by identical recombination modules.

In Chapter 4, I examined the role of quorum sensing, a form of bacterial communication, in phage biology. Specifically we examined the role of the QS regulator OpaR in regulation of the replication and excision of a filamentous phage f237 that is homologous to the filamentous phage CTX that contain the cholera toxin genes. In this work we show that luxO is a negative regulator of f237 and in a luxO mutant f237 genes are more highly expressed and phage excision is increased.

## Chapter 2

### **PATHOGENICITY ISLAND CROSS-TALK MEDIATED BY RECOMBINATION DIRECTIONALITY FACTORS FACILITATES EXCISION FROM THE CHROMOSOME**

The work in this chapter was published in the *Journal of Bacteriology*

#### **Pathogenicity Island Cross Talk Mediated by Recombination Directionality Factors Facilitates Excision from the Chromosome.**

Carpenter MR, Rozovsky S, Boyd EF.

J Bacteriol. 2015 Dec 14;198(5):766-76. doi: 10.1128/JB.00704-15.

#### **Introduction**

Mobile integrative genetic elements (MIGEs) are the vectors of horizontal gene transfer (HGT) among bacteria via the mechanisms of transformation, transduction or conjugation (Frost, Leplae et al. 2005; Boyd, Almagro-Moreno et al. 2009; Boyd 2012). Members of MIGEs include transposons, integrons, bacteriophages, plasmids, integrative conjugative elements (ICEs)/conjugative transposons and genomic/pathogenicity islands (GEIs/PAIs). Commensal bacteria can be converted into deadly pathogens through HGT of MIGEs that contain virulence factors, or a pathogen may become more virulent with the acquisition of additional virulence factors (Almagro-Moreno, Murphy et al. 2011; Boyd 2012; Boyd 2012;

Boyd, Carpenter et al. 2012). PAIs were first described in Uropathogenic *Escherichia coli* (UPEC) and have since been described in many pathogenic Proteobacteria (Carniel, Guilvout et al. 1996; Karaolis, Johnson et al. 1998; Hacker and Kaper 2000; Al-Hasani, Rajakumar et al. 2001; Jermyn and Boyd 2002; Hensel 2004; O'Shea, Finnan et al. 2004; Jermyn and Boyd 2005; Hurley, Quirke et al. 2006; Quirke, Reen et al. 2006; Murphy and Boyd 2008). PAIs range in size from 10-200 kilobases (kb) and have a guanine/cytosine (GC) content that differs from that of the core bacterial chromosome. These islands encode virulence factors, are found integrated in the chromosome at the 3' end of tRNA loci and are absent from non-pathogenic strains or closely related species (Hacker and Kaper 2000). PAIs also contain an integrase usually belonging to the tyrosine recombinase (TR) family. Based on phylogenetic analysis of PAI encoded integrases, it was demonstrated that PAIs are a distinct class of integrative elements and are not degenerate remnants of integrated plasmids, prophages, integrons or conjugative transposons (Boyd, Almagro-Moreno et al. 2009; Napolitano, Almagro-Moreno et al. 2010; Napolitano and Boyd 2012).

Integrases are necessary for both integration and excision of MIGEs by catalyzing site-specific recombination between two DNA substrates through DNA cleavage, formation of a Holiday junction and rejoining of DNA fragments (Ramsay, Sullivan et al. 2006; Lee, Auchtung et al. 2007; Flanigan and Gardner 2008; Fogg, Rigden et al. 2011). In TR integrase-encoding prophages it was demonstrated that the excision event also requires an additional protein, an excisionase or recombination directionality factor (RDF) (Sam, Papagiannis et al. 2002; Warren, Sam et al. 2003; Panis, Mejean et al. 2007; Coddeville and Ritzenthaler 2010; Singh, Plaks et al. 2013). In fact, it is believed that all TR integrases require an RDF to perform excision (Lewis

and Hatfull 2001; Fogg, Rigden et al. 2011). RDFs are small DNA-binding accessory proteins that aid TR integrases in controlling the directionality of their site-specific recombination reactions (Lewis and Hatfull 2001). These proteins govern directionality by favoring one reaction direction and interfering with the other. While the integrases act to catalyze the site-specific recombination reaction necessary for MIGE excision and integration, their cognate RDFs have been described to perform various accessory functions such as binding and bending DNA near the *att* sites, acting as positive or negative integrase transcriptional regulators as well as offering stability to their integrase protein partners at the excision site (Numrych, Gumport et al. 1992; Sam, Papagiannis et al. 2002; Swalla, Cho et al. 2003; Abbani, Papagiannis et al. 2007; Coddeville and Ritzenthaler 2010; Panis, Duverger et al. 2010). Tyrosine recombinase integrases and RDFs have not been studied extensively in PAIs. Phylogenetic analysis of *E. coli* RDFs and their corresponding integrases from PAIs suggests that they co-evolved together (Napolitano, Almagro-Moreno et al. 2010). This co-evolution was also observed in the grouping patterns of bacteriophage and transposon integrases with their partner RDFs (Lewis and Hatfull 2001).

*Vibrio cholerae* O1 serogroup El Tor, the cause of the seventh ongoing cholera pandemic, contains four PAIs that are of particular interest due to their genetic content and unique recombination modules. *Vibrio* pathogenicity island-1 (VPI-1) and VPI-2 are present in all cholera pandemic strains, both classical and El Tor biotypes (Karaolis, Johnson et al. 1998; Dziejman, Balon et al. 2002; Jermyn and Boyd 2002; O'Shea, Finnan et al. 2004; Jermyn and Boyd 2005). The presence of two additional PAIs, named the *Vibrio* seventh pandemic island (VSP)-I and VSP-II, distinguishes biotype El Tor from the classical biotype strains that caused the first six pandemics

(Karaolis, Johnson et al. 1998; Dziejman, Balon et al. 2002; Jermyn and Boyd 2002; O'Shea, Finnan et al. 2004). The genes encoding the toxin co-regulated pilus (TCP), a type IV bundle forming pilus, critical for intestinal colonization, are found within VPI-1 (Taylor, Miller et al. 1987; Herrington, Hall et al. 1988; Manning 1997; Karaolis, Johnson et al. 1998). *Vibrio* pathogenicity island-2 contains genes for the production of sialidase as well as a sialic acid transport and catabolism gene cluster (Jermyn and Boyd 2002; Almagro-Moreno and Boyd 2009; Almagro-Moreno and Boyd 2009; Chowdhury, Norris et al. 2012). Both VPI-1 and VPI-2 have a TR integrase (*intV*) located adjacent to a tRNA locus, are flanked by attachment (*att*) sites and have a low GC content compared to the *V. cholerae* core genome (Jermyn and Boyd 2002; Murphy and Boyd 2008). It was proposed that VPI-1 was actually a filamentous phage named VPI $\Phi$ , however, this has proven not to be the case and the region is indeed a PAI (Faruque, Zhu et al. 2003; Rajanna, Wang et al. 2003). Two RDF genes, *vefA* and *vefB*, were identified by bioinformatic analysis in VPI-2 and a third one *vefC* is found in VSP-II, however no RDF was identified in VPI-1 (Almagro-Moreno, Napolitano et al. 2010) (**Figure 2 and 3**).

PAI dynamics have been examined in the UPEC strain 536, which contains six PAIs (PAI I<sub>536</sub>–PAI VI<sub>536</sub>). Five of the PAIs were shown to excise from the chromosome with the exception of PAI IV<sub>536</sub> (Hochhut, Wilde et al. 2006; Dobrindt, Chowdary et al. 2010). RDFs were not identified or examined in these studies. A more recent study however found that each PAI<sub>536</sub> contained its own cognate putative RDF by bioinformatics (Napolitano, Almagro-Moreno et al. 2010). In PAIs of *Yersinia pseudotuberculosis* and *Shigella flexneri* excision from the bacterial chromosome was demonstrated and shown to require the cognate island integrase and an RDF, named

Hef and Rox respectively (Al-Hasani, Rajakumar et al. 2001; Lesic, Bach et al. 2004). In *V. cholerae*, PAI excision was examined in VPI-1, VPI-2 and VSP-II (Rajanna, Wang et al. 2003; Murphy and Boyd 2008; Almagro-Moreno, Napolitano et al. 2010). For VPI-1 it was shown that the cognate TR integrase and the transposase, VpiT, both contributed to excision, since only in the absence of both *intVI* and *vpiT*, was excision abolished (Rajanna, Wang et al. 2003). For VPI-2 the integrase and an RDF were shown to be required for excision (Murphy and Boyd 2008; Almagro-Moreno, Napolitano et al. 2010). For VSP-II the cognate integrase was required for excision and formation of a non-replicative circular intermediate (CI) (Murphy and Boyd 2008).

In this study, we examined the role of RDFs and their relationship with integrases in the excision of both VPI-1 and VPI-2 to address the question of whether RDFs are an absolute requirement for excision since the VPI-1 region does not contain an RDF. We have identified extensive RDF-mediated cross-talk between VPIs and demonstrated dual functionality of these proteins to aid in PAI excision and act as transcriptional regulators.

## Materials and Methods

### Bacterial strains, plasmids and growth conditions.

All bacterial strains and plasmids used in this study are listed in **Table 1**. Bacterial strains were grown overnight aerobically at 225 rpm in Luria broth (LB) (Fisher Scientific, Fair Lawn, NJ) at 37°C unless otherwise stated. The diaminopimelic acid (DAP) auxotroph, *Escherichia coli*  $\beta$ 2155, was grown in the presence of 0.3 mM DAP (Sigma, St. Louis, MO). When appropriate, antibiotics were

added to media in the following concentrations: 200 µg/ml streptomycin (Sm), 25 µg/ml chloramphenicol (Cm) and 100 µg/ml ampicillin (Amp) (Fisher Scientific). When selecting for the double-cross, LB agar was supplemented with 10% sucrose.

#### Mutant strain construction.

Primers were designed and purchased from Integrated DNA Technologies (Coralville, IA) in order to create in-frame deletions of genes in *V. cholerae* N16961 using splicing by overlap extension (SOE) PCR and homologous recombination (Horton, Hunt et al. 1989). The mutant strains are listed in **Table 1** and primers used for their construction are found in **Table 2**. In order to make  $\Delta intVI$ , two sets of primers were designed to amplify a region flanking the integrase gene, VC0847. Primer pairs VC0847A/VC0847B and VC0847C/VC0847D were used to generate PCR fragments 556 bp and 548 bp in size, respectively. The AB and CD purified PCR products were used as template for PCR with primers VC0847A and VC0847D to form a PCR product containing a truncated *intVI* gene, 351 bp in size. The truncated *intVI* AD product was blunt-end cloned into the vector pJET1.2 (Thermo Fisher Scientific, Pittsburgh, PA) and transformed into *E. coli* DH5 $\alpha$ . Following plasmid isolation (Qiagen, Valencia, CA) and restriction digestion with SacI and XbaI (Thermo Fisher Scientific, Pittsburgh, PA) the fragment was cloned into the suicide vector pDS132. pDS $\Delta intVI$  containing the truncated AD insert was then transformed into the DAP auxotroph strain, *E. coli*  $\beta$ 2155 which served as the donor for conjugation with *V. cholerae* N16961. Selective plating without DAP, but containing Cm, allowed for detection of single cross-over events via homologous recombination of pDS $\Delta intVI$  into the chromosome of *V. cholerae* and were confirmed by AD primer pair PCR. Next the double cross-over event was selected for by plating on LB agar

containing 10% sucrose without Cm. Colony PCR using AD and flanking primers confirmed the double cross-over event. All mutant strains were created using the same method and confirmed by sequencing. Primers used for each mutant construction are listed in **Table 2**. The strain  $\Delta intV1$  (VC0847) has a 918 bp deletion resulting in a 351 bp truncated gene;  $\Delta vpiT$  (VC0817) has an 816 bp deletion resulting in a 168 bp truncated gene;  $\Delta intV2$  (VC1758) has 1,065 bp deleted and a truncated *intV2* gene 171 bp in size. The strain  $\Delta vefC$  (VC0497) has 171 bp deleted and a 30 bp truncated *vefC* gene.

#### Excision assays.

In order to observe excision of VPI-1 and VPI-2 from the *V. cholerae* genome, two excision assays were performed to detect unique sites in the chromosome (*attB1* and *attB2*) and in the circular intermediate (CI) (*attP1* and *attP2*) formed following the excision event (**Figure 4**). Genomic DNA was isolated from overnight liquid incubations using the GNOME Kit (MP Biomedicals, Santa Ana, CA) following manufacturer's procedure and used as template in the *attB* assays to detect the empty chromosomal site following island excision. Additionally, plasmid DNA was isolated using Qiagen Plasmid Isolation kit per manufacturer's instructions and used as template for the *attP* assay in order to observe the island containing *attP* site following site specific recombination from the *V. cholerae* chromosome. Due to the low rate of island excision under normal laboratory growth conditions, a two stage nested PCR was performed to observe excision products in the form of *attB* and *attP*. Primers were designed based on the *V. cholerae* N16961 genome in order to amplify the island containing *attP* site or chromosomal *attB* site for VPI-1 and VPI-2 (**Figure 4**) and are listed in **Table 2**. PCR was performed in 20  $\mu$ l reactions with Taq DNA Polymerase

(Denville, Holliston, MA) under the following conditions: 95°C for 5 min followed by 30 cycles of 94°C 30 s, 60°C 30 s or 90 s, 72°C 30 s and 5 min at 72°C. Genomic or plasmid DNA (10 ng) was used as template for the first round of PCR and 1 µl PCR product from this reaction was used as template for the second nested round of PCR. PCR products were analyzed on a 2% agarose gel and stained with ethidium bromide for visualization under UV light. At least two biological replicas and three technical replicas were performed for each assay.

#### Mutant strain complementation.

Mutant strains which were defective for excision or which showed reduced excision were complemented with expression plasmids containing full intact genes. The integrase, *intVI*, and its ribosomal binding site was amplified using primers VC0847F and VC0847R with HotStar HiFidelity Polymerase (Qiagen) per manufacturer's instructions. The PCR product was cloned into pJET1.2 and transformed into *E. coli* DH5α. Following plasmid isolation and restriction digestion with XbaI and SacI the fragment was cloned into the arabinose inducible plasmid pBAD33. The resulting plasmid, pBA*intVI*, was transformed into *E. coli* β2155 and used as the donor strain in conjugation with Δ*intVI*. The resulting strain Δ*intVI* *pintVI* was used to observe excision. Each complementation strain was created in this manner using primers listed in **Table 2** except pBAVefA which was cloned from pBBR1MCSVefA into pBAD33 using XbaI and KpnI restriction enzymes. All complementation strains were induced with 0.02 % (w/v) arabinose unless otherwise stated.

### Quantitative real-time PCR.

RNA was extracted from wild-type and  $\Delta vefA\Delta vefB\Delta vefC$  *V. cholerae* (N16961) cells grown to mid-logarithmic stage ( $OD_{600} = 0.5 - 0.6$ ) using Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. RNA was subsequently treated with Turbo DNase (Invitrogen) following manufacturer's instructions and quantified with a Nanodrop spectrophotometer. cDNA was synthesized starting with 500 ng of DNase treated RNA as template using Superscript III reverse transcriptase (Invitrogen). A 1:10 dilution of each sample was used as template for quantitative real-time PCR (qPCR). PCR reactions were performed using Fast SYBR Green Master Mix (Applied Biosystems, Carlsbad, CA) and run on an Applied Biosystems 7500 fast real-time PCR system. Cycling conditions were: holding stage at 95°C for 25 s and cycling stage at 95°C for 3 s and 60°C for 30 s. Primers for *intV1*, *intV2* and *topI* are listed in **Table 2**. Data was analyzed using the Applied Biosystems 7500 software. Relative gene expression was determined using the  $\Delta\Delta C_t$  method (Livak and Schmittgen 2001).

### Protein purification of VefA and VefB.

*Vibrio cholerae* VefA and VefB were separately cloned into the pMALc5x-His6x expression plasmid using primers BamHIVC1785R/NcoIVC1785F and NcoIVefBFwd/BamHIVefBRev, respectively (**Table 2**), in which a His6x tagged Maltose Binding Protein (MBP) is fused to VefA/VefB, separated by a Tobacco Etch Virus (TEV) protease cleavage site. The Q5 Site Directed Mutagenesis Kit (New England BioLabs, Ipswich, MA) was used to remove a 4 bp sequence in the pMALVefA construct allowing in-frame translation of the fusion protein, following manufacturer's protocol. Primers VefAmutRev/VefAmutFwd used in this reaction are

listed in **Table 2**. Plasmid DNA was confirmed by sequencing and transformed into *E. coli* BL21(DE3) using standard CaCl<sub>2</sub> method.

The pMALc5x-His6x-VefA and pMALc5x-His6x-VefB were expressed in *E. coli* BL21(DE3). Ten milliliters of overnight culture was inoculated into 1 L Terrific Broth (TB) supplemented with glucose at 37°C and induced with 1 mM isopropyl-1-thio-β-d-galactopyranoside (IPTG) at OD<sub>600</sub> 0.5. Growth continued overnight at 18°C. Cells were harvested by centrifugation (5,000 xg for 20 min at 4°C) and were resuspended in Amylose Wash Buffer (50 mM sodium phosphate, 200 mM NaCl, pH 7.5) supplemented with 1 mM phenylmethanesulfonyl fluoride (PMSF) and 1 mM benzimidazole. Bacterial cells were lysed on ice using a high-pressure homogenizer (EmulsiFlex-C5, Avestin, Ottawa, Canada). Cell debris was removed by centrifugation (15,000 xg for 1 h at 4°C). The supernatant was passed through a column containing 30 mL amylose resin (New England BioLabs). The column was washed with 10 column volumes (CV) of Amylose Wash Buffer. The fusion protein, either MBP-VefA or MBP-VefB was eluted with 3 CV Amylose Elution Buffer (50 mM sodium phosphate, 200 mM NaCl, 30 mM maltose, pH 7.5).

A hexahistidine-tagged TEV protease was added to the eluent in a 1:10 molar ratio (TEV:MBP-VefX) and the cleavage reaction proceeded for 45 min for MBP-VefA and overnight for MBP-VefB, at 4°C. Further TEV cleavage reaction time resulted in significant precipitation and loss of VefA. The cleavage mixture was centrifuged, adjusted to 30 mM imidazole and subject to immobilized metal affinity chromatography (IMAC) using nickel resin to remove the His-tagged TEV and MBP as well as any remaining un-cleaved fusion protein. The flow through and first two CV of wash with IMAC Wash Buffer (50 mM sodium phosphate, 200 mM NaCl, 30 mM

imidazole, pH 7.5) contained either VefA or VefB. VefB IMAC flow through and wash contained equal parts of MBP and was subject to further purification. Following a 2-fold dilution, VefB was captured on a heparin column (GE Healthcare Life Sciences, Pittsburgh, PA) using 20 mM Tris, 100 mM NaCl, 1mM EDTA, 1 mM DTT, pH 8.0. A linear gradient up to 1M NaCl was performed to elute VefB from the column. Fractions that contained VefB (56 – 60% NaCl) were pooled. Successive concentrations and dilutions using 3.5 kDa cut-off concentrators (EMD Millipore, Billerica, MA) with 20 mM Tris pH 8.0 were performed to reduce NaCl concentration to 115 mM. Protein molecular weight was confirmed by mass spectrometry and its purity was determined to be higher than 95% by a 16% SDS-PAGE Tris-Tricine gel.

#### Electrophoretic Mobility Shift Assays.

DNA fragments *attR1* and *attL2* were amplified using primers sets attR1Fwd/attR1Rev and attL2Fwd/attL2Rev, respectively (**Table 2**). Varying concentrations of purified VefA or VefB were incubated with 30 ng of target DNA in binding buffer (10 mM Tris, 150 mM KCl, 0.1 mM dithiothreitol, 0.1 mM EDTA, 5% PEG, pH 7.4) for 20 min at room temperature and 10  $\mu$ L were loaded onto a pre-run (200 V for 2 h at 4°C) 8% native acrylamide gel. The gel was run at 200 V for 2.5 h in 1x Tris-acetate-EDTA (TAE) buffer at 4°C. Following electrophoresis, gels were stained in an ethidium bromide bath (0.5  $\mu$ g/ml) for 20 min, washed with water and imaged.

## Results

All three RDFs VefA, VefB and VefC can induce VPI-2 excision.

VPI-2 contains a TR integrase and two RDFs (**Figure 2**). A two stage nested PCR approach was used to confirm excision of VPI-2 in several *V. cholerae* wild-type strains that contained these genes through detection of both the circular intermediate *attP2* and the empty *attB2* site left in the chromosome following excision (**Figure 4**). Excision was detected through both the *attP2* and *attB2* PCR assays in the second round of PCR for *V. cholerae* N16961, O395 and MO2 strains (**Figure 6A and 6B**). In strain SG-7, which is devoid of VPI-2, *attB2* is present in every copy of its genome, and thus the empty site was detected in the first and second round of PCR (**Figure 6B**).

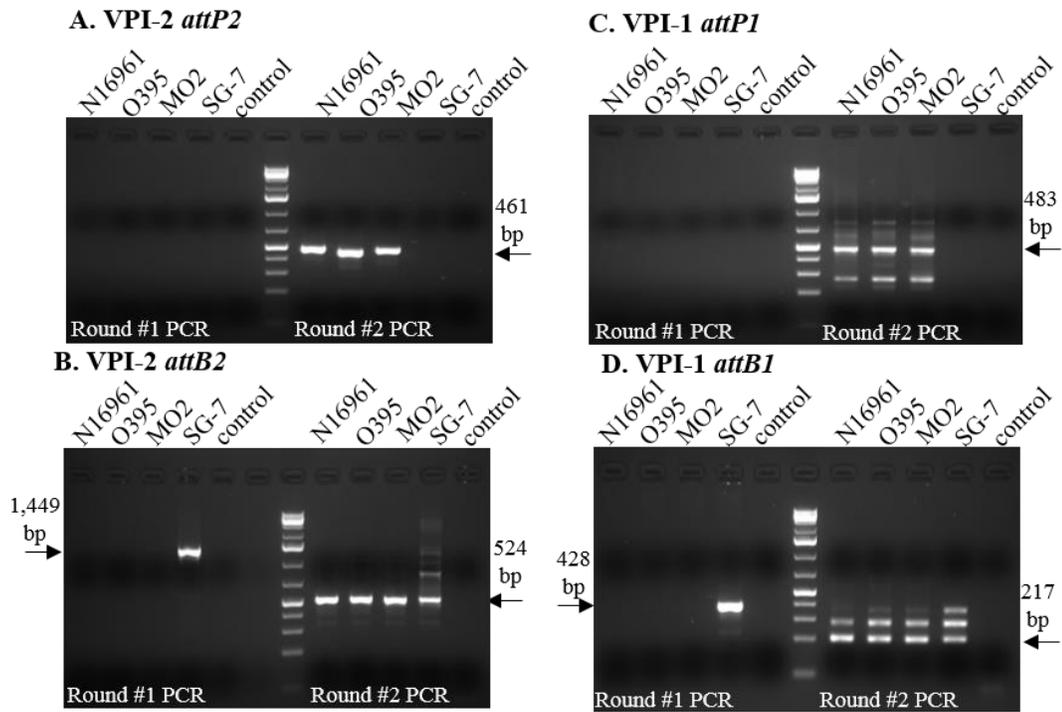


Figure 6 ***Vibrio cholerae* wild-type excision.** N16961 (serogroup O1 biotype El Tor), O395 (O1 classical) and MO2 (O139). *Vibrio cholerae* SG-7 (O56) does not contain VPI-1 or VPI-2 and served as a negative control.

Two RDFs are present on VPI-2 (*vefA* and *vefB*) and a third RDF is present in VSP-II (*vefC*) (Almagro-Moreno, Napolitano et al. 2010). These Vef protein sequences are diverse, containing a shared sequence identity of <50%, however their predicted secondary structure is similar (**Figure 5**). Each RDF contains three  $\alpha$ -helices with the second and third separated by a  $\beta$ -sheet (**Figure 5**). This conserved helix-turn-helix motif suggests that the Vefs are capable of binding DNA (Aravind, Anantharaman et al. 2005). The RDFs share 21 conserved amino acids, with the majority of them near the winged helix of the C-terminus (**Figure 5**).

To determine the role of all three RDFs in excision of VPI-2, the two-stage nested PCR assay was used to detect the empty *attB2* site in single, double and triple RDF mutant strains (**Figure 7**). We constructed in-frame non polar deletions of VC0497 (*vefC*), VC1785 (*vefA*) and VC1809 (*vefB*) in *V. cholerae* N16961 and examined excision using the *attB* excision assay. A wild-type excision product (524 bp) for VPI-2 was observed in the second round of the *attB2* PCR assay for each single RDF mutant and a double mutant strain lacking *vefA* and *vefB* (**Figure 7**). Only the triple RDF mutant ( $\Delta vefA\Delta vefB\Delta vefC$ ) gave no excision product which demonstrates that at least one RDF is required for excision of VPI-2 (**Figure 7**). The triple RDF mutant was complemented with a functional copy of each RDF individually. Each RDF restored the VPI-2 excision phenotype, which was detected in the first round of PCR, evidenced by a 1,449 bp product, indicative of excision above the wild-type level that we termed super excision (**Figure 7**). These data demonstrate that an RDF is required for VPI-2 excision and that all three RDFs can induce excision of VPI-2. The multiple bands seen in the second round of *attB1* PCR represents mispriming from round one. To confirm this, the round one PCR products were treated with an enzyme cocktail EXOSAP-IT (USB Corporation, Cleveland, OH) to degrade unwanted dNTPs and primers. Using the EXOSAP-IT treated product as template for round two of the *attB* assay resulted in a single product of the appropriate size (data not shown).

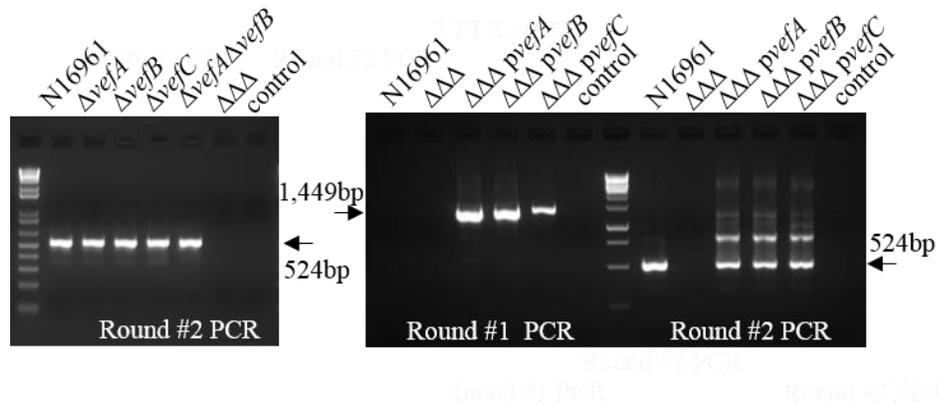


Figure 7 **RDFs promote VPI-2 excision.**  
 A. Excision of VPI-2 was examined using the *attB2* excision assay in the in single, double and triple ( $\Delta\Delta\Delta$ ) RDF mutant backgrounds. Only round #2 is shown as no excision products were detected in round #1. B. The triple RDF mutant was complemented with each RDF, *vefA*, *vefB* and *vefC*, individually (*pvefA*, *pvefB* or *pvefC*) and examined for VPI-2 excision. Excision of the complements was detected in round #1 and referred to as a super-excision phenotype. Control indicates sample with no template DNA.

*intV2* can mediate VPI-1 excision in the absence of *intV1*.

VPI-1, which contains a cognate integrase *IntV1* but lacks any RDFs, was examined for excision similarly to VPI-2, using *attP1* and *attB1* two-stage PCR assays (**Figure 8C and 8D**). In the *attP1* PCR assay, no product was detected in the first round of PCR indicating a low excision rate as previously suggested (Rajanna, Wang et al. 2003), however a 483 bp excision product was detected in the second round of PCR (**Figure 8C**). Similarly the *attB1* assay gave an *attB1* PCR product in the second round of PCR for all strains examined (**Figure 8D**), indicating VPI-1 is capable of excision from these strains.

Next we reexamined the role of *intV1* (VC0847) in excision by constructing an in-frame deletion in VC0847 in *V. cholerae* N16961. The excision of VPI-1 was nearly abolished in the *intV1* mutant evidenced by a faint 217 bp PCR band present in the second round of PCR (**Figure 8A**). This suggests that there is low level excision occurring. Next we constructed a *vpiT* (VC0817) deletion mutant similar to *intV1*. In the *vpiT* single mutant strain VPI-1 excision occurred at a similar level to that seen for wild-type (**Figure 8A**). However, no excision was observed in the  $\Delta intV1\Delta vpiT$  using our *attB1* assay, consistent with previous findings (**Figure 8A**) (Rajanna, Wang et al. 2003). Interestingly, in the double integrase mutant,  $\Delta intV1\Delta intV2$  excision was abolished, which suggest IntV2 from VPI-2 plays a role in VPI-1 excision (**Figure 8A**). To further investigate this result, we complemented the  $\Delta intV1\Delta intV2$  double mutant with a functional *intV2* gene and found that VPI-1 excision was restored in the *attB1* excision assay (**Figure 8A**). These data indicate cross-talk between the VPIs mediated by integrases. Analysis of IntV1 and IntV2 amino acid sequences revealed significant homology (53% amino acid identity) displaying a strongly conserved C-terminal domain, typical of TR integrases, which houses the catalytic domain. Both IntV1 and IntV2 contain the essential tyrosine (Y) nucleophile surrounded by the highly conserved “RHR triad” (in tertiary structure)(Grainge and Jayaram 1999), revealing highly similar active sites (**Figure 9**). Three conserved domains were identified in both integrases: the N-terminus contained a DUF4102 (pfam13356) domain as well as a SAM-like domain (pfam14659) common to a wide variety of phage integrases, and the C-terminal domain contained a P4 catalytic domain (pfam00589) responsible for the breaking and rejoining of DNA fragments during recombination (**Figure 9**). The data suggest that both integrases can catalyze excision

of VPI-1, which suggests either the integrases have a more promiscuous *att* site or there is a common *att* site that both integrases act on. RDFs are not essential for VPI-1 excision but can promote excision.

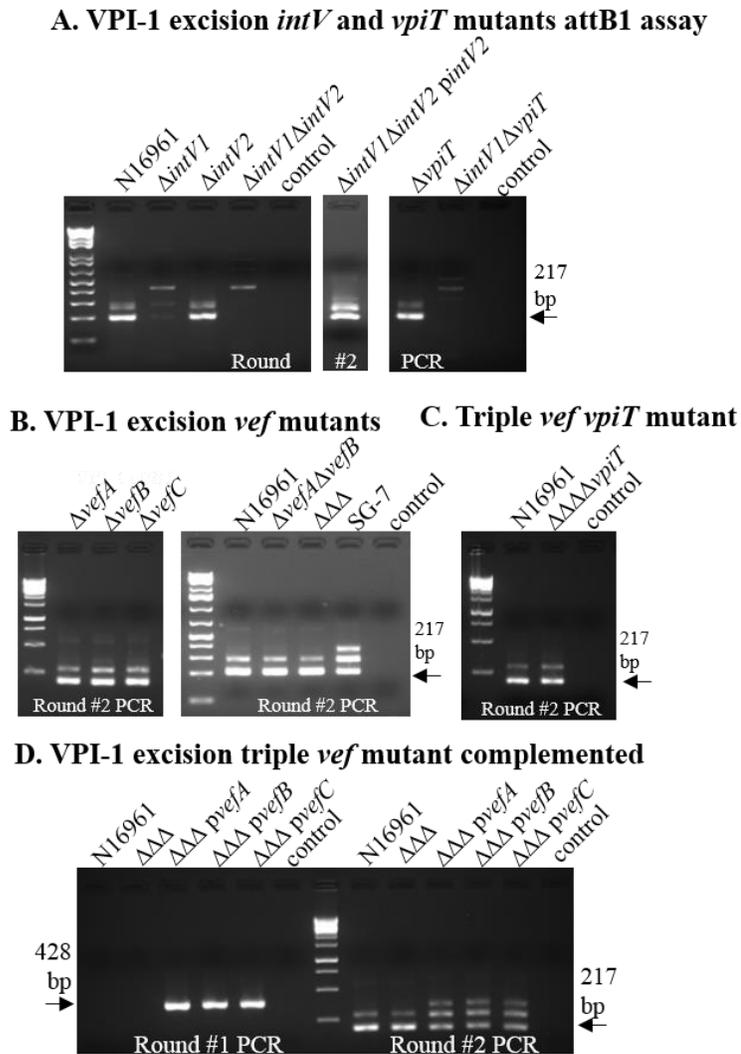


Figure 8

**Role of integrases and RDFs in excision of VPI-1.**

A. VPI-1 excision in WT (N16961), *intV* and *vpiT* single and double mutants. The double *intV* mutant was complemented with *intV2* to confirm the role of *intV2* in VPI-1 excision B. Excision of VPI-1 using the attB1 excision assay in single, double and triple ( $\Delta\Delta\Delta$ ) RDF mutant backgrounds. C. Excision of VPI-1 in the quadruple mutant ( $\Delta\Delta\Delta\Delta vpiT$ ), which lacks all three RDFs as well as the transposase, *vpiT*. C. Only round #2 is shown as no excision products were detected in round #1. D. VPI-1 excision in  $\Delta\Delta\Delta$  mutant complemented with each RDF individually. Excision of the complements was detected in round #1 and referred to as a super-excision phenotype. Control indicates sample with no template DNA.

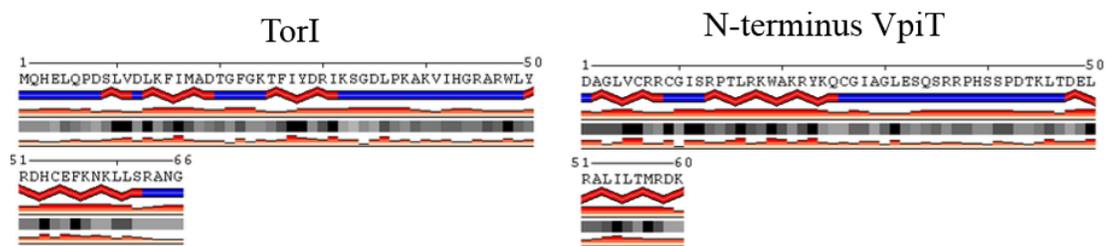


**Figure 9 Conserved domains of IntV1 and IntV2.** Amino acid alignment of IntV1 (gi: 9655299) and IntV2 (gi: 15641762) using the MUSCLE algorithm in Jalview. Both integrases share three conserved domains (represented by blocks in sequential order): DUF4102 (pfam13356) domain (IntV1: 5-94, IntV2: 8-97), SAM-like domain (pfam14659) (IntV1: 103-151, IntV2: 106-157) and the C-terminal domain contained a P4 catalytic domain (pfam00589) (IntV1: 213-390, IntV2: 215-393). The highly conserved active site residues are marked by triangles.

RDFs are not essential for VPI-1 excision but can promote excision.

Although we were able to observe excision of VPI-1 both in our *attP1* and *attB1* excision assays, this island contains no annotated RDF (**Figure 6C and 6D**). To explore this further, VPI-1 excision was examined in the RDF mutant strains using the *attB1* excision assay, and excision was observed in the single, double and triple mutants (**Figure 8B**). While no RDF has been annotated within VPI-1, it is possible that one with little sequence homology exists. In order to search for potential new RDFs we used a method similar to one that Fogg et al. used to identify the *E. coli*  $\Phi 24_B$  Xis (Fogg, Rigden et al. 2011). The DNA sequence of the entire VPI-1 region and flanking open reading frames (ORFs) (VC0848 – VC0816, genome coordinates 872764 – 915211) was analyzed for all ORFs using the bacterial code which includes

start codons GTG, TTG, CTG and ATT. This analysis identified 35 ORFs between 50 - 110 amino acids in length (typical RDF protein length). These were further analyzed using the HHPred server, a more sensitive analysis tool based on hidden Markov models (HMMs) (Soding, Biegert et al. 2005). None of the 35 identified ORFs yielded any similarity to previously described excisionases. However, when the *vpiT* (VC0817) amino acid sequence was submitted to HHPred, 52 amino acids near the N-terminus displayed a distant relationship to TorI, a previously characterized excisionase (Panis, Mejean et al. 2007; Panis, Duverger et al. 2010; Panis, Franche et al. 2012). Although the shared identity in this region was 12% and an e-value 0.0061, the predicted secondary elements of these regions were near identical (**Figure 10**). To eliminate the possibility of the VpiT aiding in the excision of VPI-1, a quadruple mutant was constructed in which the three RDFs and *vpiT* were deleted. The quadruple mutant was examined and the excision of VPI-1 was still detected (**Figure 8C**), suggesting the VpiT is not involved and that RDFs are not necessary for integrase mediated VPI-1 excision.



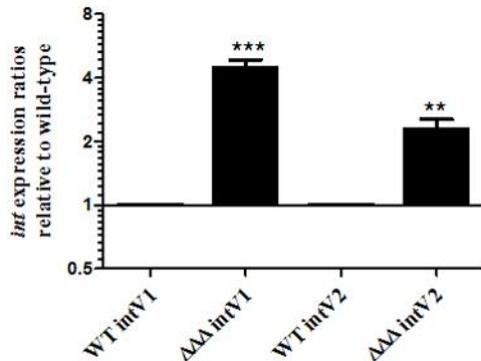
**Figure 10 Predicted secondary structure of N-terminus VpiT and TorI.** In search for an unannotated RDF, the HHPred server identified a 52 amino acid region near the N-terminus of VpiT as having a distant relationship to TorI, a previously characterized excisionase. Although the shared identity in this region was 12% and the e-value of 0.0061, the predicted secondary structures of these regions were near identical.

We next examined excision of VPI-1 in the triple RDF mutant complemented strains as described for VPI-2. Here we also observed the super-excision phenotype with each RDF (**Figure 8D**). Complementation with empty pBAD33 had no effect on excision of either island (data not shown). This data suggests that although the RDFs are not essential for VPI-1 excision, they do function to promote excision of VPI-1.

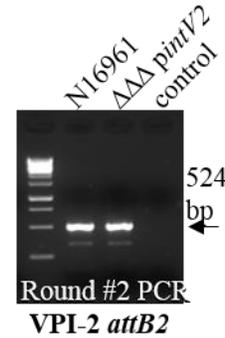
RDFs are negative regulators of *intV1* and *intV2*.

In order to determine if the RDFs play a role in controlling transcription of the integrase, real-time quantitative PCR was used to determine the transcript levels of *intV1* and *intV2* in wild-type and  $\Delta\text{vefA}\Delta\text{vefB}\Delta\text{vefC}$  cells. We found that *intV1* and *intV2* are 4.5-fold ( $p < 0.005$ ) and 2.5-fold ( $p < 0.05$ ) higher in the triple RDF mutant compared to wild-type (**Figure 11A**), indicating that the RDFs act as negative transcriptional regulators of *intV1* and *intV2*.

**A. QPCR *intV* triple *vef* mutant**



**B. *intV2* over-expression in  $\Delta\Delta\Delta$**



**Figure 11 RDFs act as transcriptional repressors of the integrases *intV1* and *intV2*.**

Expression analysis of *intV1* and *intV2* in wild-type vs.  $\Delta$ *vefA* $\Delta$ *vefB* $\Delta$ *vefC* logarithmic cells using real-time quantitative PCR (QPCR). Fold change values were extracted using the  $\Delta\Delta$ Ct method and are represented on a log 10 scale. A. *intV* expression in WT compared to triple Vef mutant. B. VPI-2 attB2 excision assay of *intV2* over-expressed in the triple Vef mutant background. Only round #2 is shown as no excision products were detected in round #1. Control indicates sample with no template DNA.

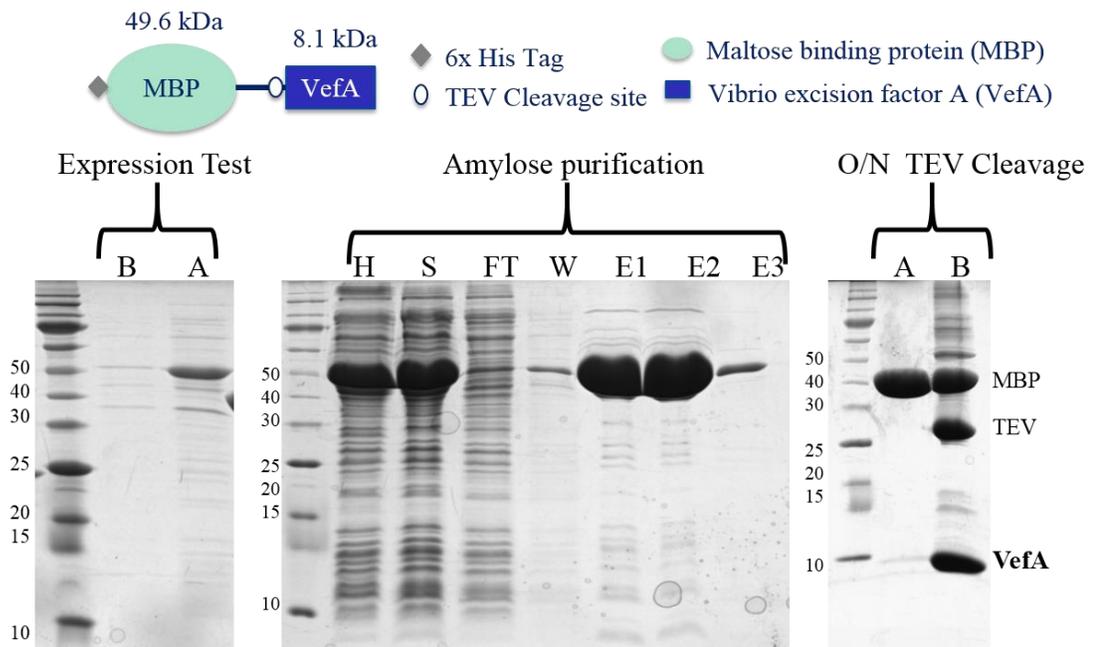
Over-expression of *intV2* can compensate for the absence of RDFs in VPI-2 excision.

The transcriptional data revealed higher levels of *intV1* than *intV2* in  $\Delta$ *vefA* $\Delta$ *vefB* $\Delta$ *vefC*. This higher level of integrase observed in the triple RDF mutant may explain the VPI-1 excision phenotype still observed in this strain (**Figure 8B and 11A**). To determine if increasing the levels of *intV2* can restore VPI-2 excision, in the absence of RDFs, the *intV2* gene was ectopically expressed under control of an arabinose promoter in the  $\Delta$ *vefA* $\Delta$ *vefB* $\Delta$ *vefC* background. The *attB2* assay was used to

examine the excision phenotype of this strain. Over-expression of *intV2* in the RDF negative background did indeed result in restoration of VPI-2 excision (**Figure 11B**). These data suggest that if integrase levels are high enough they do not require a RDF to aid in their excision function and may explain why VPI-1 excision still occurs in an RDF negative background.

#### Protein purification of VefA and VefB

VefA and VefB proteins were purified as described in the materials and methods. The initial expression test revealed high induction of MBP-VefA (**Figure 12**). A four liter culture of BL21(DE3) pMALVefA yielded approximately 515 mg of the MBP-VefA fusion protein (49.6 kDa) after elution from the amylose column (**Figure 12**). Next, the TEV protease was added to cleave the MBP from VefA, and after overnight incubation at 4°C (standard incubation time) significant precipitation of VefA occurred (**Figure 12**). In order to minimize this protein aggregation, the TEV cleavage reaction was carried out under varying concentrations (10, 5 and 1 and 0.5mg/ml) with or without glycerol or DTT. Glycerol is a known inhibitor of protein aggregation, thought to function by inhibiting protein unfolding and through interactions with hydrophobic surface regions to stabilize aggregation-prone intermediates(Vagenende, Yap et al. 2009). The addition of DTT served to prevent disulfide bonds between VefA monomers since a cysteine is found in the protein sequence. The lower protein dilutions nor the additives provided any reduction in VefA aggregation. However, allowing cleavage to occur at 4°C for only 45 minutes resulted in reduced precipitation (**Figure 13**). When 200 mg of the MBP-VefA fusion protein is treated with TEV protease for 45 minutes and subject to IMAC, approximately 60 mg of cleaved VefA was recovered (**Figure 13**).



**Figure 12 SDS-PAGE of MBP-VefA purification: Expression test, Amylose column and TEV cleavage.**  
 Expression test: Cell homogenate before (B) and after (A) IPTG induction. Amylose purification: bacterial cell homogenate (H), supernatant (S), amylose column flow-through (FT), wash (W), and elutions (E) 1-3. TEV cleavage of MBP-VefA after overnight incubation at 4°C after (A) and before (B) centrifugation

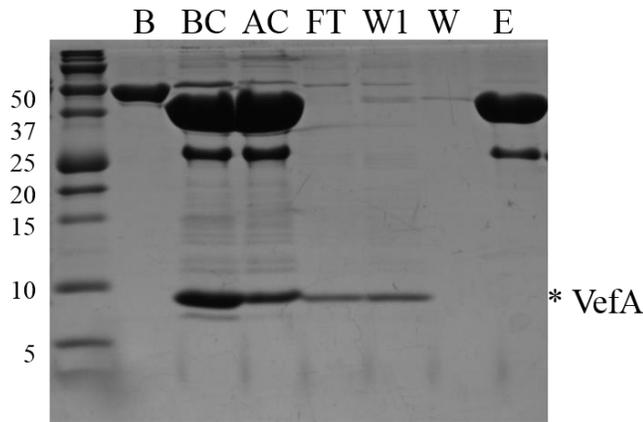


Figure 13 **SDS-PAGE of MBP-VefA purification: TEV cleavage and IMAC.** MBP-VefA protein before cleavage (B), MBP-VefA post cleavage (4C for 45 mins) before centrifugation (BC) and after centrifugation (AC), IMAC flow-through (FT), 1 column volume (CV) wash (W1), remaining wash and 1 M imidazole elution (E).

The MBP-VefB was purified in a similar manner as MBP-VefA. The fusion protein was first separated using an amylose column (**Figure 14**). A two liter harvest yielded approximately 94 mg of MBP-VefB post amylose elution, significantly lower than the MBP-VefA elution. This lower yield could be attributed to the lower expression levels compared to MBP-VefA revealed in the initial expression tests (**Figure 12 and 14**) Additionally, the  $OD_{600}$  of the bacterial expression culture after overnight growth was only 1.1, significantly lower than that for MBP-VefA ( $OD_{600}$  10). This indicates that MBP-VefB may be toxic to the *E. coli* cells. After TEV cleavage and IMAC cleanup to remove His-tagged MBP and TEV, there were equal amounts of MBP and VefB, indicating associated of the two proteins. A heparin column was utilized to separate MBP from VefB. Heparin is a sulfated glucosaminoglycan consisting of alternating units of uronic acid and D-glucosamine which mimics the DNA backbone and is commonly used in the purification of DNA-

binding proteins. VefB bound to the heparin column under low salt conditions and was eluted in a 56-60% NaCl elution without MBP association. Elutions were concentrated and the buffer was exchanged prior to use in DNA binding assays.

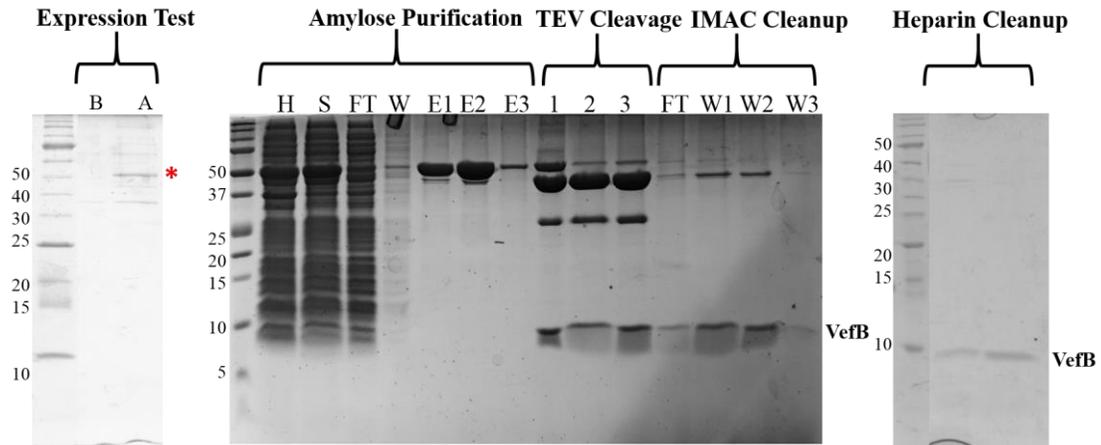


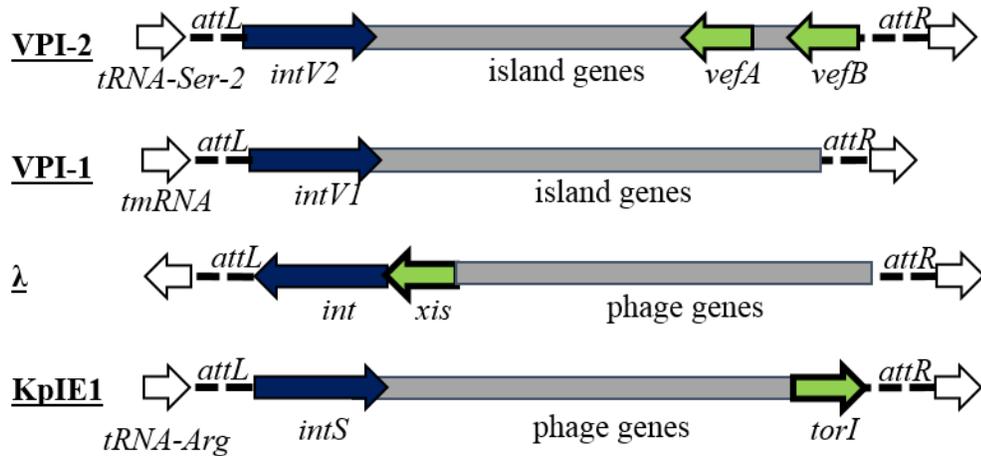
Figure 14 **SDS-PAGE of MBP-VefB purification: Expression Test, TEV cleavage, IMAC and Heparin Cleanup.**

Expression test: Cell homogenate before (B) and after (A) IPTG induction. The star (\*) denotes MBP-VefB fusion. Amylose Purification: bacterial cell homogenate (H), supernatant (S), amylose column flow-through (FT), wash (W), and elutions (E) 1-3. TEV cleavage of MBP-VefB after incubation at 4°C after 1.5 hours (1) and after overnight TEV cleavage, before (2) and after (3) centrifugation. IMAC flow-through (FT) and 1 column volume (CV) washes (W1-W3). The IMAC Washes were subject to a heparin column to separate VefB from MBP. VefB was eluted at approximately 56-60% NaCl in two fractions and concentrated to 0.1 mg/ml. Theoretical pI/Mw of MBP-VefB: 5.24 / 50739.72 Da and VefB post TEV cleavage: 10.44 / 9219.60 Da.

VefA and VefB can bind the *att* sites of VPI-1 and VPI-2.

The RDF-mediated mechanism of integrase control we observed has also been proposed for the *E. coli* KpIE1 prophage (Panis, Duverger et al. 2010), which contains

a recombination module similar to that of VPI-1 and VPI-2, in that the integrase promoter is within the *att* site (**Figure 15**).



**Figure 15 Mobile integrative genetic element (MIGE) recombination modules.** A comparison of the genetic organization of the two *V. cholerae* VPIs examined in this study with the bacteriophage  $\lambda$  and the *E. coli* prophage KpIE1. Contrary to bacteriophage  $\lambda$  the integrases of the VPIs and KpIE1 possess promoter regions that overlap with attachment sites, making them accessible for regulation by intasome binding proteins. Additionally, the genetic organization reveals that the RDF(s) and integrase of VPI-2 and KpIE1 are not co-regulated as they are in bacteriophage  $\lambda$ . The position of *vefB* also exposes the gene for possible regulation by intasome binding proteins.

We next sought to determine whether the *V. cholerae* RDF VefA can bind at the *attR1* and *attL2* sites of VPI-1 and VPI-2 respectively. The DNA fragments *attR1* and *attL2* encompass the region between the *att* site and translational start site of the integrases IntV1 and IntV2, respectively (**Figure 16A**). Electrophoretic mobility shift assays (EMSAs) were used to determine whether VefA can bind these *att* sites. Increasing amounts of VefA were incubated with 30 ng of each DNA fragment. A

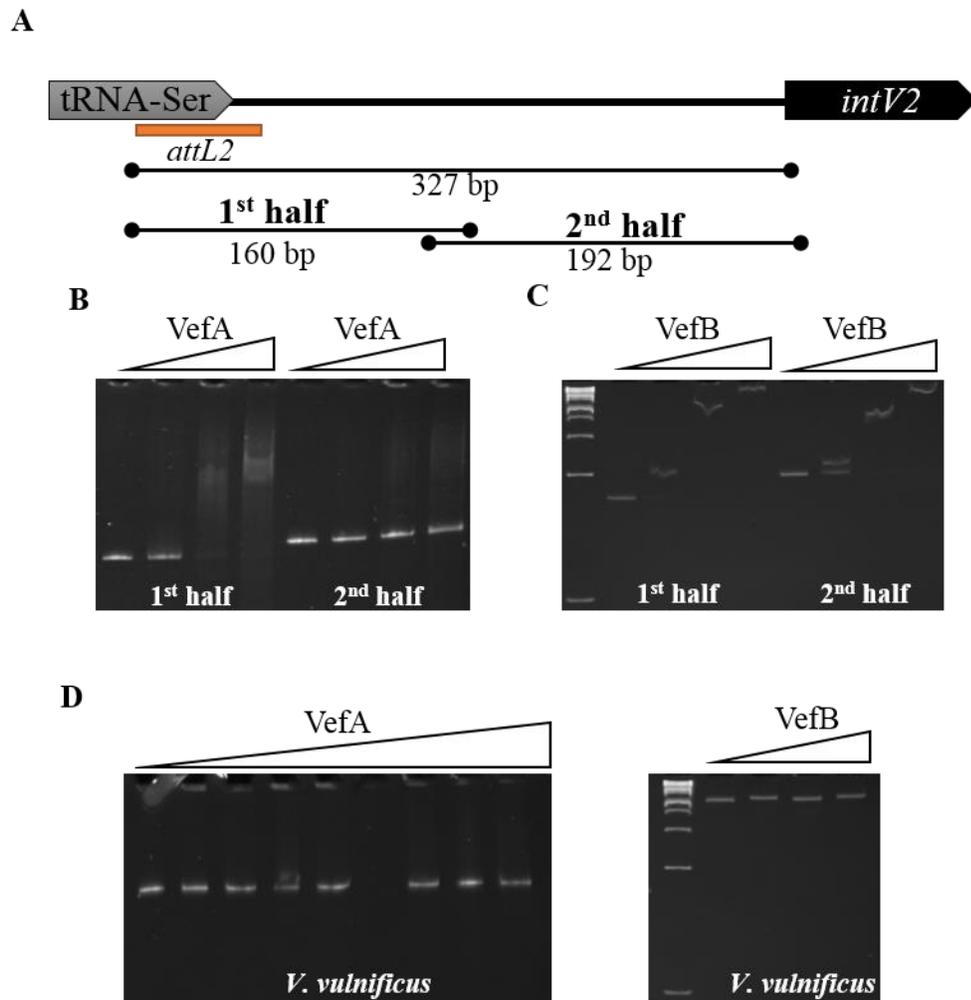
shift was observed for *attR1* and *attL2* DNA fragments near 10  $\mu$ M and 4.1  $\mu$ M VefA, respectively (**Figure 16B**). No VefA binding was detected when using *V. vulnificus* DNA amplified from VV1\_0809 as a control for EMSAs under the same ratios (**Figure 17D**). A DNA alignment of the *attR1* and *attL2* DNA fragments using ClustalW identified 76% conservation and repeats up to 14 bp were identified (**Figure 16C**). Several shared motifs were also identified in the *attR1* and *attL2* DNA fragments near the core *att* sites suggesting these may be RDF binding sites (**Figure 16C**). Additionally, the *attL2* sequence possessed five repeated motif sequences in the predicted 5' UTR of *intV2* (**Figure 16C**).

We next determined whether VefB can also bind at the attachment sites of VPI-1 and VPI-2. Although only sharing 40% identity at the amino acid level, the predicted secondary elements of VefA and VefB are near identical (**Figure 5**). However, the number and location of charged residues is distinctly different; the isoelectric point of VefA is 5.31 while VefB's is 10.44. Binding of VefB to the DNA fragments using EMSAs was observed at 2.5  $\mu$ M for the *att* site and 1.4  $\mu$ M for *attR1* and *attL2*, respectively (**Figure 16E**). Both VefA and VefB bind to the VPI-2 *att* site at a higher affinity than the *attR1* site of VPI-1 (**Figure 16B and 16E**). To further narrow down where these RDFs may be binding, two DNA fragments were generated from the *attL2* sequence, named 1<sup>st</sup> half (includes core *att* site) and 2<sup>nd</sup> half (includes *intV2* start codon) and were used for EMSAs (**Fig. 17A**). Both VefA and VefB were capable of binding the first half, however only VefB was able to bind the second half of the *attL2* DNA fragment (**Fig. 17B and 17C**).



Figure 16 **VefA and VefB bind the att sites of VPI-1 and VPI-2.**

A. Representation of the DNA sequences, *attR1* and *attL2*, which span from the core att site to the translational start site of intV1 and intV2 of VPI-1 and VPI-2, respectively. B. Electrophoretic mobility shift assays (EMSAs) of the DNA fragments *attR1* (30 ng) and *attL2* (30 ng) with varying concentrations of VefA (0 - 70  $\mu$ M). Binding of VefA is observed at 10  $\mu$ M and 4.1  $\mu$ M for *attR1* and *attL2* respectively. C. Alignment of *attR1* and *attL2* DNA fragments used in EMSAs. Exact repeats shared between the two sequences are boxed. Core att site sequences are boxed and highlighted. Repeated motif sequences identified by MEME are represented with arrows and their corresponding p-values. D. Logo sequences of the three motifs identified in the att sequences in C. E. EMSA of the DNA fragments *attR1* (30 ng) and *attL2* (30 ng) with varying concentrations of VefB (0 – 20.5  $\mu$ M). Binding of VefB is observed at 2.5  $\mu$ M and 1.4  $\mu$ M for *attR1* and *attL2* respectively.



**Figure 17 VefA and VefB bind the 1st half *attL2* site at a higher affinity.**  
 A. Representation of the 1st and 2nd half fragments of the *attL2* DNA sequence, amplified with primer pairs attL2Fwd/attL2BRev and attL2CFwd/attL2Rev, respectively (Table 2), used for electrophoretic mobility shift assays (EMSAs). B. Varying concentrations of VefA (0 – 46  $\mu$ M) with attL2 1st half (30 ng) and 2nd half (30 ng). Binding is observed at 23  $\mu$ M for the first half. C. Varying concentrations of VefB (0 - 23.1  $\mu$ M) with attL2 1st half (30 ng) and 2nd half (30 ng). Binding is observed at 2.8  $\mu$ M and 2.4  $\mu$ M for the 1st and 2nd half, respectively. D. EMSA using control *V. vulnificus* DNA (30 ng) with varying concentrations of VefA (0 – 12  $\mu$ M) and VefB (0 - 4.19  $\mu$ M). No binding was observed.

## Distribution of RDFs among Vibrio

BLAST analysis revealed over 3,000 representatives (among 105 genera and 283 species) with equal to or greater than 70% query cover and 40% identity to the *V. cholerae* RDFs across five bacterial classes: Acidobacteria, Alpha-proteobacteria, Beta-proteobacteria, Cyanobacteria and Gamma-proteobacteria (**Figure 18**). The vast majority were members of the order Enterobacteriales: *E. coli* (787), *Klebsiella pneumoniae* (276), *Salmonella enterica* (166), *Yersinia tuberculosis* (118) and *Enterobacter cloacae* (109), bacterial groups that are also disproportionately represented in the genome database. These data demonstrate that the occurrence of these RDFs is widespread among bacteria.

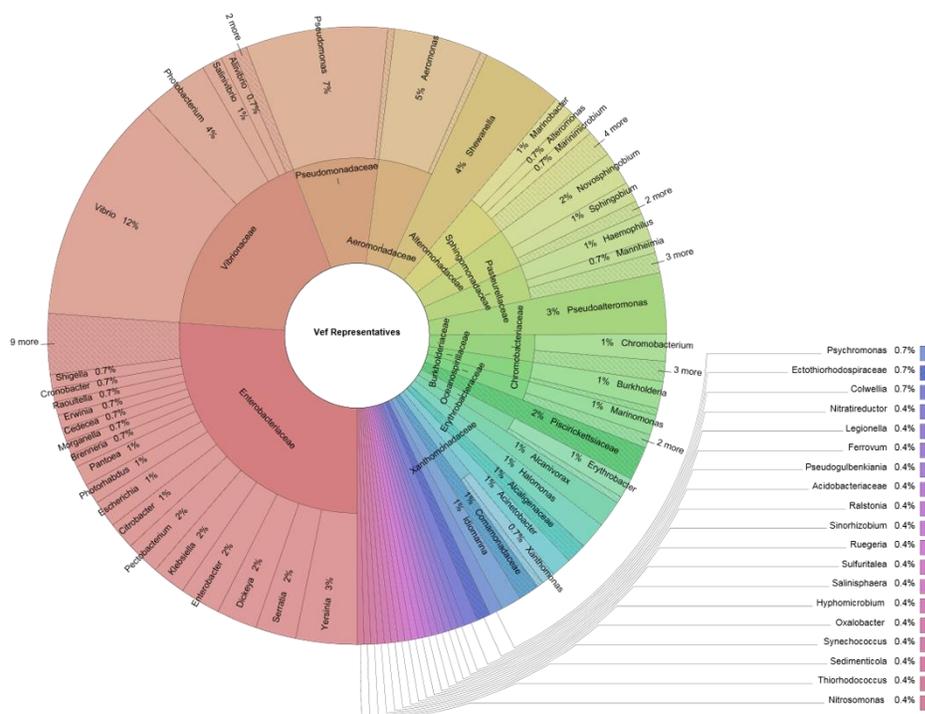


Figure 18 **Distribution of Vefs in Bacteria.**

A BLAST analysis revealed over 3,000 representatives (among 105 genera and 283 species) with equal to or greater than 70% query cover and 40% identity to the *V. cholerae* RDFs.

In addition, we examined *Vibrio* species in detail to identify island regions with IntV1 homologues and associated RDFs. From this analysis, IntV1 was found in association with homologs of VefA, VefB and VefC in numerous other *V. cholerae* strains (**Figure 19**), indicating that at one time VPI-1 may have contained an RDF. For example, *V. cholerae* O1 strain 12129(1) contains VPI-1 with both VefA and VefB and a non-O1/non-O139 strain *V. cholerae* HE-09 contains a VPI-1 variant with VefC. Several other examples of PAIs containing IntV1 and Vef homologs exist within *V. cholerae* as well as in other *Vibrionaceae* (**Figure 19**). These findings demonstrate the association, at least in proximity, of IntV1 and all 3 RDFs found in *V. cholerae* N16961, suggesting a shared evolutionary history, supporting their ability to cross-talk.

*Vibrio cholerae* sv Inaba 12129(1)



*Vibrio cholerae* CT 5369-93



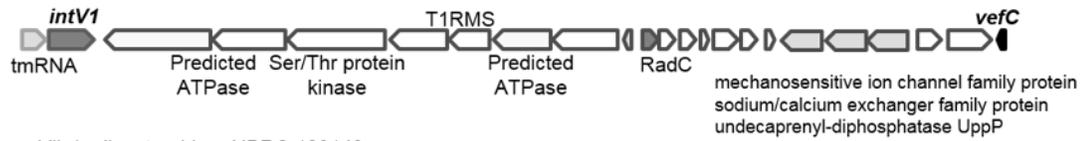
*Vibrio cholerae* sv. O135 RC385



*Listonella anguillarum* sv. O1 96F



*Vibrio cholerae* HE-09



*Vibrio diazotrophicus* NBRC 103148



*Vibrio vulnificus* YJ016 VVI-V



Strain	IntV1	VefA	VefB	VefC
<i>V. cholerae</i> sv Inaba 12129(1)	93%	99%	83%	
<i>V. cholerae</i> CT 5369-93	81%			98%
<i>V. cholerae</i> sv. O135 RC385	100%	91%	97%	
<i>Listonella anguillarum</i> sv. O1 96F	82%			94%
<i>V. cholerae</i> HE-09	81%			100%
<i>V. diazotrophicus</i> NBRC 103148	81%			94%
<i>V. vulnificus</i> YJ016	81%	88%		

Figure 19 **Vibrio island regions inserted at tmRNA genes containing RDFs.**  
A) Island regions of *V. cholerae* and other Vibrionaceae which contain IntV1 homologs inserted at tmRNA genes as well as Vef homologs. Note the genetic variation of the gene content of each island. *Vibrio vulnificus* island (VVI) - 5. B) Strain name with % homology to *V. cholerae* N16961 IntV1 and RDFs VefA, VefB or VefC. These observations suggest an evolutionary relationship between IntV1 and the Vefs, supporting the VPI-1 cross-talk hypothesis.

## Discussion

Pathogenicity islands, which possess a wide range of virulence genes, are widespread amongst bacteria but are the least examined class of MIGEs (Carniel, Guilvout et al. 1996; Karaolis, Johnson et al. 1998; Hacker and Kaper 2000; Al-Hasani, Rajakumar et al. 2001; Jermyn and Boyd 2002; Hensel 2004; O'Shea, Finnan et al. 2004; Jermyn and Boyd 2005; Hurley, Quirke et al. 2006; Quirke, Reen et al. 2006; Murphy and Boyd 2008). Excision of PAIs from the bacterial chromosome is the first step in the transfer and spread of these elements. Thus determining the excision mechanism is essential in understanding the distribution potential of the mobilome and its role in the evolution of bacteria. *Vibrio cholerae* was selected as a model organism for this study as it contains two PAIs important for its virulence. Given the ability of VPIs to excise from the bacterial genome (Rajanna, Wang et al. 2003; Murphy and Boyd 2008; Almagro-Moreno, Napolitano et al. 2010), we sought to decipher this mechanism, particularly how VPI-1 could excise without containing an RDF. Previous studies which reported that an RDF was not necessary for TR integrase mediated excision, later identified RDFs bioinformatically that was required (Fogg, Rigden et al. 2011). Thus, the VPI-1 region was analyzed for unannotated

ORFs using a similar method but none were identified with any homology to excisionases (see *Results*).

We determined that all three RDFs, *vefA*, *vefB* and *vefC*, have redundant roles in promoting excision of VPI-1 and VPI-2, as over-expression of each individual RDF resulted in super excision of these islands. These data indicate extensive cross-talk between three different islands: *vefA* and *vefB* found on VPI-2 can promote excision of VPI-1 and *vefC* found on VSP-II can aid in excision of both VPI-1 and VPI-2. This is the first report of RDF-mediated cross-talk in any MIGE. The cross-talk was not limited to the RDFs; we found that the cognate integrase of VPI-2 can also play a role in excision of VPI-1. This is not the first report of integrase mediated cross-talk between pathogenicity islands - in UPEC, the cognate integrase of PAI II<sub>536</sub> was capable of mediating excision of PAI V<sub>536</sub>, an island located elsewhere on the genome and this too was one-directional cross-talk (Hochhut, Wilde et al. 2006).

Once the role of RDFs in PAI excision was established, we next examined the molecular mechanism by which these proteins act. Some RDFs have been proposed to act as transcriptional activators of the integrase. This mechanism has been reported for the Cox protein of phage HPI1 and HPI2 (Esposito and Scocca 1997) and AlpA of the K-12 *E. coli* cryptic P4-like prophage (Kirby, Trempey et al. 1994; Trempey, Kirby et al. 1994). To date, there is no data on the mechanism of RDF function in *V. cholerae*. We performed real-time quantitative PCR to determine whether the RDFs control integrase transcription and found that the Vefs were negative transcriptional regulators. The unique configuration of the VPI recombination modules (*att* sites, *intV* and *vef*) can provide insight into this negative regulation. The promoter regions of *intV1* and *intV2* are located at the attachment sites of the PAI, whereas in the

bacteriophage  $\lambda$ , in which the majority of RDF studies have been done, the integrase promoter lies within the phage genome (**Figure 15**). The position of the PAI integrase promoters allows for potential regulation by proteins that bind at the *att* sites to aid in formation of the excisive intasome (protein-DNA complex necessary for excision), which we propose are the Vefs. In the *E. coli* KpIE1 prophage, which has a similar integrase and *att* site arrangement as the VPIs (**Figure 15**), Panis and colleagues also found that the RDF, TorI, negatively regulates expression of the *intS* gene (Panis, Duverger et al. 2010). Based on the recombination module, it can be suggested that the integrases may also be self-regulating transcription of their own gene. It can be postulated that the same is true for *vefB*, whose promoter overlaps with the *attR2* site. Others have found RDFs to act as integrase repressors, such as the RDF Vis from the P4 bacteriophage, which was found to repress the P4 integrase at transcription initiation as well as post-transcriptionally (Piazzolla, Cali et al. 2006).

Our transcriptional analysis of integrase expression shed light onto the question of why we observe an excision product for VPI-1 and not VPI-2 in the absence of the RDFs. Expression levels of *intV1* and *intV2* were both increased in  $\Delta vefA\Delta vefB\Delta vefC$  compared to wild-type, however this affect was more pronounced for *intV1* resulting in more cognate integrase to facilitate the site-specific recombination reaction, thus allowing excision of VPI-1. The increased levels of *intV2* in this genetic background may also aid in the excision of VPI-1, as we have shown there is cross-talk with this integrase in VPI-1 excision. From this observation, we next wanted to determine if VPI-2 excision could occur in an RDF negative background if more *intV2* was present. We over-expressed *intV2* in the  $\Delta vefA\Delta vefB\Delta vefC$  background and found that VPI-2 excision was detectable at wild-

type levels. These data indicate that if enough integrase is provided *intV1* and *intV2* can facilitate site-specific recombination and excision of VPIs without the presence of RDFs. This data highlights the importance of the RDFs in regulating the amount of integrase necessary for controlling VPI excision.

The majority of RDFs that have been described are from prophages and were shown to control directionality of recombination by playing an architectural role by binding at the *att* sites, to each other and/or to the integrase (Lewis and Hatfull 2001; Warren, Sam et al. 2003; Abbani, Iwahara et al. 2005; Flanigan and Gardner 2008; Coddeville and Ritzenthaler 2010; Fogg, Rigden et al. 2011; Singh, Plaks et al. 2013). We performed EMSAs with purified Vef proteins and *attR1* and *attL2* DNA fragments from VPI-1 and VPI-2 respectively. VefA and VefB bound the *att* sites of both islands, although VefB bound both sites at a higher affinity, causing a shift with four times less protein than VefA. VefA and VefB have similar predicted secondary elements and their basic residues, which are essential for DNA binding (Jones, van Heyningen et al. 1999), are highly conserved. However, VefA contains additional charged residues, including acidic residues in exposed loops. Furthermore, VefA contains a sole Cys at position 70 (**Figure 5**) whose contribution to function remains unknown and warrants further investigation. Thus, the higher affinity of VefB for the *att* site DNA may be due to its higher positive charge density, allowing for favorable interaction with DNA. The shift of the protein-bound DNA changes as the amount of protein increases, suggesting multiple Vef binding sites and formation of protein oligomers. Cooperative binding of the lambda Xis protein at the *att* DNA sites has been reported and formation of this micronucleoprotein filament has shown to drive DNA bending (Sam, Papagiannis et al. 2002; Abbani, Papagiannis et al. 2007; Singh,

Plaks et al. 2013). Therefore, we speculate that the multiple Vef monomers bind at the *att* site and act similarly to bend the DNA, thereby driving the excision reaction.

Based on our findings, we have proposed a model of RDF function in the control of PAI excision in *V. cholerae*. In the absence of the RDFs, *intV1* and *intV2* expression increases, however as RDF levels increase the repression of *int* is enhanced. We showed that RDFs bind at the *att* sites of VPI-1 and VPI-2 to aid in excision. Consequently, binding, specifically at the *attR1* and *attL2* sites resulted in a decrease of *intV1* and *intV2* transcription, respectively. The negative regulation of *int* transcription may be caused by preventing promoter recognition either by altering DNA topology or direct RDF binding at the *int* promoters. It may be possible that the integrase also plays a role in regulating its own transcription and that VefB, whose promoter also overlaps with the *attR2* site, may also self-regulate.

This is the first report of RDF-mediated cross-talk in PAI excision. We have identified Vef homologs in over 100 genera and 280 species, confirming their prevalence among bacteria. Due to their low sequence homology this list is most likely vastly under-represented. Understanding the mechanism of PAI excision is important in the evolution and spread of these elements to their non-pathogenic counterparts. This study reveals extensive cross-talk, through RDFs and integrases, between multiple PAIs in a single bacterium, to control the excision and probable spread of these elements to other bacteria. It is likely that other species which contain numerous MIGEs are also capable of cross-talk. Thus it is important to consider the genetic content outside of an individual element when determining the inputs necessary for excision and its impact on the bacterial mobilome.

Table 1 Bacterial strains and plasmids used in this study.

Strain or plasmid	Description	Reference
<b><i>V. cholerae</i> strains</b>		
N16961	O1 El Tor, VPI-1, VPI-2, VSP-I, VSP-II, SmR	(Heidelberg, Eisen et al. 2000)
O395	classical, VPI-1, VPI-2	Laboratory strain
MO2	O139, VPI-1, VPI-2	Laboratory strain
SG-7	Environmental	Laboratory strain
MC0817 ( $\Delta$ vpiT)	N16961 $\Delta$ VVC0817, SmR	This study
MC0847 ( $\Delta$ intV1)	N16961 $\Delta$ VVC0847, SmR	This study
CN1758 ( $\Delta$ intV2)	N16961 $\Delta$ VVC1758, SmR	This study
MC4758 ( $\Delta$ intV1 $\Delta$ intV2)	$\Delta$ intV1 $\Delta$ VVC1758, SmR	This study
$\Delta$ intV1pintV1	$\Delta$ intV1 harboring pBAintV1	This study
$\Delta$ intV2pintV2	$\Delta$ intV2 harboring pBAintV2	This study
SAM1785 ( $\Delta$ vefA)	N16961 $\Delta$ VVC1785, SmR	(Almagro-Moreno, Napolitano et al. 2010)
SAM1809 ( $\Delta$ vefB)	N16961 $\Delta$ VVC1809, SmR	(Almagro-Moreno, Napolitano et al. 2010)
MC0497 ( $\Delta$ vefC)	N16961 $\Delta$ VVC0497, SmR	This study
MC8509 ( $\Delta$ vefA $\Delta$ vefB)	$\Delta$ vefA $\Delta$ VVC1809, SmR	This study
MC850997 ( $\Delta$ vefA $\Delta$ vefB $\Delta$ vefC)	$\Delta$ vefA $\Delta$ vefB $\Delta$ VVC0497, SmR	This study
MC85099717 ( $\Delta$ vefA $\Delta$ vefB $\Delta$ vefC $\Delta$ vpiT)	$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC $\Delta$ VVC0817, SmR	This study
$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC pvefA	$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC harboring pBBvefA	This study
$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC pvefB	$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC harboring pBAvefB	This study
$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC pvefC	$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC harboring pBAvefC	This study
$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC pintV2	$\Delta$ vefA $\Delta$ vefB $\Delta$ vefC harboring pBAintV2	This study
<b><i>E. coli</i> strains</b>		
Dh5 $\alpha$ pir		Laboratory strain
$\beta$ 2155pir	Donor for bacterial conjugation, DAP auxotroph	Laboratory strain
BL21(DE3)	Expression strain	

Table 1 (Continued)

<b>Strain or plasmid</b>	<b>Description</b>	<b>Reference</b>
<b>Plasmids</b>		
pJET1.2	Cloning vector, AmR	Fermentas
pDS132	Suicide vector, SacB, CmR	(Philippe, Alcaraz et al. 2004)
pBAD33	Arabinose promoter, CmR	(Guzman, Belin et al. 1995)
pDSΔvpiT	pDS132 harboring truncated vpiT gene	This study
pDSΔintV1	pDS132 harboring truncated intV1 gene	This study
pDSΔintV2	pDS132 harboring truncated intV2 gene	This study
pDSΔvefA	pDS132 harboring truncated vefA gene	(Almagro-Moreno, Napolitano et al. 2010)
pDSΔvefB	pDS132 harboring truncated vefB gene	(Almagro-Moreno, Napolitano et al. 2010)
pDSΔvefC	pDS132 harboring truncated vefC gene	This study
pBAintV1	pBAD33 with full copy of intV1	This study
pBAintV2	pBAD33 with full copy of intV2	
pBAvefA	pBAD33 with full copy of vefA	This study
pBAvefB	pBAD33 with full copy of vefB	This study
pBAvefC	pBAD33 with full copy of vefC	This study
pMALc5x-His6x	Expression vector, TEV site, AmR	(Liu, Srinivasan et al.)
pMALc5x-His6x-VefA	pMALc5x-His6x harboring VefA	This study
pMALc5x-His6x-VefB	pMALc5x-His6x harboring VefB	This study

Table 2 Primers used in this study.

Primer Name	Sequence (5' → 3')	Product size (bp)
<b>Mutant Construction</b>		
VC0817A	tctagaAAGCTGCGATAGGGAGCT	520
VC0817B	ACCACAGCGTCGACATACGAG	
VC0817C	CTCGTATGTCGACGCTGTGGT	561
	ATCAAGAAAGAGCGGTTTCAAG	
VC0817D	gagctcGGCTAACGTCCCTATAGTC	
VC0817FF	AGCTCTTCCATCGACAACACCTCT	2,200
VC0817FR	TCGCCAGTGGAAGCAGAGCA	
VC0847A	tctagaTAGCCATTCGTTAGCGTGTG	556
VC0847B	ATCGTCTTGTGGATCGATGC	
VC0847C	GCATCGATCCACAAGACGAT	548
	AGAGGTAGCCTATCTGTGAC	
VC0847D	gagctcTGAGGATCCATGATATCCGC	
VC0847FF	CTAGCTTCCGCTTGTAAGAC	2,287
VC0847FR	TACTAACGGGTATCGAACTC	
cnVC1758A	tctagaGATTCGGTGAGTTGTCCGAG	531
cnVC1758B	TTGCCATGAGCGAGAATTGC	
cnVC1758C	GCAATTCTCGCTCATGGCAA	578
	GGAAGTTACAGTGTGGCTGG	
cnVC1758D	gagctcTCAGTAACAGAAAGGCTGCC	
cnVC1758FF	AAGCAAACGCACTCAATGCG	2,285
cnVC1758FR	CTGGTCGACAAGATTGTCTG	
VC0497A	GCtctagaGTACTCTTGCATGCGTTTGG	541
VC0497B	TCGTAGAAATCTCATGGGAATTGTc	
VC0497C	GACAATCCCATGAGATTTCTACGA	554
	GAATCCTTTCAGTAACACAAACATC	
VC0497D	CgagctcCGCTTCAACCGCTAACAGGT	
VC0497FF	CGAACACAACCCACCTTCCAACC	1,345
VC0497FR	GAGTCTGAACAACCTGTATTGG	
<b>Excision Assays (attP)</b>		
InvVC0847B	ATCGTCTTGTGGATCGATGC	744
Inv/NestVC0817R2	GCATCACCACATTCCTCATAAC	
NestVC0847F3	TTTCTCTCTAGGTTTGGAGG	483
NestVC0817R3	CTCTGTCCATAGACACCCAG	
VPI-2 InvVC1808F	AGCTAGACAGATTAGCTAACC	1,352
CnVC1758B	TTGCCATGAGCGAGAATTGC	
NestVC1758comR	AGAATTGCTTGGACGTACGC	461
NestVC1809comF	GCGTTAACTGAGAAAGTGTG	
<b>Excision Assays (attB)</b>		
VPI-1 attBF1	ATAGGGAGCTGGGCGTTAAT	428
VPI-1 attBR1	TGTAAGACGGGGAAATCAGG	
VPI-1 attBF2	TTGATGAGACGCTCTGAACC	217
VPI-1 attBR2	ATTCGTTAGCGTGTCCG	
VPI-2 fk1	CTGGCTATGAGCTGATTTG	1,449

Table 2 (Continued)

Primer Name	Sequence (5' → 3')	Product size (bp)
VPI-2 fk2	AGGGATTGGCTTTGAGG	
VPI-2attF	AGAGTGAAAGTCGCCAAAGC	524
VPI-2attR	GGGTGCAATTTTCGCATGTTGC	
<b>Complementation</b>		
VC0847F	gagctcAGGTACACTACAAGGTACAC	1,385
VC0847R	tctagaAGCGTATTCCACTGACAACC	
VC1758F	ggtaccGAGCTCGCTTTGAATATAGG	1,303
VC1758R	gagctcGAGTCCTCATGCTCTAGCCAG	
VC1785F	tctagaGGCATGCTGGTGTGTTACTAC	414
VC1785R	tctagaGACGCATGTATAATCACGC	
VC1809F	gagctcGTGGTTGATAGGCAATTGCAC	414
VC1809R	ctgcagGATCACACAAGTAAACCAGC	
VC0497F	gagctcCAGTGGCGGCTTATCCATGG	293
VC0497R	ctgcagCAGCGCCCTGTTGATGATGT	
<b>QPCR</b>		
VC0847QF	GCTACCTTTGGCTTCAATCG	193
VC0847QR	TGAGCACACCTTTGAGAACG	
McVC1758QF	AGAAGACACTGGAAGGGAATTG	199
McVC1758QR	ACACCGCTATTTACCGCATAG	
VC1730topIQF	ACTTTCACATAGCCACGGTC	151
VC1730topIQR	TCAGCCTAGATCCGAAACAG	
<b>Protein Expression</b>		
BamHIVC1785R	CTCGGATCCTCATCGCTGTTTAAGAC	225
NcoIVC1785F	TGCCCATGGATGGGACAACAAGACATAG	
VefAmutRev	CATGGACTGAAAATACAGG	
VefAmutFwd	GGACAACAAGACATAGGAG	
NcoIVefBFwd	TGCCCATGGGATTGCAACTAACCAACCTGAG	251
BamHIVefBRev	CTCGGATCCCTAATCCATTTCGACGATTGG	
<b>Binding Assays</b>		
attL2Fwd	TCTCACTCACCGCCACATTCAA	327
attL2Rev	GCCATTTCAATCCCTCAAGT	
attR1Fwd	GATGGGTACACCAGATTAACACG	190
attR1Rev	CCCCAGCTCCACCAAATCATAG	
attL2BRev	GTGGGTAGCTGGGTATTACTT	
attL2CFwd	AAGTAATACCCAGCTACCCAC	
VV1_0809A	tctagaAGCCAGGGTTTAACTACCGC	440
VV1_0809B	GGCTGAGTAAGCACGTTGA	

## Chapter 3

### CONSERVED RECOMBINATION MODULES CARRIED ON EXCISABLE PATHOGENICITY ISLANDS WITH NOVEL CARGO GENES IN VIBRIO CHOLERAЕ

The work presented in this Chapter was completed in collaboration with undergraduate researcher Molly Peters and is currently under review in the *Journal of Bacteriology*.

#### Introduction

*Vibrio cholerae* is the causative agent of the severe diarrheal disease cholera. The factors necessary for the manifestation of this disease were acquired horizontally. The cholera toxin (CT), a potent enterotoxin whose effects are responsible for the copious rice-water diarrhea associated with the disease, is encoded on the lysogenic filamentous phage CTX $\Phi$  (De 1959; Waldor and Mekalanos 1996). The toxin co-regulated pilus (TCP), a type IV pilus necessary for colonization of the small intestine is found on a pathogenicity island (PAI) named *Vibrio* pathogenicity island-1 (VPI-1) (Taylor, Miller et al. 1987; Herrington, Hall et al. 1988; Karaolis, Johnson et al. 1998). Furthermore, another PAI, named VPI-2, encodes a sialidase which cleaves the amino-sugar sialic acid from terminal glycoproteins to expose the cell surface receptor for CT (Jermyn and Boyd 2002). VPI-2 also contains a sialic acid catabolism and transport cluster, offering the bacterium a novel nutrient source *in vivo* that has been shown to give the bacterium a competitive advantage in the intestinal environment (Jermyn and

Boyd 2002; Almagro-Moreno and Boyd 2009; McDonald, J.B. Lubin et al. 2016). The pathogenic O1 and O139 serogroups of *V. cholerae*, which are responsible for epidemic cholera contain CTX $\Phi$ , VPI-1 and VPI-2, all regions that are absent from non-pathogenic strains. Thus, the acquisition of these mobile integrative genetic elements (MIGEs) has allowed *V. cholerae* to become one of the most successful enteric pathogens.

We recently published details on the mechanism of VPI-1 and VPI-2 site-specific excision from the bacterial chromosome (Carpenter, Rozovsky et al. 2015). The recombination module of VPI-1 and VPI-2 consists of a cognate tyrosine recombinase (TR) family integrase, IntV1 and IntV2, respectively, which share less than 50% amino acid identity and each island contains flanking *attL* and *attR* attachment sites marking the site of recombination. However, VPI-2 contains two excisionases (also known as recombination directionality factors (RDFs)), VefA and VefB, while VPI-1 contains none. (Carpenter, Rozovsky et al. 2015). RDFs have been shown to be essential for the excision function of TR integrases. The cognate TR integrases for both VPI-1 and VPI-2 are essential for excision of these islands from the bacterial chromosome (Murphy and Boyd 2008; Carpenter, Rozovsky et al. 2015)The integrase IntV2 found on VPI-2 can also aid in the excision of VPI-1 and the VPI-2 encoded RDFs VefA and VefB, are required in VPI-2 excision and can also facilitate VPI-1 excision (Carpenter, Rozovsky et al. 2015). These data demonstrate that there is significant crosstalk between VPIs. Various combinations of identical components of these recombination modules have been found in other members of Vibrionaceae suggesting a shared evolutionary history, supportive of their cross-talk abilities. In addition to the conserved recombination modules, several strains were identified

which possessed diverse cargo genes (Carpenter, Rozovsky et al. 2015). Of particular note was an island region that carried genes encoding for a type III secretion systems (T3SS) and a second island that contained a type VI secretion systems (T6SS) and a CRISPR-Cas region.

The T3SS is a contact dependent bacterial secretion system, an apparatus that spans the inner and outer bacterial membrane and the eukaryotic cell membrane to inject effector proteins into these eukaryotic cells, resulting in a variety of effects on host cell structures and signaling pathways that have significant effects on cell physiology (Hueck 1998). T3SSs are present in a wide range of *Vibrio* species. T3SSs are found *V. cholerae* non-choleraenic non-O1 and non-O139 serogroups, which have been attributed to significant inflammatory gastrointestinal infections worldwide (Dziejman, Serruto et al. 2005; Tam, Serruto et al. 2007; Shin, Tam et al. 2011).

*Vibrio cholerae* NRT36S serogroup O31 strain, isolated from a Japanese patient with travelers' diarrhea, contains a T3SS but does not contain the genes that encode CT and TCP (Chen, Johnson et al. 2007; Murphy and Boyd 2008). The T3SS in *V. cholerae* strains is found within a pathogenicity island, which we named VPI-3, inserted at a tRNA-serine (Ser) gene, the same location as VPI-2 (Chen, Johnson et al. 2007; Murphy and Boyd 2008). In VPI-3, the T3SS replaces the restriction modification system found in VPI-2 and the Mu phage region is absent from these strains but contains the genes for the scavenging, uptake and catabolism of sialic acid that are present in VPI-2 (**Figure 20**) (Chen, Johnson et al. 2007; Murphy and Boyd 2008). The T3SS containing island VPI-3 includes a recombination module identical to VPI-2, which suggests that this region can excise from the chromosome, the first step in its spread to other bacteria (Chen, Johnson et al. 2007; Murphy and Boyd 2008).

Type VI secretion systems are yet another mechanism for the translocation of proteins out of the bacterial cell directly into prokaryotic or eukaryotic cells and are structurally similar to the T4 bacteriophage needle (Leiman, Basler et al. 2009). We identified a *Vibrio* strain RC385, which contains portions of both the VPI-2 and VPI-1 recombination modules, contains cargo genes encoding for production of a T6SS apparatus and a CRISPR-Cas region (**Figure 19**) (Carpenter, Rozovsky et al. 2016).

In this study, we characterized several novel pathogenicity islands which possess conserved recombination modules but contain novel cargo genes, such as a T3SS or a T6SS. Furthermore, we showed that these islands can excise from the bacterial chromosome using a similar mechanism to that previously described for toxigenic *V. cholerae*. We examined the distribution and evolutionary history among *V. cholerae* isolates and show that the T3SS region is highly divergent among strains and were likely acquired independently in two sets of strains. These data suggest that acquisition of novel secretion systems on pathogenic islands may be more common than previously appreciated. It also suggests a mechanism by which different combinations and numbers of secretion systems can arise in different isolates of the same species.

## **Materials and Methods**

### **Bacterial Strains, Plasmids, & Growth Conditions.**

All bacterial strains and plasmids used in this study are listed in **Table 3**. Bacterial strains were grown in lysogeny broth (LB) (Fisher Scientific, Fair Lawn, NJ) overnight, at 37° C aerobically at 225 rpm unless otherwise noted. The diaminopimelic acid (DAP) auxotroph, *E. coli*  $\beta$ 2155 $\lambda$ *pir*, was supplemented with 0.3

mM DAP. When appropriate, media was supplemented with the following antibiotics: 200 µg/ml streptomycin (Sm), 25 µg/ml chloramphenicol (Cm) and 100 µg/ml ampicillin (Amp). LB agar was supplemented with 10% sucrose for double cross screening.

#### Mutant Construction.

An integrase deletion mutant in *V. cholerae* NRT36S was constructed using Splicing by Overlap-Extension (SOE) and homologous recombination (Horton, Hunt et al. 1989). Primers designed to create an in-frame deletion of the VPI-3 *intV2* in strain NRT36S are listed in **Table 4**. Briefly, *V. cholerae* NRT36S genomic DNA was used as template for PCR with VC1758A/NRTintv2B and NRTintv2C/NRTintv2D primer pairs to generate AB and CD products. Purified AB and CD products were used as template for PCR with the primer pair VC1758A/ NRTintv2D to create the truncated *intV* gene, utilizing the overlapping sequences found on NRTintv2B and NRTintv2C (**Table 4**). This truncated 1,161 base pair product was blunt-end cloned into pJET1.2 (Fisher Scientific) and transformed into *Escherichia coli* DH5 $\alpha$ . Plasmid DNA (pJET1.2  $\Delta$ *intV*) was isolated using the QIAGEN Plasmid Mini Kit (Qiagen, Valencia, CA) and digested with SacI and XbaI restriction enzymes (Fisher Scientific). The digested  $\Delta$ *intV* fragment was cloned into the suicide vector pDS132, forming pDS132 $\Delta$ *intV* and transformed into the DAP-auxotroph *E. coli*  $\beta$ 2155. *Escherichia coli*  $\beta$ 2155 pDS132 $\Delta$ *intV* served as the donor strain in conjugation with *V. cholerae* NRT36S. Transconjugants were selected for on LB agar supplemented with Sm and Cm and colony PCR with primer pair VC1758A/ NRTintv2D was used to detect the single-crossover event. Single-cross colonies were grown in the presence of 10% sucrose without Cm to detect homologous recombination and replacement of the

full *intV* gene with the truncated *intV* in *V. cholerae* NRT36S. The double cross-over event was screened for with AD and flanking primer sets and confirmed by sequencing.

#### Excision Assays.

Two-stage nested PCR assays were used to detect a unique site in the chromosome (*attB*) following excision of VPI-3 from the *V. cholerae* NRT36S chromosome. Primer pairs used for the excision assay are listed in **Table 4**. Since the excision event is rare under normal laboratory conditions, two-stages of PCR are necessary to detect the *attB* product. For round one, ten ng of *V. cholerae* NRT36S genomic was used as template for the *attB* as previously described (Carpenter, Rozovsky et al. 2016). One  $\mu$ L of round one PCR product was used as template for round two of PCR and products were analyzed on a 2% agarose gel. *Vibrio cholerae* SG7 was used as a negative control as it does not contain any PAIs and *V. cholerae* N16961 was used as a positive control as VPI-2 is inserted at the same tRNA-Serine locus and was previously detected (Carpenter, Rozovsky et al. 2016). At least two biological replicates and three technical replicates of each strain were examined.

#### Comparative Genomic and phylogenetic analysis.

The VPI-3 region of *V. cholerae* NRT36S was assembled from raw whole genome sequence data generated as previously described (Chen, Johnson et al. 2007) and obtained from Colin Stine's lab. The NRT36S contigs were aligned to the VPI-3 region found in AM-19226 (accession no. DS265226.1) and assembled in the following order (contig# (position #): contig00524 (59156 – 79436), contig01386 (1 – 28770), contig01319 (1 – 11182), contig01380 (1 – 8600), contig00036 (192-36803).

The following gaps exist: 107 bp between contig1386 and contig1319, 1,478 bp between contig01319 and contig01380 and 879 bp between contig1380 and contig00036.

The FASTA sequence of VPI-3 for each of the strains examined (*V. cholerae* N16961, AM-19226, TMA21, 12129, EM-1676A, NHCC008d, VC35, HE-25, P-18785, CP1110, 1587, and 623-39) was downloaded from either the NCBI or IMG databases. The nucleotide sequence for each strain was compared to NRT36S using the Artemis comparison tool (Rutherford, Parkhill et al. 2000). Using A33\_1669 that encodes the ATPase from *V. cholerae* strain AM-19226 as a seed the NCBI genome database was examined for homologues (Altschul, Madden et al. 1997). Sequences were obtained from either NCBI or IMG databases and uploaded into MEGA where they were aligned using the ClustalW algorithm (Thompson, Higgins et al. 1994; Higgins, Thompson et al. 1996; Tamura, Stecher G. et al. 2013). After alignment, MEGA was used to generate a neighbor-joining tree (Saitou and Nei 1987).

## Results

Recombination modules and cargo genes of VPI-3 and VPI-6.

Comparative analysis of VPI-3 from *V. cholerae* NRT36S with VPI-2 from *V. cholerae* N16961 showed significant homology in many regions (**Figure 20**). VPI-3 in NRT36S is a 68 kb region that is integrated into chromosome 1 at a tRNA-serine locus, the same site as VPI-2 in N16961. The *attL* and *attR* attachment sites of VPI-3 share 100% homology with the *att* sites of VPI-2 and the TR integrase present on VPI-3 shares 98% identity with the VPI-2 integrase. Two excisionases are also present on VPI-3 *vefA* and *vefB* and share 96% homology with those present on VPI-2. These data indicate that the recombination module of VPI-3 is homologous to VPI-2 and

should function similarly. Downstream of the integrase in VPI-3 are the genes encoding a T3SS, which share significant homology with the T3SS present in *V. cholerae* AM-19226 (**Table 5**). The T3SS appears to replace the type I restriction modification genes present in VPI-2 (**Figure 20**). The 12 kb region comprised of sialic acid transport and catabolism genes are present in both islands, share 97% sequence homology and include the sialic acid scavenging enzyme sialidase. Downstream of this region is *vefA* and just prior to the *attR* sites is *vefB* (**Figure 20**).

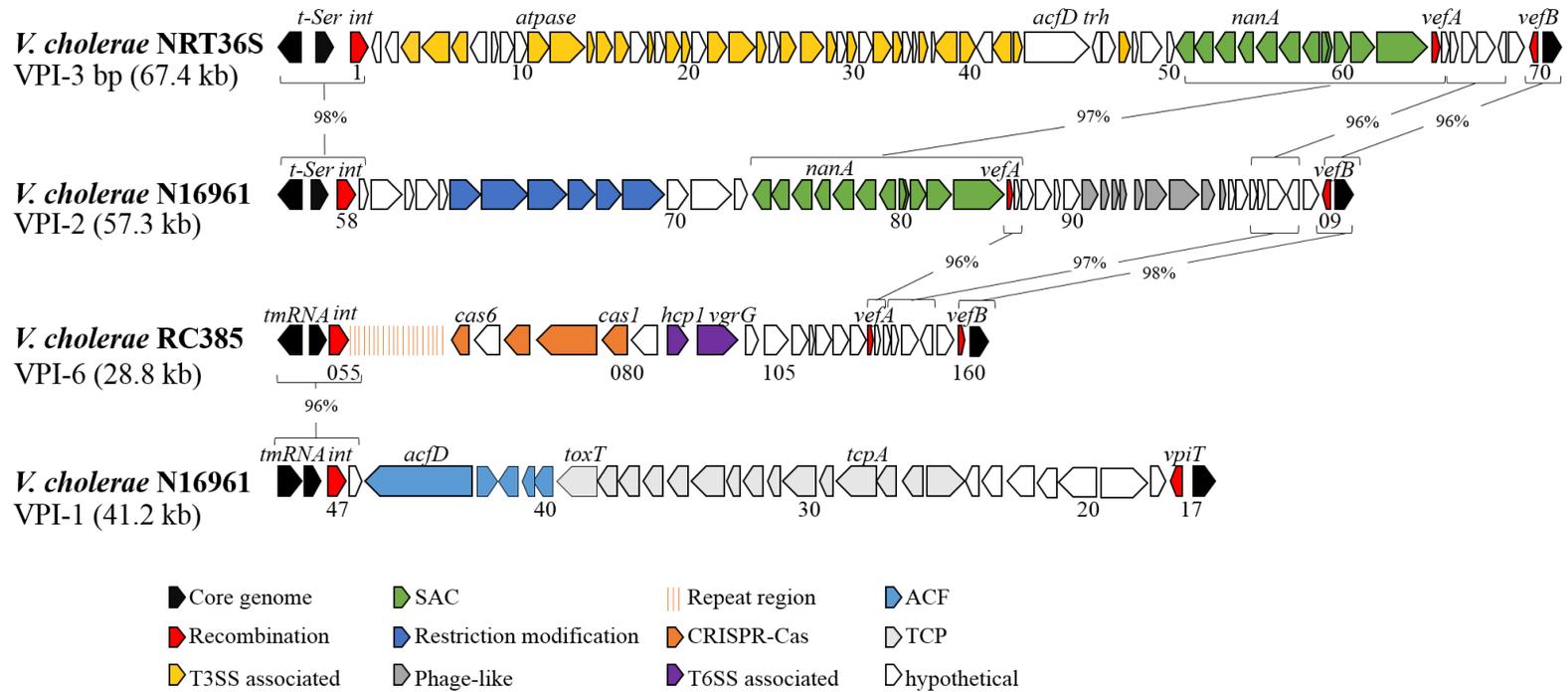


Figure 20 **Gene organization and recombination modules of VPI-3 and VPI-6.** Arrows indicate ORFs found in each corresponding island. Similarities of VPI-3 and VPI-6 to previously studied VPIs (VPI-1 and VPI-2) are shown by brackets and lines with % shared DNA homology.

The association of secretion systems with a mobile integrated genetic element is not unprecedented. In our previous study characterizing the recombination modules of VPI-1 and VPI-2, we showed that the recombination module of VPI-1 was associated with different cargo genes in different strains of *V. cholerae* and in different *Vibrio* species (Carpenter, Rozovsky et al. 2016). For example, in *V. cholerae* strain RC385 at the tmRNA locus, a type VI secretion system (T6SS) gene cluster and a type IIF Clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins (CRISPR-Cas) system is adjacent to an integrase (VCRC385\_RS05055) with 97% homology to IntV1 (VC0847) present on VPI-1 in *V. cholerae* N16961 (**Figure 20**). Further examination of this island revealed the presence of VefA and VefB homologs which share 91% (VCRC385\_RS05125) and 97% (VCRC385\_RS05160) identity to those of N16961, respectively. Additionally, the attachment sites of VPI-1 were found flanking the RC385 island with *attL* (ccccagctccacca) adjacent to *intV1* and *attR* (ccccagcccacca) adjacent to a *vefB* (**Figure 20**). Several genes between the *vefA* and *vefB* homologs in RC385 are also homologous to those found in VPI-2, however no other similarities are present to previously identified islands (**Table 6**). Thus, the T6SS CRISPR-Cas island is flanked on one end by the VPI-1 integrase and at the other end with VPI-2 excisionase *vefB*, and is flanked with *att* sites identical to VPI-1. Previously, we demonstrated that both *vefA* and *vefB* can facilitate excision of both VPI-1 and VPI-2 suggesting a similar function in RC385 (Carpenter, Rozovsky et al. 2015).

T6SSs are contact dependent secretion systems that directly inject effector proteins from the bacterial cytosol into a bacterial prey cell. They function similar to T3SSs and T4SSs but are structurally very different from these systems (Basler,

Pilhofer et al. 2012; Ho, Dong et al. 2013). The RC385 region contains *hcp* (VCRC385\_RS05090) with 97% similarity to that of Hcp1 found in N16961. Another prominent feature of T6SS, VgrG, is encoded by VCRC385\_RS05095 and shares 69% similarity with N16961. VasK has also shown to be essential for T6SS export, however no vasK homolog was identified within the region. Both *hcp1* and *vgrG* in N16961 are present on chromosome II and are not associated with a genomic island in this strain. In RC385, upstream of the T6SS genes, separating the region from the CRISPR-Cas region is VCRC385\_RS05085 encoding a hypothetical protein which shows homology to VC1772 of *V. cholerae* N16961 containing a helix-turn-helix domain presumed to be a transcriptional regulator (**Table 6**).

CRISPR-Cas systems are a bacterial immunity mechanism against phage infection that is adaptive. The repeat region is located after the integrase (VCRC385\_RS05055) and before the beginning of the *cas* genes. Based on the *cas1* (VCRC385\_RS05080) gene found in the RC385 island, this CRISPR-Cas system can be classified as subtype I-F. RC385 also contains *cas3* (VCRC385\_RS05075) and *cas6* (VCRC385\_RS05060) but lacks a *cas4* gene, common characteristics of the I-F subtype. Traditional BLAST searches of VCRC385\_RS05065 and VCRC385\_RS05070 yielded no significant homology to known proteins in the database, however an HHPred search found that VCRC385\_RS05070 was homologous to a Cas7 of the I-C subtype. To date, only a type I-E CRISPR-Cas system has been identified in *V. cholerae* in the classical O395 strain. Recently a type I-F has been described in the ICP1 phage isolated from cholera stool samples which was found to be necessary for lytic infection of *V. cholerae* strains (Seed, Lazinski et al. 2013).

VPI-3 can excise from the *V. cholerae* NRT36S bacterial genome.

It has previously been demonstrated that VPI-2 of *V. cholerae* N16961 can excise from the chromosome, the likely first step in the transfer of this island (Murphy and Boyd 2008; Almagro-Moreno, Napolitano et al. 2010; Carpenter, Rozovsky et al. 2016). The excision of a T3SS containing island has not yet been established in any species and thus, we sought to determine if VPI-3 could excise from the chromosome of NRT36S. VPI-3 excision was examined through detection of two distinct sites formed following the excision event: *attP* and *attB* (**Figure 21**). In order to detect the *attP* site, an inverse PCR primer pair was designed within the *intV* (the first gene within VPI-3) and *vefB* (the last gene within VPI-3) homologues in NRT36S such that a PCR product will only be obtained if VPI-3 excises and forms a circular intermediate (CI) (**Figure 21A**). Inverse PCR was performed on a plasmid DNA preparation from *V. cholerae* NRT36S (VPI-3) and N16961 (VPI-2), and no product was detected in this first round of PCR (**Figure 21B**). A second nested PCR assay was then carried out using the product of the first PCR reaction as template with primers designed within an internal fragment of the inverse PCR product (**Figure 21B**). A product of 461 bp was detected in NRT36S as well as in the control strain N16961, in which excision has previously been shown, indicating the presence of a VPI-3 CI and confirmation of island excision (**Figure 21B**). A two-stage nested PCR was also performed using genomic DNA to detect the empty *attB* site left in the *V. cholerae* chromosome following excision of VPI-3. An excision product of the expected size (524 bp) was observed in the second round of the *attB* assay for *V. cholerae* strain NRT36S as well as the control strain N16961 (**Figure 21C**). This data indicated that VPI-3 is capable of excision from the genome of *V. cholerae* NRT36S and this excision can be detected through both the *attP* and *attB* PCR assays.

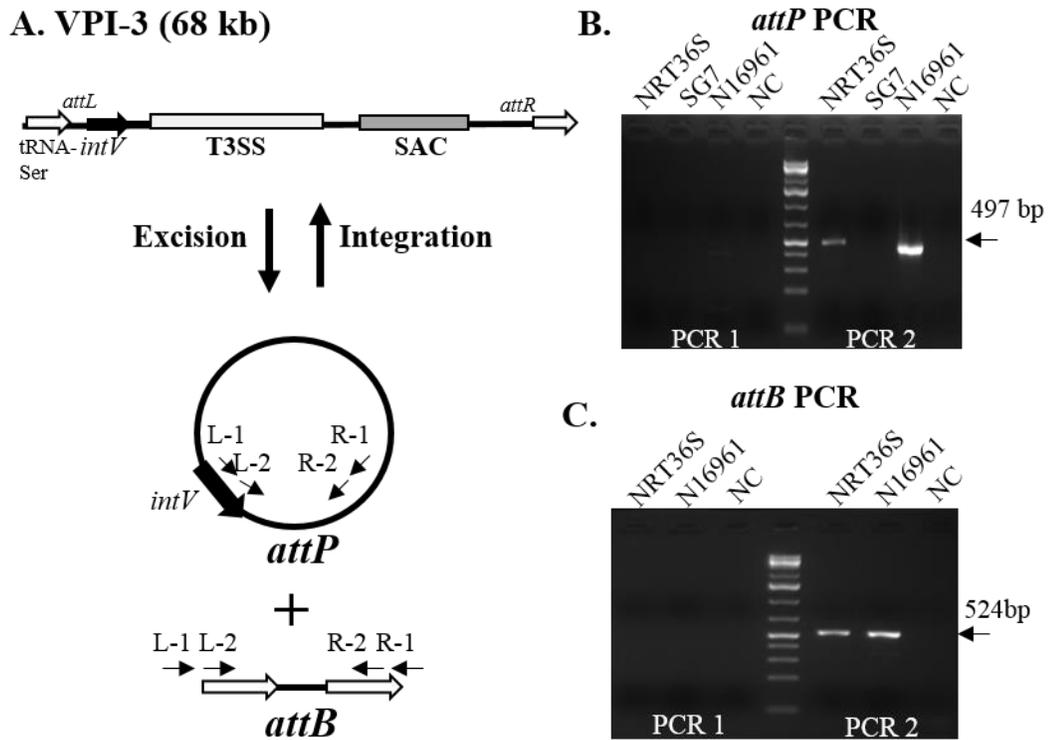


Figure 21 **VPI-3 can excise from chromosome 1 of *V. cholerae* NRT36S.**  
**A.** Primer design for PCR amplification of the *attP* and *attB* sites of *V. cholerae* N16961 and NRT36S strains. Due to the low excision rate of this island, no product from PCR1 was detected; therefore, product from PCR 1 was used as template for PCR 2 to amplify **B.** *attP* and **C.** *attB*. *Vibrio 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 *attP* and *attB* assays.

VPI-6 can excise from the RC385 *V. cholerae* genome

The excision of VPI-1 from *V. cholerae* N16961 has been demonstrated previously (Rajanna, Wang et al. 2003; Carpenter, Rozovsky et al. 2015), however the excision of a PAI containing a CRISPR-Cas system or T6SS related genes has never been demonstrated. Thus we sought to determine if the CRISPR-Cas, T6SS-containing

island VPI-6 could excise from the chromosome of *V. cholerae* RC385 using a similar approach as for VPI-3 (**Figure 21A**). Primers were designed to amplify the attP and attB sites of VPI-6 present in the CI and left in the chromosome, respectively, following site-specific excision of the island from RC385 (**Figure 22A**) and are listed in **Table 4**. The 198 bp attP PCR product was detected in round 2 of PCR in *V. cholerae* RC385, indicating VPI-6 is capable of excision from the chromosome. No product was detected in round 1 indicating the excision event is quite rare, similar to VPI-3 and other VPIs (Rajanna, Wang et al. 2003; Murphy and Boyd 2008; Carpenter, Rozovsky et al. 2015). *Vibrio cholerae* strain SG-7 was used as a negative control as it does not contain a VPI at the tmRNA locus and as expected, no excision product was observed (**Figure 22B**). An attB PCR product of 210 bp was detected in round 2 PCR, supporting the ability of VPI-6 to excise from the chromosome (**Figure 22C**). A non-specific band greater than 500 bp of the incorrect size was detected in round one of the attB PCR (**Figure 22C**). The correct empty attB site (427 bp) was detected in the negative control strain SG-7 which does not contain an island at this site (**Figure 22C**). This data indicated that VPI-6 can excise from RC385 and this excision can be detected through both the *attP* and *attB* PCR assays

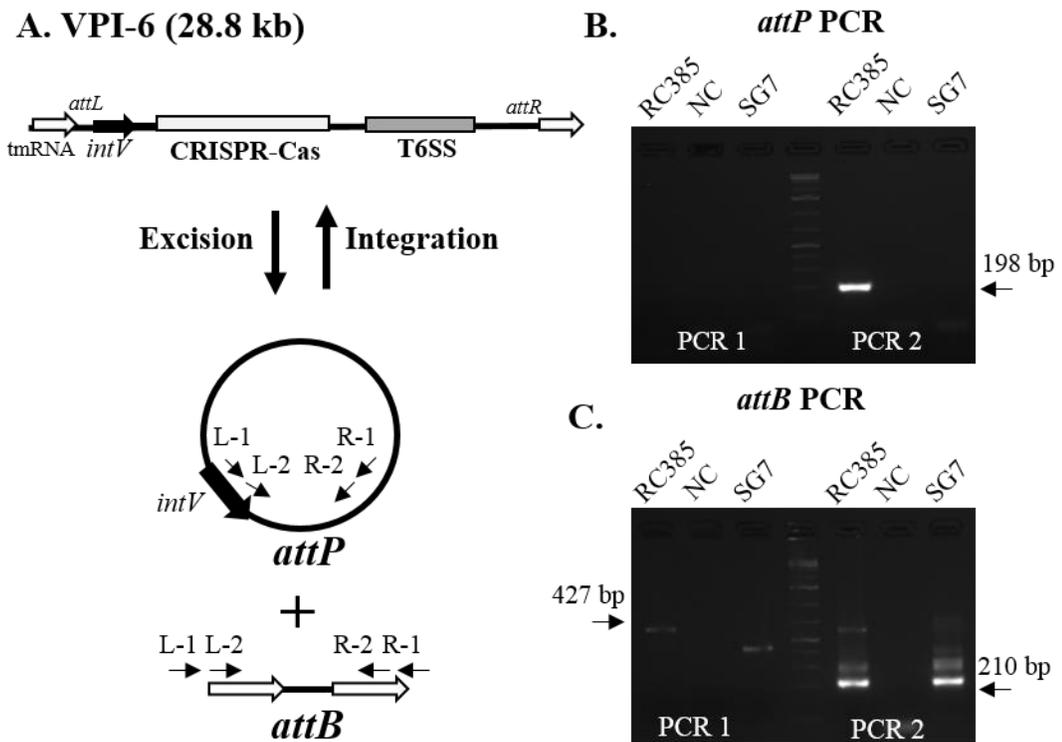


Figure 22 **VPI-6 can excise from chromosome 1 of *V. cholerae* RC385.**  
**A.** Primer design for PCR amplification of the *attP* and *attB* sites of *V. cholerae* RC385. Due to the low excision rate of this island, no product from PCR1 was detected; therefore, product from PCR 1 was used as template for PCR 2 to amplify **B.** *attP*. The **C.** *attB*. Excision product was detected in the first round of PCR. *Vibrio cholerae* SG7 was included as a negative control, as SG7 does not contain VPIs. These results demonstrated that VPI-6 of RC385 can excise from the chromosome and can be detected through our *attP* and *attB* assays.

The cognate integrase of VPI-3 is necessary for island excision.

It was previously shown that VPI-2 of *V. cholerae* N16961 excises from the core chromosome and that this excision is dependent upon the presence of the cognate integrase encoded on the island (Murphy and Boyd 2008; Almagro-Moreno, Napolitano et al. 2010; Carpenter, Rozovsky et al. 2016). In order to examine whether

the cognate integrase of NRT36S VPI-3 (*intV2*) is also necessary for the excision of VPI-3, we constructed an in-frame deletion mutation (strain MRCIntV) using SOE PCR (**Figure 23A**) and performed the *attB* excision assay to determine whether excision occurs (**Figure 23B**). No PCR product was observed in either round of the *attB* assay for strain MRCIntV, indicating that the cognate VPI-3 integrase is necessary for excision of this island (**Figure 23B**).

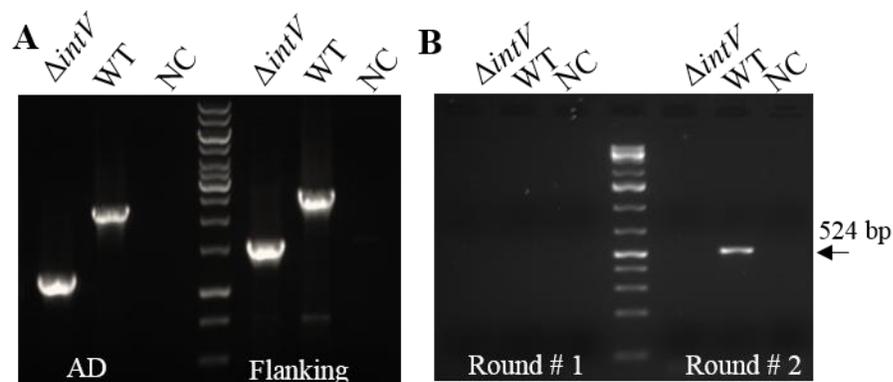
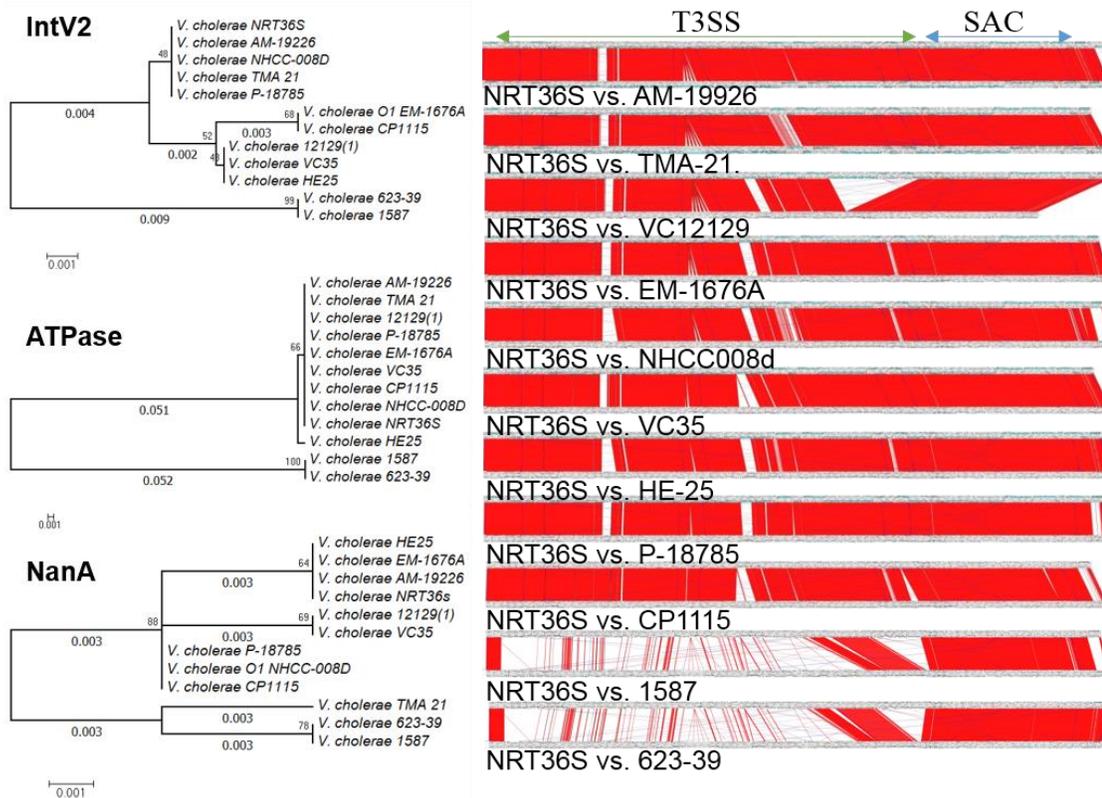


Figure 23 **The integrase (*intV2*) is essential for VPI-3 excision.**  
**A.** Confirmation of *intV2* deletion mutant in *V. cholerae* NRT36S using Flanking and AD primer sets. **B.** The two-stage *attB* PCR amplification assay was used to detect excision of VPI-3 in  $\Delta intV2$  and wild-type 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.

Comparative and phylogenetic analyses of VPI-3 and VPI-6.

VPI-3 of NRT36S was compared to 11 additional T3SS containing strains of *V. cholerae* (**Figure 24**). While the main components of the T3SS and the sialic acid metabolism region were conserved amongst 9 strains, there were 2 strains whose T3SS

region showed little homology to the others (**Figure 24**). In strains 1587 and 623-39, the T3SS is highly divergent from the other isolates examined and are unrelated. This suggests that the T3SS region was acquired independently in the two sets of strains. Interestingly, among nine isolates that contained near identical T3SS, there was one small region that showed significant sequence differences, which appears to be driven by duplication events. VopM encodes an effector protein first identified in strain AM-19226, which contains three repeat regions (Chaand, Miller et al. 2015). Variations in the number of repeats at this locus was identified among the *V. cholerae* strains examined here. In *V. cholerae* strains TMA-21 and VC12129, 4 repeats are present, 2 repeats are present in EM-1676A and one repeat in strains NHCC008d, VC35, P-18785. In *V. cholerae* strains HE-25 and CP1115 VopM was absent. ORF A33\_1690, which encodes effector protein VopW was absent in seven strains examined. Other differences with the T3SS regions were noted among *V. cholerae* strains TMA-21, NHCC008d and NRT36S. Strain VC12129 lacks a 10 kb region present in all other T3SS-containing strains, which contains homologs of A33\_1700 to A33\_1706 that includes two effector proteins VopY and VopZ as well as the virulence related proteins AcfD and TDH related hemolysin (TRH).



**Figure 24 Evolutionary relationship between T3SS-containing VPIs.**  
**A.** The evolutionary history of IntV2, ATPase and NanA of 12 *V. cholerae* strains was inferred using the Neighbor-Joining method with bootstrap values (1000 replicates) shown next to branches (Saitou and Nei 1987). The tree is drawn to scale, with branch lengths indicated. The evolutionary distances were computed using the p-distance method (Nei and L. 1989). Evolutionary analyses were conducted in MEGA6 (Tamura, Stecher G. et al. 2013) **B.** ACT was used to compare VPI-3 islands in 11 *V. cholerae* strains. Red blocks represent regions of homology and white block indicate no homology between the islands examined.

Phylogenetic trees of the integrase IntV2, T3SS ATPase, and NanA proteins of the 12 *V. cholerae* strains were used to infer the evolutionary history of these T3SS islands (**Figure 24**). Analysis of IntV2 from the 12 *V. cholerae* strains shows that they

are highly homologous to one another as indicated by the small scale bar. However, IntV2 from both 623-39 and 1587 do cluster separately from the other integrases. Similarly, the phylogenetic analysis of NanA shows that they are highly related to one another with small branch lengths indicating limited divergence. Also, the NanA protein from both 623-39 and 1587 cluster separately from the other NanA proteins (**Figure 24**). The T3SS ATPase in strains NRT36S, AM-19226, CP1110, EM-1676A, V51, TMA21, 12129, VC35, HE-25 and P-18785 are all highly homologous whereas the ATPase from strains 1587 and 623-39 cluster together separately from the other ATPases (**Figure 24**). However, it should be noted that although the ATPases appear to be conserved, the majority of the T3SS region shows no homology between these two strains and the others examined. These data suggest that the recombination module and sialic acid region are core to the island while the T3SS region was acquired separately. To further determine the evolutionary history of the T3SS ATPase from VPI-3, we examined the phylogenetic relationship between the *V. cholerae* T3SS ATPases and those from other enteric species (**Figure 25**). This data shows that the T3SS from *V. cholerae* is most closely related to T3SSs present in *V. parahaemolyticus*. Pathogenic isolates of *V. parahaemolyticus* contain two T3SSs, T3SS-1 on chromosome I that is present in all isolates and T3SS-2 on chromosome 2 that is present only in pathogenic isolates (Makino, Oshima et al. 2003). In *V. parahaemolyticus*, there are two T3SS-2 variants described depending on the strain, designated T3SS-2 $\alpha$  and T3SS-2 $\beta$  (Okada, Iida T et al. 2009). The T3SS ATPase from *V. cholerae* is highly related to T3SS-2 from *V. parahaemolyticus*, with 10 *V. cholerae* strains (two representatives displayed NRT36S and AM-19226) showing homology to T3SS-2 $\alpha$  (**Figure 25**). These ATPase proteins showed significant sequence homology

to an ATPase from an isolate of *Grimontia hollisae*, which also contains a T3SS region inserted at the tRNA-ser locus. However, no IntV2 is present but several transposases are associated with this T3SS. The two *V. cholerae* ATPases from strains 1587 and 623-39 showed homology to the T3SS-2 $\beta$  ATPase from *V. parahaemolyticus* (Okada, Iida T et al. 2009) (**Figure 25**).

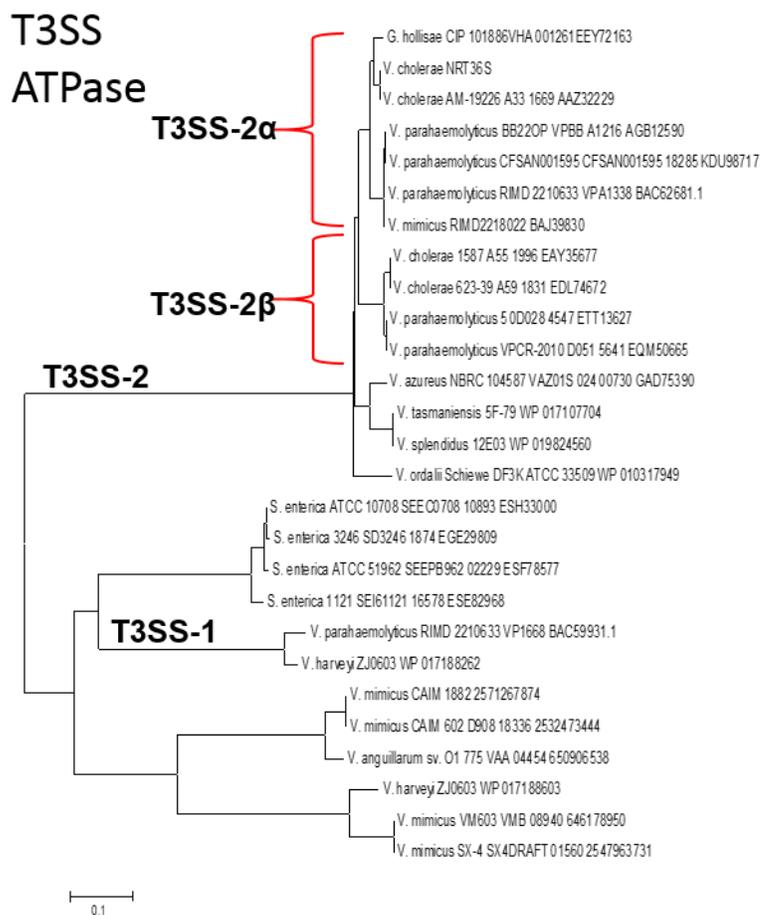


Figure 25 **T3SS ATPase phylogeny.**  
 The evolutionary history of ATPase outside of *V. cholerae* was inferred using the Neighbor-Joining method (Saitou and Nei 1987) and evolutionary analyses were conducted in MEGA6 (Tamura, Stecher G. et al. 2013). Numbers following strain names indicate IMG gene ID or protein locus tags. T3SS-1 and T3SS-2 groupings are indicated.

The 28 kb T6SS CRISPR-Cas/T6SS island is also present in *V. cholerae* RC586. Additional strains which have both CRISPR-Cas and T6SS regions clustered together at the tmRNA locus, such as TM11079-80 and HC36A1 were examined (**Figure 26**). Sequence alignments of these four strains reveal significant homology

among the tmRNA insertion site and the integrase. Part of the 5' region of *intV1* is missing from strains TM11079-80 and HC36A1, whose significance is unknown. Interestingly, these two strains also have putative transposase ORFs, not found in either RC385 or RC586.

Although the CRISPR-Cas region of HC36A1 is similar to the I-F type, little homology is observed between this CRISPR-Cas region and that of the other 3 strains. The *hcp* and *vgrG* ORFs of a putative T6SS are however highly conserved among all four strains. An additional putative transposase is found in between *cas1* and *cas3* in TM1079-80. There is also a gap in homology of the hypothetical ORF in between *cas6* and *cas7*. Although TM11079-80 lacks *vefA* and *vefB* it does contain intact *attL* and *attR* sites indicating it is capable of integrase-mediated site specific excision and could potentially utilize an RDF found elsewhere on the genome to facilitate excision. Furthermore, a partial sequence of *V. mimicus* VM603 indicating the presence of this island outside of *V. cholerae*, however the contigs have not been aligned and the complete island cannot be determined.

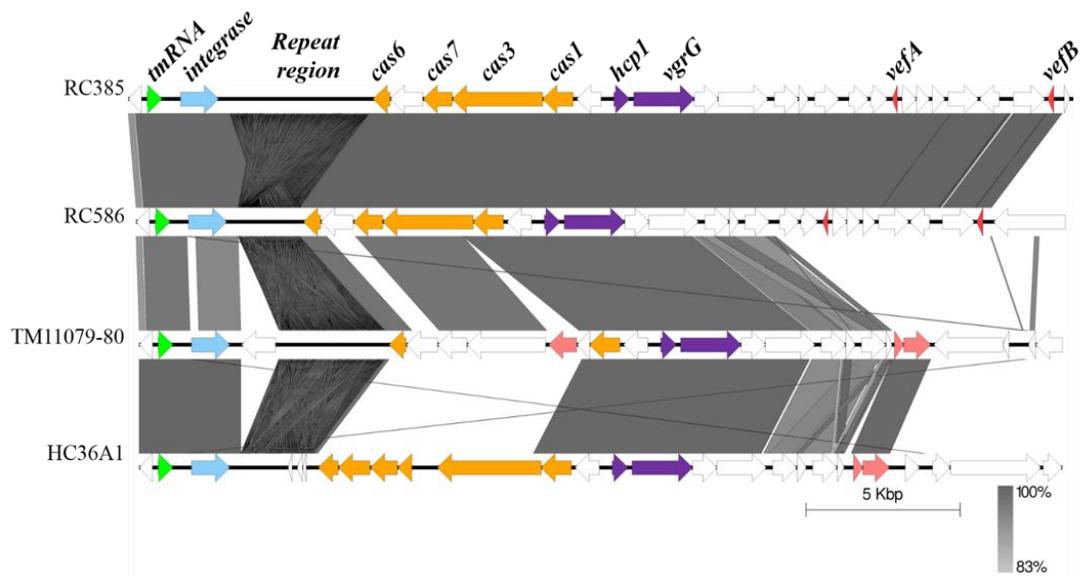


Figure 26 **Comparison of VPI-6 like islands in *V. cholerae* strains.** ORFs of each strain are represented by labeled colored arrows. Regions of DNA homology are shown via grey boxes.

## Discussion

This is the first study to show the presence of contact dependent secretion systems, T3SS and T6SS on mobile genetic elements. Both T3SSs and T6SSs were carried on pathogenicity islands that shared identical recombination modules with previously described pathogenicity islands. We demonstrated that the T3SS-containing island VPI-3 as well as the T6SS and CRISPR-Cas containing island VPI-6 can excise from the bacterial chromosome, the likely first step in the transfer and spread of these elements. Additionally, we constructed an in-frame deletion of VPI-3's cognate integrase and could no longer detect excision, indicating its requirement for excision.

Some variations in sequence homology were observed when VPI-3 of NRT36S was compared to other *V. cholerae* strains containing a T3SS-2 $\alpha$  PAI. Several of these

differences were either deletions, duplications or sequence divergence of effector proteins. For example in strains that contained T3SS-2 $\alpha$ , the effector VopW characterized in *V. cholerae* as well as *V. parahaemolyticus* was absent in seven strains (Zhou 2012; Chaand, Miller et al. 2015). In addition, *V. cholerae* VC12129 isolated from a water sample, lacked a 10 kb region that includes effectors VopY and VopZ. Thus, it appears that the T3SS-2 $\alpha$  region is evolving predominantly by deletion and insertion events. To explore the evolutionary history of these T3SS islands further we compared phylogenetic trees of different components of the 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. PAIs may be acquired and evolve as one defined unit over time, or the islands can be more fluid in nature, possessing a mosaic-like modular structure where regions are acquired at different time points (Napolitano and Boyd 2012). We compared the phylogenetic trees of the integrase, ATPase and NanA proteins of VPI-3 and found similar clustering patterns, indicating similar evolutionary paths. Although the ATPase present on T3SS-2 $\alpha$  and T3SS-2 $\beta$  share significant homology, the rest of the region is not conserved. As others have noted, of the 49 ORFs within the T3SS region, 36 show no homology between AM-19226 (T3SS-2 $\alpha$ ) and 1587 (T3SS-2 $\beta$ ) (Haley, Choi et al. 2014). These data indicate that the T3SS was acquired at a different point in the course of the evolution of VPI-3, providing evidence to support the theory that PAIs can evolve in a mosaic-like manner (Napolitano and Boyd 2012). Examination of the ATPase outside of *V. cholerae* noted the presence of an T3SS-2 $\alpha$  ATPase in a *V. mimicus* and a T3SS has been described in this species (Okada, Matsuda S et al. 2010). In addition, many strains of *V. mimicus* contain a T3SS-1-like ATPase. It will

be of interest to determine whether this T3SS-2 $\alpha$  is associated with a PAI given that *V. mimicus* is one of the closest related species to *V. cholerae*. A recent study describing another close relative of *V. cholerae*, *V. metoecus*, showed that most strains of this species contained a VPI-2 region that lacked the type 1 restriction modification system, but contained genes encoding for DNA phosphorothioation, a novel DNA modification system involved in incorporating sulfur into the DNA backbone (Orata, Kirchberger et al. 2015).

A predicted T6SS has also been identified on an integrative conjugative element (ICE) in Bacteroidales and later predicted to be found in 25% of the human colon microbiota, suggesting the horizontal transfer of this apparatus plays a crucial role in the evolution of this community (Coyne, Zitomersky et al. 2014; Coyne, Roelofs et al. 2016).

The broader phylogenetic analysis of the T3SS ATPase indicates that the T3SS region has been horizontally transferred among species and the data presented in this study suggests a mechanism by which this can occur. In addition, we have also presented evidence that a T6SS and a CRISPR-Cas are present on a pathogenicity island suggesting a possible mechanism of acquisition. Horizontal transfer of T6SSs has been suggested previously in *V. parahaemolyticus* (Salomon, Klimko et al. 2015). The data presented here show that pathogenicity islands are playing an important role in the transfer of contact dependent secretion systems among bacteria.

Table 3 Bacterial strains and plasmids used in this study

Strain or Plasmid	Description	Source
<b><i>V. cholerae</i> strains</b>		
NRT36S	O31, VPI-3, partial VSP-II, Sm <sup>R</sup>	
N16961	O1, El Tor strain, VPI-1, VPI-2, VSP-I, VSP-II, Sm <sup>R</sup>	
RC385	O135, VPI-6	
MRCIntV	NRT36S, $\Delta intV2$ , Sm <sup>R</sup>	(Taviani, Grim et al. 2010)
<b><i>E. coli</i> strains</b>		
DH5 $\alpha$ pir		This study
$\beta$ 2155 $\lambda$ pir	Donor for bacterial conjugation, DAP auxotroph	Lab collection
<b>Plasmids</b>		
pJET1.2	cloning vector, Am <sup>R</sup>	Lab collection
pJET1.2 $\Delta intV2$	pJET1.2 harbouring truncated intV2 gene	Fermentas
pDS132	Suicide Vector, SacB, Cm <sup>R</sup>	This study
pDS132 $\Delta intV2$	pDS132 harboring truncated <i>intV2</i> gene	Lab collection
		This study

Table 4 Primers used in this study

<b>Primer Name</b>	<b>Sequence (5' → 3')</b>	<b>Product size (bp)</b>
<b>Mutant Construction</b>		
VC1758FF	AAGCAAACGCACTCAATGCG	
NRTintV2FR	GGATTTCTGCCTACTACCGT	2,269
VC1785A	tctagaGATTCGGTGAGTTGTCCGAG	
NRTintV2B	CATGAGCGAGAATTACTTGG <i>CCAAGTAATTCTCGCTCATG</i>	523
NRTintV2C	AAGCTACAGTGTCTGCTGGTG	
NRTintV2D	gagctcGCCTGTGGCGAGATAGAGAC	678
<b>Excision Assays</b>		
<b>VPI-3</b>		
VPI-2fk1	CTGGCTATGAGCTGATTTG	
VPI-2fk2	AGGGATTGGCTTTGAGG	1,449
VPI-2attF	AGAGTGAAAGTCGCCAAAG	
VPI-2attR	GGTGCAATTTTCGCATGTTGC	524
VPI-2 InvVC1808F	AGCTAGACAGATTAGCTAACC	
CnVC1758B	TTGCCATGAGCGAGAATTGC	1,352
NestVC1758comR	AGAATTGCTTGGACGTACGC	
NestVC1809comF	GCGTAACTGAGAAAGTGTG	461
<b>VPI-6</b>		
VPI-1 attBR1	TGTAAGACGGGGAAATCAGG	
RC385attBF1	TTTGTTGATGAGCAGGATGG	427
VPI-1 attBR2	ATTCGTTAGCGTGTCCG	
RC385attBF2	AGTGAATCTTGATGAGACGC	210
NestVC0847F3	TTTCTCTCTAGGTTTGGAGG	
RC385attPR1	CTACTGCTTTATCAGGACCC	413
RC385attPF2	GCTGCTATGGAATCTTGTGG	
RC385attPR2	GGTACACTAAAAGGTACACC	198

Table 5 ORFs in *V. cholerae* NRT36S VPI-3

	Gene Name	Length (aa)	AM-19226 Homolog	N16961 Homolog	Other Homolog
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) <sup>1</sup>
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)
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)		

Table 5 (Continued)

	Gene Name	Length (aa)	AM-19226 Homolog	N16961 Homolog	Other Homolog
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) <sup>1</sup>		
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) <sup>2</sup>
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-acetylneuraminate	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)	
58	Hypothetical protein	132		VC_1780 (97) <sup>2</sup>	
59	N-acetylmannosamine-6- phosphate 2-epimerase	240	A33_1714 (100)	VC_1781 (99)	
60	N-acetylmannosamine kinase	287	A33_1715 (100)	VC_1782 (100)	
61	N-acetylglucosamine-6- phosphate deacetylase	378	A33_1716 (100)	VC_1783 (100)	
62	Neuraminidase	803	A33_1717 (99)	VC_1784 (99)	

<sup>1</sup> 85% Query Cover<sup>2</sup> 73% Query Cover

Table 5 (Continued)

	<b>Gene Name</b>	<b>Length (aa)</b>	<b>AM-19226 Homolog</b>	<b>N16961 Homolog</b>	<b>Other Homolog</b>
<b>63</b>	VefA	68	A33_1718 (100)	VC_1785 (100)	
<b>64</b>	RadC	158	A33_1719 (100)	VC_1786 (100)	
<b>65</b>	Hypothetical protein	77	A33_1720 (100)		
<b>66</b>	Hypothetical protein	148	A33_1721 (100)		
<b>67</b>	GTPase, hypothetical protein	328	A33_1722 (100)		
<b>68</b>	Conserved hypothetical protein	213	A33_1723 (100)		
<b>69</b>	CopG family transcriptional regulator	340	A33_1724 (99)		
<b>70</b>	VefB	76	A33_1725 (100)	VC_1809 (100)	

Table 6 ORFs in *V. cholerae* RC385 VPI-6 and predicted gene products based on BLAST homology search.  
 Note: a Type I-F CRISPR-Cas system was experimentally verified in *Pectobacterium atrosepticum* SCRI1043.

RC385 Locus Tag	Size (aa)	Gene product	Organism	Homolog	ID (%)
VCRC385_RS05055	412	Integrase (IntV1)	<i>V. cholerae</i> N16961	NP_230494	97
VCRC385_RS05060	183	Type I-F Cas6 endoribonuclease	<i>P. atrosepticum</i> SCRI1043	CAG76582	38
VCRC385_RS05065	351	Hp			
VCRC385_RS05070	315	Type I-C Cas7 *			19
VCRC385_RS05075	978	Type I-F Cas3	<i>P. atrosepticum</i> SCRI1043	CAG76578	37
VCRC385_RS05080	329	Type I-F Cas1	<i>P. atrosepticum</i> SCRI1043	CAG76577	58
VCRC385_RS05085	264	Hp (VC1772)	<i>V. cholerae</i> N16961	NP_231407	38
VCRC385_RS05090	172	Hcp1	<i>V. cholerae</i> N16961	NP_232418	97
VCRC385_RS05095	658	VgrG	<i>V. cholerae</i> N16961	NP_232419	69
VCRC385_RS05100	262	Hp			
VCRC385_RS05105	536	Hp			
VCRC385_RS0318635	262	Hp			
VCRC385_RS05110	99	Hp (VCA0105)	<i>V. cholerae</i> N16961	NP_232506	53
VCRC385_RS0318630	262	Hp			
VCRC385_RS05115	209	Hp			
VCRC385_RS05120	179	Hp			
VCRC385_RS05125	58	VefA	<i>V. cholerae</i> N16961	NP_231420	91
VCRC385_RS05130	158	RadC	<i>V. cholerae</i> N16961	NP_231421	99
VCRC385_RS05135	142	Hp (VC1804)	<i>V. cholerae</i> N16961	NP_231439	91
VCRC385_RS05140	136	Hp (VC1805)	<i>V. cholerae</i> N16961	NP_231440	98
VCRC385_RS05145	328	GTPase	<i>V. cholerae</i> N16961	NP_231441	97
VCRC385_RS05150	213	Hp (VC0494)	<i>V. cholerae</i> N16961	NP_230148	31
VCRC385_RS05155	344	Hp (VC1808)	<i>V. cholerae</i> N16961	NP_231442	30
VCRC385_RS05160	66	VefB	<i>V. cholerae</i> N16961	NP_231443	97

## Chapter 4

### QUORUM SENSING REGULATION OF EXCISION OF THE FILAMENTOUS PHAGE F237 IN VIBRIO PARAHAEMOLYTICUS

The work presented in this Chapter was completed in collaboration with graduate student researcher Joe Borowski.

#### Introduction

Horizontal gene transfer plays an integral role in the evolution of bacteria, allowing bacteria to obtain new traits and phenotypes in a single event. The acquisition of mobile integrated genetic elements (MIGEs) such as plasmids, bacteriophages, pathogenicity islands (PAIs), integrons, and integrated conjugative transposons (ICEs) are some of the key vectors of HGT. These MIGEs can make up a significant portion of the bacterial genome. For example *Escherichia coli* O157:H7 contains nearly 1 Mb more DNA than K12, with much of this extra DNA comprised of bacteriophages and PAIs.

Filamentous phages represent 5% of known phages and have contributed significantly to the evolution of bacteria (Mai-Prochnow, Hui et al. 2015). Filamentous phage belong to the genus *Inovirus*, contain a single-stranded circular DNA genome and are surrounded by a long thin filament of coat protein (Russel 1995). Filamentous phage genomes range in size from 4 to 12 kb and dictate the size of the phage particle (Rakonjac, Bennett et al. 2011). They are predominantly found in Gram-negative bacteria. The most widely studied filamentous phages are the F-pilus phages (Ff) from *Escherichia coli*, first isolated in the 1960s which include: M13, f1

and fd, none of which integrate into the host chromosome (Loeb 1960; Hoffmann-Berling, Marvin et al. 1963; Hofschneider and Preuss 1963). The next most extensively studied filamentous phage is the CTX $\phi$  from *Vibrio cholerae*, which encodes the genes necessary for production of cholera toxin, the causative agent of the secretory diarrheal disease cholera (Waldor and Mekalanos 1996). When originally described CTX $\phi$  was unusual in that it was integrated into the host chromosome. Filamentous phage can assemble and release phage particles without harming their bacterial hosts. Many filamentous phage encode their own secretion protein, however in the CTX $\phi$  genome the secretion gene appears to have been replaced by the *ctxAB* genes that encode cholera toxin (**Figure 27**). CTX $\phi$  uses the host cell type II secretion system to release phage particles (Waldor and Mekalanos 1996; Davis, Lawson et al. 2000). Some filamentous phages can produce turbid plaques on bacterial lawns that are attributed to large phage productivity slowing host cell growth (Rakonjac, Bennett et al. 2011). Filamentous phages have been found to replicate either exclusively episomally (as does the M13 and Ff) or exist as a temperate chromosomally integrated phage which may be induced to start episomal replication (as does CTX $\Phi$ ) (Waldor, Rubin et al. 1997; Rakonjac, Bennett et al. 2011; Mai-Prochnow, Hui et al. 2015) .

CTX $\Phi$  has a 7-kb genome and cholera toxin is the only known toxin to be present on a filamentous phage. Filamentous phage genomes have functionally distinct modules involved in replication, structure and assembly of the phage. CTX $\Phi$  has a DNA replication module named RS2 that encodes three genes, *rstR*, *rstA*, and *rstB*, and a core module that contains *psh*, *cep*, *orfU* (*gIII*), *ace*, *zot*, and *ctxAB* (**Figure 27**) (Waldor and Mekalanos 1996). The RstA and RstB proteins are required for CTX $\Phi$  DNA replication, and RstR is a repressor protein (Waldor, Rubin et al. 1997). The

core genes *cep*, *orfU*, *ace*, and *zot* correspond to gVIII, gIII, gVI, and gI of phage M13 (**Figure 27**) (Waldor and Mekalanos 1996). The Psh, Cep, OrfU, and Ace proteins are phage structural proteins and Zot is an assembly protein (Waldor and Mekalanos 1996). The *ctxAB* genes are unique to CTX $\Phi$  but are not essential for phage production, and replace gIV from M13 a secretin protein, required for phage secretion from the bacterial cell. CTX $\phi$  integrates at the *dif* site on the host chromosomes via host XerCD recombinases required for resolution of dimeric chromosomes during replication (Huber and Waldor 2002; McLeod and Waldor 2004). Integration of CTX $\phi$  into the *V. cholerae* genome is usually irreversible and CTX prophages are vertically transmitted along with the rest of the genome. However, CTX prophage can act as template for the production of ssDNA phage, and this process requires the presence of either a tandem array of CTX prophages or a CTX-RS1 array (Davis, Moyer et al. 2000; Davis and Waldor 2000; Moyer, Kimsey et al. 2001). In addition, CTX prophage can occasionally produce virions; treatment of CTX lysogens with mitomycin C or UV light induces the SOS response and results in virion production. The SOS induction of CTX $\Phi$  production required RecA-stimulated autocleavage of LexA, a global regulator in *E. coli* (Waldor and Mekalanos 1996; Quinones, Kimsey et al. 2005; Quinones, Kimsey et al. 2005).

The filamentous phage f237 examined here is found integrated into the *dif* site of the *V. parahaemolyticus* RIMD2210633 chromosome, likely through the same XerCD mechanism described for CTX $\phi$  (Nasu, Iida et al. 2000; Iida, Hattori et al. 2001; Iida, Makino et al. 2002). Phage f237 is found predominantly in epidemic and pandemic *V. parahaemolyticus* strains a leading cause of seafood-related gastroenteritis in the United States. Originally, f237 was a marker for the pandemic *V.*

*parahaemolyticus* serogroup O3:K6 clone that emerged in Kolkata, India in 1996 and was found in Europe, the United States, Mexico and South American countries by 2006, marking the start of the first *V. parahaemolyticus* pandemic (Nair, Ramamurthy et al. 2007). The f237 filamentous phage was first isolated by Nasu et al. from a patient with traveler's diarrhea in Osaka, Japan and shown to form plaques on an indicator strain (Nasu, Iida et al. 2000). The f237 genome is 8.6 kb, contains 10 ORFs and can exist in a circular form f237 (pO3K6) that was suggested to be replicative (Nasu, Iida et al. 2000). Several genes show similarity to the CTX $\Phi$  (Waldor and Mekalanos 1996) and another *V. parahaemolyticus* phage Vf33 (Chang, Taniguchi et al. 1998) in both sequence homology and genetic structure (Nasu, Iida et al. 2000). However, one ORF named ORF8 did not show homology, either through sequence similarity or gene location, to other phage genes. This ORF was determined to be distinct to the f237 genome, found exclusively in pandemic O3:K6 or near genetically identical O4:K6 and O1:KUT *V. parahaemolyticus* strains and speculated to confer an advantageous phenotype to these strains (Nasu, Iida et al. 2000; Iida, Hattori et al. 2001).

Similar to other MIGEs, once filamentous phages are incorporated into the host chromosome, their gene expression must be tightly controlled by the host in order not to disrupt the normal physiology of the bacterial cell. The bacterial histone-like nucleoid structuring protein (H-NS) is a global regulator that controls the expression of 100s of host genes and has been identified as a general silencer of AT-rich horizontally acquired DNA such as PAIs in several enteric pathogens such as *Salmonella enterica*, *Escherichia coli* and *Pseudomonas aeruginosa* (Navarre, Porwollik et al. 2006; Oshima, Ishikawa et al. 2006; Navarre, McClelland et al. 2007;

Castang, McManus et al. 2008; Fang and Rimsky 2008). Activation of the SOS response has also been implicated in the dissemination of lambdoid-like bacteriophages and ICEs demonstrating that LexA plays an important role in repressing gene expression in these MIGEs (Eisen, Brachet et al. 1970), (Maiques, Ubeda et al. 2006), (Beaber, Hochhut et al. 2004). However, for the vast majority of MIGEs, no host regulatory control mechanism has been established. In this study, we re-examined the genome of phage f237 and demonstrate that in addition to forming circular intermediates (CIs), f237 can site-specifically excise from the *V. parahaemolyticus* genome. We demonstrate that production of CI as well as phage gene expression is increased in a *luxO* mutant strain, in which OpaR, the quorum sensing high cell density master regulator, is constitutively active. We also show OpaR can bind to promoter regions in the f237 phage genome suggesting its role as a positive regulator of phage production. This is the first study directly specifically linking a quorum sensing regulator to phage gene regulation.

## Materials and Methods

### Bacterial strains used and growth conditions.

Bacterial strains and plasmids used in this study are listed in **Table 7**. The clinical *Vibrio parahaemolyticus* strain RIMD2210633 was used as wild-type as well as the background strain for mutant construction. All bacterial strains were grown aerobically (225 rpm) at 37°C in either lysogeny broth (LB) medium (Fisher Scientific, Fair Lawn, NJ), adjusted to 3% NaCl for *V. parahaemolyticus* strains, or M9 minimal medium (Sigma) supplemented with 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 20 mM glucose (M9G medium), unless stated otherwise. The following antibiotics

were used: ampicillin (100 µg/ml), chloramphenicol (Cm) (25 µg/ml) and streptomycin (Sm) (200 µg/ml).

#### Characterization of the f237 genome and its attachment *att* sites

Since the genes which encode filamentous phage proteins are not well conserved, structural modules of f237 were predicted based on size, location and membrane topology of previously described filamentous phages (Mai-Prochnow, Hui et al. 2015). *Escherichia coli* Ff (M13) and *V. cholerae* CTX genetic module comparisons were made and used in this identification (**Fig. 27**). A 12 bp *dif* (*att*) site (cataatgcgcac) was found in identical copies at either end of the integrated f237 phage – one copy being present within ORF10 (VP1549) and the other after ORF9 (VP1562).

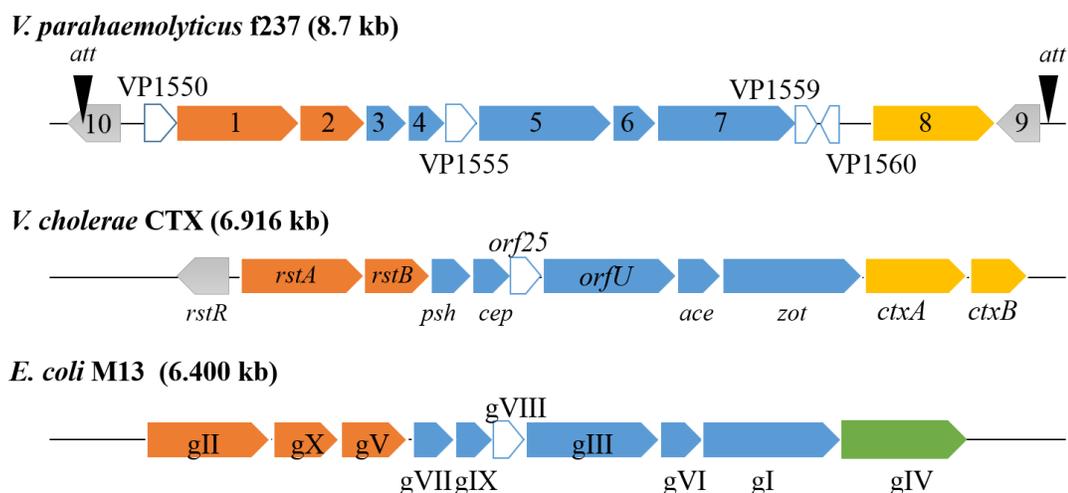


Figure 27 **f237 and other well-known filamentous phages.** Comparison of the genetic structure of f237 phage from *V. parahaemolyticus* to well-known filamentous phage CTX from *V. cholerae* and M13 from *E. coli*. Orange (replication), blue (structure and assembly), yellow (accessory/unknown) and green (secretion arrows represent ORFs).

#### Excision assays for f237 (Detection of replicative form of f237)

In order to observe production of circular intermediates (CIs) and excision of the f237 phage from the *V. parahaemolyticus* genome, *attP* and *attB* were detected through PCR amplification, respectively, in wild-type and a *luxO* deletion mutant strains. Primers were designed based on the *V. parahaemolyticus* RIMD2210633 genome and are listed in **Table 8**. Genomic DNA and plasmid DNA was isolated from overnight cultures grown in LB at 37°C shaking at 225 rpm using the NucleoSpin® Tissue Kit (Macherey-Nagel, Bethlehem, PA) and NucleoSpin® Plasmid DNA Purification Kit (Macherey-Nagel), respectively, following manufacturer's procedure. A 10 ng aliquot of genomic or plasmid DNA was used as template for *attB* and *attP* PCR assays. PCR was performed in 20 µl reactions with Taq DNA Polymerase

(Denville, Holliston, MA) under the following conditions: 95°C for 5 min followed by 29 cycles of 94°C 30 s, 64°C 30 s, 72°C 30 s or 45 s and afterwards 7 min at 72°C. PCR was also performed using primer pairs attBFwd/attPRev and attPFwd/attBRev for detection of the flanking sites at each end of the integrated phage, *attL* and *attR*, respectively. For *attL* and *attR* the conditions were: 94°C 30 s, 62°C 30 s, 72°C 30 s or 45 s and afterwards 7 min at 72°C. PCR products were analyzed on a 1% agarose gel and stained with ethidium bromide for visualization under UV light. Gel electrophoresis was done for 30 min at 100 volts.

#### RNA isolation and cDNA synthesis

Bacterial cells were grown in LB at 37°C shaking at 225 rpm for 12 hours and adjusted to OD<sub>600</sub> of 1.0 using. RNA was isolated using TRIzol® Reagent (Invitrogen, Carlsbad, CA) following the manufacturer's recommended protocol in a sterile RNA-free work station and subsequently treated with TURBO DNase (Invitrogen) following manufacturer's procedure. The DNase-treated RNA (500 ng) was used to synthesize cDNA using SuperScript® III reverse transcriptase (Invitrogen) following the manufacturer's protocol.

#### Real-time quantitative PCR

To determine the expression levels of f237 phage genes in multiple strains, a 1:25 dilution of cDNA samples were used as template for real-time quantitative PCR with primer pairs for the following genes VP1551, VP1552, VP1558, VP1561, and 16S rRNA (**Table 8**). Quantification of the *attP* site associated with production of phage circular intermediates was detected using the primer pair attPf237QFwd/attPf237QRev and 20 ng of plasmid DNA was used as template.

Quantification of the *attB* site associated with phage excision was detected using the primer pair attBf237QFwd/attBf237QRev and 20 ng of genomic DNA was used as template.

The qPCRs were performed using Fast SYBR green master mix (Applied Biosystems, Carlsbad, CA) on an Applied Biosystems 7500 fast real-time PCR system using the following cycling conditions: holding stage at 95°C for 25 s and cycling stage at 95°C for 3 s and 60°C for 30s. At least two biological replicates and four technical replicates were performed. The gene expression and *att* quantification data were analyzed using the Applied Biosystems 7500 software and relative gene expression or relative *attP/attB* detection was determined using the  $\Delta\Delta C_T$  method (Livak and Schmittgen 2001).

#### Protein purification of OpaR.

*Vibrio parahaemolyticus* RIMD 2210633 OpaR was cloned into the pProExHta expression plasmid in which a 6x His tag is fused to OpaR, separated by a Tobacco Etch Virus (TEV) protease cleavage site. The primer pair SfoIVP2516Fwd/SacIVP2516Rev (**Table 8**) was used to amplify *opaR* (VP2516) from the *V. parahaemolyticus* RIMD genome using Accura HiFidelity Polymerase (Lucigen, Middleton, WI) following manufacturer's instruction. The *opaR* PCR product was gel cut purified using the Nucleospin Gel and PCR cleanup kit (Macherey-Nagel) and cloned into pJET1.2 using the blunt end ligation protocol. The subsequent pJETOpaR was transformed into *E. coli* Dh5 $\alpha$  using standard CaCl<sub>2</sub> transformation protocol. Plasmid DNA was isolated from colonies containing proper inserts and subject to restriction digestion with EheI (SfoI isochizomer) and XbaI and subsequently ligated into EheI/XbaI cut pProExHta. The ligation product was

transformed into *E. coli* DH5 $\alpha$  and colonies were selected for the proper plasmid by plating on LB + Amp and colony PCR. Plasmid DNA was isolated and confirmed by sequencing before being transformed into *E. coli* BL21(DE3) using standard CaCl<sub>2</sub> method.

The pProExHtaOpaR plasmid was expressed in *E. coli* BL21(DE3). Ten milliliters of overnight culture was inoculated into 1 L Luria Broth (LB) at 37°C and induced with 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) at OD<sub>600</sub> 0.5. Growth continued overnight at 18°C. Cells were harvested by centrifugation (5,000 xg for 20 min at 4°C) and were re-suspended in IMAC Wash Buffer (50 mM sodium phosphate, 200 mM NaCl, 20 mM imidazole pH 7.6) supplemented with 1 mM phenylmethanesulfonyl fluoride (PMSF) and 1 mM benzimidazole. Bacterial cells were lysed on ice using a high-pressure homogenizer (EmulsiFlex-C5, Avestin, Ottawa, Canada). Cell debris was removed by centrifugation (15,000 xg for 1 h at 4°C). The supernatant was passed through a column containing 5 mL nickel resin (New England BioLabs). The column was washed with 10 column volumes (CV) of IMAC Wash Buffer. The fusion protein, 6xHis-OpaR was eluted with 3 CV IMAC Elution Buffer (50 mM sodium phosphate, 200 mM NaCl, 500 mM imidazole, pH 7.6).

A hexahistidine-tagged TEV protease was added to the eluent in a 1:10 molar ratio (TEV: 6xHis-OpaR) and the cleavage reaction proceeded overnight at 4°C. The cleavage mixture was centrifuged, adjusted to 20 mM imidazole and subject to immobilized metal affinity chromatography (IMAC) using nickel resin to remove the His-tagged TEV and any remaining un-cleaved fusion protein. The flow through and one CV of wash with IMAC Wash Buffer contained OpaR. Protein molecular weight

was confirmed by mass spectrometry and its purity was determined to be higher than 95% by SDS-PAGE.

### Electrophoretic Mobility Shift Assays

DNA fragments *attR* and *attL* were amplified using primer sets attR1Fwd/attR1Rev and attL2Fwd/attL2Rev, respectively (**Table 8**). Varying concentrations of purified OpaR were incubated with 30 ng of target DNA in binding buffer (10 mM Tris, 150 mM KCl, 0.1 mM dithiothreitol, 0.1 mM EDTA, 5% PEG, pH 7.4) for 20 min at room temperature and 10  $\mu$ L were loaded onto a pre-run (200 V for 2 h at 4°C) 8% native acrylamide gel. The gel was run at 200 V for 2.5 h in 1x Tris-acetate-EDTA (TAE) buffer at 4°C. Following electrophoresis, gels were stained in an ethidium bromide bath (0.5  $\mu$ g/ml) for 20 min, washed with water and imaged.

## Results

### Genome organization of phage f237

The f237 prophage is inserted in the *V. parahaemolyticus* RIMD2210633 genome on chromosome 1 at the highly conserved *dif* site, which is the site XerCD recombinase acts on in host chromosome dimer resolution. The f237 phage genome is 8,796 bp and encodes 14 ORFs, of which many we could predict putative functions based on sequence homology, size and location within the phage (**Table 9**). The 14 ORFs encompass VP1549 to VP1562 and ORFs range in size from 38 amino acid to 504 amino acids. Since filamentous phage genes share little similarity at the nucleotide or amino acid level, homology searches have limited success in identifying function of these phage genes, instead their function can be predicted based on their relative location, size and secondary structure. The f237 genome contains all the core putative

genes necessary for infection of host, replication and assembly of the filamentous phage (described in the review Mai-Prochnow et al.).

The f237 filamentous phage genome was compared to the well-studied filamentous phages *E. coli* M13 and *V. cholerae* CTX $\Phi$  genomes (**Figure 27**). M13 strictly replicates episomally upon cell infection, however CTX, like f237, is an integrated filamentous phage and contains an encoded transcriptional regulator, RstR, involved in repression of *rstA*, which encodes for a replication protein. Two putative transcriptional regulators are found at either end of f237, suggesting a more complex transcriptional regulatory mechanism. Unpublished RNA-seq data of *V. parahaemolyticus* RIMD2210633 cells grown in M9 plus glucose as the sole carbon source, indicates that ORF VP1550 is the first ORF in a 10 ORF operon. ORFs VP1551 and VP1552 show homology to *rstA* and *rstB* found in CTX $\Phi$ , which represent pII and pV of the M13 filamentous phage classification, essential for rolling circle phage replication and binding to ssDNA during replication, respectively (**Figure 27**). The putative minor virion proteins are encoded by ORFs VP1553 and VP1555 (gVII and gIX) and are found at one end of the filament involved in the viral export process. VP1554 is the putative major coat protein (gVIII). A second set of putative minor coat proteins are encoded by VP1556 and VP1557 (gVI and gI) which are involved in entry by binding the host cell receptor during infection in addition to liberation from the host post-assembly. VP1558 is putatively classified as gI, was shown to participate in the assembly/secretion process with gIV, a member of the secretin family, which encodes for production of a gated channel in the outer membrane, allowing phage particles to exit (Feng, Model et al. 1999). No gIV homolog was identified in f237. Although VP1561 (*orf8*) is similar in size and is in a

suggestive location to be predicted to be gIV, but it does not contain any putative internal helices or outer membrane motifs. The psortB (bacterial protein subcellular localization prediction) tool rated Orf8 with a cytoplasmic score of 8.96, well above the threshold of 7.5 (Yu, Wagner et al. 2010). Both the *E. coli* M13 gIV and *Pseudomonas* phage Pf3 gIV received outer membrane scores of 10 and 9.93, respectively. Several filamentous phages have lost gIV, and resort to using host cell processes for phage secretion such as the CTX phage which utilizes the host type II secretion system as mentioned previously. It is interesting that chromosomally integrated phage such as f237 and CTX appear to be better adapted to their hosts through the presence of transcriptional regulators and loss of their own secretion machinery, while strict episomal phages such as M13 lack transcriptional regulators and are not regulated by hosts but have retained their own secretion machinery.

The *attL* and *attR* sites at either end of the integrated phage are identical (cataatgcgac). The *attL* site is within ORF10 (VP1549), a putative transcriptional regulator, 25 bp from the translational stop sequence. Another *att* sequence has also been reported within a putative transcriptional regulator in VGJ phage (Campos, Martinez et al. 2003). The *attR* sequence is located in an intergenic region 80 bp downstream of the start codon of ORF9 (VP1562), another putative transcriptional regulator, and upstream of the host core protein VP1563.

The f237 filamentous phage contains a GC content of 45%, identical to that of chromosome I of *V. parahaemolyticus*. The genomic region of VP1550 – VP1559 is 48% GC and VP1560 to VP1562 is 40% GC, suggesting these modules may have been separately acquired.

### Detection of the replicative form of f237

Filamentous phages have previously been shown to replicate off the chromosome in a mechanism similar to rolling circle replication, resulting in formation of a circular intermediate which contains an *attP* site (**Figure 28**). Nasu and colleagues detected CI of f237 after plasmid isolation. To confirm that f237 could be detected in the circular intermediate form, PCR primers were designed within VP1549 and VP1562 (f237attPFwd and f237attPRev) to amplify the *attP* site found in the phage circular intermediate of the wild-type *V. parahaemolyticus* strain (**Figure 28B**). An *attP* PCR product was obtained using plasmid DNA as template, confirming the presence of a circular intermediate (**Figure 28B**). The *attP* PCR product was sequenced and confirmed the predicted 12 bp *att* site (cataatgcgcac) (**Figure 28B**).

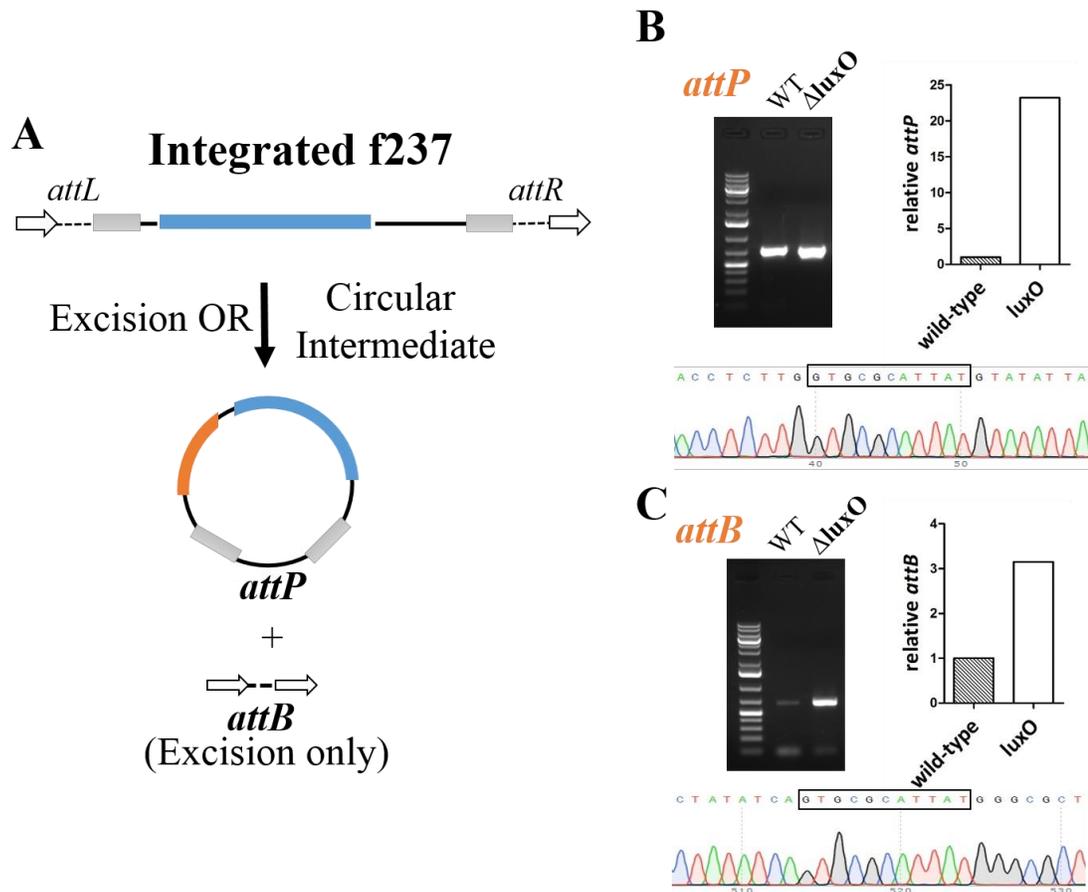


Figure 28 **f237 phage excision confirmed using attP and attB detection PRC assays.**

A. schematic representing formation of the circular intermediate (CI) with unique *attP* site and empty chromosomal *attB* site following excision from the chromosome. Detection of *attP* (B) and *attB* (C) through conventional PCR and qPCR in wild-type and a *luxO* mutant. Sequence of *attP* and *attB* results are shown to confirm site-specific recombination at *att* sites.

Higher levels of f237 circular intermediates detected in a *luxO* mutant

When compared to a *luxO* mutant strain (locked in HCD gene expression), higher levels of the *attP* site were detected relative to wild-type as shown in **Figure 28B**. Quantification of the *attP* site in these two strains was examined using the primer

set f237attPQFwd and f237attPQRev and revealed 23 times more *attP* in the *luxO* mutant relative to wild-type (**Figure 28B**).

#### Higher levels of f237 *attB* products detected in a *luxO* mutant

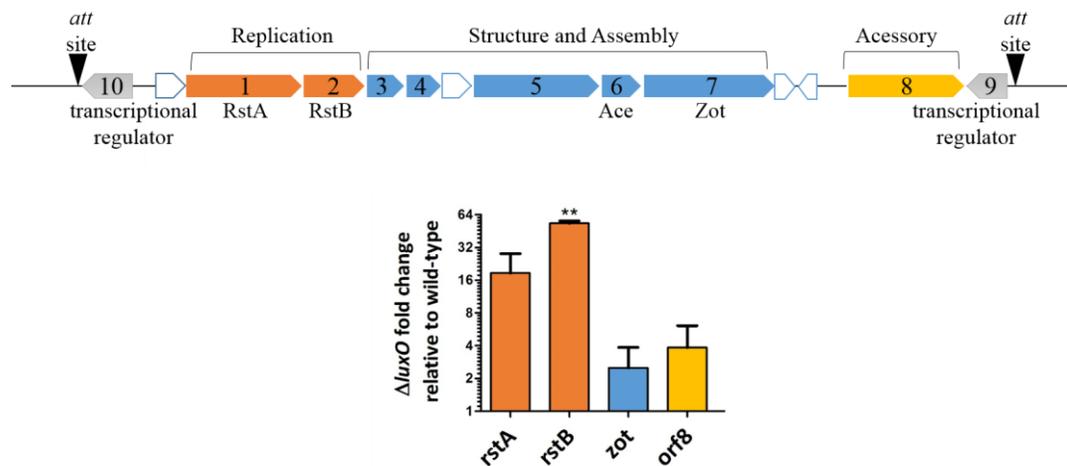
In addition to higher levels of *attP* detection in the *luxO* mutant, we also detected more *attB*, suggesting increased excision of this phage from the bacterial chromosome (**Figure 28C**). The *attB* site was detected with primers f237attBFwd and f237attBRev (**Table 8**) located in the core chromosome, outside the f237 phage region. Quantification using qRT-PCR revealed 3 times more *attB* in *luxO* relative to wild-type (**Figure 28C**).

#### Expression of f237 genes is increased in a *luxO* mutant

RNA-Seq analysis of cells grown statically at 37°C in M9+ 20 mM glucose to an OD<sub>600</sub> of 0.1 showed an increase in all genes encoded by the f237 filamentous phage (VP1549 – VP1562) in a *luxO* mutant strain compared to wild-type. Based on the genetic organization of the f237 ORFs, it is likely that VP1550 – VP1559 consist of an operon, while VP1549, VP1560, VP1561 and VP1562 appear to have their own promoter regions (**Figure 29**). The database of prokaryotic operons (DOOR<sup>2</sup>) identified VP1550-VP1555 and VP1556-VP1559 as two distinct operons. There is a 45 bp intergenic region between these two operons. There is a 156 bp intergenic region between VP1549 and VP1550, a likely promoter region for controlling transcription of both VP1549, a transcriptional regulator as well as the VP1550 operon. There are 439 bp of intergenic sequence between VP1559 and VP1560.

To confirm these results we examined the expression of VP1551 (*rstA*) and VP1552 (*rstB*) at the beginning of the operon and VP1558 (*zot*) at the end of the

operon as well as VP1560 (*orf8*), which showed the highest induction in RNA-seq and appears to be controlled by a separate promoter. Quantitative real time PCR (qRT-PCR) showed induction of all four genes in a *luxO* mutant relative to wild-type, validating the RNA-Seq analysis (**Figure 29**). The largest induction was seen for *rstA* and *rstB* which showed 18 and 53 times more transcript in the  $\Delta luxO$  compared to wild-type (**Figure 29**). A 2.5 and 3.8 fold increase in the transcript levels of *zot* and *orf8* were seen in the *luxO* mutant (**Figure 29**).



**Figure 29 The filamentous phage f237 and phage gene expression.**  
**A.** Depiction of f237 integrated into the *V. parahaemolyticus* RIMD genome. The phage is 8,796 bp, contains 14 ORFs and is flanked by attachment sites *attL* and *attR* (cataatgcgcac). Genes predicted to be involved in phage replication (orange), structure and assembly (blue), transcriptional regulation (grey) and genes of unknown function are shown. **B.** qRT-PCR examining f237 genes in wild-type and  $\Delta luxO$  cells grown to stationary phase (adjusted to OD<sub>600</sub> 1.0), shaking in LB. Fold change values were calculated using the  $\Delta\Delta C_t$  method. Bioinformatic identification of OpaR binding sites

Given that the OpaR is constitutively active in the *luxO* mutant, we suspected that OpaR may be directly involved in the increase of f237 gene expression and circular intermediate production. We identified putative OpaR binding sites in the DNA sequence of the f237 phage using the consensus sequence (WDHWDWBHDHDHWVWHHDHW) published by Zhang et al. (Zhang, Qiu et al. 2012). However several of these were within ORFs, therefore we focused on the two promoter binding sites within potential promoter regions of the f237 phage. Putative OpaR binding sites were identified between VP1549 and VP1550 in a 156 bp intergenic region. Upstream of VP1550 is an ORF on the negative strand VP1549 (Orf10) and downstream on the positive strand is the operon VP1550-VP1559, these ORFs could be subject to regulation by OpaR (**Figure 30**). Putative OpaR binding sites were also identified upstream of the VP1561 (orf8) promoter region, in the 308 bp intergenic region indicating a potential role in regulation of this ORF (**Figure 30**).

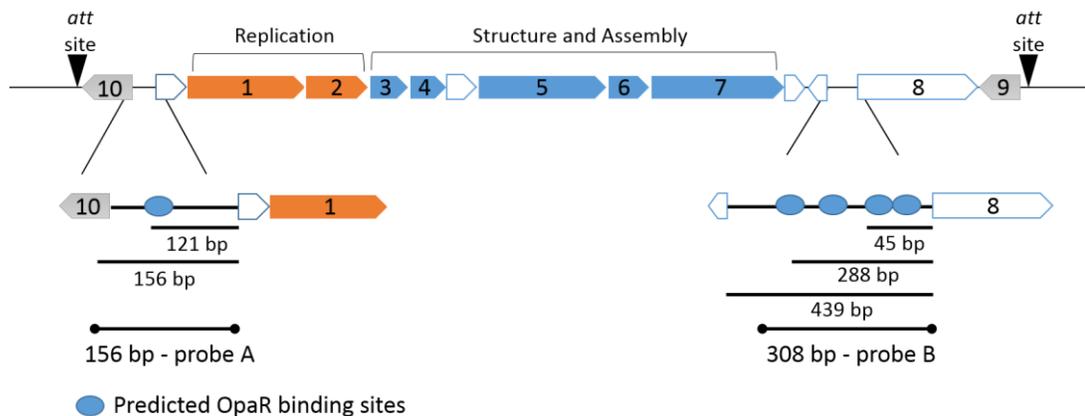
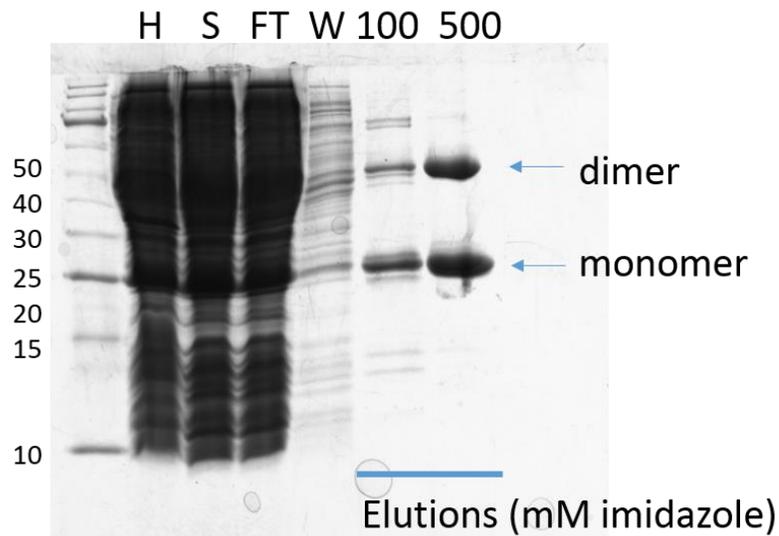


Figure 30 **Bioinformatics identification of putative OpaR binding sites.** OpaR in the f237 phage are indicated by blue circles in two promoter regions of the phage.

## Protein purification of OpaR

The OpaR protein was purified utilizing a 6X-His tag as described in *Materials and Methods*. The purified OpaR protein was detected as a monomer and dimer as shown in **Figure 31**. This is not unexpected as OpaR is a member of the TetR family of transcriptional regulators which function as dimers to recognize DNA sequences, similar to its homolog LuxR in *V. harveyi* (van Kessel, Ulrich et al. 2013). This dimer appears to be due to disulfide bonds as addition of 1 mM DTT results in a higher proportion of monomer by SDS-PAGE (data not shown). The His tag was cleaved off using the TEV protease as described in *Materials and Methods* and the size of the cleaved OpaR was confirmed by mass spectrometry before use for electrophoretic mobility shift assays (EMSAs).



**Figure 31 SDS-PAGE of OpaR protein purification.** Homogenate (H), supernatant (S), flow-through (FT) wash (W) and 100 mM and 500 mM imidazole elutions. The theoretical weight of the his-tagged OpaR protein is 26.7 kDa. OpaR can be seen as a monomer and a dimer

OpaR can bind to predicted promoter regions in f237 phage

The DNA sequence of both promoter regions described in **Figure 30** were amplified using the primer pairs VP1550Fwd/ VP1550Rev and VP1561Fwd/ VP1561Rev, respectively and used in EMSA assays to determine if OpaR can bind these sequences. Partial OpaR binding was observed at a 1: 10 (DNA: Protein) molar ratio for probe A (upstream of VP1550) and probe B (upstream of orf8), equivalent to .295  $\mu\text{M}$  and .149  $\mu\text{M}$  of OpaR, respectively (**Figure 32**). All of the DNA probe was bound to OpaR at 2.95  $\mu\text{M}$  (probe A) and 1.49  $\mu\text{M}$  (probe B) (1:100 molar ratio) (**Figure 29**). DNA amplified from the *V. parahaemolyticus* 16S RNA gene was used as a negative control, confirming that OpaR's binding is site specific (**Figure 33**).

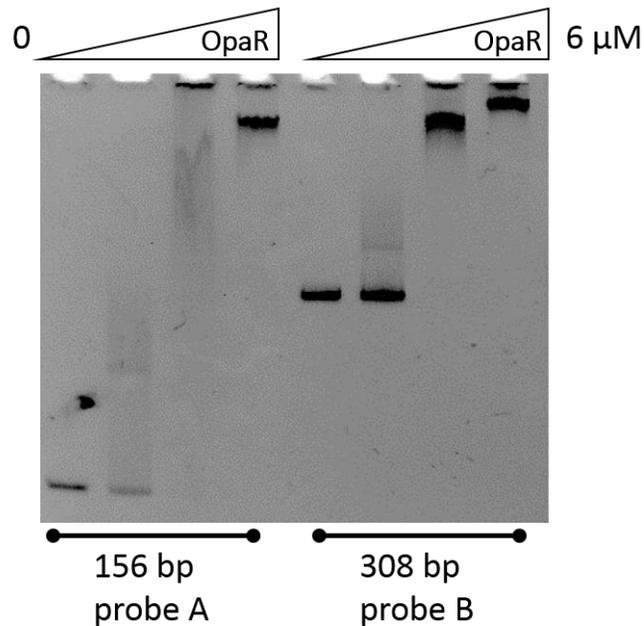


Figure 32 **OpaR can bind promoter regions in f237 phage.** EMSA utilizing OpaR and probe A and probe B shown in **Figure 30**. OpaR can bind both DNA fragments at concentrations of .295  $\mu\text{M}$  and .149  $\mu\text{M}$ .

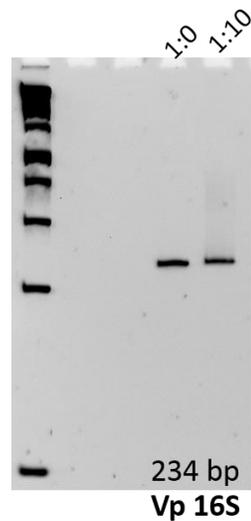


Figure 33 **OpaR EMSA negative control.**  
 16S DNA was amplified from the *V. parahaemolyticus* genome and used as a negative control probe to show specific OpaR binding. No shift was observed upon addition of 10-times the amount of OpaR to 16S DNA probe.

## Discussion

In this study, we present evidence for the direct regulation of phage f237 by the high cell density quorum sensing regulator OpaR. First, f237 contains all the putative genes necessary for phage infection, replication, assembly and secretion, indicating this phage is fully functional. In a *luxO* mutant in which cells are locked in a high cell density state and OpaR is constitutively expressed, we showed that several f237 genes (*rstA*, *rstB*, *zot* and *orf8*) are more highly expressed, ranging from 3-fold to over 60-fold, relative to wild-type. We next determined that this increase in f237 expression also correlated with a > 20-fold increase in production of phage circular intermediates.

Additionally, we found an increase in the amount of the *attB* excision product in a *luxO* mutant relative to wild-type. The function of this excision remains unknown and its effects on the physiology of the cell, if any. Given that increased f237 expression and CI production were observed in a *luxO* mutant, we determined whether OpaR may be responsible for these phenotypes. We confirmed that OpaR could bind to these DNA regions using EMSA assays. All these data support the model that OpaR is directly and positively regulating the transcription of f237 genes as well as production of f237 circular intermediates.

Interestingly, a recent paper has linked lytic phage induction with high cell density, Rossmann and colleagues showed quorum sensing inducer AI-2 promoted the release and expression of the lytic *pp5* prophage genes in the Gram-positive bacterium *Enterococcus faecalis* V583 $\Delta$ ABC and implicated this phage induction in biofilm dispersion (Rossmann, Racek et al. 2015). Additionally, it was found that AHLs can induce phage production in soil and groundwater bacteria and that this induction process is SOS-independent but dependent on *sdiA* (AHL receptor) and *rcaA* (EPS synthesis) (Ghosh, Roy et al. 2009). Finally, a *luxS* mutant (deficient in AI-2 production) resulted in a decrease in the production of the Stx genes of enterohemorrhagic *E. coli* (EHEC) compared to a wild-type strain. This study found that genes involved in the SOS response to also be downregulated in this mutant and suggested the implication of this pathway in phage regulation. Stx is encoded on a lambda like phage and inhibition of the production of the quorum sensing molecule AI-2, resulted in decreased transcription of phage genes (Sperandio, Torres et al. 2001).

Given that constitutively active OpaR is correlated with increased f237 production we can speculate that as cell density increases so too do the production of phage. The role of the f237 filamentous phage in *V. parahaemolyticus* remains to be answered. Based on findings implicating filamentous phage from *Pseudomonas aeruginosa* in biofilms, we can speculate a function. Specifically it was found that filamentous phage induction of cell death and release of bacterial DNA aided in the progression of the biofilm matrix (Rice, Tan et al. 2009) and more recently Pf phages were found to form highly ordered liquid crystals in *P. aeruginosa* biofilms which function to prevent desiccation, increase adhesion and increase antimicrobial tolerance (Secor, Sweere et al. 2015). Although production of f237 phage does not result in cell death, export of the phage could aid in formation of biofilm during stationary phase. Future studies to determine if biofilm production or stability is related to this observed production of f237 phage in absence of *luxO* should be carried out.

There have been several studies linking prophage induction with stationary phase (Ghosh, Roy et al. 2009; Rossmann, Racek et al. 2015), however this is the first study directly linking a global quorum sensing regulator, OpaR, to induction of a filamentous phage. This filamentous phage shows homology to the CTX phage, which carries the cholera toxin genes responsible for the secretory diarrheal disease cholera.

Table 7 Strains used in this study.

Strain or plasmid	Strain Description	Reference
<b>Strains</b>		
<i>Vibrio parahaemolyticus</i> RIMD 2210633	O3:K6 clinical isolate, Sm <sup>R</sup>	(Makino, Oshima et al. 2003)
$\Delta luxO$	$\Delta V P 2099$	S. Kalburge
$\Delta opaR$	$\Delta V P 2516$ , Sm <sup>R</sup>	S. Kalburge
$\Delta phaA$	$\Delta V P 2762$ , Sm <sup>R</sup>	S. Kalburge
WT pBBOpaR	RIMD2210633 harboring pBBOpaR	S. Kalburge
<i>E. coli</i> Dh5 $\alpha$		Laboratory strain
<i>E. coli</i> BL21(DE3)	Expression strain	Laboratory strain
<b>Plasmids</b>		
pJET1.2	Cloning vector; Am <sup>R</sup>	Fermentas
pJET1.2 OpaR	pJET1.2 harboring OpaR	This study
pProExHta	Expression vector, TEV site; Am <sup>R</sup>	Invitrogen
pProOpaR	pProExHta harboring OpaR; Am <sup>R</sup>	This study
pBBR1MCS	Expressin vector, Cm <sup>R</sup> , IPTG inducible	Laboratory strain
pBBOpaR	pBBR1MCS harboring OpaR	S. Kalburge

Table 8 Primers used in this study.

Name	Sequence ( 5' → 3' )	Tm (°C)	Product size (bp)
f237attBFwd	GATAAGTGTTTGGACGCTGG	55	
f237attBRev	AACCGTACCATCAAGCCAACG	59	659
f237attPFwd	AAGTTTGGCAGCTTCTTCGACG	59	
f237attPRev	TGTAGTCAGGTAGCAAGTCG	54	707
VP1551qFwd	GGCTCTTGAGCAGAACTCG	57	
VP1551qRev	CAGAGGCCAGCAGGTCTAAC	58	206
VP1552qFwd	TGGCATGGATATCACTTGGA	54	
VP1552qRev	AGCAGGGCATATTCATGGTC	56	174
VP1558qFwd	GGTTGCGGGAATGAGTCTTA	55	
VP1558qRev	GACGCCCAACTTGTTTAGGA	56	173
VP1561qFwd	GATGCATGCAGAAAGCTGAA	55	
VP1561qRev	AGCCCCTCGACTTAAACGAT	57	222
f237attBQFwd	CAGCTTCCTAAAGTTAGCTC	51	
f237attBQRev	ACGGCTTGGTTATCAGACAC	56	182
f237attPQFwd	GCAATCGCATTCAACATAAGCG	57	
f237attPQRev	TCTGATTGGCTTAGCGATGG	56	195
VP1550Fwd	GTAGCATCCTCAATTCACGTC	54	
VP1550Rev	CATGAAATCGCCTTGACCAG	55	156
VP1561Fwd	TGCGAGTGTCGAGCAATGCT	61	
VP1561Rev	TCCCTTCATAGCCCAACATG	55	308
SfoIVP2516Fwd	TACGGCGCCATGGACTCAATTGCAAAGAG	54	
SacIVP2516Rev	CACGAGCTCTTAGTGTTCGCGATTGTAGA	52	615

Table 9 f237 ORF descriptions and predicted function.

<b>ORF</b>	<b>Strand</b>	<b>Size (aa)</b>	<b>*Ff Nomenclature</b>	<b>Function (based on homology, gene order, size and/or membrane topology)</b>
<b>VP1549 (ORF10)</b>	-	122		Transcriptional regulator
<b>VP1550</b>	+	71		Unknown <sup>1</sup>
<b>VP1551 (ORF1)</b>	+	402	pII	RstA (CTXΦ 36%) replication, recombination and repair
<b>VP1552 (ORF2)</b>	+	119	pV	RstB (CTXΦ) required for integration
<b>VP1553 (ORF3)</b>	+	76	pVII	Inovirus G7P protein – viral transmembrane, assembly
<b>VP1554 (ORF4)</b>	+	81	pVIII	Major phage coat protein
<b>VP1555</b>	+	36	pIX	Unknown function
<b>VP1556 (ORF5)</b>	+	504	pIII	Tail protein
<b>VP1557 (ORF6)</b>	+	114	pVI	Ace (CTXΦ) structural protein
<b>VP1558 (ORF7)</b>	+	461	pI	Zot (CTXΦ) assembly protein
<b>VP1559</b>	+	47		Unknown
<b>VP1560</b>	-	38		Unknown
<b>VP1561 (ORF8)</b>	+	497		Unknown
<b>VP1562 (ORF9)</b>	-	121		Transcriptional regulator

<sup>1</sup> – sterile alpha domain (SAM) identified

\*Ff (M13, f1 and d1) nomenclature based on gene order and size (aa)

Genes labeled based on predictive protein function as grey (transcriptional regulation), orange (replication), blue (phage structure and assembly) and white (unknown).

## Chapter 5

### PERSPECTIVES AND FUTURE DIRECTIONS

The research described in this dissertation centers on excision mechanisms and host control of mobile genetic element transfer. Chapter two examined the mechanism of pathogenicity island (PAI) excision in *Vibrio cholerae*, a critical first step in the transfer and spread of virulence-carried genetic material to non-pathogenic bacteria. Chapter three assessed the evolution and excision capability of two novel PAIs which contain a type III secretion system (T3SS), named VPI-3, and a type IV secretion system (T6SS) and CRISPR-Cas region named VPI-6. Finally, chapter 4 investigated the role of host regulatory mechanisms, specifically quorum sensing, in the excision of a filamentous bacteriophage.

#### Cross-talk Mediated Pathogenicity Island Excision

The role of integrases and excisionases in PAI excision from two PAIs in *V. cholerae* was determined. Vibrio pathogenicity island (VPI) -1 harbors the genes encoding the toxin co-regulated pilus, essential for colonization of the small intestine as well as the virulence regulator ToxT. The second island, VPI-2 contains genes enabling production of sialidase and sialic acid transport and catabolism. Both of these islands contain their own cognate integrase named *intV1* and *intV2*, respectively. VPI-2 contains two excisionases named vibrio excision factor (*vef*) A and *vefB*, while VPI-1 contains none. Only one other excisionase, *vefC*, has been identified in the *V. cholerae* genome and is located on the Vibrio seventh pandemic (VSP) island II.

The question of how a PAI without an excisionase could excise from the bacterial chromosome has been answered in this work. It was found that the three *V. cholerae* excisionases are redundant in function and that any of them, *vefA*, *vefB* or *vefC* can facilitate excision of VPI-1. Interestingly, in the absence of all three excisionases ( $\Delta vefA\Delta vefB\Delta vefC$ ) only VPI-1 excision can be detected, but not VPI-2 excision. This phenotype was attributed to transcript levels of integrases – *intV1* was 2 times more highly expressed in the triple *vef* mutant compared to *intV2*. This was confirmed through restoration of VPI-2 excision when *intV2* was over expressed in a triple *vef* mutant background. In addition to *vefA*, *vefB* and *vefC* being capable of facilitating excision of VPI-1 and VPI-2, the cognate integrase of VPI-2, *intV2* can facilitate VPI-1 excision. Extensive cross-talk between these PAIs was observed and is depicted in **Figure 33**. This is the first study to show cross-talk, mediated via excisionases to facilitate excision of PAIs or any mobile genetic element for that matter.

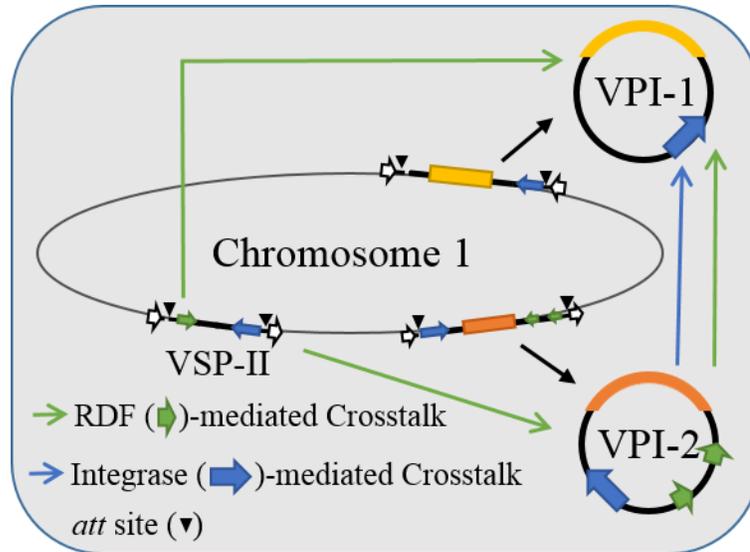


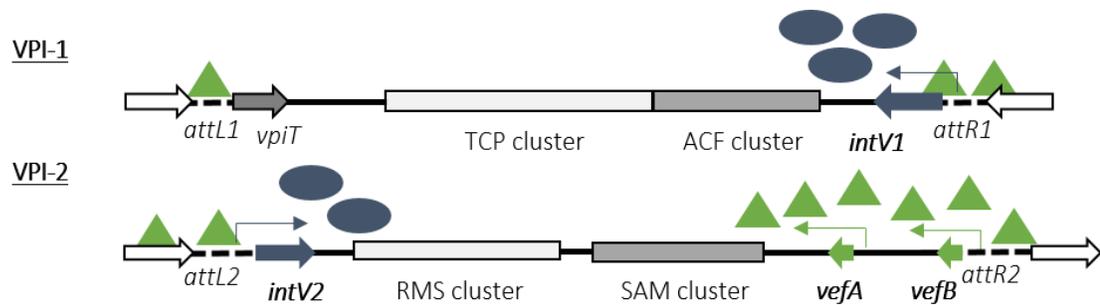
Figure 34 **Vibrio pathogenicity island (VPI) crosstalk.** The three VPIs, VPI-1, VPI-2 and VSP-II are depicted integrated into chromosome 1 of *V. cholerae*. Excision (black arrow) of VPI-1 and VPI-2 was examined in this study and their circular intermediates (CIs) are depicted. Yellow and orange blocks represent island cargo genes. Integrases are represented by blue arrows and RDFs/excisionases are represented by green arrows on the integrated and CIs. Arrows point from one island to another island indicated integrase (blue) or RDF (green) mediated cross-talk.

Findings from this study may be implicated in predicting the mobility of uncharacterized genetic elements. This study confirms that the presence of a canonical integrase and excisionase are not necessarily essential to facilitate PAI excision. This is evidenced by the excision of VPI-1 which lacks an excisionase, but can borrow those found on other PAIs. Even in the absence of excisionases from the *V. cholerae* genome, VPI-1 can still excise. We found that increased levels of *intVI* in this genetic background contributed to maintaining the excision phenotype. Furthermore, these findings can aid in determining the mobilome contribution of other bacteria with multiple MIGEs. The ability of PAIs who have an incomplete recombination module

to borrow excision components from other PAIs may still render them capable of excision and thus able to contribute to the pool of mobile bacterial genetic material available for transfer, aiding in the rapid evolution of bacteria.

### Dual Functionality of VPI Excisionases

In chapter two we also determined that the excisionases had multiple functions. Firstly, they acted to promote PAI excision. Over expression of each individual excisionase resulted in the “super-excision” phenotype of VPI-1 and VPI-2, referring to detection of excision in the first round of our assay, not seen in any wild-type strain. The excisionases were also found to repress expression of the integrases *intV1* and *intV2*. Expression of the integrases in the triple *vef* mutant relative to wild-type showed 4.5 and 2 fold higher levels of *intV1* and *intV2*, respectively. It may seem counterintuitive for excisionases, whose primary reported role is to aid in excision, to repress integrases, the protein possessing the enzymatic activity essential to facilitate the excision event. This is certainly not seen in the bacteriophage  $\lambda$ , one of the most well studied MIGEs. This drastic difference in mechanism can be attributed to the genetic organization of these two element’s recombination modules. The promoter regions of the integrases found on VPI-1 and VPI-2 overlap with the attachment sites of the islands – the same location where excisionases commonly bind to aid in the excision event. In the bacteriophage  $\lambda$ , the integrase and excisionase are located internally and their promoter regions are not overlapping with the attachment sites. We found that both VefA and VefB can bind the *att* sites upstream of the *intV1* and *intV2* promoter regions, suggesting this binding is the cause of Vef-mediated integrase repression. These findings led to a proposed VPI excision model depicted in **Figure 34**.



**Figure 35 Model for VPI Excision.**  
 The recombination module and island contents of VPI-1 and VPI-2 are depicted. Based on transcriptional analysis and observation of VefA and VefB DNA binding we predict that the RDFs possess dual function in a) binding at *att* sites to promote VPI excision and b) controlling *int* transcription. This tight control of integrase expression by the cells is likely essential for precise control of VPI excision. Triangles represent RDFs VefA, VefB or VefC and ovals represent integrase IntV1 or IntV2.

To confirm this proposed mechanism, it will be essential to determine the precise binding sites of the Vefs within the attachment site region. This could be done through DNase foot printing (Zianni, Tessanne et al. 2006). Once essential Vef binding sites are identified, site-directed mutagenesis to either delete or modify the binding sequence should be performed followed by an EMSA showing lack of Vef binding. Then transcript levels of the integrase can be examined and compared to a strain which contains a wild-type promoter to observe any change in integrase expression. Likewise, the promoter region can be cloned upstream of a reporter gene such as *lacZ* and promoter activity measured by  $\beta$ -galactosidase production.

This data suggests that the mechanisms of controlling expression of genes involved in excision and the excision event itself are tightly regulated. We can speculate that the purpose of RDF negative regulation of integrase transcription may be tied to the excision event itself. If integrase are expressed constitutively at low

levels, as our qPCR data suggests, then production and binding of Vefs at the *att* sites may be a signal that sufficient integrase is present to facilitate that excision event, and therefore halt integrase production. Others have found that a fine-tuned amount of integrase is required for excision of other MIGES such as bacteriophage and a particular threshold exists. If there is too much integrase present, excision can be inhibited and only the integration event is favored. Therefore the RDFs may be acting selfishly to promote their own cause.

The reason for higher amounts of *intV1* relative to *intV2* has yet to be determined. Over the course of evolution, it is likely that VPI-1 contained its own RDF at one point and has since lost it. Perhaps this is because the genetic contents of the island, the toxin co-regulated pilus and ToxT regulator, offer a superb fitness advantage to *V. cholerae* in the human intestine and thus the island is on its way to becoming stably maintained. Without a cognate RDF, the island's integrase, *intV1*, is no longer subject to tight control by its RDF. Our EMSA data does show the VefA and VefB, of VPI-2, can bind at the *att* site upstream of *intV1*, but at a much lower affinity than the VPI-2 sites. This may be attributed to mutations in the promoter region of *intV1* preventing tight Vef binding and allowing higher *int* expression activity rendering the PAI excisable, at low rates, without an excisionase.

The core recombination modules discussed in chapter 2 were found associated with a variety of other cargo genes such as T3SS, T6SS, CRISPR-Cas systems, toxin-antitoxin systems, chemotaxis genes and *ci-di* GMP genes. In chapter 3 we examined the excision potential of novel islands VPI-3 (T3SS-containing) and VPI-6 (CRISPR-Cas and T6SS-containing). We found that both of these islands can excise from the chromosome, the first step in transfer to other bacterial strains. Phylogeny

demonstrated that PAIs are modular in structure and each module of the island evolved separately rather than the island evolving as a single unit. This work suggests that similar recombination modules can associate with different cargo genes that can convert a strain into a pathogen with very different disease pathologies. Similar to phage evolution, pathogenicity islands have a modular structure where different functional regions are acquired by identical recombination modules. This is the first work demonstrating the excision of a MIGE containing a T3SS and T6SS, presenting a novel mechanism for the transfer of these complex systems. Future directions for this project involve examining transfer of the islands, post-excision, to island negative strains. *Vibrios* are naturally competent in the presence of chitin, a compound found in copious amounts in their natural environment. The VPIs can be marked with a selectable marker and examined for transfer to a VPI-negative strain using chitin induced competence, a protocol that has been previously described (Meibom, Blokesch et al. 2005; Marvig and Blokesch 2010).

## REFERENCES

- Abbani, M., M. Iwahara, et al. (2005). "The structure of the excisionase (Xis) protein from conjugative transposon Tn916 provides insights into the regulation of heterobivalent tyrosine recombinases." J Mol Biol 347(1): 11-25.
- Abbani, M. A., C. V. Papagiannis, et al. (2007). "Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly." Proc Natl Acad Sci U S A 104(7): 2109-14.
- Al-Hasani, K., K. Rajakumar, et al. (2001). "Genetic organization of the she pathogenicity island in *Shigella flexneri* 2a." Microb Pathog 30(1): 1-8.
- Almagro-Moreno, S. and E. F. Boyd (2009). "Insights into the evolution of sialic acid catabolism among bacteria." BMC Evol Biol 9: 118.
- Almagro-Moreno, S. and E. F. Boyd (2009). "Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine." Infect Immun 77(9): 3807-16.
- Almagro-Moreno, S., R. A. Murphy, et al. (2011). "How Genomics Has Shaped Our Understanding of the Evolution and Emergence of Pathogenic *Vibrio cholerae*." Genomes of Foodborne and Waterborne Pathogens: 85-99.
- Almagro-Moreno, S., M. G. Napolitano, et al. (2010). "Excision dynamics of *Vibrio* pathogenicity island-2 from *Vibrio cholerae*: role of a recombination directionality factor VefA." BMC Microbiol 10: 306.
- Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res 25(17): 3389-402.
- Aravind, L., V. Anantharaman, et al. (2005). "The many faces of the helix-turn-helix domain: Transcription regulation and beyond." Fems Microbiology Reviews 29(2): 231-262.
- Bachmann, V., B. Kostiuk, et al. (2015). "Bile Salts Modulate the Mucin-Activated Type VI Secretion System of Pandemic *Vibrio cholerae*." PLoS Negl Trop Dis 9(8): e0004031.

- Ball, C. A. and R. C. Johnson (1991). "Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein." J Bacteriol 173(13): 4027-31.
- Basler, M., M. Pilhofer, et al. (2012). "Type VI secretion requires a dynamic contractile phage tail-like structure." Nature 483(7388): 182-6.
- Beaber, J. W., B. Hochhut, et al. (2004). "SOS response promotes horizontal dissemination of antibiotic resistance genes." Nature 427(6969): 72-4.
- Boyd, E. F. (2012). "Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions." Adv Virus Res 82: 91-118.
- Boyd, E. F. (2012). Mobile and integrative genetic elements encoding cargo genes that are important for bacterial virulence Foodborne and Waterborne Bacterial Pathogens: Epidemiology and Molecular Biology. S. M. Faruque, Horizon Scientific Press.
- Boyd, E. F., S. Almagro-Moreno, et al. (2009). "Genomic islands are dynamic, ancient integrative elements in bacterial evolution." Trends Microbiol 17(2): 47-53.
- Boyd, E. F., M. R. Carpenter, et al. (2012). "Mobile effector proteins on phage genomes." Bacteriophage 2(3): 139-148.
- Boyd, E. F., M. R. Carpenter, et al. (2015). "Post-Genomic Analysis of Members of the Family Vibrionaceae." Microbiol Spectr 3(5).
- Campos, J., E. Martinez, et al. (2003). "VGJ phi, a novel filamentous phage of *Vibrio cholerae*, integrates into the same chromosomal site as CTX phi." J Bacteriol 185(19): 5685-96.
- Carniel, E., I. Guilvout, et al. (1996). "Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*." J Bacteriol 178(23): 6743-51.
- Carpenter, M. R., S. Rozovsky, et al. (2015). "Pathogenicity Island Cross Talk Mediated by Recombination Directionality Factors Facilitates Excision from the Chromosome." J Bacteriol 198(5): 766-76.
- Castang, S., H. R. McManus, et al. (2008). "H-NS family members function coordinately in an opportunistic pathogen." Proc Natl Acad Sci U S A 105(48): 18947-52.

- Chaand, M., K. A. Miller, et al. (2015). "Type 3 Secretion System Island Encoded Proteins Required for Colonization by Non-O1/non-O139 Serogroup V. cholerae." Infect Immun.
- Chang, B., H. Taniguchi, et al. (1998). "Filamentous bacteriophages of *Vibrio parahaemolyticus* as a possible clue to genetic transmission." J Bacteriol 180(19): 5094-101.
- Chatterjee, S. N. and K. Chaudhuri (2003). "Lipopolysaccharides of *Vibrio cholerae*. I. Physical and chemical characterization." Biochim Biophys Acta 1639(2): 65-79.
- Chen, Y., J. A. Johnson, et al. (2007). "The genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from those of O1 *Vibrio cholerae*." Infect Immun 75(5): 2645-7.
- Chin, C. S., J. Sorenson, et al. (2011). "The origin of the Haitian cholera outbreak strain." N Engl J Med 364(1): 33-42.
- Chowdhury, N., J. Norris, et al. (2012). "The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen *Vibrio cholerae*." Microbiology 158(Pt 8): 2158-67.
- Coddeville, M. and P. Ritzenthaler (2010). "Control of directionality in bacteriophage mv4 site-specific recombination: functional analysis of the Xis factor." J Bacteriol 192(3): 624-35.
- Coyne, M. J., K. G. Roelofs, et al. (2016). "Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements." BMC Genomics 17(1): 58.
- Coyne, M. J., N. L. Zitomersky, et al. (2014). "Evidence of extensive DNA transfer between bacteroidales species within the human gut." MBio 5(3): e01305-14.
- Craig, N. L. and H. A. Nash (1984). "*E. coli* integration host factor binds to specific sites in DNA." Cell 39(3 Pt 2): 707-16.
- Davies, B. W., R. W. Bogard, et al. (2012). "Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence." Cell 149(2): 358-70.
- Davis, B. M., E. H. Lawson, et al. (2000). "Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi." Science 288(5464): 333-5.

- Davis, B. M., K. E. Moyer, et al. (2000). "CTX prophages in classical biotype *Vibrio cholerae*: functional phage genes but dysfunctional phage genomes." J. Bacteriol. 182: 6992-6998.
- Davis, B. M. and M. K. Waldor (2000). "CTX $\phi$  contains a hybrid genome derived from tandemly integrated elements." Proc Natl Acad Sci U S A 97: 8572-8577.
- De, S. N. (1959). "Enterotoxicity of bacteria-free culture-filtrate of *Vibrio cholerae*." Nature 183: 1533-4.
- Dobrindt, U., M. G. Chowdary, et al. (2010). "Genome dynamics and its impact on evolution of *Escherichia coli*." Med Microbiol Immunol 199(3): 145-54.
- Dziejman, M., E. Balon, et al. (2002). "Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease." Proc Natl Acad Sci U S A 99(3): 1556-61.
- Dziejman, M., D. Serruto, et al. (2005). "Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system." Proc Natl Acad Sci U S A 102(9): 3465-70.
- Eisen, H., P. Brachet, et al. (1970). "Regulation of repressor expression in lambda." Proc Natl Acad Sci U S A 66(3): 855-62.
- Esposito, D. and J. J. Scoocca (1997). "The integrase family of tyrosine recombinases: evolution of a conserved active site domain." Nucleic Acids Res 25(18): 3605-14.
- Fang, F. C. and S. Rimsky (2008). "New insights into transcriptional regulation by H-NS." Curr Opin Microbiol 11(2): 113-20.
- Faruque, S. M. and J. J. Mekalanos (2012). "Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*." Virulence 3(7): 556-65.
- Faruque, S. M., J. Zhu, et al. (2003). "Examination of diverse toxin-coregulated pilus-positive *Vibrio cholerae* strains fails to demonstrate evidence for *Vibrio* pathogenicity island phage." Infect Immun 71(6): 2993-9.
- Feng, J. N., P. Model, et al. (1999). "A trans-envelope protein complex needed for filamentous phage assembly and export." Mol Microbiol 34(4): 745-55.
- Flanigan, A. and J. Gardner (2008). "Structural prediction and mutational analysis of the Gifsy-I Xis protein." BMC Microbiol 8: 199.

- Fogg, P. C., D. J. Rigden, et al. (2011). "Characterization of the relationship between integrase, excisionase and antirepressor activities associated with a superinfecting Shiga toxin encoding bacteriophage." Nucleic Acids Res 39(6): 2116-29.
- Frerichs, R. R., P. S. Keim, et al. (2012). "Nepalese origin of cholera epidemic in Haiti." Clin Microbiol Infect 18(6): E158-63.
- Frost, L. S., R. Leplae, et al. (2005). "Mobile genetic elements: the agents of open source evolution." Nat Rev Microbiol 3(9): 722-32.
- Galen, J. E., J. M. Ketley, et al. (1992). "Role of *Vibrio cholerae* neuraminidase in the function of cholera toxin." Infect Immun 60(2): 406-15.
- Ghosh, D., K. Roy, et al. (2009). "Acyl-homoserine lactones can induce virus production in lysogenic bacteria: an alternative paradigm for prophage induction." Appl Environ Microbiol 75(22): 7142-52.
- Grad, Y. H., P. Godfrey, et al. (2013). "Comparative genomics of recent Shiga toxin-producing *Escherichia coli* O104:H4: short-term evolution of an emerging pathogen." MBio 4(1): e00452-12.
- Grad, Y. H., M. Lipsitch, et al. (2012). "Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011." Proc Natl Acad Sci U S A 109(8): 3065-70.
- Grainge, I. and M. Jayaram (1999). "The integrase family of recombinases: organization and function of the active site." Molecular Microbiology 33(3): 449-456.
- Guzman, L. M., D. Belin, et al. (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." J Bacteriol 177(14): 4121-30.
- Hacker, J. and J. B. Kaper (2000). "Pathogenicity islands and the evolution of microbes." Annu Rev Microbiol 54: 641-79.
- Haley, B. J., S. Y. Choi, et al. (2014). "Genomic and phenotypic characterization of *Vibrio cholerae* non-O1 isolates from a US Gulf Coast cholera outbreak." PLoS One 9(4): e86264.
- Hazen, T. H., L. Pan, et al. (2010). "The contribution of mobile genetic elements to the evolution and ecology of Vibrios." FEMS Microbiol Ecol 74(3): 485-99.

- Heidelberg, J. F., J. A. Eisen, et al. (2000). "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*." Nature 406(6795): 477-83.
- Hensel, M. (2004). "Evolution of pathogenicity islands of *Salmonella enterica*." Int J Med Microbiol 294(2-3): 95-102.
- Herrington, D. A., R. H. Hall, et al. (1988). "Toxin, toxin-coregulated pili, and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans." J Exp Med 168(4): 1487-92.
- Higgins, D. G., J. D. Thompson, et al. (1996). "Using CLUSTAL for multiple sequence alignments." Methods Enzymol 266: 383-402.
- Ho, B. T., T. G. Dong, et al. (2013). "A view to a kill: the bacterial type VI secretion system." Cell Host Microbe 15(1): 9-21.
- Hochhut, B., C. Wilde, et al. (2006). "Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic *Escherichia coli* strain 536." Mol Microbiol 61(3): 584-95.
- Hoffmann-Berling, H., D. A. Marvin, et al. (1963). "[a Filamentous DNA Phage (Fd) and a Spherical Rna Phage (Fr), Host-Specific for the Male Strain of *E. Coli*. 1. Preparation and Chemical Properties of Fd and Fr]." Z Naturforsch B 18: 876-83.
- Hofschneider, P. H. and A. Preuss (1963). "M 13 Bacteriophage Liberation from Intact Bacteria as Revealed by Electron Microscopy." J Mol Biol 7: 450-1.
- Horton, R. M., H. D. Hunt, et al. (1989). "Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension." Gene 77(1): 61-8.
- Huber, K. E. and M. K. Waldor (2002). "Filamentous phage integration requires the host recombinases XerC and XerD." Nature 417(6889): 656-9.
- Hueck, C. J. (1998). "Type III protein secretion systems in bacterial pathogens of animals and plants. ." Microbiol Mol Biol Rev 62:379-433. 62: 379-433.
- Hurley, C. C., A. Quirke, et al. (2006). "Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates." BMC Genomics 7: 104.
- Iida, T., A. Hattori, et al. (2001). "Filamentous phage associated with recent pandemic strains of *Vibrio parahaemolyticus*." Emerg Infect Dis 7(3): 477-8.

- Iida, T., K. Makino, et al. (2002). "Filamentous bacteriophages of vibrios are integrated into the dif-like site of the host chromosome." J Bacteriol 184(17): 4933-5.
- Jermyn, W. S. and E. F. Boyd (2002). "Characterization of a novel *Vibrio* pathogenicity island (VPI-2) encoding neuraminidase (*nanH*) among toxigenic *Vibrio cholerae* isolates." Microbiology 148(11): 3681-93.
- Jermyn, W. S. and E. F. Boyd (2005). "Molecular evolution of *Vibrio* pathogenicity island-2 (VPI-2): mosaic structure among *Vibrio cholerae* and *Vibrio mimicus* natural isolates." Microbiology 151(Pt 1): 311-22.
- Jones, S., P. van Heyningen, et al. (1999). "Protein-DNA interactions: A structural analysis." J Mol Biol 287(5): 877-96.
- Kaper, J. B., J. P. Nataro, et al. (2004). "Pathogenic *Escherichia coli*." Nat Rev Microbiol 2(2): 123-40.
- Karaolis, D. K., J. A. Johnson, et al. (1998). "A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains." Proc. Natl. Acad. Sci. U S A 95(6): 3134-9.
- Karaolis, D. K., J. A. Johnson, et al. (1998). "A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains." Proc Natl Acad Sci U S A 95(6): 3134-9.
- Karaolis, D. K., S. Somara, et al. (1999). "A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria." Nature 399(6734): 375-9.
- Kirby, J. E., J. E. Trempy, et al. (1994). "Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*." J Bacteriol 176(7): 2068-81.
- Kitts, P. A. and H. A. Nash (1988). "An intermediate in the phage lambda site-specific recombination reaction is revealed by phosphorothioate substitution in DNA." Nucleic Acids Res 16(14B): 6839-56.
- Lantagne, D., G. Balakrish Nair, et al. (2014). "The Cholera Outbreak in Haiti: Where and how did it begin?" Curr Top Microbiol Immunol.
- Lee, C. A., J. M. Auchtung, et al. (2007). "Identification and characterization of *int* (integrase), *xis* (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*." Mol Microbiol 66(6): 1356-69.

- Leiman, P. G., M. Basler, et al. (2009). "Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin." Proc Natl Acad Sci U S A 106(11): 4154-9.
- Lesic, B., S. Bach, et al. (2004). "Excision of the high-pathogenicity island of *Yersinia pseudotuberculosis* requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor." Mol Microbiol 52(5): 1337-48.
- Lewis, J. A. and G. F. Hatfull (2001). "Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins." Nucleic Acids Res 29(11): 2205-16.
- Liu, J., P. Srinivasan, et al. "Expression and purification of the membrane enzyme selenoprotein K." Protein Expr Purif 86(1): 27-34.
- Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods 25(4): 402-8.
- Loeb, T. (1960). "Isolation of a bacteriophage specific for the F plus and Hfr mating types of *Escherichia coli* K-12." Science 131(3404): 932-3.
- Mai-Prochnow, A., J. G. Hui, et al. (2015). "'Big things in small packages: the genetics of filamentous phage and effects on fitness of their host'." FEMS Microbiol Rev 39(4): 465-87.
- Maiques, E., C. Ubeda, et al. (2006). "beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*." J Bacteriol 188(7): 2726-9.
- Makino, K., K. Oshima, et al. (2003). "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*." Lancet 361(9359): 743-9.
- Manning, P. A. (1997). "The tcp gene cluster of *Vibrio cholerae*." Gene 192(1): 63-70.
- Marvig, R. L. and M. Blokesch (2010). "Natural transformation of *Vibrio cholerae* as a tool--optimizing the procedure." BMC Microbiol 10: 155.
- McDonald, N. D., J.B. Lubin, et al. (2016). "Host-derived sialic acids are an important nutrient source that is required for optimal bacterial fitness in vivo." mBio In Press.

- McLeod, S. M. and M. K. Waldor (2004). "Characterization of XerC- and XerD-dependent CTX phage integration in *Vibrio cholerae*." Mol Microbiol 54(4): 935-47.
- Meibom, K. L., M. Blokesch, et al. (2005). "Chitin induces natural competence in *Vibrio cholerae*." Science 310(5755): 1824-7.
- Miller, H. I. and D. I. Friedman (1980). "An E. coli gene product required for lambda site-specific recombination." Cell 20(3): 711-9.
- Moyer, K. E., H. H. Kimsey, et al. (2001). "Evidence for a rolling-circle mechanism of phage DNA synthesis from both replicative and integrated forms of CTXphi." Mol Microbiol. 41: 311-323. .
- Murphy, R. and E. Boyd (2008). "Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates." J Bacteriol. 190(2): 636-647
- Nair, G. B., T. Ramamurthy, et al. (2007). "Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants." Clin Microbiol Rev 20(1): 39-48.
- Napolitano, M. G., S. Almagro-Moreno, et al. (2010). "Dichotomy in the evolution of pathogenicity island and bacteriophage encoded integrases from pathogenic *Escherichia coli* strains." Infect Genet Evol 11(2): 423-36.
- Napolitano, M. G. and E. F. Boyd (2012). Pathogenicity Island Evolution: A distinct new class of integrative element or a mosaic of other elements? Bacterial Integrative Mobile Genetic Elements. A. P. Roberts and P. Mullany, Landes Bioscience Press.
- Nasu, H., T. Iida, et al. (2000). "A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains." J Clin Microbiol 38(6): 2156-61.
- Navarre, W. W., M. McClelland, et al. (2007). "Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA." Genes Dev 21(12): 1456-71.
- Navarre, W. W., S. Porwollik, et al. (2006). "Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*." Science 313(5784): 236-8.
- Nei, M. and J. L. (1989). "Variances of the average numbers of nucleotide substitutions within and between populations." Mol. Biol. Evol. 6: 290-300.

- Numrych, T. E., R. I. Gumport, et al. (1992). "Characterization of the bacteriophage lambda excisionase (Xis) protein: the C-terminus is required for Xis-integrase cooperativity but not for DNA binding." EMBO J 11(10): 3797-806.
- Nunes-Duby, S. E., L. Matsumoto, et al. (1987). "Site-specific recombination intermediates trapped with suicide substrates." Cell 50(5): 779-88.
- O'Shea, Y. A. and E. F. Boyd (2002). "Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction." FEMS Microbiol Lett 214(2): 153-7.
- O'Shea, Y. A., S. Finnan, et al. (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(Pt 12): 4053-63.
- Okada, N., Iida T, et al. (2009). "Identification and characterization of a novel type III secretion system in trh-positive *Vibrio parahaemolyticus* strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level." Infect Immun. 77: 904-13.
- Okada, N., Matsuda S, et al. (2010). "Presence of genes for type III secretion system 2 in *Vibrio mimicus* strains." BMC Microbiol. 10: 302.
- Orata, F. D., P. C. Kirchberger, et al. (2015). "The Dynamics of Genetic Interactions between *Vibrio metoecus* and *Vibrio cholerae*, Two Close Relatives Co-Occurring in the Environment." Genome Biol Evol 7(10): 2941-54.
- Oshima, T., S. Ishikawa, et al. (2006). "*Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase." DNA Res 13(4): 141-53.
- Panis, G., Y. Duverger, et al. (2010). "Tight regulation of the intS gene of the KpIE1 prophage: a new paradigm for integrase gene regulation." PLoS Genet 6(10).
- Panis, G., N. Franche, et al. (2012). "Insights into the functions of a prophage recombination directionality factor." Viruses 4(11): 2417-31.
- Panis, G., V. Mejean, et al. (2007). "Control and regulation of KpIE1 prophage site-specific recombination: a new recombination module analyzed." J Biol Chem 282(30): 21798-809.
- Philippe, N., J. P. Alcaraz, et al. (2004). "Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria." Plasmid 51(3): 246-55.

- Piazzolla, D., S. Cali, et al. (2006). "Expression of phage P4 integrase is regulated negatively by both Int and Vis." J Gen Virol 87(Pt 8): 2423-31.
- Quinones, M., H. H. Kimsey, et al. (2005). "LexA cleavage is required for CTX prophage induction." Mol Cell 17(2): 291-300.
- Quirke, A. M., F. J. Reen, et al. (2006). "Genomic island identification in *Vibrio vulnificus* reveals significant genome plasticity in this human pathogen." Bioinformatics 22(8): 905-10.
- Rajakumar, K., C. Sasakawa, et al. (1997). "Use of a novel approach, termed island probing, identifies the *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins." Infect Immun 65(11): 4606-14.
- Rajanna, C., J. Wang, et al. (2003). "The vibrio pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products." J Bacteriol 185(23): 6893-901.
- Rajeev, L., K. Malanowska, et al. (2009). "Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions." Microbiol Mol Biol Rev 73(2): 300-9.
- Rakonjac, J., N. J. Bennett, et al. (2011). "Filamentous bacteriophage: biology, phage display and nanotechnology applications." Curr Issues Mol Biol 13(2): 51-76.
- Ramsay, J. P., J. T. Sullivan, et al. (2006). "Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS." Mol Microbiol 62(3): 723-34.
- Rasko, D. A., D. R. Webster, et al. (2011). "Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany." N Engl J Med 365(8): 709-17.
- Rice, S. A., C. H. Tan, et al. (2009). "The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage." ISME J 3(3): 271-82.
- Robins, W. P. and J. J. Mekalanos (2014). "Genomic Science in Understanding Cholera Outbreaks and Evolution of *Vibrio cholerae* as a Human Pathogen." Curr Top Microbiol Immunol.

- Rossmann, F. S., T. Racek, et al. (2015). "Phage-mediated dispersal of biofilm and distribution of bacterial virulence genes is induced by quorum sensing." PLoS Pathog 11(2): e1004653.
- Russel, M. (1995). "Moving through the membrane with filamentous phages." Trends Microbiol 3(6): 223-8.
- Rutherford, K., J. Parkhill, et al. (2000). "Artemis: sequence visualization and annotation." Bioinformatics 16(10): 944-5.
- Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Mol. Biol. Evol. 4: 406-425.
- Sakellaris, H., S. N. Luck, et al. (2004). "Regulated site-specific recombination of the she pathogenicity island of *Shigella flexneri*." Mol Microbiol 52(5): 1329-36.
- Salomon, D., J. A. Klimko, et al. (2015). "Type VI Secretion System Toxins Horizontally Shared between Marine Bacteria." PLoS Pathog 11(8): e1005128.
- Sam, M. D., C. V. Papagiannis, et al. (2002). "Regulation of directionality in bacteriophage lambda site-specific recombination: structure of the Xis protein." J Mol Biol 324(4): 791-805.
- Secor, P. R., J. M. Sweere, et al. (2015). "Filamentous Bacteriophage Promote Biofilm Assembly and Function." Cell Host Microbe 18(5): 549-59.
- Seed, K. D., D. W. Lazinski, et al. (2013). "A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity." Nature 494(7438): 489-91.
- Shin, O. S., V. C. Tam, et al. (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.
- Singh, S., J. G. Plaks, et al. (2013). "The structure of Xis reveals the basis for filament formation and insight into DNA bending within a mycobacteriophage intasome." J Mol Biol 426(2): 412-22.
- Soding, J., A. Biegert, et al. (2005). "The HHpred interactive server for protein homology detection and structure prediction." Nucleic Acids Res 33(Web Server issue): W244-8.

- Sperandio, V., A. G. Torres, et al. (2001). "Quorum sensing is a global regulatory mechanism in enterohemorrhagic *Escherichia coli* O157:H7." J Bacteriol 183(17): 5187-97.
- Sun, Y., E. E. Bernardy, et al. (2013). "Competence and natural transformation in vibrios." Mol Microbiol 89(4): 583-95.
- Swalla, B. M., E. H. Cho, et al. (2003). "The molecular basis of co-operative DNA binding between lambda integrase and excisionase." Mol Microbiol 50(1): 89-99.
- Tam, V. C., D. Serruto, et al. (2007). "A type III secretion system in *Vibrio cholerae* translocates a formin/spire hybrid-like actin nucleator to promote intestinal colonization." Cell Host Microbe 1(2): 95-107.
- Tamura, K., Stecher G., et al. (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. ." Molecular Biology and Evolution 30: 2725-2729.
- Taviani, E., C. J. Grim, et al. (2010). "Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis." FEMS Microbiol Lett 308(2): 130-7.
- Taylor, R. K., V. L. Miller, et al. (1987). "Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin." Proc. Natl. Acad. Sci. U. S. A 84(9): 2833-7.
- Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res 22(22): 4673-80.
- Trempey, J. E., J. E. Kirby, et al. (1994). "Alp suppression of Lon: dependence on the *slpA* gene." J Bacteriol 176(7): 2061-7.
- Vagenende, V., M. G. Yap, et al. (2009). "Mechanisms of protein stabilization and prevention of protein aggregation by glycerol." Biochemistry 48(46): 11084-96.
- van Kessel, J. C., L. E. Ulrich, et al. (2013). "Analysis of activator and repressor functions reveals the requirements for transcriptional control by LuxR, the master regulator of quorum sensing in *Vibrio harveyi*." MBio 4(4).
- Waldor, M. K. and J. J. Mekalanos (1996). "Lysogenic conversion by a filamentous phage encoding cholera toxin." Science 272(5270): 1910-4.

- Waldor, M. K., E. J. Rubin, et al. (1997). "Regulation, replication, and integration functions of the *Vibrio cholerae* CTX $\Phi$  are encoded by region RS2." Mol. Microbiol 24(5): 917-26.
- Warren, D., M. D. Sam, et al. (2003). "Identification of the lambda integrase surface that interacts with Xis reveals a residue that is also critical for Int dimer formation." Proc Natl Acad Sci U S A 100(14): 8176-81.
- Weisberg, R. A., L. W. Enquist, et al. (1983). "Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda." J Mol Biol 170(2): 319-42.
- Wilde, C., D. Mazel, et al. (2008). "Delineation of the recombination sites necessary for integration of pathogenicity islands II and III into the *Escherichia coli* 536 chromosome." Mol Microbiol 68(1): 139-51.
- Yu, N. Y., J. R. Wagner, et al. (2010). "PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes." Bioinformatics 26(13): 1608-15.
- Zhang, Y., Y. Qiu, et al. (2012). "Transcriptional regulation of opaR, qrr2-4 and aphA by the master quorum-sensing regulator OpaR in *Vibrio parahaemolyticus*." PLoS One 7(4): e34622.
- Zhou, X., J. M. Ritchie, H. Hiyoshi, T. Iida, B. M. Davis, M. K. Waldor, and T. Kodama. (2012). "The hydrophilic translocator for *Vibrio parahaemolyticus*, T3SS2, is also translocated." Infect Immun 80:(2940-7.).
- Zianni, M., K. Tessanne, et al. (2006). "Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument." J Biomol Tech 17(2): 103-13.

## Appendix A

### PERMISSION LETTER



RightsLink®

Home

Create Account

Help



AMERICAN  
SOCIETY FOR  
MICROBIOLOGY

**Title:** Pathogenicity island cross-talk mediated by recombination directionality factors (RDFs) facilitates excision from the chromosome  
**Author:** Megan R. Carpenter, Sharon Rozovsky, E. Fidelma Boyd et al.

**Publication:** Journal of Bacteriology

**Publisher:** American Society for Microbiology

**Date:** Dec 14, 2015

Copyright © 2015, American Society for Microbiology

LOGIN

If you're a [copyright.com user](#), you can login to RightsLink using your [copyright.com credentials](#). Already a [RightsLink user](#) or want to [learn more?](#)

#### Permissions Request

Authors in ASM journals retain the right to republish discrete portions of his/her article in any other publication (including print, CD-ROM, and other electronic formats) of which he or she is author or editor, provided that proper credit is given to the original ASM publication. ASM authors also retain the right to reuse the full article in his/her dissertation or thesis. For a full list of author rights, please see: [http://journals.asm.org/site/misc/ASM\\_Author\\_Statement.xhtml](http://journals.asm.org/site/misc/ASM_Author_Statement.xhtml)

## Appendix B

### THE ROLE OF BACTERIAL HOST-ENCODED FACTORS AND ENVIRONMENTAL STIMULI ON VPI EXCISION

Some work presented in this Appendix was completed in collaboration with undergraduate researcher Molly Peters.

#### A.1 Host Encoded Factors

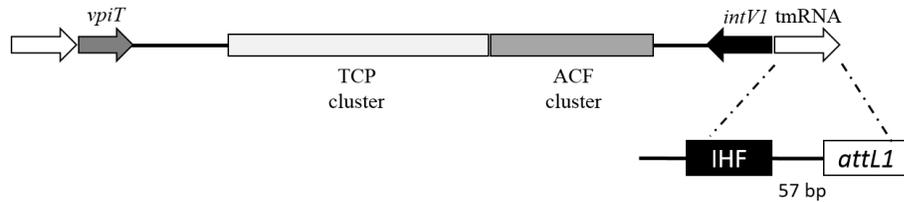
Several bacterial host-encoded genes have been implicated in control of mobile genetic elements. For example, two nucleoid proteins named integration host factor (IHF) and factor for inversion stimulation (Fis) have been implicated in  $\lambda$  phage biology; IHF is required for integration and excision and Fis stimulates excision (Miller and Friedman 1980; Ball and Johnson 1991).

First, we examined VPI-1 and VPI-2 regions for putative IHF binding sites using the consensus sequence WATCAANNNTTR (Craig and Nash 1984). Each island contains one putative IHF binding site upstream of its cognate integrase in the att region and VPI-2 has two additional putative IHF sites at the attR2 region of the island (**Figure 35**). Given this data, we next examined if IHF was playing a role in excision of VPI-1 or VPI-2. We examined the excision potential of the strain Sm463 in which *himA*, the IHF $\alpha$  subunit, is deleted. We found that VPI-1 and VPI-2 are still able to excise from the *V. cholerae* genome despite not having this architectural factor, indicating IHF is not necessary for excision of these islands (**Figure 36 and 37**). A *fis* mutant, Sm463, was also examined for VPI-1 and VPI-2 excision (**Figure 36 and 37**). No excision defect was observed, indicating neither IHF nor Fis are essential for VPI excision from the chromosome, as they are for bacteriophage  $\lambda$ . This does not however eliminate the possibility that they may aid in excision.

## IHF Binding Sites on VPI-1 and VPI-2

Consensus: WATCAANNNTTR (W = A or T, R = A or G)

### VPI-1 VC0817 – VC0847 (41.2 kb)



### VPI-2 VC1758 – VC1809 (57.3 kb)

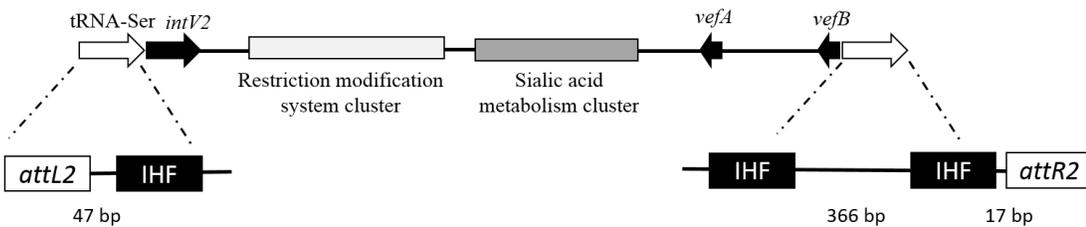


Figure 36 **Schematic of putative IHF binding sites on VPI-1 and VPI-2.** The att sites of each island are zoomed in and direct att repeats are represented as white boxes. The 13 bp putative IHF binding sites are represented as black boxes. Length between IHF and att sites in bp are indicated.

We also confirmed that the VPI excision mechanism is via site-specific recombination mediated by an integrase and not *recA*-mediated homologous recombination. VPI-1 and VPI-2 excision can indeed occur in the *recA* mutant strain, Bah-3, indicating that this excision is independent of *recA* (Figure 36 and 37).

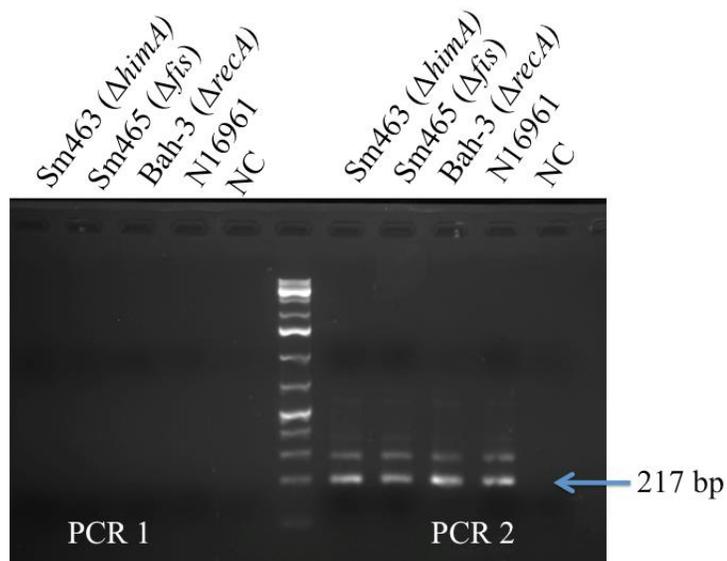


Figure 37 **VPI-1 attB1 Excision Assay Results.**  
 Excision occurs in all deletion mutants, indicating that the presence of the examined genes is not necessary for the excision of *V. cholerae* pathogenicity island VPI-1

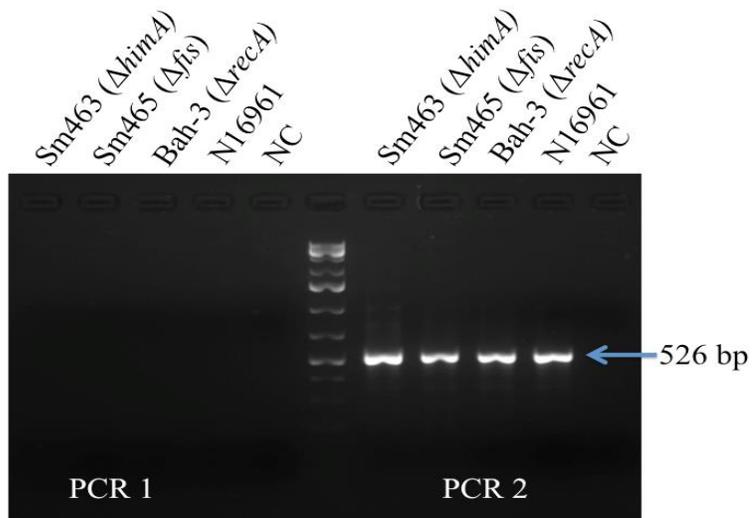
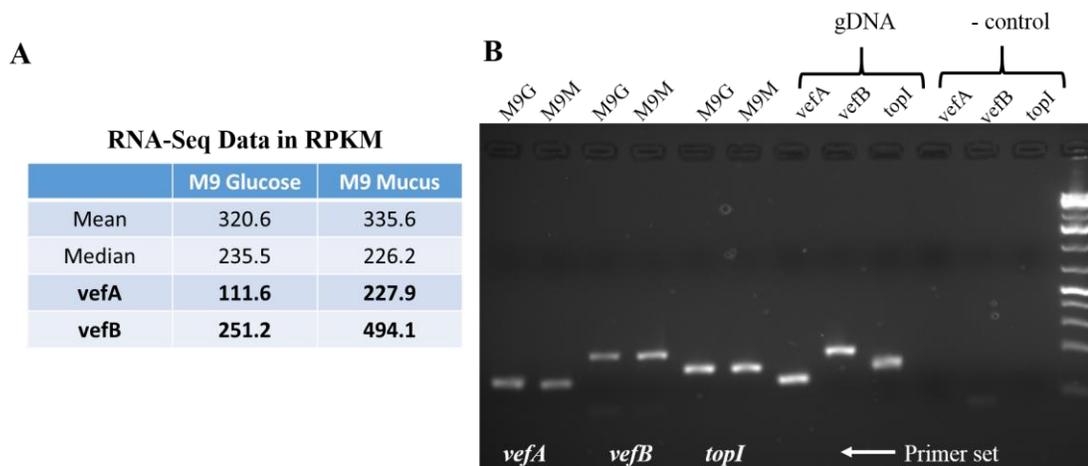


Figure 38 **VPI-2 attB2 Excision Assay results.**  
 Excision occurs in all deletion mutants, indicating that the presence of the examined genes is also not necessary for the excision of *V. cholerae* PAI VPI-2.

## A.2 Intestinal Mucus and RDF Expression

RNA-Seq analysis performed by another lab member (Nathan McDonald) revealed higher expression of *vefA* and *vefB* in the minimal media M9 supplemented with mucus (M9M) isolated from the murine intestinal tract, compared to M9 supplemented with glucose (M9G). Bacterial cells were grown statically to an OD<sub>600</sub> of 0.1. The RPKM (reads per million kilobases) for *vefA* and *vefB* are 227.9 and 494.1 in M9M compared to 111.6 and 251.2 in M9G, respectively. cDNA derived from RNA used in the aforementioned experiment was used for semi-quantitative reverse transcriptase PCR (**Figure 38**). While the change in *vefA* expression appears unchanged there appears to be an increase in *vefB* expression in M9M compared to M9G. The expression levels of the housekeeping gene encoding topoisomerase (*topI*) are unchanged between the two samples (**Figure 38**).



**Figure 39 Vef Expression in M9 Glucose and M9 Mucus.**

A. The mean and median reads per million kilobases (RPKM) values are shown for RNA-Seq analysis of *V. cholerae* N16961 cells grown under static conditions to an OD<sub>600</sub> of 0.1 at 37°C in either M9 minimal media with glucose or mucus isolated from the murine intestinal track. RPKM values are shown for *vefA* and *vefB*, which are higher in the M9 Mucus samples. B. RNA-Seq analysis was confirmed by semi-quantitative reverse transcriptase PCR using primer sets for *vefA* and *vefB* as well as the housekeeping gene *topI* in undiluted cDNA samples (M9G (M9 Glucose) and M9M (M9 Mucus)) used for RNA-Seq Genome DNA (gDNA) and water were used as a positive and negative control, respectively.

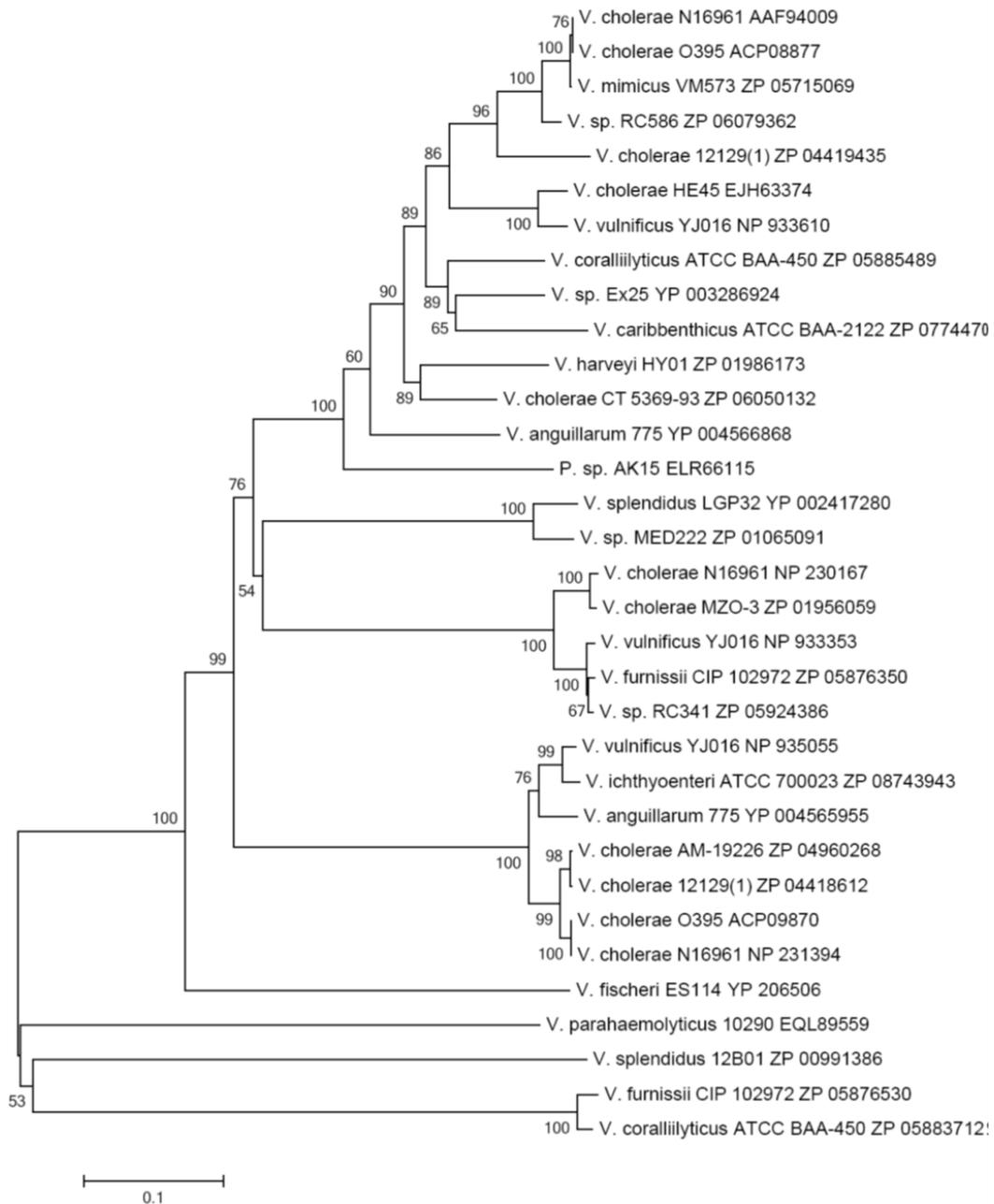


Figure 40 ***Vibrionaceae* PAI Integrase phylogeny.** Phylogeny was inferred using the Neighbor-Joining method (Saitou and Nei 1987) and evolutionary analyses were conducted in MEGA6 (Tamura, Stecher G. et al. 2013). Bootstrap values (1000 replicates) shown next to branches (Saitou and Nei 1987). Numbers following strain names indicate IMG gene ID or protein locus tags.

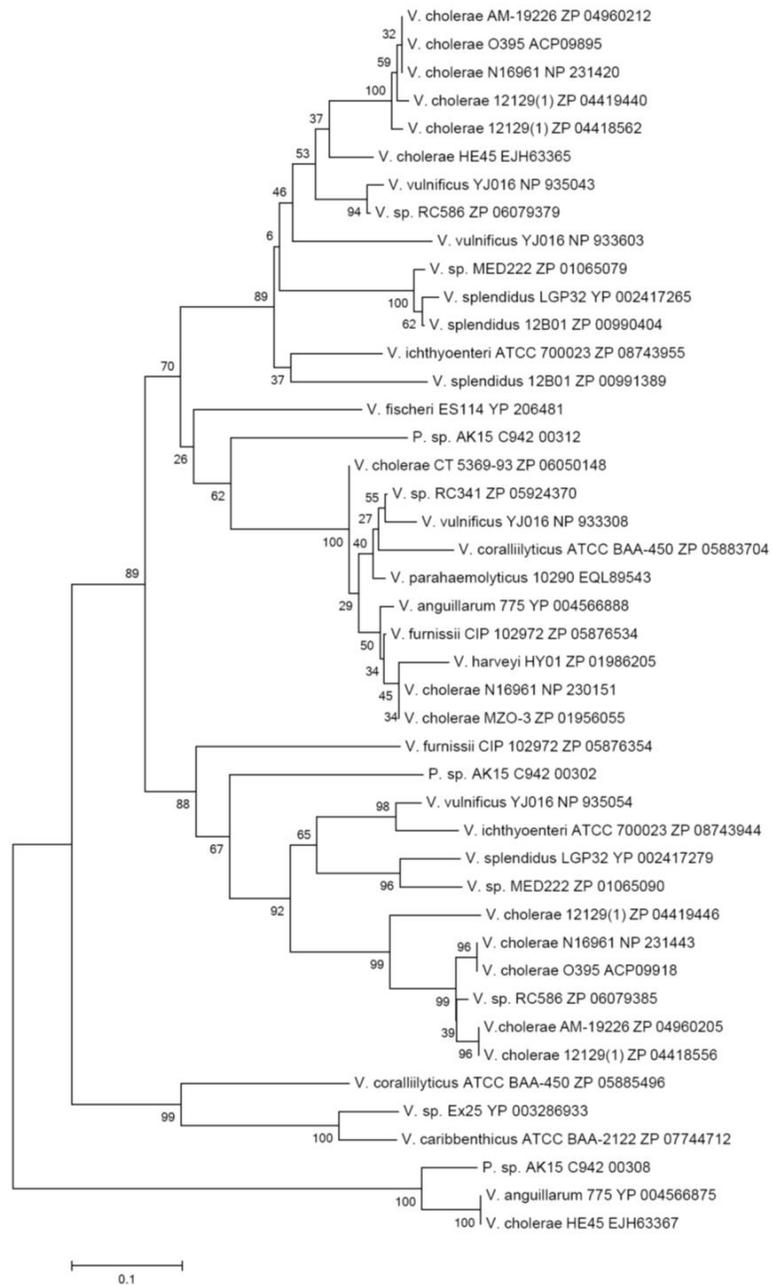


Figure 41 ***Vibrionaceae* PAI RDF phylogeny.** Phylogeny was inferred using the Neighbor-Joining method (Saitou and Nei 1987) and evolutionary analyses were conducted in MEGA6 (Tamura, Stecher G. et al. 2013). Bootstrap values (1000 replicates) shown next to branches (Saitou and Nei 1987). Numbers following strain names indicate IMG gene ID or protein locus tags.

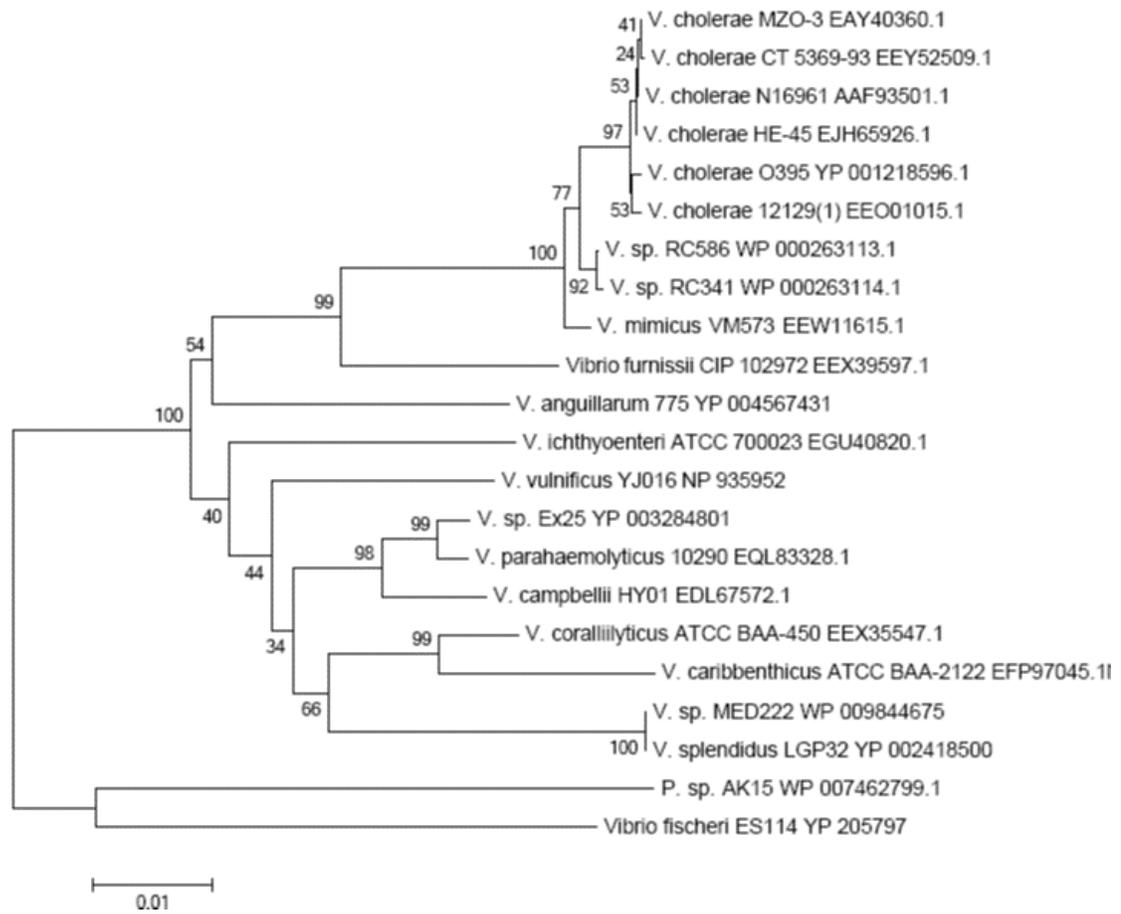


Figure 42 ***Vibrionaceae* RNA polymerase B phylogeny.** Phylogeny was inferred using the Neighbor-Joining method (Saitou and Nei 1987) and evolutionary analyses were conducted in MEGA6 (Tamura, Stecher G. et al. 2013). Bootstrap values (1000 replicates) shown next to branches (Saitou and Nei 1987). Numbers following strain names indicate IMG gene ID or protein locus tags.

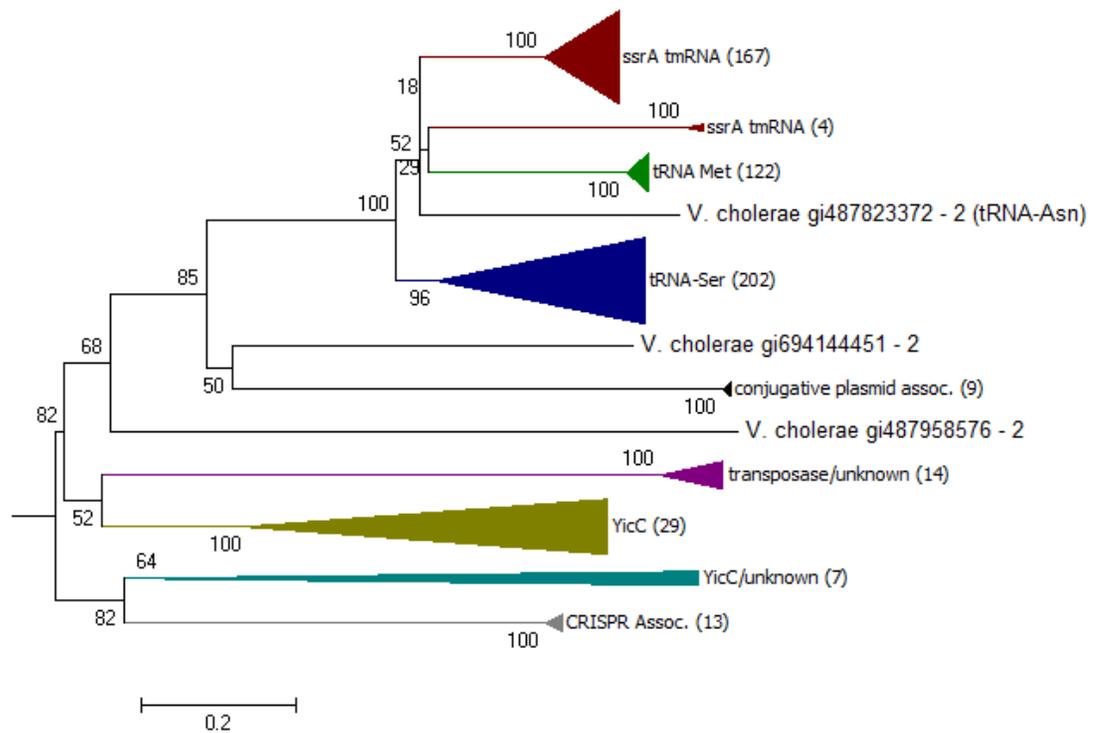


Figure 43 **Phylogeny of Vibrio MIGE Integrases (Condensed).** 105 *V. cholerae* sequences generated from blastp search of NP\_231394 (VC1758 intV2) with 90% or greater query cover and 22% or greater identity. Sequences were aligned using ClustalW and phylogeny deduced using the Neighbor joining method with bootstrap values (1000 replicates) shown next to branches.