PHENOTYPIC, GENETIC, AND REGULATORY ANALYSES OF CARBOHYDRATE METABOLISM SYSTEMS IN *VIBRIO PARAHAEMOLYTICUS*

by

Abish Regmi

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

Spring 2019

PHENOTYPIC, GENETIC, AND REGULATORY ANALYSES OF

CARBOHYDRATE METABOLISM SYSTEMS IN

VIBRIO PARAHAEMOLYTICUS

by

Abish Regmi

Approved: _____
Velia Fowler, Ph.D.
Chair of the Department of Biological Sciences

Approved: _____
John Pelesko, Ph.D.
Interim Dean of the College of Arts and Sciences

Approved: _____
Douglas J. Doren, Ph.D.
Interim 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: _____

Ramona Neunuebel, 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: _____

Catherine L. Grimes, 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: _____

Diane S. Herson, 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

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. E. Fidelma Boyd for believing in me and providing me with the opportunity to pursue my graduate career under her mentorship. From the beginning of my graduate career, Prof. Boyd, taught me how to manage my research, teaching and personal life, which led me to become successful in all fronts. Under her mentorship, I have learned various skill sets, especially, the ability to use any of my skill sets to solve any problem that comes my way. She has mentored me to have the ability to identify, critically analyze and solve complex scientific problems. Prof. Boyd gave me the opportunity to handle lab procurement related business, and opportunities to manage undergraduate as well as graduate students. I would not have been in this position if it was not for her support and mentorship. I would like to once again thank Prof. Boyd for an incredible and productive time with her.

I would like to thank my committee members, Drs. Ramona Neunuebel, Catherine L. Grimes, Diane Herson and John McDonald for the times they spent listening and reading about my research. Their ideas and suggestions were always valuable and made my work better and stronger. I am thankful to them for seeing a potential candidate in me and guiding me and my research towards success.

I would like to thank my mentor, Dr. Sai Siddarth Kalburge for believing in my potential, always pushing me towards success, and being there when I needed any help. Also, special thanks to my colleague Dr. Nathan McDonald, who was there with me from the beginning of this journey and helped me as a friend and a valuable

critique. His support from a personal and professional front has created a strong bond that I believe will last a long time. My past lab member Dr. Megan Carpenter guided me through my career development path, so thank you very much Megan. I also would like to thank the current Boyd's lab members, Gwen Gregory, Ji Kent Kwah, Jessica Tague, Daniel Morreale and John Rosenberger for making a pleasant and interesting work environment for me. Special thanks to Gwen for helping me with my rotation work. Daniel Morreale was a lot of fun to work with, with whom I shared authorship on a publication. Thanks to John Rosenberger for sharing with me his industrial job experience during my job search.

I would like to thank the members of BISC300 teaching group, especially Tisa Della-Volpe for helping me throughout my teaching assistant tenure. Also, special thanks to Dr. Carlton Cooper for nominating me for the Best Teaching Assistant Award in 2017. Also, I would like to especially thank Betty Cowgill for being there when help was needed and Prof. Melinda K. Duncan for all her professional support. Ginny Redmond's help during the formatting of this dissertation document cannot go unnoticed, and I really appreciate her help.

I would like to thank my family members who were there with me from the beginning of my educational journey. I will always remember my mother asking me, "When are you getting out?" and my father asking me, "Did you publish your work?" every time I had a phone call with them. Both questions were very important for me to stay focused and get things done. My wife, Ichha Pokharel, has been a pillar to my success. Her support throughout my graduate career has played a vital role in my success. She understands me in and out, personally and professionally. This dissertation would not have been possible without her support. My brother, sister-in-

law, sister and brother-in-law were always there for me with love and support. I would also like to thank my father-in-law and mother-in-law for supporting me throughout this journey. Lastly, I would like to thank all my friends everywhere, especially members of Nepalese Student Association (NSA) for creating a homey environment at the University of Delaware.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Utilizing unique carbohydrates or utilizing them more efficiently help bacteria expand and colonize new niches. Horizontal gene transfer (HGT) of catabolic systems is a powerful mechanism by which bacteria can acquire new metabolic traits that can increase survival and fitness in different niches. In this dissertation work, we investigated horizontal acquisition of catabolism systems and two major global regulators of metabolism in the survival and fitness of *Vibrio parahaemolyticus*, a marine species that is also an important human and fish pathogen. In humans, *V. parahaemolyticus* is the leading cause of bacterial seafood-related gastroenteritis worldwide.

In chapter two, we examined carbon catabolism diversity within *Vibrio parahaemolyticus* and among members of the family *Vibrionaceae* in general. Bioinformatics analysis showed that the ability to utilize D-galactose was present in all *V. parahaemolyticus* but at least two distinct transporters were present; a major facilitator superfamily (MFS) transporter and a sodium/galactose transporter (SGLT). Growth and genetic analyses demonstrated that SGLT was a more efficient transporter of D-galactose and was the predominant type among strains. The ability to utilize D-gluconate was universal within the species, however deletion of *eda* (VP0065), which encodes aldolase, a key enzyme in the Entner-Doudoroff (ED) pathway, reached a similar biomass to wild type when grown on D-gluconate as a sole carbon source. Two additional *eda* genes were identified, VPA1708 (*eda2*) associated with a D-glucuronate cluster and VPA0083 (*eda3*) that clustered with an oligogalacturonide

(OGA) metabolism cluster that were both induced by D-gluconate and both of these regions were variably present among strains. A metabolism island was identified that contained citrate fermentation, L-rhamnose and OGA metabolism gene clusters as well as a CRISPR-Cas system. Phylogenetic analysis showed that CitF (a key enzyme in the citrate fermentation pathway) and RhaA (a key enzyme in the L-rhamnose pathway) had a limited distribution among *V. parahaemolyticus*. Our analysis showed that the L-rhamnose pathway was acquired at least three times in this species. Our data suggest that horizontal transfer of metabolic systems among *Vibrionaceae* is an important source of metabolic diversity. This work describes previously uncharacterized metabolic islands that were hotspots for the gain and loss of functional modules likely mediated by transposons.

In chapter 3, we investigated the role of an important global regulator of metabolism, cAMP receptor protein (CRP) in *V. parahaemolyticus* for the first time. In bacteria CRP is activated by cAMP, a secondary messenger synthesized from ATP by adenylate cyclase (AC). We constructed Δ*crp* and Δ*cyaA* mutant strains by deleting the *crp* and *cyaA* genes to inactivate CRP. The Δ*crp* mutant was unable to grow in minimal media supplemented with non-glucose carbon sources demonstrating a requirement for CRP. Unexpectedly, the *cyaA* mutant behaved similar to wild-type, which suggested the presence of an uncharacterized adenylate cyclase. Bioinformatics analysis identified CyaB (VP1760), a class IV adenylate cyclase present in all strains, which complemented an *Escherichia coli* Δ*cyaA* mutant. The Δ*cyaB* mutant reached a similar biomass when grown on alternative carbon sources, which suggested that CyaB and CyaA can complement each other. A Δ*cyaB*/Δ*cyaA* double mutant confirmed this, as this mutant behaved identical to the Δ*crp* strain in all growth pattern

analyses. Capsule polysaccharide (CPS) production and biofilm formation showed a defect in the Δ*crp* and Δ*cyaB* mutants, whereas the Δ*cyaA* behaved like wild type. This indicated a specific requirement of CRP activation by CyaB. In an *Artemia franciscana* (brine shrimp) model of virulence, the Δ*crp* and Δ*cyaB*/Δ*cyaA* strains showed attenuated killing demonstrating that cAMP-CRP is required for virulence. Phylogenetic analysis of *cyaB* from *V. parahaemolyticus* showed that it is highly divergent from other members of the Campbellii, which contained this gene and not phylogenetically widespread. This is the first study to determine the physiological role of a class IV adenylate cyclase.

In chapter four, we examined the role of FIS (factor for inversion stimulation) in *V. parahaemolyticus* biology for the first time. FIS has been demonstrated to be a global regulator controlling genes belonging to multiple functional categories in several species. We constructed a deletion of *fis* (VP2885) in *V. parahaemolyticus* and demonstrated a motility defect in both swimming and swarming assays suggesting FIS is an important positive regulator of these phenotypes. Bioinformatics analysis identified putative FIS binding sites in the regulatory region of both polar and lateral flagellum biosynthesis gene clusters which was confirmed by electrophoretic mobility shift assays (EMSAs). In *in vitro* growth competition assays, the Δ*fis* mutant was outcompeted by wild type in minimal media supplemented with L-arabinose, D-glucosamine, or D-gluconate. The effects of FIS were direct as purified FIS protein bound to the regulatory regions of these catabolism gene clusters. The Δ*fis* mutant was more sensitive to antimicrobial peptides such as polymyxin B. In *in vivo* competition assays in a streptomycin pretreated adult mouse colonization model the *fis* deletion mutant had a colonization defect compared to wild type. Thus, similar to other enteric

pathogens, FIS plays a central role as a regulator of global gene regulation in this species.

Chapter 1

INTRODUCTION

Horizontal Gene Transfer (HGT) and Genomic Islands

Horizontal gene transfer is a mechanism by which bacterial cells obtain foreign DNA, which may contain genes encoding novel functions that can help bacteria adapt to new niches. Horizontal gene transfer is a strong driving force for bacterial evolution as it leads to "evolution in quantum leaps" (Hacker and Carniel 2001). HGT occurs via three major processes: transformation, a process by which bacteria can uptake DNA from the surrounding environment. Transduction, a process by which bacteriophage injects its genetic materials into a host bacterial cell, and conjugation, a mechanism by which two bacterial cells make contact with each other and one cell, the donor, transfers DNA to a recipient cell. The known vectors of HGT are bacteriophages, plasmids, integrative and conjugative elements (ICEs), integrons, transposons and genomic islands (GIs) (Frost, Leplae et al. 2005).

A genomic island is a large region of DNA (~10-kb to ~200-kb), which usually integrates at a tRNA locus of the host genome and is flanked by direct repeats. Genomic islands contain an integrase, which mediate site specific insertion into the host genome and also aids excision from the chromosome. Another feature of GI is their GC content, which is different than the GC content of the host genome, a signature of horizontally acquired DNA (Hacker, Blum-Oehler et al. 1997, Hacker and Kaper 2000, Dobrindt, Hochhut et al. 2004, Boyd, Almagro-Moreno et al. 2009, Napolitano, Almagro-Moreno et al. 2011). Genomic islands carry genes encoding

diverse functions and can be further classified into pathogenicity island (carrying virulence genes), degradation island (carrying xenobiotic degradation genes), resistance island (carrying antibiotic and heavy metal resistance genes), defense island (carrying restriction modification and/or CRISPR-Cas systems) and metabolism island (carrying metabolic genes) (Hacker, Blum-Oehler et al. 1997, Hacker and Kaper 2000, Dobrindt, Hochhut et al. 2004, Boyd, Almagro-Moreno et al. 2009, Napolitano, Almagro-Moreno et al. 2011).

In many pathovars of *Escherichia coli* such as Enteropathogenic and Enterohemorrhagic *Escherichia coli* (EPEC and EHEC), causative agents of persistent diarrhea and bloody diarrhea respectively, contain a pathogenicity island named locus of enterocyte effacement (LEE). The LEE pathogenicity island contains a type 3 secretion system that delivers effector proteins into host cells causing attaching and effacing lesions, a mechanism important for the virulence of both EPEC and EPEC (Franzin and Sircili 2015). *Vibrio cholerae,* a gastrointestinal pathogen, horizontally acquired the *Vibrio* Pathogenicity Island I (VPI-1), containing genes encoding a type IV bundle forming pilus, named the toxin-co-regulated pilus (TCP), which is the major intestinal colonization factor for *V. cholerae* (Taylor, Miller et al. 1987, Herrington, Hall et al. 1988, Manning 1997, Karaolis, Johnson et al. 1998). A second pathogenicity island named VPI-2 was identified, which carries the genes for sialic acid uptake and catabolism (Jermyn and Boyd 2002). Sialic acid is a 9-carbon amino sugar that is present on all mucus membranes and *V. cholerae* was shown to use sialic acid as a sole carbon source. In addition, it was shown this metabolism island was essential for *in vivo* colonization (Jermyn and Boyd 2002, Almagro-Moreno and Boyd 2009, McDonald, Lubin et al. 2016). All these genomic islands contained similar

recombination modules that are genes required for integration of the island, but they carried different cargo genes. The recombination module consists of an integrase, attachment sites and an excisionase, which differentiate them from other mobile genetic elements (Napolitano, Almagro-Moreno et al. 2011).

Vibrio parahaemolyticus: An Emerging Pathogen

A member of the *γ- proteobacteria* family*, V. parahaemolyticus* is a moderate halophile, ubiquitous in the marine and estuarine environments, associated with sediments, shellfish and zooplankton (Kaneko and Colwell 1973, Joseph, Colwell et al. 1982, Nordstrom and DePaola 2003, Zimmerman, DePaola et al. 2007). Consumption of raw and undercooked shellfish is the primary route of *V. parahaemolyticus* into the human host and causes gastroenteritis. In humans, it is the leading cause of bacterial seafood-related gastroenteritis worldwide (Nair, Ramamurthy et al. 2007, Su and Liu 2007). In the United States alone 45,000 cases of *V. parahaemolyticus* illnesses are reported each year, but the number of cases is increasing and there is much under reporting (CDC, 2018). Gastroenteritis caused by this bacterium includes symptoms such as vomiting, diarrhea, headache, nausea, low-grade fever and abdominal cramps and are commonly self-limiting (Nair, Ramamurthy et al. 2007), however, if the individual is immunocompromised, it can cause severe inflammatory diarrhea, septicemia and even death (Hondo, Goto et al. 1987, Daniels, MacKinnon et al. 2000).

V. parahaemolyticus was first isolated from a patient after a gastroenteritis outbreak in Japan in 1950 (Fujino, Okuno et al. 1953). The first pandemic strain of *V. parahaemolyticus* belonging to the O3:K6 serotype emerged in 1995 (Okuda, Ishibashi et al. 1997). The presence of *tdh* and/or *trh* genes, genes encoding the

thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) respectively have been used as a marker for virulence identification (Miyamoto, Kato et al. 1969, Okuda, Ishibashi et al. 1997), however, isolation of hemolysin-negative strains from patients suggested the role of other virulence factors in pathogenesis (Nishibuchi, Fasano et al. 1992). The first complete genome sequence of *V. parahaemolyticus* was for the pandemic O3:K6 strain RIMD2210633 (Makino, Oshima et al. 2003). Genome analysis of this strain identified a type III secretion system (T3SS) on each chromosome, designated T3SS-1 and T3SS-2 (Makino, Oshima et al. 2003). T3SS-1 is present in both clinical and environmental isolates and has a percent G+C content similar to the rest of the genome indicating that this region is ancestral (Makino, Oshima et al. 2003). The T3SS-2 region contains Tdh1, Tdh2, a G+C content of 40% (less than the overall genome) and is absent from non-pathogenic isolates indicating it was acquired by horizontal gene transfer (Makino, Oshima et al. 2003, Hurley, Quirke et al. 2006). Recent data suggest the T3SS-2 region is present within a novel Tn7-like transposon that is likely involved in its mobilization (McDonald, Regmi et al. 2019). T3SS-1 plays a role in cytotoxicity (Hiyoshi, Kodama et al. 2010), whereas T3SS-2 has been shown to be required for enterotoxicity using different animal models (Park, Ono et al. 2004, Hiyoshi, Kodama et al. 2010, Pineyro, Zhou et al. 2010, Ritchie, Rui et al. 2012). Besides being a human gastrointestinal pathogen, various strains of *V. parahaemolyticus* have also been associated with acute hepatopancreatic necrosis disease (AHPND) in shrimp from Asian and Central American regions (Joshi, Srisala et al. 2014, Soto-Rodriguez, Gomez-Gil et al. 2015). Genomic and functional analysis has demonstrated the presence of PirA and PirB toxin encoded in a plasmid in *V. parahaemolyticus* strains causing AHPND (Lee, Chen et al. 2015).

Metabolic Islands and Their Role in Bacterial Fitness

Horizontal gene transfer is a powerful tool for bacteria to acquire metabolic systems which helps them to adapt into new niches and gain a competitive advantage (Rohmer, Hocquet et al. 2011). For example, sucrose uptake and catabolism genes were horizontally acquired in a 100-kb ICE region in *Salmonella senftenberg* 5494-57 which expanded their metabolic capability (Hochhut, Jahreis et al. 1997). The meningitis/sepsis associated *E. coli* (MNEC) acquired glyoxylate and glycerol metabolism clusters in a 20 kb metabolic island increasing the pool of their nutrient utilization diversity (Schmidt and Hensel 2004). The tetrathionate respiration genes are carried on *Salmonella* pathogenicity island 2 (SPI2), and utilization of tetrathionate gives *Salmonella* a competitive advantage over the gut microbiota in the intestinal lumen (Rohmer, Hocquet et al. 2011). The marine bacterium *Pseudoalteromonas haloplanktis* ANT/505 horizontally acquired functional pectin catabolism genes carried within a genomic island (Hehemann, Truong et al. 2017). *Vibrio cholerae,* a gastrointestinal pathogen horizontally acquired sialic acid transport and utilization system in a genomic island, VPI-2, which was demonstrated to be an important carbon and energy source for the *in vivo* survival of the pathogen (Jermyn and Boyd 2002, Almagro-Moreno and Boyd 2009, McDonald, Lubin et al. 2016).

The Role of Metabolism in Bacterial Fitness

Vibrio parahaemolyticus is a marine species that has adapted and evolved to colonize and establish niches in the intestine of humans as well as fish and shellfish species. In these environments, the bacterium must compete with endogenous microbiota and other pathogenic bacteria for available nutrients. Freter's nutrient-niche hypothesis proposes that any species of bacteria should utilize at least one

limiting nutrient better than another species for successful colonization (Freter 1988). Using a phenotypic array of 190 carbon sources, it was shown that *V. parahaemolyticus* could utilize 71 different carbon sources under static growth conditions (Kalburge, Carpenter et al. 2017). Studies have shown that strains have also adapted to utilize intestinal mucus and mucus sugars L-arabinose, D-galactose, D-gluconate, D-glucuronate, D-mannose, D-glucosamine, *N*-acetyl-D-glucosamine and D-ribose as sole carbon sources (Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017). Previous work has also shown that disruption of global regulation of metabolism has a profound effect on fitness (Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017). For example, deletion of the global transcription factor sigma 54 encoded by *rpoN* enhanced *in vivo* fitness and this correlated with increased rates of carbon metabolism (Whitaker, Richards et al. 2014). Whereas deletion of the quorum sensing response regulator *luxO* inhibited carbon metabolism and caused growth defects *in vitro* and *in vivo* (Kalburge, Carpenter et al. 2017).

Dissertation Work

Utilizing unique carbohydrates or utilizing them more efficiently help bacteria expand and colonize new niches. HGT of catabolic systems is a powerful mechanism by which bacteria can acquire new phenotypic traits that can increase survival and fitness in different niches. In chapter two, we examined carbon catabolism diversity among *Vibrio parahaemolyticus*, a marine species that is also an important human and fish pathogen. Phenotypic differences in carbon utilization between *Vibrio parahaemolyticus* strains led us to examine genotypic differences in this species and the family *Vibrionaceae* in general. Bioinformatics analysis showed that the ability to utilize D-galactose was present in all *V. parahaemolyticus* but at least two distinct

transporters were present; a major facilitator superfamily (MFS) transporter and a sodium/galactose transporter (SGLT). Growth and genetic analyses demonstrated that SGLT was a more efficient transporter of D-galactose and was the predominant type among strains. Phylogenetic analysis showed that D-galactose gene *galM* was acquired multiples times within the family *Vibrionaceae* and was transferred between distantly related species. The ability to utilize D-gluconate was universal within the species. Deletion of *eda* (VP0065), which encodes aldolase, a key enzyme in the Entner-Doudoroff (ED) pathway, resulted in cultures that reached a similar biomass to wild type when grown on D-gluconate as a sole carbon source. Two additional *eda* genes were identified, VPA1708 (*eda2*) associated with a D-glucuronate cluster and VPA0083 (*eda3*) that clustered with an oligogalacturonide (OGA) metabolism cluster. EDA2 and EDA3 were variably distributed among the species. A metabolic island was identified that contained citrate fermentation, L-rhamnose and OGA metabolism clusters as well as a CRISPR-Cas system. Phylogenetic analysis showed that CitF and RhaA had a limited distribution among *V. parahaemolyticus*, and RhaA was acquired at least three times. Within *V. parahaemolyticus*, two different regions contained the gene for L-arabinose catabolism and most strains had the ability to catabolize this sugar. Our data suggest that horizontal transfer of metabolic systems among *Vibrionaceae* is an important source of metabolic diversity. This work identified four EDA homologues suggesting that the ED pathway plays a significant role in metabolism. We describe previously uncharacterized metabolic islands that were hotspots for the gain and loss of functional modules likely mediated by transposons.

Cyclic AMP (cAMP) receptor protein (CRP) is an essential global transcription factor in bacteria. CRP is activated by cAMP, a secondary messenger

synthesized from ATP by adenylate cyclase (AC). In many bacteria, in the presence of the preferred carbon source glucose, cAMP levels are low and CRP is inactivate; when glucose is absent, cAMP levels are high and CRP can activate alternative carbon catabolism genes. In chapter 3, we investigated the role of CRP in *V. parahaemolyticus* as a global regulator by constructing deletions of the *crp* and *cyaA* genes creating Δ*crp* and Δ*cyaA* mutant strains. The Δ*crp* mutant was unable to grow in minimal media supplemented with non-glucose carbon sources demonstrating a requirement for CRP. Unexpectedly, the *cyaA* mutant behaved similar to wild-type, which suggested the presence of an uncharacterized adenylate cyclase. Bioinformatics analysis identified CyaB (VP1760), a class IV adenylate cyclase present in all strains, which complemented an *Escherichia coli* Δ*cyaA* mutant. The Δ*cyaB* mutant reached a similar biomass when grown on alternative carbon sources, which suggested that CyaB and CyaA can compensate for each other. A Δ*cyaB*/Δ*cyaA* double mutant confirmed this, as this mutant behaved identical to the Δ*crp* strain in growth assays. Capsule polysaccharide (CPS) production and biofilm formation showed a defect in the Δ*crp* and Δ*cyaB* mutants, whereas the Δ*cyaA* behaved like wild-type. This indicated a specific requirement of CRP activation by CyaB. In an *Artemia franciscana* (brine shrimp) model of virulence, the Δ*crp* and Δ*cyaB*/Δ*cyaA* strains showed attenuated killing demonstrating that cAMP-CRP is required for virulence. Phylogenetic analysis of *cyaB* from *V. parahaemolyticus* showed that it is highly divergent from other members of the Campbellii and not phylogenetically widespread.

FIS (factor for inversion stimulation) belongs to the family of nucleoid associated proteins that binds and bends DNA for chromosome organization and is also a global regulator. In many bacterial species, FIS has been demonstrated to

regulate genes belonging to multiple functional categories. In chapter four, we examined the role of FIS in *Vibrio parahaemolyticus* biology for the first time. FIS showed 100% identity to FIS from *V. cholerae*. We constructed a deletion of *fis* (VP2885) in *V. parahaemolyticus* RIMD2210633 and demonstrated a motility defect in both swimming and swarming assays suggesting FIS is an important positive regulator of these phenotypes. Bioinformatics analysis identified putative FIS binding sites in the regulatory region of both polar and lateral flagellum biosynthesis gene clusters, which was confirmed by electrophoretic mobility shift assays (EMSAs). In *in vitro* growth competition assays, the Δ*fis* mutant was outcompeted by wild type in minimal media supplemented with L-arabinose, D-glucosamine, or D-gluconate, however, in D-mannose the mutant outcompeted wild type. The effects of FIS were direct as purified FIS protein bound to the regulatory regions of these catabolism gene clusters. The Δ*fis* mutant was more sensitive to polymyxin B and in *in vivo* competition assays in a streptomycin pretreated adult mouse colonization model had a fitness defect. Thus, similar to other enteric pathogens, FIS plays a central role as a regulator of global gene regulation in this species.

Chapter 2

CARBOHYDRATE METABOLIC SYSTEMS PRESENT ON GENOMIC ISLANDS ARE LOST AND GAINED IN *VIBRIO PARAHAEMOLYTICUS*

This work is under review in the journal *BMC Microbiology*

Regmi, A. and E.F. Boyd. 2019. Carbohydrate metabolism systems present on genomic islands that are lost and gained in *Vibrio parahaemolyticus*.

BMC Microbiology

Introduction

Vibrio parahaemolyticus is a Gram-negative, moderate halophile that is ubiquitous in the marine and estuarine environments, and present in sediments, shellfish and zooplankton (Kaneko and Colwell 1973, Joseph, Colwell et al. 1982, Nordstrom and DePaola 2003, Zimmerman, DePaola et al. 2007). Consumption of raw and undercooked shellfish is the primary route of *V. parahaemolyticus* into the human host, where it is the leading cause of bacterial seafood-related gastroenteritis worldwide (Nair, Ramamurthy et al. 2007, Su and Liu 2007). According to CDC, in the United States alone 45,000 cases of *V. parahaemolyticus* illnesses are estimated each year (2018).

In 1995, a highly virulent *V. parahaemolyticus* serogroup O3:K6 strain emerged in Asia and has since disseminated globally (Nair, Ramamurthy et al. 2007). This pandemic clone was sequenced, strain RIMD2210633, and two type three secretion systems (T3SS), one on each chromosome, designated T3SS-1 and T3SS-2

were identified (Makino, Oshima et al. 2003). T3SS-1 is present in all isolates and is ancestral to the species, whereas T3SS-2 is present on a pathogenicity island and is found in pathogenic strains with a number of variants of this region described (Makino, Oshima et al. 2003, Hurley, Quirke et al. 2006, Noriea, Johnson et al. 2010, Gavilan, Zamudio et al. 2013, Xu, Gonzalez-Escalona et al. 2017). Studies have identified at least three variant T3SS-2 systems; T3SS-2α present on chromosome 2 and the non-homologous T3SS-2β system present on chromosome 1 or chromosome 2, and T3SS-2γ on chromosome 2 that is closely related to T3SS2β (Okada, Iida et al. 2009, Chen, Stine et al. 2011, Xu, Gonzalez-Escalona et al. 2017). T3SS-2γ was identified in strains that are members of the clonal complex ST631, a recent pathogenic clone to emerge in the Northeast USA (Xu, Ilyas et al. 2015, Xu, Gonzalez-Escalona et al. 2017, Xu, Gonzalez-Escalona et al. 2017). A recent study has proposed that T3SS-2α is part of a novel Tn-7 like transposon that has co-opted a mini CRISPR-Cas system to potentially mobilize the entire region (McDonald, Regmi et al. 2019).

Vibrio parahaemolyticus is a marine species that has adapted and evolved to colonize and establish niches in the intestine of humans as well as fish and shellfish species. In these environments, the bacterium must compete with endogenous microbiota and other pathogenic bacteria for available nutrients. Freter's nutrient-niche hypothesis proposes that any species of bacteria should utilize at least one limiting nutrient better than another species for successful colonization (Freter 1988). Previous work has shown that like most *Vibrio* species, *V. parahaemolyticus* can catabolize *N*-acetyl-D-glucosamine, a monomer of chitin highly prevalent in the marine environment. Using a phenotypic array of 190 carbon sources, it was shown

that *V. parahaemolyticus* could utilize 71 different carbon sources under static growth conditions (Kalburge, Carpenter et al. 2017). Studies have shown that strains have also adapted to utilize mouse intestinal mucus and mucus sugars L-arabinose, D-galactose, D-gluconate, D-glucuronate, D-mannose, D-glucosamine, and D-ribose as sole carbon sources (Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017). Previous work has also shown that disruption of global regulation of metabolism has a profound effect on fitness (Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017). For example, deletion of the global transcription factor sigma 54 encoded by *rpoN* enhanced *in vivo* fitness and this correlated with increased rates of carbon metabolism (Whitaker, Richards et al. 2014). Whereas deletion of the quorum sensing response regulator *luxO* inhibited carbon metabolism and caused growth defects (Kalburge, Carpenter et al. 2017).

In the present study, differences among *V. parahaemolyticus* strains in their ability to utilize different carbohydrate sources was determined using bioinformatics, phenotypic and genetic analyses. First, phenotypic differences between a clinical strain RIMD2201633 and an environmental strain UCM-V493 were examined and significant differences in growth on many carbon sources were found, of note was D-galactose utilization amongst others. Identical catabolic gene clusters were present in the two strains, but each strain had a different transporter*;* a major facilitator superfamily (MFS) transporter and a sodium/galactose transporter (SGLT). Growth analysis revealed that RIMD2201633 that contains a MFS transporter had a significant lag phase when grown in D-galactose compared to UCM-V493, an environmental strain lacking T3SS-2 that contained SGLT. Genetic complementation suggested that SGLT is a high efficiency D-galactose transporter, which was the predominant

transporter present in *V. parahaemolyticus*. Our analysis showed that D-galactose utilization is phylogenetically widespread among the *Vibrionaceae*, present in divergent genera and species, but not highly prevalent. A D-gluconate cluster was nearly universal among *V. parahaemolyticus,* but had a restricted distribution within the *Vibrionaceae*. Interestingly, deletion of *eda1* (VP0065), that encodes 2-keto-3-deoxy-6-phosphate-gluconate (KDPG) aldolase (EDA), a key enzyme in the Entner–Doudoroff (ED) pathway, did not inhibit growth on D-gluconate. Bioinformatics identified two additional putative KDPG aldolases (*eda3* (VPA0083) and *eda2* (VPA1708)) in RIMD2210633. All three genes were induced in the presence of D-gluconate. EDA2 and EDA3 were variably distributed among *V. parahaemolyticus* and EDA1 was present in all strains. Genome comparative analysis identified a 73-kb to 166-kb metabolic island that contained a citrate fermentation cluster, an L-rhamnose cluster, and an oligogalacturonide metabolism cluster with an EDA homologue. This island also contained a type I-F CRISPR-Cas system and a Type IVb pilus. These different functional systems were flanked by transposase genes within the island suggesting a mechanism of acquisition as modular units. A second L-rhamnose cluster was present in AQ3810 adjacent to the T3SS-2α region with associated transposons suggesting a mode of acquisition. Two different regions also contained L-arabinose metabolism clusters and most *V. parahaemolyticus* strains had the ability to catabolize this sugar.

Materials and Methods

Bacterial Strains, Media, and Culture Conditions

All bacterial strains and plasmids used for functional analysis in this study are listed in **Table 1**. Bacterial strains used for *in silico* analysis are listed in **Table 2**. Unless stated otherwise, all *V. parahaemolyticus* strains were grown in Lysogeny Broth (LB) medium (Fischer Scientific, Pittsburgh, PA) containing 3% NaCl at 37°C with aeration or M9 medium media (Sigma Aldrich, St. Louis, MO) supplemented with 3% NaCl and carbon sources as required. For Simmons citrate test, bacterial colonies were inoculated by stabbing the Simmons citrate agar slant and the slant was incubated at 37°C for 24 h. Antibiotics were added to growth media at the following concentration: streptomycin (Sm), 200 μg/ml and chloramphenicol (Cm), 10 μg/ml when required.

Table 1 Bacterial strains and plasmids used in this study

Bacterial strains	**Genotype or description**	**References or sources**
Vibrio parahaemolyticus		
RIMD2210633	O3:K6 clinical isolate, Sm^r	(Makino, Oshima et al. 2003)
Δ*eda1*	RIMD2210633 Δ*eda1* (VP0065), Sm^r	This study
UCM-V493	O2:K28 environmental isolate, Sm^r	(Martinez-Urtaza, Lozano-Leon et al. 2004)
Δ*sglt*	UCM-V493 Δ*sglt* (VPUCM_0844), Sm^r	This study
RIMDpSGLT	RIMD2210633 harboring pBBR*sglt*, Cm^r	This study
Escherichia coli		
BW25113	Wild-Type *E. coli* K12 strain	(Grenier, Matteau et al. 2014)
DH5αλpir		Laboratory collection
B2155λpir	*ΔdapA::emr pir*, for bacterial conjugation	Laboratory collection
***Klebsilla pneumoniae* K2**		Laboratory collection
***Vibrio cholerae* N16961**	O1 El Tor, Sm^r	(Heidelberg, Eisen et al. 2000)
Plasmids		
pJet1.2/blunt	Cloning vector/Amp^r	Fermentas
Pjet1.2-*edA1*SOEAD	pJet1.2 harboring truncated *eda1*, Amp^r	This study
pDS132	Suicide plasmid, Cm^r, SacB	(Philippe, Alcaraz et al. 2004)
pDSΔ*eda1*	pDS132 harboring truncated *eda1*, Cm^r	This study
Pjet1.2-*sglt*SOEAD	pJet1.2 harboring truncated *sglt*, Amp^r	This study
Pjet1.2-*sglt*	pJet1.2 harboring *sglt*, Amp^r	This study
pDSΔ*sglt*	pDS132 harboring truncated *sglt*, Cm^r	This study

pBBR1MCS	Expression vector, lacZ promoter, Cm^r	(Kovach, Phillips et al. 1994)
pBBR*sglt*	pBBR1MCS harboring full-length *sglt*, Cm^r	This study

Table 2 Completed genomes of *V. parahaemolyticus* strains used for whole genome analysis.

Strains	**Year**	**Country**	**Source**	**%GC**	**Accession No. (CH I)**	**Accession No. (CH II)**
ATCC17802	1951	Japan	Seafood	45.33	NZ_CP014046.1	NZ_CP014047.1
BB22OP	1980s	Bangladesh	Environment	45.33	NC_019955.1	NC_019971.1
RIMD2210633	1996	Japan	clinical	45.4	NC_004603.1	NC_004605.1
FDAARGOS_191	1996	India	clinical	45.4	NZ_CP020427.1	NZ_CP020428.1
UCM-V493	2002	Spain	Environment	45.32	NZ_CP007004.1	NZ_CP007005.1
1682	2004	Norway	Wound	45.76	NZ_CP019059.1	NZ_CP019060.1
CDC_K4557	2007	USA	clinical	45.34	NC_021848.1	NC_021822.1
FDA_R31	2007	USA	Seafood	45.33	NC_021847.1	NC_021821.1
CHN25	2011	China	Seafood	45.19	NZ_CP010883.1	NZ_CP010884.1
MAVP-Q	2011	USA	Human	45.3	NZ_CP011884.1	NZ_CP011885.1
FORC_004	2014	South Korea	Environment	45.49	NZ_CP009847.1	NZ_CP009848.1
FORC_006	2014	South Korea	Environment	45.33	NZ_CP009765.1	NZ_CP009766.1
FORC_008	2014	South Korea	Seafood	45.44	NZ_CP009982.1	NZ_CP009983.1
FORC_023	2014	South Korea	Environment	45.44	NZ_CP012950.1	NZ_CP012951.1
FORC_014	2015	South Korea	Seafood	45.35	NZ_CP011406.1	NZ_CP011407.1
FORC_018	N/A	South Korea	Seafood	45.44	NZ_CP013826.1	NZ_CP013827.1
FORC_022	N/A	South Korea	Seafood	45.26	NZ_CP013248.1	NZ_CP013249.1

Construction of *V. parahaemolyticus* Deletion Mutants

Splicing by overlapping extension (SOE) PCR and an allelic exchange method (Ho, Hunt et al. 1989) was used to construct in-frame, non-polar deletion mutants of VPUCM_0844 (*sglt*) and VP0065 (*eda1*) in *V. parahaemolyticus* strain UCM-V493 and RIMD2210633, respectively. Briefly, primers were designed to the VPUCM_0844 and VP0065 using *V. parahaemolyticus* UCM-V493 and RIMD2210633 genomic DNA as templates. All primers used in this study are listed in **Table 3**. SOE PCR was conducted to obtain a 117-bp-truncated version of 1632-bp of VPUCM_0844. The *Δsglt* PCR fragment was cloned into the suicide vector pDS132 and named pDS*Δsglt*. pDS*Δsglt* was then transformed into *E. coli* strain β2155 λ*pir* and was conjugated into *V. parahaemolyticus* UCM-V493. Conjugation was conducted by cross streaking both strains on to LB plate containing 0.3 mM DAP. Bacterial growth lawn from this plate was scrapped, resuspended in LB, serially diluted and plated on LB agar plate containing Sm and Cm. Next day, the colonies were verified for single crossover via PCR. The colonies with single crossover were grown over night in LB supplemented with 3% NaCl with no antibiotic added and were plated onto LB 3% NaCl plate containing 10% sucrose to select for double crossover deletion mutant. The gene deletion was confirmed by PCR. A similar protocol was used to create a deletion of VP0065 (*eda1)*.

Table 3 Primers used in this study

Primer name and use	Sequence (5'-3')[a]	Melting temp (°C)
SOE PCR		
VP0065 SOEA	TCTAGAAACCGTAGCGCTAACCACTA	63
VP0065 SOEB	AATGTCACTTCCGCACATGG	59
VP0065 SOEC	CCATGTGCGGAAGTGACATTTGCAATGATGGATAACGGCG	59
VP0065 SOED	GAGCTCCAGGCCCCAATTTTGAGACC	60
VP0065 SOEFF	GTGTGGTGAGTCATTGGCTC	55
VP0065 SOEFR	GCCAATGCGTACGACAAAGA	50
SGLT SOEA	TCTAGACTTCTACCCTTTGAACTTGCAG	58
SGLT SOEB	GACATAAATGGCGAAGACC	51
SGLT SOEC	GGTCTTCGCCATTTATGTCATCGCTGCTTACGGCATAAT	66
SGLT SOED	GAGCTCGGTTTGTATGCACCACCCAC	63
SGLT SOEFF	GGTGAAAACCACCAAACCTG	55
SGLT SOEFR	TGAGGATGGGCGTTCATAAT	54
Quantitative real-time PCR (qPCR) primers		
VP0065 F	ACCGGGTGTAAACAACCCAA	60
VP0065R	TACCGCCACAAGCTACAACC	60
VP1708F	CACGCCAGTAGAAGGTGACA	60
VP1708R	GTTGCCTAGGAACGCAGAAC	59
VPA0083F	GCAAGGCAGTGAAATTGGCA	60
VPA0083R	ACGTTACGCTGTTGGCAGTA	60
VPA1708F	CAGCCGCTGAAATCACCTTC	55
VPA1708R	TTTCTTGGCACGCTTTGACG	50
16S F	ACGGCCTGGGGAGTACGGTC	60
16S R	TTGCGCTCGTTGCGGGACTT	60
Complementation primers		
SGLTforward	TCTAGAGGATGATCCCCGTAATTTCC	58
SGLTreverse	GAGCTCATTTGGCAGCAAGGTTGAATC	61

[a]Underlined base pairs indicate restriction sites

Complementation of *V. parahaemolyticus* RIMD2210633 with *sglt* (VPUCM_0844)

RIMD2210633 was complemented with the putative sodium galactose transporter gene *sglt* (VPUCM_0844) from strain UCM-V493 and the resulting strain was designated as RIMDpSGLT. A pair of primers (**Table 3**) was designed to amplify a copy of VPUCM_0844 (including the endogenous promoter region) from *V. parahaemolyticus* UCM-V493, which was cloned into pJET1.2 and transformed into *E. coli* DH5α λ*pir*. Subsequently, the fragment was subcloned into the vector pBBR1MCS, creating pSGLT, which was transformed into *E. coli* β2155 λ*pir*. pSGLT was transferred into *V. parahaemolyticus* RIMD210633 via conjugation on a LB plate containing 0.3 mM diaminopimelic acid (DAP). The bacterial growth from the conjugation plate was scraped and streaked on LB Sm Cm plates. Colonies were screened to select for *V. parahaemolyticus* RIMD2210633 harboring pSGLT. 1mM isopropyl β-D-1-thiogalactoside (IPTG) was added to the culture to induce RIMDpSGLT for growth analysis.

Comparative Genomics

Whole genome comparison between the *V. parahaemolyticus* strains was conducted using BLASTn from complete genome sequences downloaded from NCBI bacterial genome database. The genome sequences were visualized by using a Python based application, Easyfig to identify regions homologous to each other and regions unique to each strain (Sullivan, Petty et al. 2011). For genes coding hypothetical proteins, putative functions were determined by using HHpred (Soding, Biegert et al. 2005). The Carbohydrate-Active enzymes (CAZymes) database was used to determine the function of many hypothetical proteins (CAZy database: http//www.cazy.org (Lombard, Ramulu et al. 2014).

CRISPR-Cas Analysis

The FASTA files for strains CDC_K5635, 3259, FORC_022 and TUMSAT_H10_H6 were used to identify CRISPR direct repeats and spacers in each sequence using the CRISPRFinder and CRISPRDetect programs (Grissa, Vergnaud et al. 2007, Biswas, Staals et al. 2016). CRISPRtionary program was used to determine how unique each spacer was to each strain and the CRISPRMap program was used to assign type and sub-type to each CRISPR-Cas region (Grissa, Vergnaud et al. 2008, Lange, Alkhnbashi et al. 2013).

RNA Extractions, cDNA Synthesis and qRT-PCR Expression Analysis

Strains were grown at 37°C overnight, with aeration, in 5 ml LBS. Cells were then pelleted at 4000 x g for 10 min, washed twice in PBS and resuspended in 5 ml PBS. The resuspended culture was diluted 1:50 into 25 ml of M9 media supplemented with either 10 mM D-glucose or 10 mM D-gluconate. Cultures were grown with aeration at 37°C until late log phase (4 h). RNA was then extracted using TRIzol according to the manufacturer's instructions (Invitrogen, Carlsbad, CA). Total RNA samples were treated with Turbo DNAse (Invitrogen) according to the manufacturer's instructions to remove contaminating genomic DNA. Subsequently, RNA samples were quantified using a Nanodrop spectrophotometer (Thermo-Fisher Scientific, Waltham, MA). cDNA was synthesized with Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's instructions using 500 ng of RNA as the template and 200 ng of random hexamers in each synthesis reaction. The cDNA samples were diluted 1:50 and used as the template for quantitative real-time PCR (qPCR). The qPCRs were performed using the fast SYBR green master mix (Applied Biosystems) according to the manufacturer's instructions and run on an Applied

Biosystems (ABI) 7500 Fast real time PCR system. Gene primers were designed using Primer 3 and *V. parahaemolyticus* RIMD2210633 genome sequence as the template and are listed in **Table 3**. Data was analyzed using the ABI 7500 software (Applied Biosciences). The expression levels of each gene, determined by their cycle threshold (CT) values, were normalized using the 16S rRNA gene to correct for sampling errors. Differences in gene expression ratios were determined using the previously described ΔΔCT method (Pfaffl 2001).

Phylogenetic Analysis

Representative proteins from each metabolism cluster were used as seed sequences in a protein BLAST (pBLAST) to identify putative homologues in *V. parahaemolyticus* as well as in other members of the family Vibrionaceae. The phylogenetic trees were constructed from the alignment of the amino acid sequences of the identified proteins of interest. Most of the proteins from each species represent multiple strains with the same protein. The software, Molecular Evolutionary Genetic Analysis version X (MEGA X), was used to infer evolutionary history (Le and Gascuel 2008, Kumar, Stecher et al. 2018).

Phenotypic Microarray Analysis of Aerobic Growth on 190 Carbon Sources of RIMD2210633 and UCM-V493

Strains were grown in M9 medium supplemented with 10mM glucose, overnight, at 37°C with aeration (250 rpm). Overnight cultures were then diluted 1:50 into 5 ml M9 medium supplemented with 10mM glucose and grown aerobically for 4 h at 37°C. Cells were then pelleted at 4000 x g for 10 min and washed twice with PBS. Cells were resuspended in 5 ml PBS and diluted 1:50 into fresh M9 medium with no carbon source and 100 µl was transferred to each well of either a PM1A or PM2A

Biolog Phenotypic MicroArray 96 well plate (Biolog, Hayward, CA). Biolog plates were subsequently incubated at 37°C for 24 h and optical density readings at 595 nm (OD_{595}) were taken hourly using a Sunrise Tecan plate reader and Magellan software. Experiments were performed using two biological replicates. To assay the growth patterns of each strain on the various carbon sources, the area under the curve (AUC) over the 24 h of growth was calculated for each carbon source. The first well on each PM plate contained M9 medium plus cells but no carbon source and the area under the curve for this well was used as the blank, which was subtracted from the AUC for each carbon sources to determine which carbon sources *V. parahaemolyticus* exhibited growth. As a cutoff, we considered average AUC below two as no growth. Unpaired student's t-test was performed using Prism GraphPad software to identify significant difference in the AUC.

Results and Discussion

Different D-Galactose Transporters Associated with Different *V. parahaemolyticus* Isolates

Comparative phenotypic growth analysis of RIMD2210633, a clinical strain and UCM-V493, an environmental strain in different carbon sources uncovered significant differences between the two strains in their ability to use different carbon sources (**Table 4**). In addition, to overall biomass differences between the strains, there were also differences in the ability to use certain carbon sources. UCM-V493 was unable to grow in L-arabinose and D-glucuronate, whereas RIMD2210633 was unable to grow in Tween 20, glycyl-L-glutamic acid, propionic acid and chondroitin sulfate C as sole carbon source (**Table 4**). One interesting difference noted between the strains was their ability to utilize D-galactose since both strains contained the catabolic operon

galM galK galT galE. D-galactose, a hexose, is a component of intestinal mucus and was demonstrated to be an essential carbohydrate for *in vivo* fitness of *E. coli* (Conway and Cohen 2015). In *E. coli,* multiple galactose transporters were identified with at least two critical for galactose transport, MglBAC, an ATP binding cassette (ABC) type transporter, and GalP, a H+/symporter type transporter (Rotman, Ganesan et al. 1968). In *V. parahaemolyticus* all strains contain the galactose catabolic cluster, however, no strain contained a homologue of the MglBAC or the GalP type transporters (**Figure 1A and 1B**). A major facilitator superfamily (MFS) transporter was identified in RIMD2210633 and a novel sodium/galactose transporter (SGLT) was present in UCM-V493 adjacent to the *gal* genes (**Figure 1A and Table 5**). The SGLT transporter shares 32% sequence identity (60% similarity) with that of human SGLT1 (hSGLT1) (Turk, Kim et al. 2000).

Table 4 Phenotypic growth analysis of *V. parahaemolyticus* strains RIMD2210633 and UCM-V493

	Significantly higher AUC (p <0.05)		**Area Under the Curve (AUC)**						
Carbon Sources	**UCM-V 493**	**RIMD2 210633**	**UCM-V493**		**RIMD2210633**		**Average AUC**		**P-Value***
			BR1	**BR2**	**BR1**	**BR2**	**UCM-V493**	**RIMD2 210633**	
α-cyclodextrin	Yes		17.66	17.66	14.81	14.97	17.66	14.89	0.001
γ-cyclodextrin	Yes		16.25	16.44	13.82	13.99	16.34	13.90	0.003
D-mannitol			14.58	17.09	11.39	11.94	15.84	11.66	0.083
D-galactose	Yes		15.43	16.08	6.93	2.93	15.76	4.93	0.033
maltotriose			14.83	15.03	12.79	14.63	14.93	13.71	0.319
α-D-glucose			14.47	15.30	13.87	14.05	14.89	13.96	0.162
D-glucosamine	Yes		14.45	14.86	10.17	8.76	14.65	9.47	0.020
D-trehalose			12.55	16.32	12.35	12.48	14.44	12.41	0.396
dextrin	Yes		14.46	13.91	11.39	11.11	14.19	11.25	0.011
maltose			13.06	15.15	12.51	12.48	14.11	12.49	0.262
D-mannose			12.36	15.30	9.11	10.62	13.83	9.86	0.139
D-fructose			12.34	14.89	11.75	11.43	13.62	11.59	0.257
glycogen	Yes		13.57	13.64	11.82	11.32	13.60	11.57	0.015
glycerol	Yes		11.68	13.98	7.09	7.67	12.83	7.38	0.044
***N*-acetyl-D-glucosamine**			11.52	13.88	11.63	11.14	12.70	11.39	0.390
inosine			10.10	11.13	9.43	9.33	10.62	9.38	0.139
D-gluconic acid			9.97	10.95	8.07	8.78	10.46	8.43	0.079

uridine			8.48	11.83	5.60	3.82	10.15	4.71	0.103
D-ribose			9.06	11.10	8.56	6.49	10.08	7.52	0.221
gelatin	Yes		10.21	9.18	3.47	2.93	9.69	3.20	0.008
D-glucose-6-phosphate			9.60	9.72	7.60	8.77	9.66	8.19	0.129
L-glutamine			8.28	10.60	6.51	9.77	9.44	8.14	0.581
thymidine			8.79	9.70	5.76	3.73	9.25	4.75	0.056
L-malic acid			8.39	9.94	6.54	6.69	9.17	6.62	0.082
D, L-malic acid			7.97	10.01	6.38	6.42	8.99	6.40	0.127
L-glutamic acid			7.78	9.96	9.13	10.28	8.87	9.71	0.567
hydroxy- L-proline	Yes		9.05	8.67	7.02	7.37	8.86	7.20	0.023
D-fructose-6-phosphate			8.61	8.99	7.20	8.69	8.80	7.95	0.383
L-proline			7.84	9.66	6.63	7.01	8.75	6.82	0.174
fumaric acid			8.19	9.15	7.31	6.88	8.67	7.10	0.096
β-methyl-D-glucoside	Yes		7.82	8.24	5.56	5.77	8.03	5.67	0.010
putrescine	Yes		8.21	7.74	4.17	4.17	7.97	4.17	0.004
L-threonine			7.10	8.57	4.76	4.90	7.84	4.83	0.055
L-asparagine			6.91	8.65	4.50	2.94	7.78	3.72	0.073
D-malic acid			6.11	9.10	4.43	3.25	7.60	3.84	0.144
L-aspartic acid			6.62	8.01	6.45	3.32	7.32	4.88	0.291
succinic acid			6.93	7.68	6.29	5.67	7.30	5.98	0.114
L-serine			6.04	8.36	3.80	3.34	7.20	3.57	0.092
2-deoxy adenosine			6.36	7.51	5.59	5.36	6.94	5.48	0.130
pyruvic acid	Yes		7.09	6.23	3.09	4.08	6.66	3.59	0.043
D-alanine			5.07	7.89	4.40	4.29	6.48	4.34	0.268
adenosine			5.72	6.92	5.06	5.90	6.32	5.48	0.370
L-alanyl-alycine			5.45	7.17	2.72	3.95	6.31	3.33	0.106
acetic acid	Yes		6.38	6.06	2.57	1.82	6.22	2.20	0.010
α-keto-glutaric acid			6.04	6.01	3.97	5.40	6.03	4.69	0.201
L-histidine	Yes		6.06	5.89	3.44	2.65	5.98	3.04	0.019
L-lactic acid			4.52	7.24	6.94	3.76	5.88	5.35	0.824
methyl pyruvate			5.09	4.86	4.45	3.19	4.98	3.82	0.214

β-cyclodextrin		Yes	4.93	4.68	10.36	9.62	4.80	9.99	0.006
glycyl- L-proline			4.03	5.43	2.42	1.90	4.73	2.16	0.075
Tween 40	Yes		4.39	4.95	3.19	3.43	4.67	3.31	0.047
bromo succinic acid	Yes		4.25	4.55	2.99	3.19	4.40	3.09	0.018
D, L-α-glycerol-phosphate			3.34	4.95	4.37	2.71	4.15	3.54	0.654
glycine			3.72	3.98	4.50	4.97	3.85	4.73	0.082
L-ornithine			3.44	4.19	3.26	3.11	3.82	3.18	0.238
Tween 80	Yes		3.74	3.63	2.19	2.36	3.68	2.28	0.005
L-leucine			3.11	3.48	2.99	4.35	3.29	3.67	0.645
glycyl- L-aspartic acid			4.19	1.46	1.98	5.24	2.83	3.61	0.748
α-hydroxyl glutaric acid-γ-lactone			1.32	3.83	5.27	4.59	2.57	4.93	0.211
L-alanine			2.28	1.89	2.92	4.14	2.09	3.53	0.152
Carbon sources only RIMD2210633, but not UCM-V493 can utilize									
L-arabinose		Yes	0.00	0.00	9.30	9.77	0.00	9.53	ND
D-glucuronic acid		Yes	0.00	0.00	5.73	2.05	0.00	3.89	ND
L-arginine		Yes	0.00	0.00	3.67	3.65	0.00	3.66	ND
Carbon sources only UCM-V493, but not RIMD2210633 can utilize									
Tween 20	Yes		4.95	3.02	0.00	0.00	3.98	0.00	ND
glycyl-L-glutamic acid	Yes		3.75	1.67	0.00	0.00	2.71	0.00	ND
propionic acid	Yes		2.52	5.65	0.00	0.00	4.08	0.00	ND
chondroitin sulfate C	Yes		1.98	2.14	0.00	0.00	2.06	0.00	ND

***, Unpaired student's t-test was performed using Prism GraphPad software, ND, not determined, BR, Biological replicate**

Table 5 Pathogenic and non-pathogenic strains of *V. parahaemolyticus* examined in detail in this study

Strain Name	**Source**	**Year**	**Country**	**Serotype**	**T3SS-2**	**D-galactose**	**D-gluconate**	**D-glucuronate**	**OGA***	**L-arabinose**
BB22OP	N/A	1980	Bangladesh	O4:K8	T3SS-2α	YES/SGLT	YES	NO	YES	YES
AQ3810	Human	1983	Singapore	O3:K6	T3SS-2α	YES (MFS)	YES	YES	YES	YES
SO18	Human	1993	China	O3:K29	T3SS-2α	YES (MFS)	YES	YES	YES	NO
RIMD2210633	Human	1995	Japan	O3:K6	T3SS-2α	YES (MFS)	YES	YES	YES	YES
FDAARGOS_191	Human	1996	India	O3:K6	T3SS-2α	YES (MFS)	YES	YES	YES	YES
Peru-466	Human	1996	Peru	O4:K68	T3SS-2α	YES (MFS)	YES	YES	YES	YES
S073	Human	1997	Thailand	N/A	T3SS-2α	YES (MFS)	YES	YES	YES	NO
AN-5034	Human	1998	Bangladesh	O3:K6	T3SS-2α	YES (MFS)	YES	YES	YES	YES
CFSAN018752	Human	1998	Spain	N/A	T3SS-2α	YES (MFS)	YES	YES	YES	YES
Peru-288	Human	2001	Peru	O1:KIV	T3SS-2α	YES (MFS)	YES	NO	YES	YES
S132	Human	2005	China	N/A	T3SS-2α	YES (MFS)	YES	YES	YES	YES
F30368	Human	2006	Canada	N/A	T3SS-2α	YES (MFS)	YES	YES	YES	YES
3259	Human	2007	USA	O3:K6	T3SS-2α	YES (MFS)	YES	YES	YES	YES
EKP-008	Environment	2007	Bangladesh	N/A	T3SS-2α	YES (MFS)	YES	YES	YES	YES
CFSAN025052	Human	2009	Peru	O3:K59	T3SS-2α	YES (MFS)	YES	YES	YES	NO

FORC_014	Seafood	2015	South Korea	N/A	T3SS-2α	YES/SGLT	YES	YES	YES	YES
ATCC 17802	Seafood	1951	Japan	N/A	T3SS-2β	YES/SGLT	YES	NO	YES	YES
AQ4037	N/A	1985	Maldives	O3:K6	T3SS-2β	YES (MFS)	YES	YES	YES	YES
A4EZ703	Human	2004	Canada	O3:KUT	T3SS-2β	YES (MFS)	YES	YES	YES	YES
10_4243	Human	2006	Canada	OUT:KUT	T3SS-2β	YES (MFS)	YES	YES	YES	YES
10_4244	Human	2006	Canada	O8:KUT	T3SS-2β	YES (MFS)	YES	YES	YES	YES
GCSL_R144	Seafood	2007	USA	O5:KUK	T3SS-2β	YES/SGLT	YES	YES	YES	YES
CDC_K5073	Human	2007	USA	O3:K56	T3SS-2β	YES/SGLT	YES	YES	YES	NO
GCSL_R60	Seafood	2007	USA	O10:KUK	T3SS-2β	YES (MFS)	YES	NO	YES	YES (MFS)
CFSAN007459	Seafood	2010	USA	N/A	T3SS-2β	YES (MFS)	YES	YES	YES	YES
VP2007-095	Human	2007	USA	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
09_4436	Human	2009	Canada	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
MAVP-Q	Human	2011	USA	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
CFSAN007430	Human	2012	USA	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
S487-4	Seafood	2013	Canada	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
CFSAN007434	Human	2013	USA	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
MAVP-E	N/A	2015	USA	N/A	T3SS-2γ	YES/SGLT	YES	NO	YES	NO
UCM-V493	Environment	2002	Spain	O2:K28	Absent	YES/SGLT	YES	NO	YES	NO
1682	Wound	2004	Norway	N/A	Absent	YES/UK	YES	YES	YES	NO

S141	Seafood	2006	China	N/A	Absent	YES/SGLT	YES	YES	YES	YES (MFS)
CDC_K4762	Human	2006	USA	O5:K17	Absent	YES (MFS)	YES	YES	YES	NO
CDC_K4557	Human	2007	USA	O1:K33	Absent	YES (MFS)	YES	NO	YES	YES
CDC_K5635	Human	2007	USA	O5:K30	Absent	YES (MFS)	YES	YES	YES	YES
KVp10	Seafood	2007	Sweden	N/A	Absent	YES (MFS)	YES	YES	YES	YES
CHN25	Seafood	2011	China	O5:KUT	Absent	YES (MFS)	YES	YES	YES	YES
TUMSAT_H10_S6	Seafood	2013	Thailand	N/A	Absent	YES (MFS)	YES	YES	YES	YES (MFS)
FORC_004	Environment	2014	South Korea	N/A	Absent	YES (MFS)	YES	YES	YES	YES
FORC_006	Environment	2014	South Korea	N/A	Absent	YES/SGLT	YES	NO	YES	YES
FORC_008	Seafood	2014	South Korea	N/A	Absent	YES/SGLT	YES	NO	YES	YES (MFS)
FORC_023	Human	2014	South Korea	N/A	Absent	YES (MFS)	YES	YES	YES	NO
RM-13-3	Seafood	2015	Canada	N/A	Absent	YES (MFS)	YES	YES	YES	YES (MFS)
FORC_018	Seafood	N/A	South Korea	N/A	Absent	YES/SGLT	YES	NO	YES	YES (MFS)
FORC_022	Seafood	N/A	South Korea	N/A	Absent	YES/SGLT	YES	YES	YES	YES

YES, catabolic and transporter genes are present. NO, catabolic and transporter genes are missing.

YES /(MFS), catabolic gene are present with a major facilitator superfamily transporter.

YES/SGLT, catabolic genes are present with a sodium galactose transporter.

N/A, not available. UT, untypable. UK, unknown, Oligogalacturonide (OGA), *, OGAs cluster in RIMD 2210633

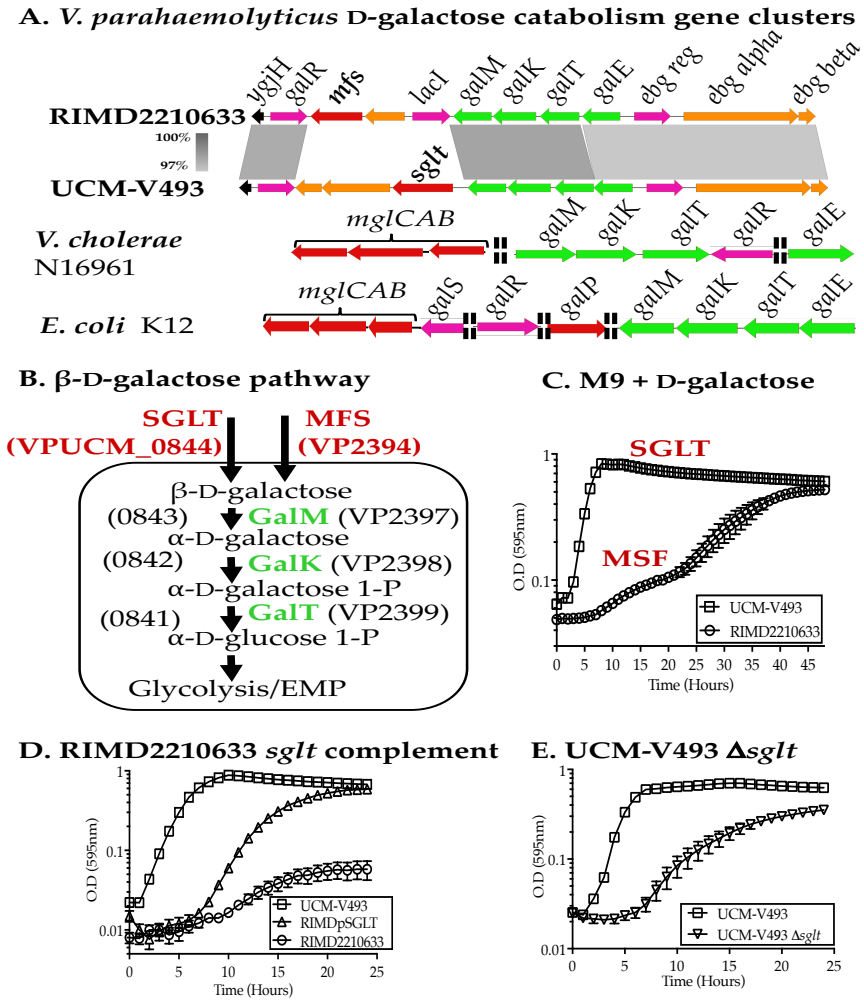


Figure 1 **Figure 1. D-galactose utilization gene cluster**. **A.** Comparative genomic analysis of the D-galactose catabolism and transport region from RIMD2210633, UCM-V493, *V. cholerae* N16961 and *E. coli* MG1655. Gray shade, region of nucleotide homology. **B.** D-galactose utilization pathway showing enzymes and open reading frames (ORFs) designations in RIMD2210633 and UCM-V493. **C.** RIMD2210633 (open circles) andUCM-V493 (open squares) were grown aerobically at 37°C for 48 h in M9 + 10mM D-galactose. **D.** Growth analysis of RIMD2210633, UCM-V493 and RIMDpSGLT (the *sglt* gene (ORF VPUCM_0844) was cloned into RIMD2210633) in M9 + 10 mM D-galactose after 24 h. **E.** Growth analysis of UCM-V493 and Δ*sglt* at 37°C in M9 + 10mM D-galactose.

Further examination of the growth patterns of RIMD2210633 and UCM-V493 in M9 minimal media supplemented with D-galactose as a sole carbon source showed RIMD2210633 had a much longer lag phase and significantly lower biomass compared to UCM-V493 after 24 h **(Figure 1C)**. However, after 48 h, RIMD2210633 reached a similar biomass to UCM-V493 (**Figure 1C)**, which suggested that there were differences in galactose transport efficiency between the two strains. To examine this further, we complemented RIMD2210633 with a functional copy of SGLT and then examined the growth of RIMDpSGLT in M9 D-galactose. RIMDpSGLT showed a significant reduction in lag phase indicating more efficient uptake of D-galactose (**Figure 1D**). To further investigate the role of SGLT in UCM-V493, a *sglt* deletion mutant was constructed in this strain. Growth analysis of the Δ*sglt* mutant in M9 D-galactose showed the mutant had a longer lag phase (around 5-hour lag) with a significantly lower biomass compared to wild type **(Figure 1E)**. This data demonstrated that SGLT is a major D-galactose transporter in UCM-V493 and that this strain harbors an additional uncharacterized D-galactose transporter. A similar finding was demonstrated in an *E. coli galP/ mgl* double mutant, which also showed growth in D-galactose. In that study, the authors showed that uptake in the double mutant was due to facilitated diffusion via PtsG, a glucose specific phosphotransferase system (PTS) permease (Kornberg and Riordan 1976).

Next, we determined the distribution of each of the D-galactose transporters within *V. parahaemolyticus*. To accomplish this, we first examined a subset of 48 strains representing pathogenic and environmental strains whose genome sequence was available in the NCBI genome database **(Table 2 and Table 5)**. The D-galactose MSF transporter was present in a range of pathogenic isolates recovered over 60 years

from different continents. The MFS transporter was also the common type found in non-pathogenic strains, whereas the SGLT system was present mainly within pathogenic strains recovered in the 2000s **(Table 5)**. Examination of the entire *V. parahaemolyticus* genome database (>700 genomes) showed that two thirds of strains contained SGLT and approximately a third of strains contained an MFS type.

Interestingly, the SGLT transporter was present in all strains that contained T3SS-2γ and further analysis showed that these strains lacked L-arabinose and D-glucuronate metabolism clusters (**Table 5**). Strains that contain T3SS-2γ belong to the newly emerged pathogenic clone ST631 identified in the USA in the late 2000s (Xu, Ilyas et al. 2015, Xu, Gonzalez-Escalona et al. 2017, Xu, Gonzalez-Escalona et al. 2017). One could speculate that acquisition of a highly efficient D-galactose transporter could give strains a competitive advantage in a low D-galactose niche or outcompete other species or strains in a D-galactose replete environment.

Phylogenetic Analysis of D-Galactose Catabolism and Transporter Proteins

In order to determine the evolutionary history of galactose catabolism and transport, we reconstructed the phylogeny of GalM, SGLT and MSF proteins. Proteins representing homologues within the *Vibrionaceae* were aligned using ClustalW and phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei 1987, Thompson, Higgins et al. 1994) (**Figure 2**).

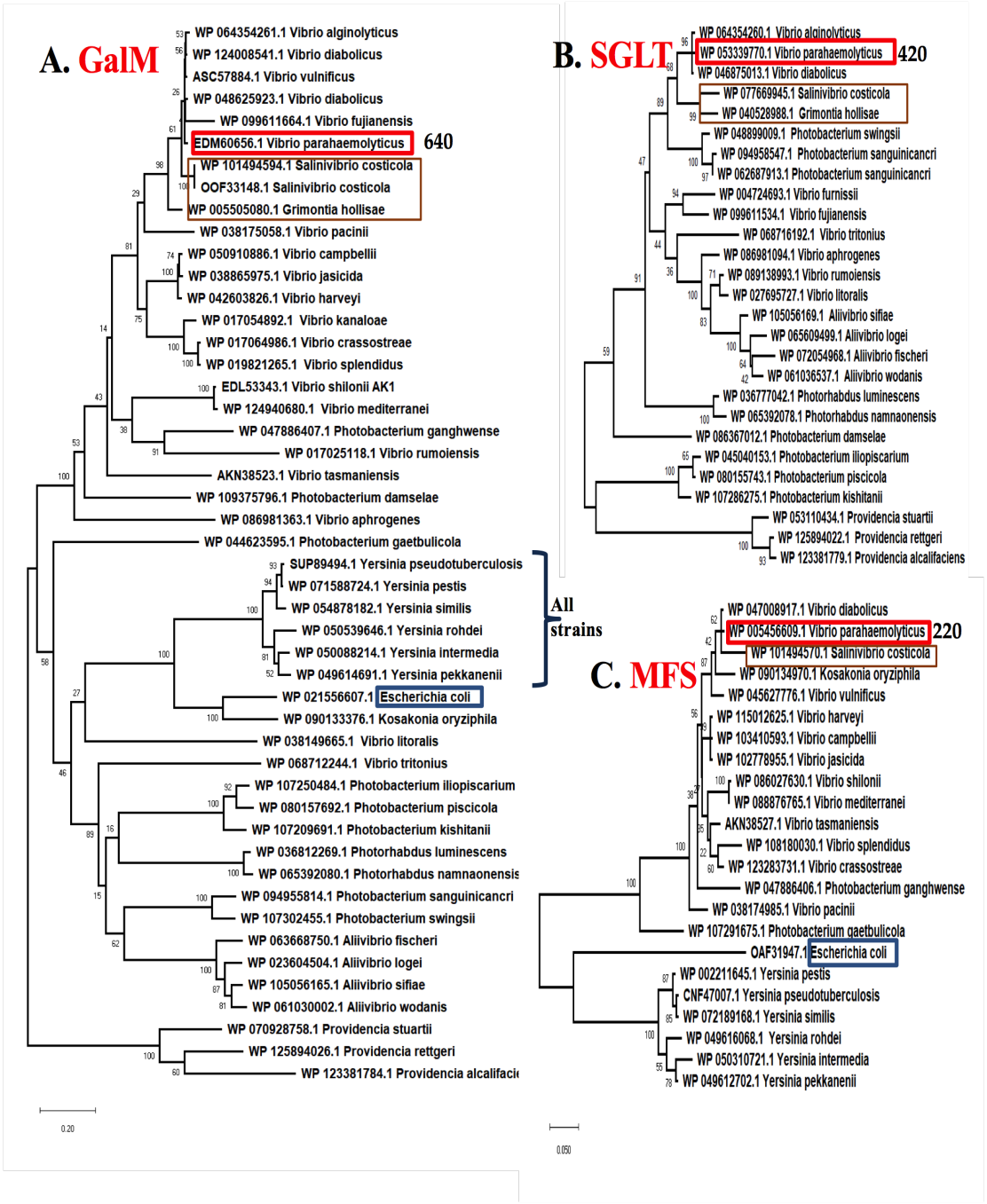


Figure 2 **Phylogenetic analysis of GalM, MFS, and SGLT.** Homologues of GalM, SGLT and MFS proteins from members of the *Vibrionaceae* were obtained from the NCBI genome database. The evolutionary history of each protein was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The percentage of replicate trees in which the associated taxa

clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). Each tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Dayhoff matrix based method (2019) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X (Kumar, Stecher et al. 2018).

Within the Campbellii clade, which consists of *V. campbellii, V. harveyi, V. alginolyticus, V. parahaemolyticus, V. diabolicus* and *V. jasicida*, all GalM proteins clustered together on the tree (**Figure 2A**). However, nested within this cluster was GalM proteins from *Salinivibrio costicola* and *Grimontia hollisae*, which are distantly related to members of the Campbellii clade, which suggests recent horizontal transfer between these groups. In addition, the GalM protein from *Alliivibrio* and *Photobacterium* species clustered divergently from most *Vibrio* species and were more closely related to the GalM from *Yersinia* and *E. coli*. These data suggest very different evolutionary origins for D-galactose clusters among the *Vibrionaceae* and D-galactose utilization is phylogenetically widespread but not highly prevalent (**Figure 2A**). Phylogenetic analysis of the SGLT protein associated with the *gal* operon in *V. parahaemolyticus* clustered closely with SGLT from *V. alginolyticus* and *V. diabolicus* strains and highly related to these was SGLT from *Salinivibrio costicola* and *Grimontia hollisae* (**Figure 2B**). SGLT was also the predominant transporter present in strains of *Alliivibrio* and *Photobacterium*, although the SGLT protein from *Photobacterium* species was more closely related to SGLT from strains of *Providencia* (**Figure 2B**). The MFS protein from *V. parahaemolyticus* clustered together with MFS protein from *V. diabolicus* and *S. costicola* indicating that in these species, similar to *V. parahaemolyticus,* different transporters can be associated with the D-galactose

catabolism cluster depending on the strain. Other members of the Campbellii group (*V. campbellii, V. harveyi, V. jasicida*) contained an MFS protein and clustered tightly together (short branch lengths). Distantly related to these was MFS proteins from *Yersinia* and *E. coli* present on separate divergent branches. Similar to the GalM protein, the SGLT and MFS proteins clustering patterns suggest multiple acquisition events.

Three Entner-Doudoroff (ED) aldolases (EDA) (VP0065, VPA0083 and VPA1708)

Entner-Doudoroff (ED) pathway, an alternative to the Embden-Meyerhof-Parnas (EMP) glycolytic pathway, utilizes two major enzymes, 6-phosphogluconate dehydratase (EDD) and 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase (EDA) to convert glucose into pyruvate and glyceraldehyde-3-P (Conway 1992)**.** An intermediate of this pathway, 6-phosphogluconate, can also be generated after phosphorylation of D-gluconate by gluconokinase (*gntK*) upon uptake. D-gluconate is a six-carbon acid, which is one step more oxidized than glucose and is present in a wide variety of environments for use as a carbon and energy source for many bacterial species. Enteric bacteria typically use the ED pathway for the inducible metabolism of D-gluconate. EDA is involved in D-glucuronate, D-galactonate, gluconides, and possibly glyoxylate metabolism, some of whose metabolic pathways were previously identified in *V. parahaemolyticus* (Suvorova, Tutukina et al. 2011). Both RIMD2210633 and UCM-V493 grew on D-gluconate as a sole carbon source, although similar to other sugars, UCM-V493 grew to a higher OD (**Table 4**). In *V. parahaemolyticus*, the predicted D-gluconate catabolism genes are clustered together

within chromosome 1 with *gntK-edd* divergently transcribed from *gntT-eda*, whereas in *E. coli* these genes are more dispersed in the chromosome (**Figure 3A**).

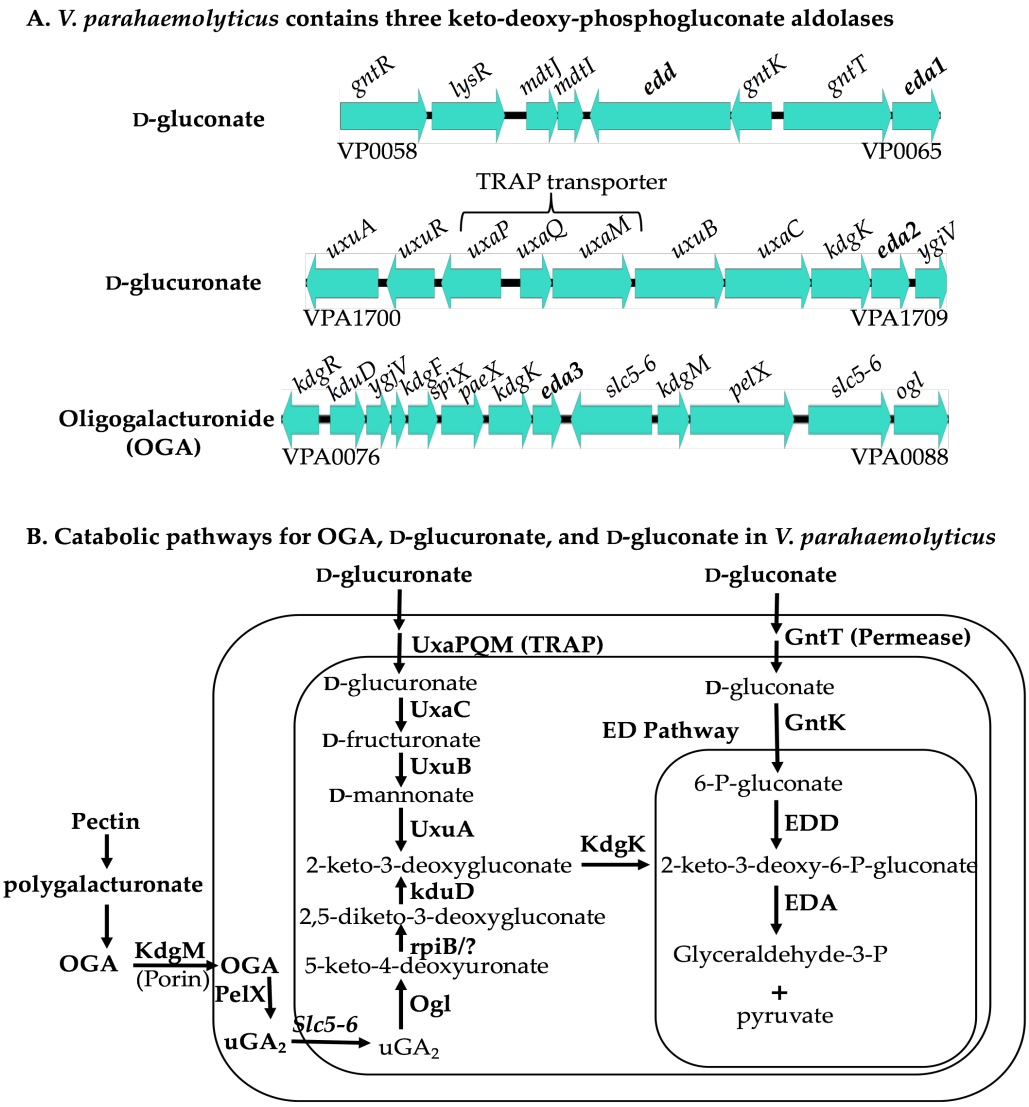
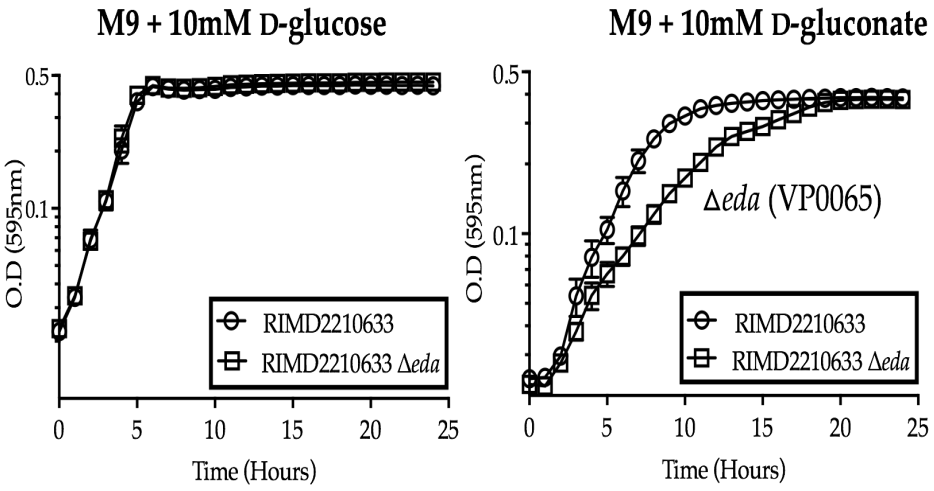


Figure 3 **Three keto-deoxy-phosphogluconate aldolases (EDA1, EDA2, and EDA3). A.** Gene cluster of D-gluconate, D-glucuronate and oligogalacturonide (OGA) present in *V. parahaemolyticus* RIMD2210633. **B**. Predicted catabolic pathway for the utilization of OGA, D-glucuronate and D-gluconate based on the ORFs present in *V. parahaemolyticus* RIMD2210633.

In *V. cholerae*, the ED pathway is an essential requirement for D-gluconate catabolism (Patra, Koley et al. 2012). We constructed an in-frame deletion of *eda* (VP0065) and examined growth in M9 minimal media supplemented with glucose (M9G) and M9 supplemented with gluconate (M9Gnt). Both mutant and wild type strains showed identical growth patterns when grown on M9G (**Figure 4A**). Surprisingly, the *eda* (VP0065) mutant was still capable of growing in M9Gnt as the sole carbon source (**Figure 4A**). Using VP0065 as a seed, BLAST analysis identified two additional EDA homologues within *V. parahaemolyticus* RIMD2210633. VPA0083 (EDA3) and VPA1708 (EDA*2*) showed 78% and 60% amino acid identity to VP0065 (EDA1), respectively. VPA0083 was within a cluster for oligogalacturonide (OGA) metabolism and VPA1708 was present within a cluster for the metabolism of D-glucuronate (**Figure 3A and 3B**). RIMD2210633 grew on D-glucuronate as a sole carbon source but UCM-V493 did not (**Table 4**). The D-glucuronate gene cluster is absent from UCM-V493.

A. Growth analysis of wild type and Δ*eda1* mutant



B. Expression analysis of wild type and Δ*eda1* mutant

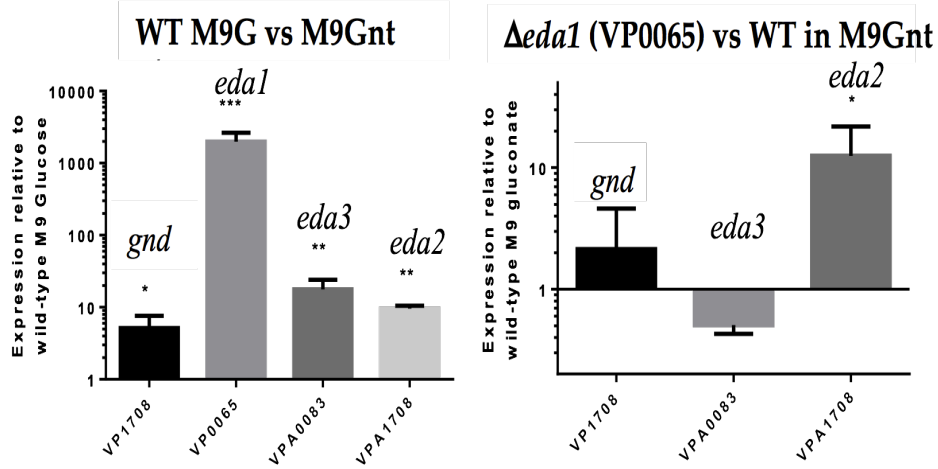


Figure 4 **D-glucoconate metabolism. A.** Growth analysis of *V. parahaemolyticus* RIMD2210633 and Δ*eda1* mutant strain in M9 supplemented with 10 mM D-glucose and 10 mM D-gluconate. **B.** Expression analysis of the first gene in the pentose phosphate pathway (VP1708) and the three putative aldolases (VP0065, VPA0083, VPA1708) in wild-type strain in gluconate relative to the expression of these genes in wild-type in glucose (Courtesy of Brandy Haines-Menges, Ph.D. Dissertation 2015). Expression analysis of VP1708 and the two putative aldolases (VPA0083, VPA1708) in the Δ*eda1* mutant strain in gluconate relative to the expression of these genes in the wild-type in gluconate (. *p<0.05, **p<0.01, ***p<0.001

The *eda1* mutant compared to wild type had a longer doubling time but reached a similar final biomass to wild type, which suggested that the *eda1* mutant may have been complemented by either *eda2* or *eda3* (**Figure 4A**). Expression analysis in M9G compared to M9Gnt showed that all three *eda* genes were induced, with *eda1* showing the highest expression levels in M9Gnt (**Figure 4B**). We also examined *gnd,* which encodes phosphogluconate dehydrogenase, a key enzyme in the Pentose Phosphate (PP) pathway, which can serve as an alternative pathway in gluconate catabolism in some species. The expression data showed a slight induction of *gnd* (VP1708), which suggests this pathway is likely not a major player in D-gluconate catabolism but does suggest the pathway is functional (**Figure 4B**). Next, we examined expression of *eda2, eda3* and *gnd* in the Δ*eda1* mutant compared to wild type in M9Gnt and found that only *eda2* (VPA1708) was significantly induced (**Figure 4B**). This suggested that *eda2* may be compensating for the lack of *eda1* when grown on D-gluconate as a sole carbon source.

Bioinformatics analysis revealed that EDA1 (VP0065) is present in all *V. parahaemolyticus* and many members of the Campbellii clade, in general. In order to determine the evolutionary relationships and history of EDA, we reconstructed the

phylogeny of EDA homologues present in the *Vibrionaceae* family (**Figure 5**). This analysis indicated that EDA1 from members of the Campbellii clade are all closely related clustering together on the tree with short branch lengths (**Figure 5**). Clustering with this group is the EDA1 protein from *V. vulnificus*, a species that is distantly related to the Campbellii clade, indicating that EDA1 has a shared evolutionary history in these species (**Figure 5**). EDA1 is also present in most *V. cholerae* strains and clustered with EDA1 from *V. metoecus*, a close relative of this species. Within one of the most divergent branches of the EDA tree are proteins from *V. fluvalis* and *V. furnissii,* which are member of the Cholerae clade. This suggests the EDA1 was acquired independently within these closely related species, although alternatively in *V. fluvalis* and *V. furnissii*, EDA1 could be evolving faster. Overall, phylogenetic analysis suggests that EDA1 is phylogenetically widespread in the *Vibrionaceae* but is only present in a small subset of members of the family.

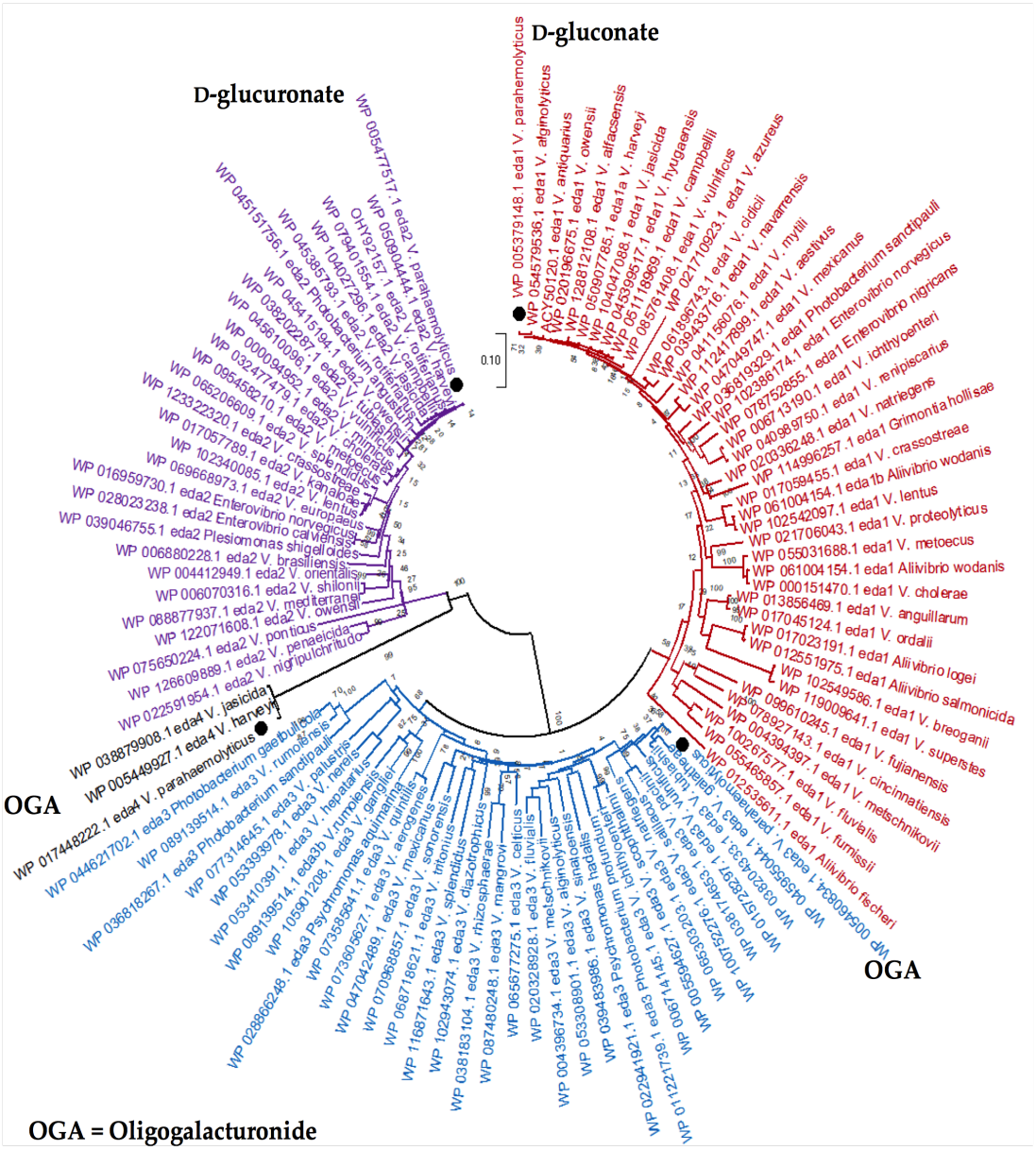


Figure 5 **Phylogenetic analysis of EDA1, EDA2, EDA3 and EDA4.** This analysis involved 108 amino acid sequences representing EDA1, EDA2, EDA3 and EDA4 homologues present in the *Vibrionaceae*. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 6.21447650 is shown. The percentage of replicate trees in which the

associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Dayhoff matrix based method (2019) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 213 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar, Stecher et al. 2018).

EDA2 was present within the D-glucuronate metabolism cluster in *V. parahaemolyticus* (**Figure 3A**). Similar to EDA1, EDA2 is prevalent among the Campbellii clade members, however it is not always present in all members of a species (**Figure 5**). For example, about half the *V. parahaemolyticus* genomes (>700 genomes) in the database contained EDA2. EDA2 was present in a handful of *V. cholerae* strains and the phylogeny of EDA2 suggests it was lost from *V. cholerae* given it is present in the closely related species *V. metoecus, V. mimicus,* and *V. fluvalis,* which all cluster together and most representatives contained the protein. EDA2 had a limited distribution among *Vibrionaceae* in general and was distantly related to EDA1. EDA3 is associated with an OGA metabolism cluster, which is involved in pectin metabolism and part of the KdgR regulon in *V. parahaemolyticus* (Rodionov, Gelfand et al. 2004). EDA3 is more closely related to EDA1 than EDA2. EDA3 is present in all strains of this species and in all strains of *V. alginolyticus,* but these closely related species are present on divergent branches on the EDA3 tree. EDA3 is present in all *V. vulnificus* isolates and clustered closely with EDA3 from *V. parahaemolyticus* with no other major representatives of the Campbellii clade containing this protein (**Figure 5**).

Whole Genome Comparative Analysis

The differences amongst strains in the presence of hexuronate catabolic clusters led us first to perform whole genome comparative analysis among 17 *V. parahaemolyticus* completed genomes (**Table 2**). These 17 strains represented clinical strains that contain T3SS-2 and environmental strains that do not recovered from a range of different sources between 1951 and 2015 from Asia, Europe, South and North America (**Table 2**). A summary of the regions of difference is found in (**Table 6**). We identified several distinct metabolic regions that were variably present among the strains. In particular, a metabolic island was present within chromosome 2 that contained a citrate fermentation cluster, an L-rhamnose cluster, and a second OGA metabolism cluster that also contained a KDPG aldolase that we named EDA4. A second region was identified that contained a non-homologous L-rhamnose metabolism cluster. A limited number of strains contained an island with an L-arabinogalactan metabolism cluster.

Table 6 Unique regions (10-kb or >) present in the 16 *V. parahaemolyticus* strains that are absent in RIMD2210633.

Strain	Chr.	Locus Tag of the region	ORFs flanking the region	Size (kb)	Int /Tran	Region Type	%GC
BB22OP	Chr 1	VPBB_RS00865 - VPBB_RS00945	VP0175-VP0176	13.5	Integrase	UN	43
	Chr 1	VPBB_RS01050 - VPBB_RS01130	VP0195-VP0213	15.3	-	LPS biosynthesis	40
	Chr 1	VPBB_RS01165-VPBB_RS01265	VP0219-VP0236	25.6	-	LPS biosynthesis	39
	Chr 1	VPBB_RS08315-VPBB_RS08725	VP1783-VP1865	61.8	Integrase	Integron	40.4
	Chr 1	VPBB_RS10940-VPBB_RS11015	VP2342-VP2343	24.9	-	RM system	41.8
	Chr 1	VPBB_RS13075-VPBB_RS13285	VP2727-VP2729	32.7	Integrase	Phage	49.9
	Chr 1	VPBB_RS15025 - VPBB_RS15095	VP3056-VP3058	25.2	Integrase	UN	41.2
UCM-V493	Chr 1	VPUCM_RS00970-VPUCM_RS01065	VP0189-VP0212	19.7	-	LPS biosynthesis	42.2
	Chr 1	VPUCM_RS01075-VPUCM_RS01340	VP0212-VP0237	63.5	-	LPS biosynthesis	40.6
	Chr 1	VPUCM_RS03050-VPUCM_RS03115	VP0617-VP0618	16.4	Integrase	Conjugal transfer	40.4
	Chr 1	VPUCM_RS03500-VPUCM_RS04030	VP2440-VP2441	110	Integrase	ICE	47.5
	Chr 1	VPUCM_RS05330-VPUCM_RS05400	VP0953-VP0954	19.9	Transposase	CRISPR system	39.4
	Chr 1	VPUCM_RS05810-VPUCM_RS05895	VP1070-VP1088	35.6	Integrase	RM system	42.7
	Chr 1	VPUCM_RS24170-VPUCM_RS07490	VP1856-VP1462	103	Integrase	Integron	40.3
	Chr 1	VPUCM_RS08200-VPUCM_RS08285	VP1547-VP1587	10.2	Integrase	Phage	43
	Chr 1	VPUCM_RS11985-VPUCM_RS12020	VP0805-VP0806	14.6	-	Cellulose biosynthesis	44.7
	Chr 1	VPUCM_RS13835-VPUCM_RS13875	VP2649-VP2650	11.1	Integrase	UN	42
FDA_R31	Chr 1	HK86_RS03000-HK86_RS03085	VP0218-VP0236	19.9	-	LPS biosynthesis	36.8
	Chr 1	HK86_RS03750-HK86_RS03795	VP0379-VP0404	13.3	Integrase	UN	39.5
	Chr 1	HK86_RS04845-HK86_RS04900	VP0617-VP0618	14.8	Integrase	Conjugal transfer	42.7
	Chr 1	HK86_RS05840-HK86_RS05875	VP805-VP0806	14.6	-	Cellulose biosynthesis	44.7
	Chr 1	HK86_RS09445-HK86_RS09520	VP1523-VP1524	13.9	Transposases	UN	42
	Chr 1	HK86_RS10185-HK86_RS24975	VP1783-VP1865	119	Integrase	Integron	41.5
	Chr 1	HK86_RS24110-HK86_RS11345	VP1883-VP1892	57	Integrase	Phage	43.7
	Chr 1	HK86_RS12400-HK86_RS12460	VP2108-VP2109	14.6	Integrase	UN	39.4
	Chr 1	HK86_RS12680-HK86_RS12880	VP2168-VP2169	33.6	Integrase	Phage	45.5
	Chr 2	HK86_RS18585-HK86_RS18660	VPA0195-VPA0196	17.5	Integrase	UN	42.8
	Chr 2	HK86_RS19120-HK86_RS19140	VPA0296-VPA0297	13.2	-	α-1,2-mannosidase	49.8
	Chr 2	HK86_RS20855-HK86_RS20950	VPA0712-VPA0713	15.5	-	UN	41.5
	Chr 2	HK86_RS23745-HK86_RS15915	VPA0403-VPA1317	80	Transposase	T3SS-2γ	40
FORC_014	Chr1	FORC14_RS00845-FORC14_RS00910	VP0067-VP0068	16.8	Integrase	UN	40

	Chr 1	FORC14_RS01530-FORC14_RS01590	VP0196-VP0211	13.3	-	LPS biosynthesis	41.4
	Chr 1	FORC14_RS01615-FORC14_RS01745	VP0214-VP0237	32.7	-	LPS biosynthesis	38.8
	Chr 1	FORC14_RS04365-FORC14_RS04400	VP0805-VP0806	14.6	-	Cellulose biosynthesis	44.7
	Chr 1	FORC14_RS05365-FORC14_RS05555	VP1003-VP1004	34.3	Integrase	Phage	45.8
	Chr 1	FORC14_RS08365-FORC14_RS08400	VP1552-VP1565	10	-	Phage	42
	Chr 1	FORC14_RS09390-FORC14_RS09985	VP1355-VP1865	88.4	Integrase	Integron	40.5
	Chr 1	FORC14_RS10085-FORC14_RS10435	VP1883-VP1892	65.2	Transposase	Phage	43.4
	Chr 1	FORC14_RS14765-FORC14_RS14830	VP3028-VP3029	14.5	Integrase	UN	41.5
	Chr 2	FORC14_RS17495-FORC14_RS17515	VPA0296-VPA0297	13.3	-	α-1,2-mannosidase	49.9
	Chr 2	FORC14_RS20065-FORC14_RS20155	VPA0888-VPA0895	11.8	-	Phage	43.8
	Chr 2	FORC14_RS21095-FORC14_RS21145	VPA0635-VPA0636	10.8	-	UN	45.7
	Chr 2	FORC14_RS22010-FORC14_RS22065	VPA0432-VPA0449	15.1	Integrase	UN	42.2
	Chr 2	FORC14_RS25860-FORC14_RS22845	VPA1507-VPA1509	27.8	Integrase	UN	40.42
ATCC17820	Chr 1	AL464_RS02600-AL464_RS02710	VP0218-VP0236	26.3	-	LPS biosynthesis	37.4
	Chr 1	AL464_RS23435-AL464_RS02955	VP0175-VP0176	12.8	Transposase	UN	39.9
	Chr 1	AL464_RS05340-AL464_RS05420	VP2638-VP2639	21.9	Integrase	RM system	43.5
	Chr 1	AL464_RS10065-AL464_RS10745	VP1784-VP1865	106.6	Integrase	Integron	40.7
	Chr 2	AL464_RS25385-AL464_RS17080	VPA1308-VPA1397	93	Transposase	T3SS-2β	39
	Chr 2	AL464_RS17725-AL464_RS24675	VPA0798-VPA0801	42.1	Integrase	UN	40.7
	Chr 2	AL464_RS20095-AL464_RS20115	VPA0296-VPA0297	13.2	-	α-1,2-mannosidase	50
FORC_004	Chr 1	FORC4_RS01065-FORC4_RS01120	VP0195-VP0213	14.1	-	LPS biosynthesis	39.4
	Chr 1	FORC4_RS01155-FORC4_RS01260	VP0218-VP0236	25.2	-	LPS biosynthesis	38.7
	Chr 1	FORC4_RS02980-FORC4_RS03050	VP0617-VP0618	16.6	Integrase	Conjugal transfer	44.09
	Chr 1	FORC4_RS06570-FORC4_RS06670	VP1357-VP1369	25.7	-	lipid, AA metabolism, efflux	46.2
	Chr 1	FORC4_RS07540-FORC4_RS07655	VP1549-VP1565	16.4	-	Phage	45
	Chr 1	FORC4_RS24005-FORC4_RS09260	VP1783-VP1865	89.3	Integrase	Integron	40.3
	Chr 1	FORC4_RS08665-FORC4_RS09260	VP2638-VP1639	17.9	Integrase	RM system	41.7
	Chr 2	FORC4_RS13055-FORC4_RS13135	VPA0296-VPA0297	13.2	-	α-1,2-mannosidase	49.8
FORC_006	Chr 1	FORC6_RS03945-FORC6_RS03980	VP0805-VP0806	14.6	-	Cellulose biosynthesis	44.7
	Chr 1	FORC6_RS08430-FORC6_RS09645	VP1783-VP1865	188.1	Integrase	Integron	41.4
	Chr 1	FORC6_RS11335-FORC6_RS11395	VP879-VP0880	21.7	-	Sulphur modification system	41.1
	Chr 1	FORC6_RS13725-FORC6_RS13820	VP0375-VP0404	17.6	Integrase	Conjugal transfer	45.6

	Chr 1	FORC6_RS14450-FORC6_RS14830	VP0189-VP0236	82.1	-	LPS and T2SS	41.4
	Chr 2	FORC6_RS17325-FORC6_RS17345	VPA0296-VPA0297	13.3	-	α-1,2-mannosidase	50.1
FORC_008/	Chr 1	FORC8_RS05985-FORC8_RS06625	VP1783-VP1865	104.4	Integrase	Integron	39.9
FORC_018	Chr 1	FORC8_RS11985-FORC8_RS12080	VP0636-VP0644	21.7	Integrase	RM system	40.4
	Chr 1	FORC8_RS13850-FORC8_RS13945	VP0218-VP0239	22.4	-	LPS biosynthesis	37.5
	Chr 1	FORC8_RS13990-FORC8_RS14050	VP0196-VP0211	10.9	-	LPS biosynthesis	39.2
	Chr 2	FORC8_RS17205-FORC8_RS17250	VPA0441-VPA0449	13.2	-	L-arabinofuranose	48.8
	Chr 2	FORC8_RS20560-FORC8_RS20790	VPA1158-VPA1159	36.7	Integrase	Phage	43.9
FORC_022	Chr 1	FORC22_RS01055-FORC22_RS01115	VP0218-VP0237	16.9	-	LPS biosynthesis	36.2
	Chr 1	FORC22_RS05280-FORC22_RS05400	VP1070-VP1086	22	Integrase	UN	40.6
	Chr 1	FORC22_RS06720-FORC22_RS06790	VP1354-VP1888	11.4	Integrase	Phage	42.1
	Chr 1	FORC22_RS06820-FORC22_RS07010	VP1355-VP1891	31.7	-	UN	41.7
	Chr 1	FORC22_RS07990-FORC22_RS08080	VP1583-VP1584	13.5	-	UN	44.3
	Chr 1	FORC22_RS09055-FORC22_RS10255	VP1786-VP1865	124.3	Integrase	Integron	40.2
	Chr 2	FORC22_RS19725-FORC22_RS20270	VPA0712-VPA0713	135	Integrase	Metabolism and defense	45.3
FORC_023	Chr 1	FORC23_RS00045-FORC23_RS00140	VP0001-VP0002	21.4	Integrase	RM system	43
	Chr 1	FORC23_RS00430-FORC23_RS00485	VP3028-VP3031	11.8	Intgrase	UN	41.4
	Chr 1	FORC23_RS06330-FORC23_RS06925	VP1366-VP1865	103.2	Integrase	Integron	40.1
	Chr 1	FORC23_RS12150-FORC23_RS12205	VP0805-VP0806	14.6	-	Cellulose biosynthesis	44.7
	Chr 1	FORC23_RS12150-FORC23_RS12205	VP0617-VP0618	14.8	Integrase	Conjugal transfer	42.6
	Chr 1	FORC23_RS13900-FORC23_RS14065	VP0218-VP0239	40.3	-	LPS biosynthesis	38.6
	Chr 1	FORC23_RS14090-FORC23_RS14150	VPA0196-VPA0211	13.3	-	LPS biosynthesis	41.3
	Chr2	FORC23_RS17610-FORC23_RS17720	VPA1158-VPA1159	25.7	Transposase	L-rhamnose catabolism	43.7
CDC_K4557	Chr 1	HK81_RS24030-HK81_RS12825	VP1783-VP1865	94.4	Integrase	Integron	40.5
	Chr 1	HK81_RS16895-HK81_RS16995	VP0880-VP0879	28.5	Integrase	UN	38.8
	Chr 1	HK81_RS20140-HK81_RS20225	VP0239-VP0218	22.2	-	LPS biosynthesis	40.2
	Chr 1	HK81_RS23405-HK81_RS20360	VP0211-VP0196	14.2	-	LPS biosynthesis	42.2
	Chr 1	HK81_RS20415-HK81_RS20470	VP0175-VP0176	12.4	-	UK	40
	Chr 1	HK81_RS22085-HK81_RS22185	VP2939-VP2930	20.6	Integrase	RM system	43.9
	Chr 2	HK81_RS02645-HK81_RS02920	VPA0712-VPA0713	73	Integrase	Citrate metabolic island	42.6
	Chr 2	HK81_RS03755-HK81_RS03815	VPA0888-VPA0913	11.5	-	Phage	43.7
CHN25	Chr 1	vpachn25_RS00840-vpachn25_RS00970	VP0175-VP0176	28.6	Integrase	UN	42.6
	Chr 1	vpachn25_RS01075-vpachn25_RS01165	VP0196-VP0197	22.1	-	LPS biosynthesis	41
	Chr 1	vpachn25_RS01190-vpachn25_RS01320	VP0200-VP0239	33.4	-	LPS biosynthesis	39

	Chr 1	vpachn25_RS03110-vpachn25_RS03210	VP0636-VP0644	27.7	Integrase	RM system	40.4
	Chr 1	vpachn25_RS03730-vpachn25_RS03980	VP0747-VP0749	34.6	Transposon	Phage	46.1
	Chr 1	vpachn25_RS04240-vpachn25_RS04275	VP0805-VP0806	14.6	-	Cellulose biosynthesis	44.8
	Chr 1	vpachn25_RS05240-vpachn25_RS05445	VP1003-VP1004	35.3	Integrase	Phage	45.5
	Chr 1	vpachn25_RS05770-vpachn25_RS05890	VP1070-VP1088	36	Integrase	UN	40.7
	Chr 1	vpachn25_RS07975-vpachn25_RS08015	VP1547-VP1587	10.1	Integrase	UN	43.17
	Chr 1	vpachn25_RS08085-vpachn25_RS08395	VP1604-VP1605	51.8	-	Phase	42.8
	Chr 1	vpachn25_RS09160-vpachn25_RS09235	VP1357-VP1865	17.7	Integrase	Integron	40.3
	Chr 1	vpachn25_RS09535-vpachn25_RS09570	VP1934-VP1935	10.66	Integrase	UN	40.3
	Chr 1	vpachn25_RS11955-vpachn25_RS25475	VP2440-VP2441	88.1	Integrase/Transposase	ICE	47
	Chr 2	vpachn25_RS16500-vpachn25_RS16535	VPA0141-VPA0142	13.4	Integrase	UN	36.5
	Chr 2	vpachn25_RS17285-vpachn25_RS17300	VPA0296-VPA0297	13.3	-	α-1,2-mannosidase	50
	Chr 2	vpachn25_RS19890-vpachn25_RS19995	VP0888-VPA0913	18.4	Integrase	Phase	42.5
	Chr 2	vpachn25_RS20975-vpachn25_RS21225	VPA0611-VPA0612	61.5	Integrase	L-arabinogalactan	40.6
MAVP-Q	Chr 1	AB831_RS06405-AB831_RS07615	VP1355-VP1865	131.4	Integrase	Integron	40.43
	Chr 1	AB831_RS09455-AB831_RS09720	VP1783-VP1891	37.6	-	UN	42.1
	Chr 1	AB831_RS09750-AB831_RS09820	VP1354-VP1888	11.4	Integrase	Phage	42.2
	Chr 1	AB831_RS11700-AB831_RS11795	VP0954-VP0953	32.6	Integrase/Transposase	CRISPR systems	42.4
	Chr 1	AB831_RS12555-AB831_RS12590	VP0805-VP0806	14.6	-	Cellulose biosynthesis	44.6
	Chr 1	AB831_RS14725-AB831_RS14820	VP0379-VP0404	19.3	Integrase	Conjugal transfer	41.3
	Chr 1	AB831_RS15465-AB831_RS15625	VP0213-VP0239	42.6	-	LPS biosynthesis	40.1
	Chr 1	AB831_RS15640-AB831_RS15775	VP0189-VP0212	24.7	-	LPS biosynthesis	41.1
	Chr 2	AB831_RS18425-AB831_RS18445	VPA0296-VPA0297	13.2	-	α-1,2-mannosidase	50.2
	Chr 2	AB831_RS23330-AB831_RS23740	VPA0403-VPA1258	79.5	Transposase	T3SS-2ɣ	40.7
	Chr 2	AB831_RS25465-AB831_RS25540	VPA1737-VPA1738	14.53	-	UN	39.1

Note: FORC_008 and FORC_018 have identical genome, FDAARGOS_191 has no unique regions and 1682 has no unique regions more than 10-kb

A Metabolic Island That Harbors a Citrate Fermentation Gene Cluster

Comparative genome analysis of RIMD2210633 and CDC_K4557 identified a 73-kb genomic island on chromosome 2 between VPA0712 and VPA0713 relative to RIMD2210633, which lacked the region. The metabolic island showed all the features of a horizontally acquired region; it was flanked by an integrase at the 3' end, contained attachment sites attL and attR, which indicated site specific integration, and had a %GC content of 42.6 % lower than 45.35 % of the entire genome (**Figure 6A**) (Hacker, Blum-Oehler et al. 1997, Hacker and Kaper 2000, Dobrindt, Hochhut et al. 2004, Boyd, Almagro-Moreno et al. 2009, Napolitano, Almagro-Moreno et al. 2011). The island contained a complete citrate fermentation system (**Figure 6A and 6B**). In *V. cholerae* citrate fermentation is a hallmark physiological test for its identification and the genes required are present in all strains of the species. In *V. parahaemolyticus* RIMD2210633, only homologues of *oadA oadB oadG (*VP2543-VP2545) are present and it cannot grow on citrate anaerobically (**Figure 6C**). In *V. parahaemolyticus* CDC_K4557, the citrate fermentation cluster was similar to the cluster present in *V. cholerae,* and adjacent to the cluster was a transposase (**Figure 6B**). BLAST analysis showed that the citrate fermentation cluster was present in a total of 39 *V. parahaemolyticus* strains within the same island (**Table 7**). Of the 39 strains, 34 were pathogenic and harbored a T3SS-2α system. This island was present in clinical isolates recovered in the 1980s, 1990s and 2000s. Interestingly, all 20 isolates in the database that were isolated during the *V. parahaemolyticus* outbreak in Peru in 2009 contained this region (**Table 7**).

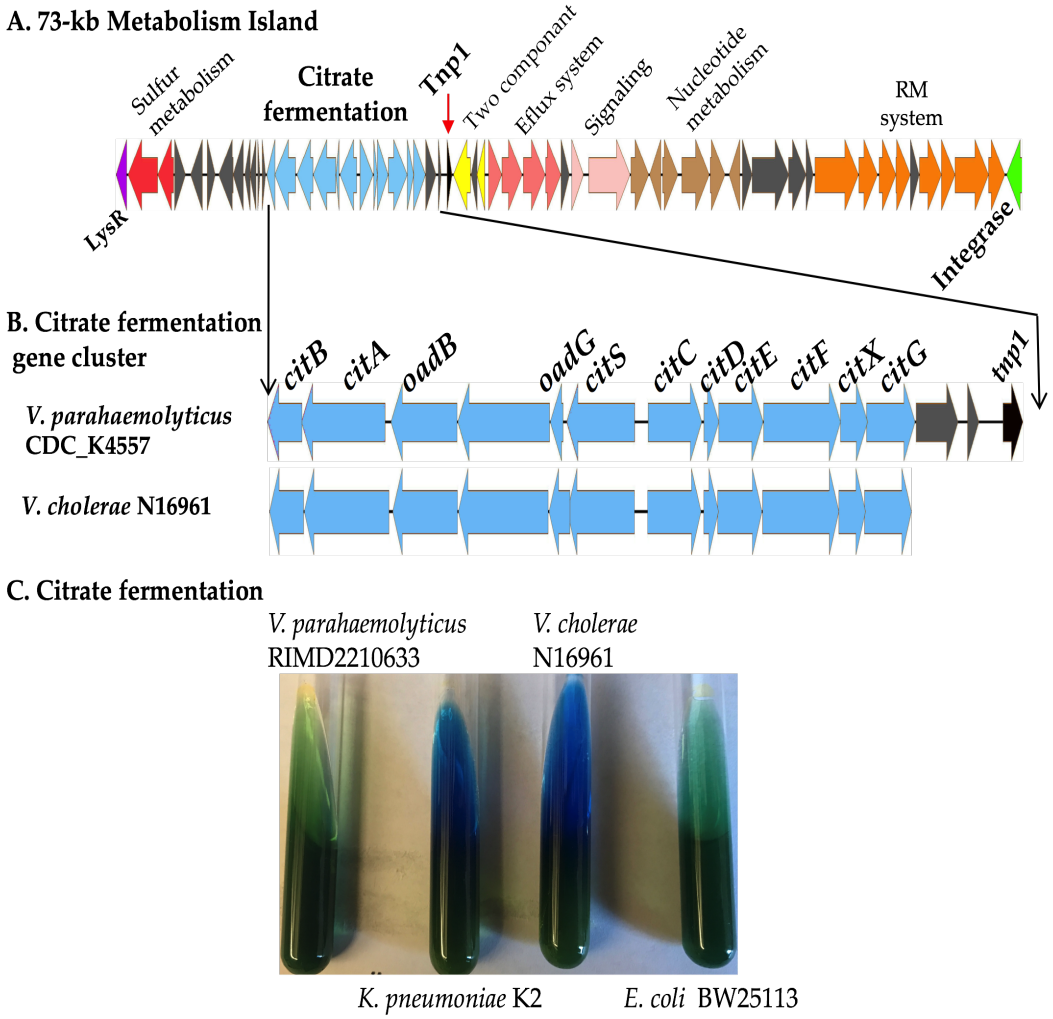


Figure 6 **A 73-kb metabolic island in *V. parahaemolyticus*. A.** Detailed analysis of 73-kb metabolic island identified in chromosome 2 of CDC_K4557, which was missing from RIMD2210633. Arrows indicate ORFs and ORFs with identical color indicate similar function. Gray arrows, genes coding hypothetical and other functional proteins. Black arrow represents a transposase (Tnp). **B.** Detailed analysis of the citrate fermentation cluster in CDC_K4557 and comparison of the same region in *V. cholerae*. **C.** Citrate fermentation in Simmons citrate slant. Conversion of green slant to blue indicates citrate utilization.

These *V. parahaemolyticus* outbreak strains belonged to ST 120 clonal group that originated in China (Gonzalez-Escalona, Gavilan et al. 2016). We speculate that the ability to utilize citrate, in part, may have provided the Peru clone with a competitive advantage, during the epidemic in 2009.

Table 7 *V. parahaemolyticus* strains containing citrate fermentation cluster.

Strains	**Serogroup**	**Country**	**Source**	**Year**	**T3SS-2**
S049	O4:K8	Japan	Clinical	1984	T3SS-2α
S035	O4:K56	Japan	Clinical	1984	T3SS-2α
S022	O5:K15	Japan	Clinical	1984	T3SS-2α
S016	O3:K29	China	Clinical	1992	T3SS-2α
S018	O3:K29	China	Clinical	1993	T3SS-2α
S033	O5:K60	China	Clinical	1994	T3SS-2α
S098	O3:K48	USA	Seafood	1997	Absent
S032	O1:K56	Phillippines	Clinical	1998	T3SS-2α
S036	N/A	Thailand	Clinical	1999	T3SS-2α
S034	N/A	India	Clinical	1999	T3SS-2α
VIP4-0219	N/A	China	Seafood	2006	Absent
CDC_K4557	O1:K33	USA	Clinical	2006	Absent
F30368	O1:KIV	Canada	Clinical	2006	T3SS-2α
CDC_K5635	O5:K30	USA	Wound	2007	Absent
EKP-008	N/A	USA	Environment	2007	T3SS-2α
3259	N/A	USA	Clinical	2007	T3SS-2α
GCSL_R13	O4:K10	USA	Seafood	2007	Absent
07_2964	O4:KUT	Canada	Clinical	2007	T3SS-2α
CFSAN025052	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025053	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025054	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025055	O3:K59	Peru	Clinical	2009	T3SS-2α

CFSAN025056	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025057	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025058	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025059	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025060	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025061	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025062	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025063	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025064	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025065	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025066	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025067	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN025068	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN029653	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN029654	O3:K59	Peru	Clinical	2009	T3SS-2α
CFSAN029656	O3:K59	Peru	Clinical	2009	T3SS-2α
GIMxtfL71-2011.05	N/A	China	Clinical	2011	T3SS-2α

N/A, not available. UT, untypable. UK, unknown.

To examine the evolutionary history of citrate fermentation within *V. parahaemolyticus* and among the *Vibrionaceae*, we performed a phylogenetic analysis of representative CitF proteins from this group (**Figure 7**). The limited distribution of CitF within *V. parahaemolyticus* (>5% of sequenced strains) indicates that it is a recent addition to the species and was likely acquired from a closely related species. Within the Cholerae clade (*V. cholerae, V. metoecus, V. mimicus, V. albensis*), CitF is present in all members of the species and these species have closely related CitF proteins indicating that the protein maybe ancestral to the group. Interestingly within this group is CitF from *V. anguillarum* strains, a species that is distantly related to this clade. This suggests that *V. anguillarum* acquired CitF by horizontal gene transfer

between these species. The CitF distribution was confined to a limited number of *Vibrionaceae* species that formed highly divergent branches with several CitF proteins from *Vibrio* species clustering with CitF from *E. coli* and *Klebsiella pneumonia* (**Figure 7**).

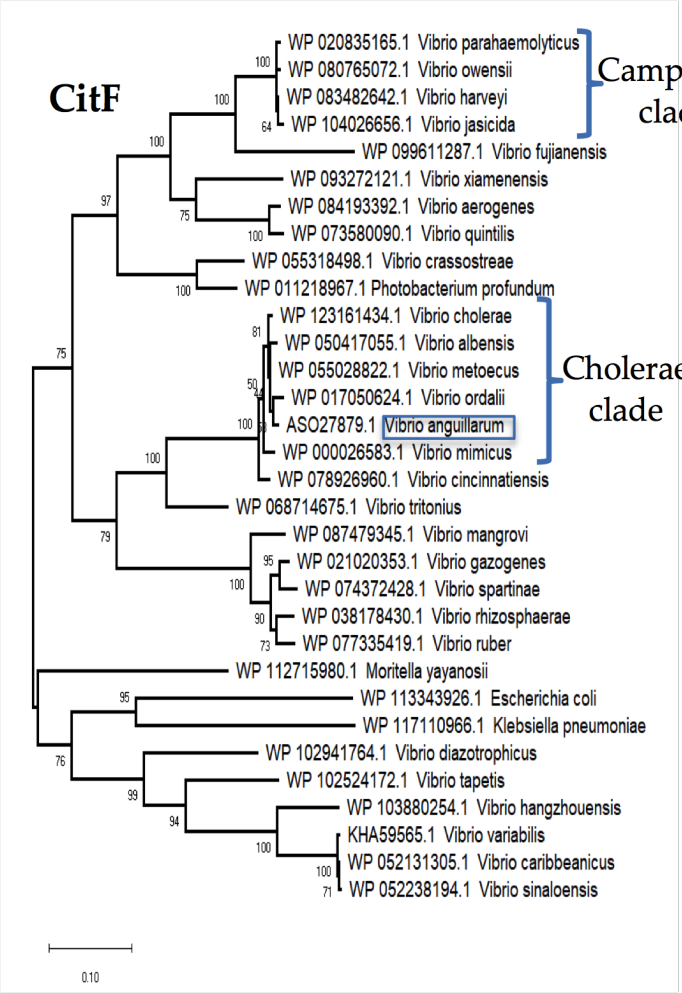


Figure 7 **Phylogenetic analysis of CitF among the *Vibrionaceae*.** This analysis involved 32 amino acid sequences representing CitF homologues present among the *Vibrionaceae*. Phylogeny was inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model (Le and Gascuel 2008). The tree with the highest log likelihood (-7580.58) is shown. The percentage of trees in which the associated taxa clustered

together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.4398)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 489 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar, Stecher et al. 2018).

Variants of the 73-kb genomic island containing the citrate fermentation cluster were identified (**Figure 8**). In strain CDC_K5635, the metabolic island contained a CRISPR-Cas system. Our analysis identified the CRISPR-Cas system as a type I-F system, which contained the *cas* operon of *cas1cas2/3cas8cas5cas7cas6f*, a leader sequence and a CRISPR array. The CRISPR array contained a typical type I-F repeat and 34 spacers, which suggested the system was functional. The CRISPR-Cas system was inserted adjacent the citrate gene cluster (**Figure 8**). In strain 3259, the metabolic island was 105-kb and in addition to the citrate cluster and CRISPR-Cas system also contained the genes for L-rhamnose and OGA catabolism (**Figure 8**). Associated with each of these function modules were transposases suggesting a possible mechanism of acquisition within the island (**Figure 8**). A previous study identified this type I-F CRISPR-Cas system in a total of 10 *V. parahaemolyticus* strains, all recovered in the mid-2000s (McDonald, Regmi et al. 2019).

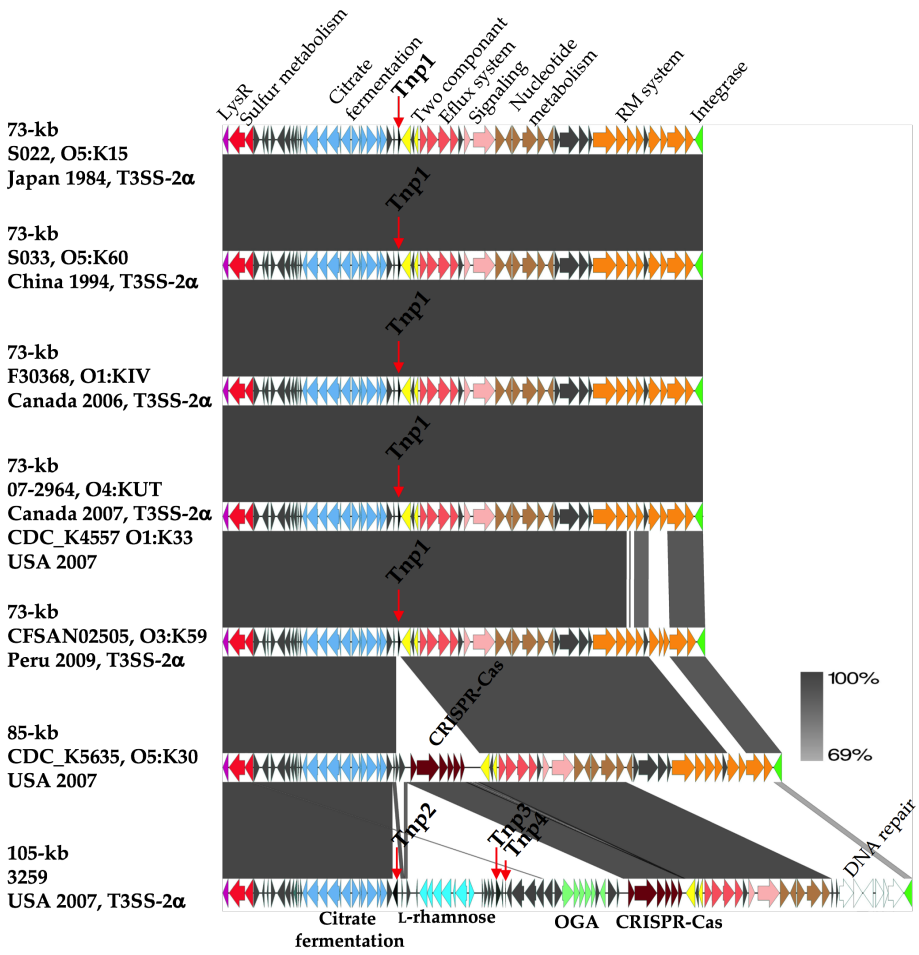


Figure 8 **Variants of the Metabolic islands containing a citrate fermentation gene cluster in *V. parahaemolyticus*.** Comparative analysis of the 73-kb metabolic island containing the citrate fermentation cluster in *V. parahaemolyticus*. Gray shade indicates homologous regions between strains. Arrows represent ORFs, ORFs with the same color represent functionally similar proteins. Gray arrows, genes coding hypothetical and other functional proteins. Black arrows represent transposases.

L-Rhamnose Utilization Clusters Within Two Different Genomic Islands

L-rhamnose, a deoxy-hexose, is commonly found in plants and the bacterial cell wall (Giraud and Naismith 2000). Genome comparative analysis of

RIMD2210633 and FORC_022, identified a 135-kb island related to the 73-kb metabolic island described above, but lacked the citrate fermentation gene cluster (**Figure 9A**). Similar to strain 3259, in FORC_022, the metabolic island contained

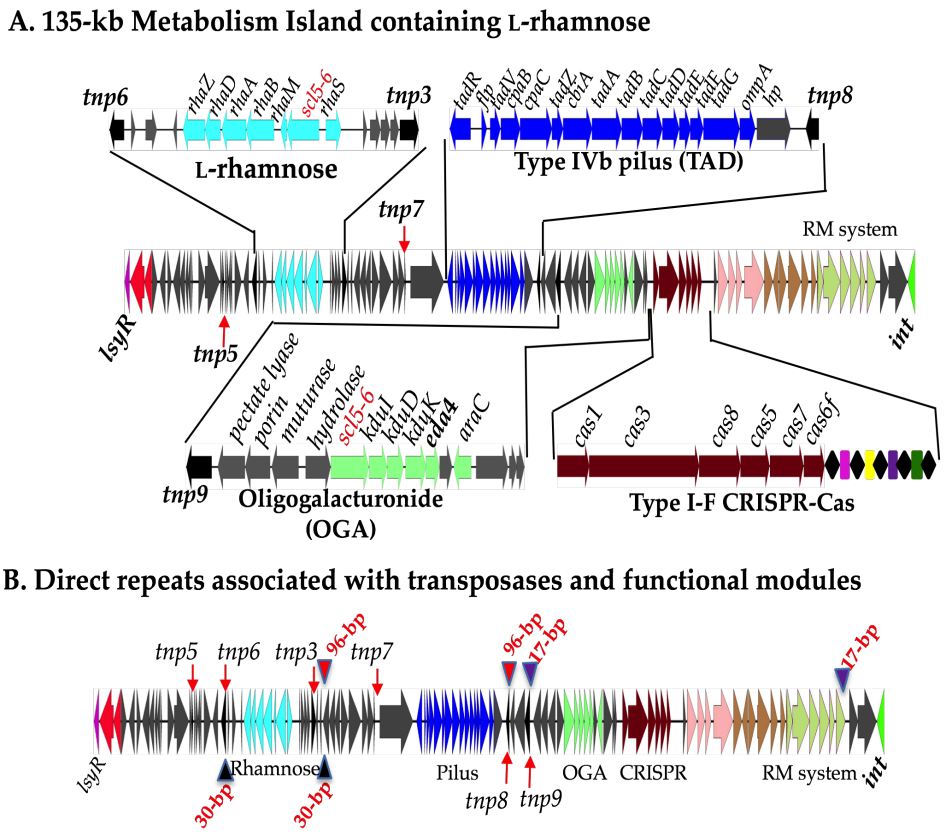


Figure 9 **A 135-kb Metabolism island in *V. parahaemolyticus*. A.** Detailed analysis of a 135-kb metabolic island identified in chromosome 2 of FORC_022 absent from RIMD2210633. Arrows indicate ORFs and ORFs with identical color indicate similar function. Gray arrows, genes coding hypothetical and other functional proteins. Black arrows represent transposases. **B.** Transposases and direct repeats identified flanking the functional modules with the 135-kb island.

homologues of *rhaM, rhaA, rhaB, rhaD,* a SCL5-6-like sodium symporter, *rhaZ,* and a regulator encoded by *rhaS,* all required for L-rhamnose utilization (**Figure 9A**). In addition, the island contained a type IV pilus gene cluster and an oligogalacturonide (OGA) catabolism cluster (**Figure 9A**). The OGA cluster contained homologues of *kduI, kduD, kdgK,* and another KDPG aldolase EDA4. The EDA4 protein shared 50% amino acid identity with EDA3 (VPA0083), 53% with EDA1 (VP0065) and 57% identity with EDA2 (VPA1708). The distribution of EDA4 was limited, present in some *V. harveyi* and *V. jasicida* strains (**Figure 5**). Within the OGA cluster, a SCL5-6-like sodium symporter, an AraC family regulator, two hydrolases, and a porin required for the uptake and catabolism of OGA and unsaturated monomers of pectin were identified (**Figure 10B**). Within this island a complete type I-F CRISPR-Cas system was present similar to the system present in strains CDC_K5635 and 3259 but with different sized CRISPR arrays and spacer sequences in each strain. Strain 3259 contained 4 spacers, FORC_022 had 28 spacers and CDC_K5635 had 34 spacers suggesting these systems were active and acquiring spacers.

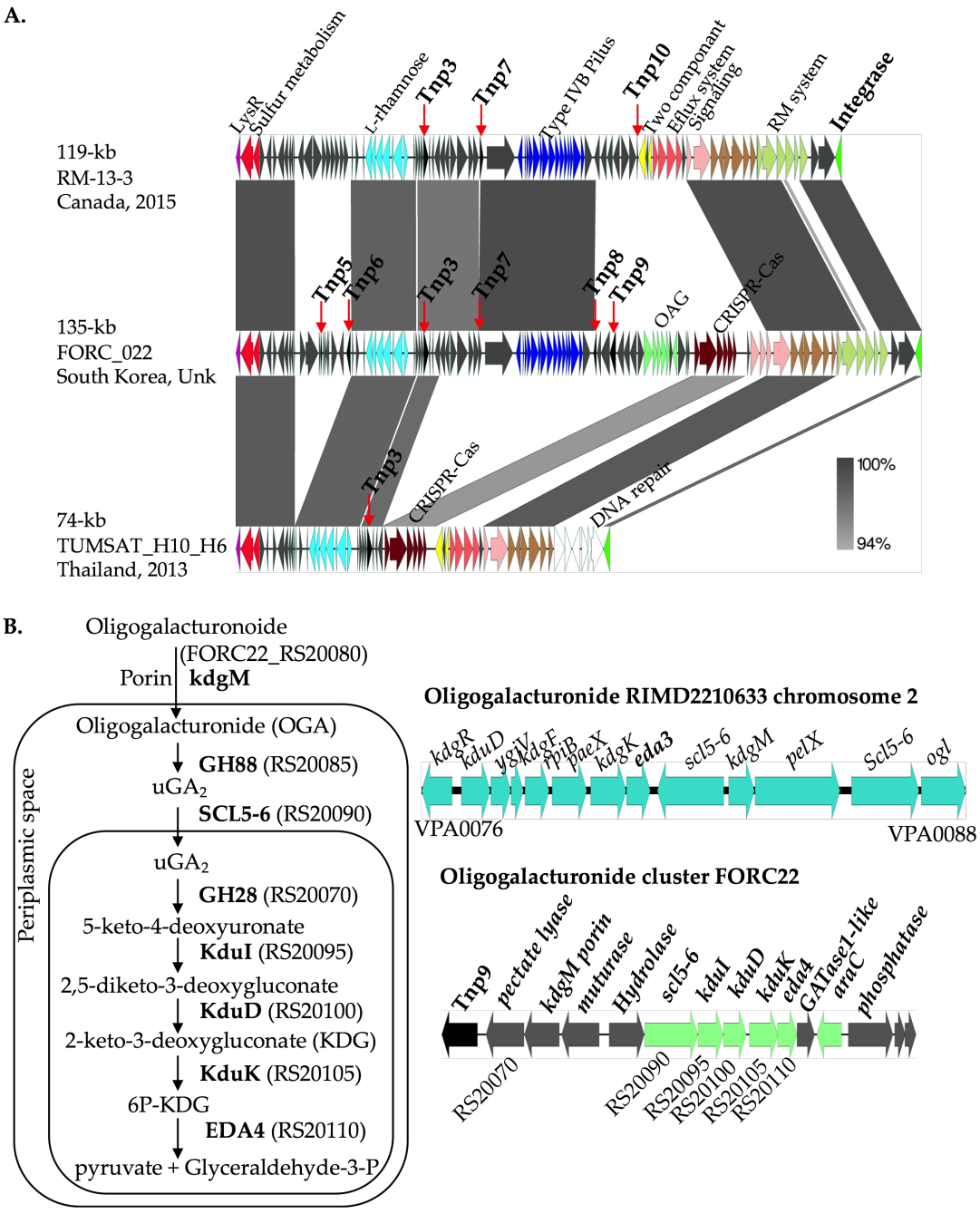


Figure 10 **Variants of the Metabolic island containing L-rhamnose utilization and OGA catabolism gene clusters. A.** Comparative genomic analysis

of islands containing the L-rhamnose gene cluster. Gray shade indicates homologous regions between strains. Arrows represent ORFs, identical colored ORFs indicate similar function. Gray arrows, genes coding hypothetical and other functional proteins. Black arrows indicate transposases. **B.** OGA metabolism pathway with enzymes involved and ORFs identified in the 135-kb metabolic island of FORC_022. OGA catabolism cluster in RIMD2210633 is also shown. uGA_2, unsaturated galacturonate dimer, GH, glycoside hydrolase.

BLAST analysis identified an additional 19 strains that harbored a homologous L-rhamnose cluster associated with variants of this island, at the same genomic location on chromosome 2 (**Table 8**). These strains were predominantly isolated from Asia from 2006 to 2013 and most strains were non-pathogenic (lacked a T3SS-2 system) (**Table 8**). This suggests that the L-rhamnose cluster was recently acquired in a subpopulation of bacteria. In strain RM-13-3, a 119-kb island similar to the 135-kb metabolic island was present but lacked the OAG and CRISPR-Cas clusters (**Figure 10A**). In strain TUMSAT_H10_H6, a 74-kb variant of the metabolic island was present that contained only the L-rhamnose and CRISPR-Cas clusters (**Figure 10A).** In TUMSAT_H10_H6, the type I-F CRISPR array contained 24 spacers, which did not share homology with spacer sequences from other type I-F systems identified in *V. parahaemolyticus*. The type IV pilus cluster, OGA cluster, and the L-rhamnose clusters all had transposase genes nearby and we identified direct repeats flanking these functional modules (**Figure 9B**). This suggested to us a mechanism of acquisition as discrete modules within the island by a transposon like mechanism. Overall, these data suggest that this metabolic island is a dynamic region with the loss and gain of functional modules occurring over evolutionarily short timescales and confined to subpopulations of bacteria.

Table 8 *V. parahaemolyticus* strains containing 26-kb (upper panel) and metabolic island (lower panel) with L-rhamnose cluster.

Strains	**Serogroup**	**Country**	**Source**	**Year**	**T3SS-2**
AQ3810	O3:K6	Singapore	Clinical	1983	T3SS-2a
S004	O3:K6	Maldives	Clinical	1985	T3SS-2β
AQ4037	O3:K6	Maldives	Clinical	1985	T3SS-2β
S008	O3:K6	Thailand	Clinical	1987	T3SS-2β
S009	O3:K6	Thailand	Clinical	1987	T3SS-2β
S012	O3:K6	Thailand	Clinical	1990	T3SS-2β
S011	O3:K6	Thailand	Clinical	1990	T3SS-2β
S013	O3:K6	China	Clinical	1996	T3SS-2β
S095	O3:K6	China	Clinical	1996	T3SS-2β
FORC_023	N/A	South Korea	Environment	2014	Absent
S005	O3:K6	Thailand	Clinical	N/A	T3SS-2β
S003	O3:K6	China	Clinical	N/A	T3SS-2β
VIP4-0219	N/A	China	Seafood	2006	Absent
S141	N/A	China	Seafood	2006	Absent
S154	N/A	China	Seafood	2006	Absent
S156	N/A	China	Seafood	2006	Absent
S155	N/A	China	Seafood	2006	Absent
S150	N/A	China	Seafood	2006	Absent
S153	N/A	China	Seafood	2006	Absent
S144	N/A	China	Seafood	2006	Absent
S158	N/A	China	Seafood	2006	Absent
S161	N/A	China	Seafood	2006	Absent
S151	N/A	China	Seafood	2006	Absent
S143	N/A	China	Seafood	2006	Absent
EKP-008	N/A	Bangladesh	Environment	2007	T3SS-2α
3259	N/A	USA	Clinical	2007	T3SS-2α
S170	N/A	China	Environment	2007	Absent

49	N/A	USA	Seafood	2007	Absent
SNUVpS-1	N/A	South Korea	Seafood	2009	Absent
TUMSAT_H10_S6	N/A	Thailand	Seafood	2013	Absent
RM-13-3	N/A	Canada	Seafood	2015	Absent
FORC_022	N/A	South Korea	Seafood	N/A	Absent

N/A, not available. UT, untypable. UK, unknown.

Interestingly, we identified 12 strains with a second L-rhamnose cluster, which shared only 74% to 86% amino acid identity with the cluster carried within the metabolic island (**Table 8 and Figure 11A**). In strain AQ3810, the L-rhamnose cluster was present within a 26-kb region that was inserted at ORF VPA1309 relative to RIMD2010633, which lacked the region. Adjacent to the 26-kb region were several transposase genes. Both strains contain the T3SS-2α region at this site (**Figure 11B and 11D**). The 26-kb L-rhamnose region contained a gene encoding rhamnosidase suggesting an ability to catabolize L-rhamnose oligosaccharides. No canonical L-rhamnose or L-rhamnose oligosaccharide transporter was identified but a PTS system was adjacent to the L-rhamnose catabolic genes (**Figure 11A**). In strains that have a T3SS-2β present on chromosome 1, the 26-kb L-rhamnose region is at VPA1309 with the same transposases as described above (**Figure 11D and Table 8**). In FORC_023 isolated in South Korea in 2014, the 26-kb region was present between VPA1158 and VPA1159 relative to RIMD2210633 (**Figure 11D**). Thus, our genomic data indicates that *V. parahaemolyticus* acquired the L-rhamnose catabolic cluster at least three times during its evolution, which suggests that L-rhamnose utilization may convey an important fitness advantage in certain niches.

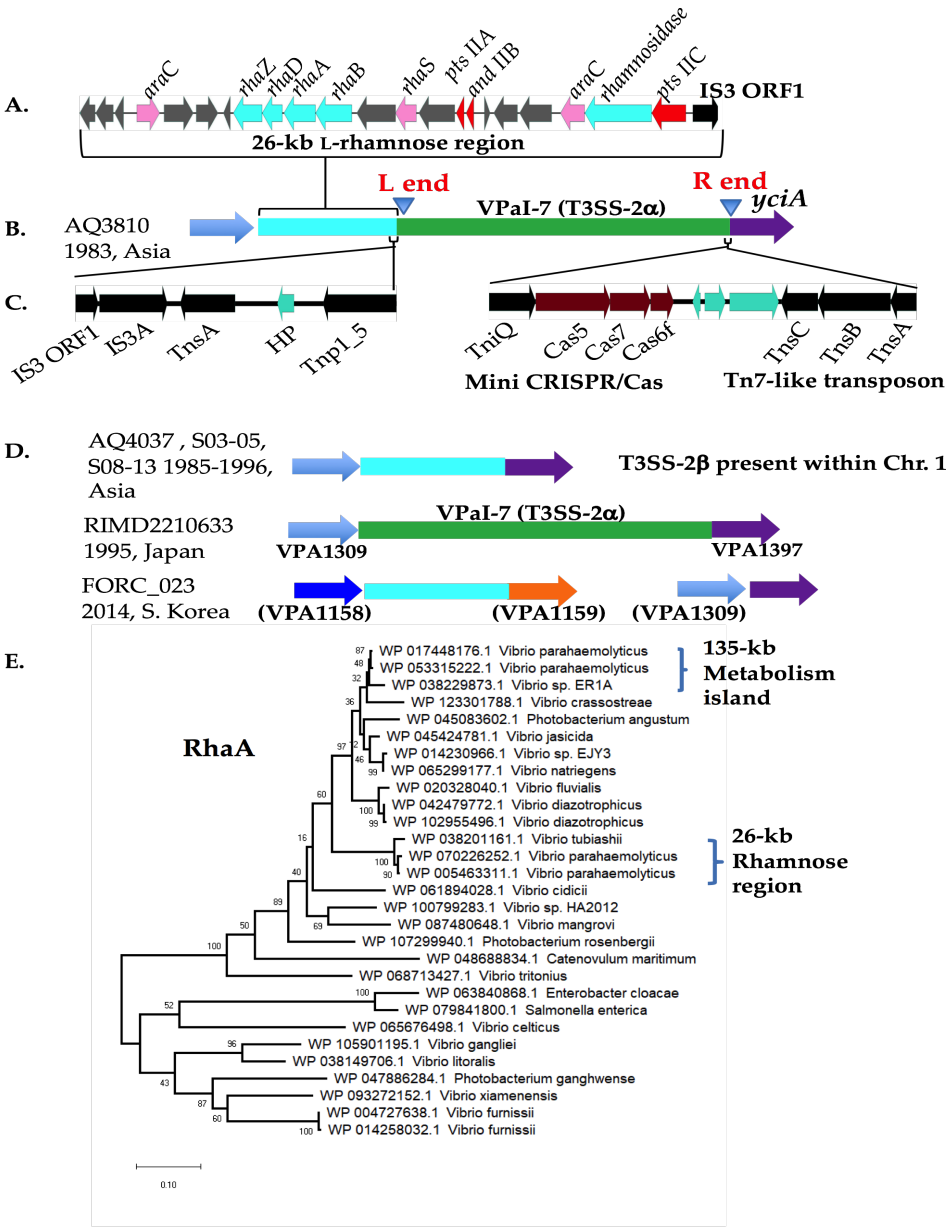


Figure 11 **Analysis of the 26-kb L-rhamnose utilization island. A.** Gene cluster of the L-rhamnose utilization in *V. parahaemolyticus* AQ3810. **B**. Genomic locus of the 26-kb L-rhamnose region in AQ3810. **C**. A Tn7-like transposon and a mini CRISPR-Cas system associated with T3SS-2α and a transposon like region associated with the 26-kb L-rhamnose region. **D.** Schematic of the L- rhamnose genomic loci in AQ4037, S03-S05, S08-S13 VPA1309 in chromosome 2 that is also the location of T3SS-2α in

RIMD2210633. Location of the 26-kb L-rhamnose island in FORC_023 and the empty site at VPA1309. Parenthesis indicates homologous ORFs. **E.** Phylogenetic analysis of RhaA protein from FORC_022 and AQ3810 among members of the family *Vibrionaceae*. The evolutionary history was inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model (Le and Gascuel 2008). The tree with the highest log likelihood (-6340.56) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.4168)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 29 amino acid sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 418 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Le and Gascuel 2008).

The phylogeny of RhaA shows two divergent branching patterns representing RhaA from the 135-kb metabolic island and the 26-kb region. RhaA from the metabolic island is present in several species belonging to the Campbellii clade, and are closely related, clustering together with short branch length **(Figure 11E)**. The RhaA from the 26-kb region forms a divergent branch related to RhaA in other Campbellii clade species. Highly divergent RhaA proteins were present in several *Vibrio* species that were more closely related to RhaA from *Salmonella enterica* and *Enterobacter cloacae* than to other *Vibrio* species **(Figure 11E)**. Overall, the data suggest that RhaA is not phylogenetically widespread and was acquired multiple times within the *Vibrionaceae*.

T3SS-2 Present Within a Tn7-Like Transposon-CRISPR-Cas Element

In strain AQ3810, adjacent to the 26-kb L-rhamnose region was the VPaI-7 region that contains T3SS-2α (**Figure 11B and 11C**). In this strain, the VPaI-7 region is flanked by Tn7-like transposon right end and left end attachment sites (named L end and R end) (**Figure 11B and 11D**). The R end and L end sequences are absent from strains that do not have a T3SS-2 at this site (**Figure 11D**). Adjacent to the R end is a three gene cluster that showed homology to Tn7 transposon genes *tnsABC*. The *tnsD* and *tnsE* genes, involved in target site insertion in Tn7, are absent but a homologue *tniQ* is present within a mini CRISPR-Cas system (*tniQcas5cas7cas6*) close to the Tn7 gene cluster (**Figure 11C**). In all *V. parahaemolyticus* strains that contained a T3SS-2α system in chromosome 2 that we examined, the mini CRISPR-Cas system and the Tn7-like region always co-occurred with T3SS-2α. The mini CRISPR-Cas system depending on the strain has either no array or a very short array consisting of one or two spacers suggesting it is unable to acquire new spacers since it lacks Cas1 and Cas2. It has been proposed that the Tn7-like transposon and mini CRISPR-Cas function together to form a function transposon that allows for target site integration (Peters, Fricker et al. 2014, Peters, Makarova et al. 2017). This suggests a novel non-canonical function for a CRISPR-Cas system, co-opted by a transposon for the mobilization of bacterial DNA.

L-Arabinogalactan Catabolism Gene Cluster Within a Novel Metabolic Island

L-arabinose is a pentose sugar and is an important nutrient source *in vivo* as it is one of the sugars found in the proteoglycan of the mucin present in the gastrointestinal track (Conway and Cohen 2015). An L-arabinose catabolic cluster is present in *V. parahaemolyticus* RIMD2210633 and this strain can grow on L-arabinose

as a sole carbon source **(Table 4)** (Whitaker, Richards et al. 2014). UCM-V493 cannot grow on L-arabinose as a sole carbon source and lacks the gene cluster required **(Table 4)**. The catabolic and ABC-type transporter genes for L-arabinose utilization are present within a 9.5-kb region present in most pathogenic *V. parahaemolyticus* strains in the database **(Figure 12C)**. A notable exception is the absent of this cluster from strains that contained the T3SS-2γ system (**Table 5**). Bioinformatics analysis identified in FORC_008, a second L-arabinose cluster that only shared ~70% amino acid identity with that from RIMD2010633 and contained an uncharacterized major facilitator superfamily (MFS) type transporter (**Figure 12A**). This cluster was carried within a 13.2-kb region inserted between VPA0441 and VPA0449 with respect to RIMD2210633 (**Figure 12A**). The region was present in 21 non-pathogenic strains, 15 pathogenic strains that contained T3SS-2β, and one pathogenic strain with T3SS-2α (**Table 9**). This region also contained α-L- and β-L-arabinofuranosidases, an uncharacterized hydrolase and a second MFS type transporter suggesting an ability to catabolize L-arabinose oligomers (**Figure 12A**). The data indicate that *V. parahaemolyticus* has acquired two non-homologous L-arabinose utilization systems during its evolution. Interestingly, in RIMD2210633 between VPA0441 and VPA0449, a putative peptide modification system, His-Xaa-Ser, was present, whereas in strain CFSAN007457 a complete D-galactitol utilization system was present at the same position (**Figure 12A and 12B**). The D-galactitol utilization system was identified in 15 strains, mostly North American isolates containing T3SS-2β isolated from 1984 to 2010 and a few non-pathogenic strains (**Table 9**).

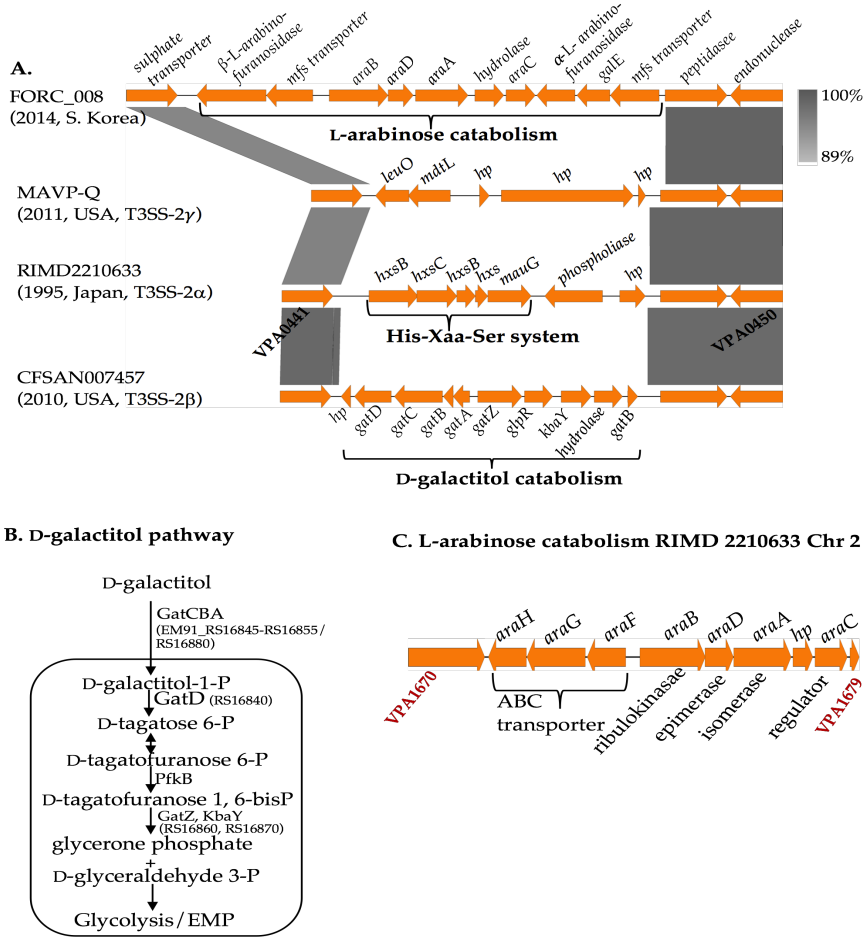


Figure 12 **Genomic analysis of L-arabinose catabolic gene cluster. A.** Comparative analysis of the genomic region between VPA0441 and VPA0450 in four *V. parahaemolyticus* strains. Gray shade, region of nucleotide homology **B.** D-galactitol pathways with proteins and ORFs identified in CFSAN007457. **C.** L-arabinose gene cluster present in *V. parahaemolyticus* strain RIMD2210633. Arrow indicated ORFs.

Table 9 *V. parahaemolyticus* strains containing D-galactitol, L-arabinose (with MFS transporter), and L-arabinogalactan utilization cluster.

Strains	Serogroup	Country	Source	Year	T3SS-2	D-galactitol	L-arabinose (MFS)	L-arabinogalactan
S129	O5:K60	Japan	Clinical	1984	T3SS-2β	**YES**	NO	NO
S104	O6:K18	USA	Clinical	1997	T3SS-2β	**YES**	NO	NO
CFSAN001617	N/A	USA	Clinical	1997	T3SS-2β	**YES**	NO	NO
AOEZ664	O6:K18	Canada	Clinical	2000	T3SS-2β	**YES**	NO	NO
AOEZ713	O6:KUT	Canada	Clinical	2000	T3SS-2β	**YES**	NO	NO
10-4287	O6:K18	Canada	Clinical	2003	T3SS-2β	**YES**	NO	NO
A3EZ770	O6:K18	Canada	Clinical	2003	T3SS-2β	**YES**	NO	NO
A3EZ634	O6:K18	Canada	Clinical	2003	T3SS-2β	**YES**	NO	NO
CFSAN001174	N/A	USA	N/A	2004	T3SS-2β	**YES**	NO	NO
04-2192	O11:KUT	Canada	Clinical	2004	T3SS-2β	**YES**	NO	NO
GCSL_R32	O10:KUK	USA	Seafood	2007	T3SS-2β	**YES**	NO	NO
CDC_K5485	O6:K18	USA	Clinical	2007	Absent	**YES**	NO	NO
S165	N/A	China	Environment	2007	Absent	**YES**	NO	NO
CFSAN007457	N/A	USA	Seafood	2010	T3SS-2β	**YES**	NO	NO
M0605	N/A	Mexico	Seafood	2013	Absent	**YES**	NO	NO
NSV 5736	N/A	USA	Clinical	N/A	T3SS-2β	**YES**	NO	NO
S117	N/A	Japan	Clinical	1984	T3SS-2β	NO	**YES**	NO
S053	O1:K69	Thailand	Clinical	1990	T3SS-2β	NO	**YES**	NO

S044	O1:K69	Thailand	Clinical	1990	T3SS-2β	NO	**YES**	NO
S045	O1:K69	Thailand	Clinical	1990	T3SS-2β	NO	**YES**	NO
S025	N/A	China	Clinical	1992	Absent	NO	**YES**	NO
S017	N/A	China	Clinical	1992	Absent	NO	**YES**	NO
S124	O3:K57	China	Clinical	1992	Absent	NO	**YES**	NO
S057	O1:K41	China	Clinical	1994	Absent	NO	**YES**	NO
S055	O1:K41	China	Clinical	1994	Absent	NO	**YES**	NO
VN-0028	N/A	Germany	Clinical	1995	T3SS-2β	NO	**YES**	NO
CFSAN018755	N/A	Spain	N/A	2002	Absent	NO	**YES**	NO
S130	N/A	N/A	Clinical	2003	Absent	NO	**YES**	NO
CAIM 1772	O5:K17	Mexico	Seafood	2005	T3SS-2β	NO	**YES**	NO
S134	N/A	China	Clinical	2005	Absent	NO	**YES**	NO
A5Z1022	O1:KUT	Canada	Clinical	2005	T3SS-2β	NO	**YES**	NO
CDC_K4764D	O8:K41	USA	Clinical	2006	T3SS-2β	NO	**YES**	NO
CDC_K4638	O10:KUK	USA	Clinical	2006	T3SS-2β	NO	**YES**	NO
CDC_K4763	O5:KUK	USA	Clinical	2006	T3SS-2β	NO	**YES**	NO
CDC_K4764L	O8:K41	USA	Clinical	2006	Absent	NO	**YES**	NO
S148	N/A	China	Seafood	2006	Absent	NO	**YES**	NO
S141	N/A	China	Seafood	2006	Absent	NO	**YES**	NO
GCSL_R60	O10:KUK	USA	Seafood	2007	T3SS-2β	NO	**YES**	NO
49	N/A	USA	Seafood	2007	Absent	NO	**YES**	NO
07-2965	O5:KUT	Canada	Clinical	2007	T3SS-2α	NO	**YES**	NO

ISF-25-6	N/A	Canada	Seafood	2010	T3SS-2β	NO	**YES**	NO
MAVP-G	N/A	Canada	Clinical	2011	T3SS-2β	NO	**YES**	NO
GIMxtfL88-2011.05	N/A	China	Clinical	2011	Absent	NO	**YES**	NO
MAVP-M	N/A	Canada	Clinical	2011	T3SS-2β	NO	**YES**	NO
MAVP-25	N/A	Canada	Clinical	2013	T3SS-2β	NO	**YES**	NO
TUMSAT_H10_S6	N/A	Thailand	Seafood	2013	Absent	NO	**YES**	**YES**
FORC_008	N/A	South Korea	Environment	2014	Absent	NO	**YES**	NO
RM-13-3	N/A	Canada	Seafood	2015	Absent	NO	**YES**	NO
MVP1	N/A	Malaysia	Seafood	2016	Absent	NO	**YES**	NO
MVP2	N/A	Malaysia	Environment	2016	Absent	NO	**YES**	NO
MVP6	N/A	Malaysia	Environment	2016	Absent	NO	**YES**	NO
S054	N/A	Thailand	Clinical	N/A	Absent	NO	**YES**	NO
FORC_018	N/A	South Korea	Environment	UN	Absent	NO	**YES**	NO
GCSL_R136	O1:K20	USA	Seafood	2007	T3SS-2γ	NO	NO	**YES**
GCSL_R137	O1:K20	USA	Seafood	2007	T3SS-2γ	NO	NO	**YES**
GCSL_R138	O1:K20	USA	Seafood	2007	T3SS-2γ	NO	NO	**YES**
PCV08-7	N/A	Malaysia	Seafood	2008	Absent	NO	NO	**YES**
NCKU_TN_SO2_SO2	N/A	Thailand	Seafood	2008	Absent	NO	NO	**YES**
CHN25	O5:KUT	China	Seafood	2011	Absent	NO	NO	**YES**
TUMSAT_H01_S4	N/A	Thailand	Seafood	2013	Absent	NO	NO	**YES**

TUMSAT_H10_S6	N/A	Thailand	Seafood	2013	Absent	NO	**YES**	**YES**
TUMSAT_DE2_S2	N/A	Thailand	Seafood	2013	Absent	NO	NO	**YES**
13-306D/4	N/A	Mexico	Seafood	2013	Absent	NO	NO	**YES**

N/A, not available. UT, untypable. UK, unknown.

YES, catabolic and transporter genes are present. NO, catabolic and transporter genes are missing.

To determine the evolutionary history of the L-arabinose clusters in *V. parahaemolyticus* RIMD2210633 and FORC_008, we reconstructed a phylogenetic tree based on AraD proteins from both clusters and examined their evolutionary relationships in *Vibrio* species. AraD from RIMD2210633 formed a distinct divergent cluster from AraD present in FORC_008 (**Figure 13**). AraD from RIMD2210633 had a limited distribution outside of *V. parahaemolyticus* with only two uncharacterized species containing a closely related protein. Branching divergently and distant from this cluster were AraD proteins from over a dozen *Vibrio* species each having long-branch lengths. AraD from FORC_008 formed a tight cluster with AraD from several *Vibrio* species and *Grimontia hollisae*. These AraD proteins formed the most divergent branch on the tree indicating that they are not related to AraD present in RIMD2210633 (**Figure 13**).

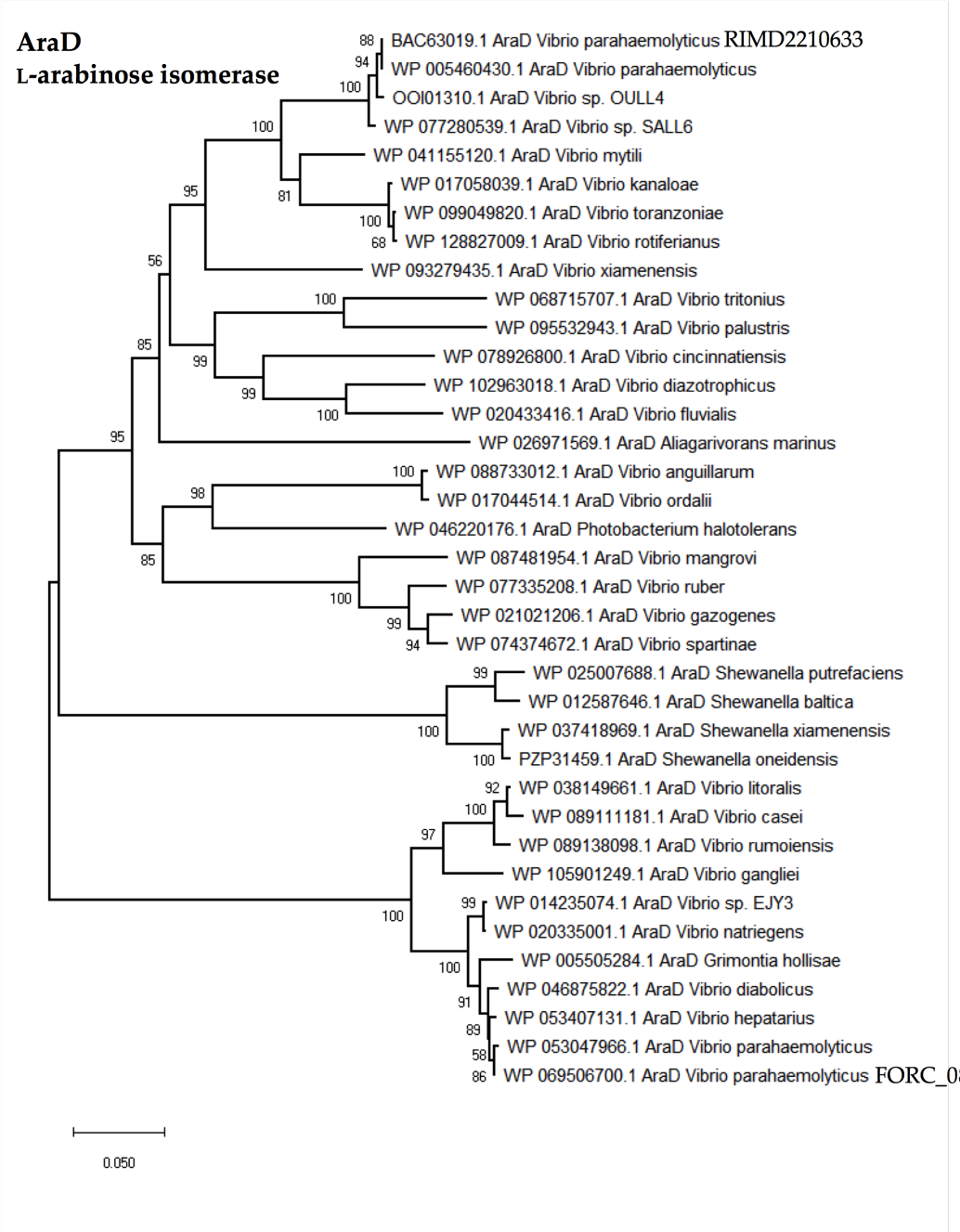


Figure 13 **Phylogenetic analysis of AraD among *Vibrionaceae*.** AraD from *V. parahaemolyticus* was used as a seed to identify homologues within the *Vibrionaceae*. This analysis involved 37 amino acid sequences of AraD

with most OTUs representing multiple strains. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The optimal tree with the sum of branch length = 2.24462315 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Saitou and Nei 1987). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Dayhoff matrix based method and are in the units of the number of amino acid substitutions per site (Saitou and Nei 1987). The rate variation among sites was modeled with a gamma distribution (shape parameter = 5). All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 507 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Saitou and Nei 1987).

Comparative genomic analysis of RIMD2210633 and CH25, a non-pathogenic strain isolated from seafood in China, identified an L-arabinogalactan catabolic cluster (**Figure 14A**). The genes for the uptake of L-arabinogalactan and its conversion to L-arabinose and D-galactose were clustered with *galETK* required for D-galactose catabolism (**Figure 14A and 14B**). The 13 gene region was present within a 62-kb island on chromosome 2 that was absent from RIMD2210633**.** An integrase was present at the 3' end of the island and the % GC of the island was 41% compared to 45% of the entire genome indicating that this island was horizontally acquired (**Figure 14A**). A gene encoding a transposase was identified next to the L-arabinogalactan cluster and the island was identified in 10 additional *V. parahaemolyticus* strains. These strains were typically from seafood samples with three strains belonging to the ST631 with a T3SS-2γ.

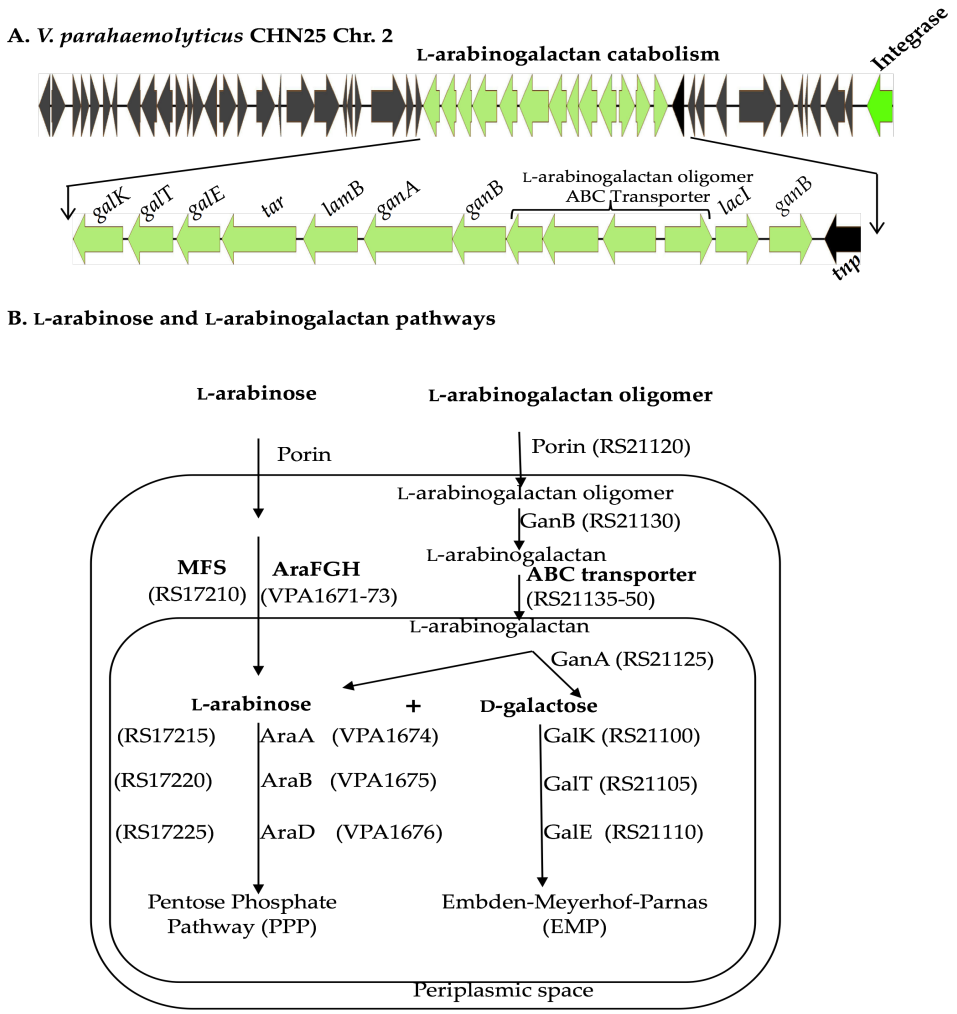


Figure 14 **L-arabinogalactan metabolism cluster within a 62-Kb island. A.** A 62-kb island containing L-arabinogalactan catabolic cluster is shown. Arrows indicate ORFs. Gray arrows, genes coding hypothetical and other functional proteins. Black arrow indicates transposase. **B**. L-arabinogalactan utilization pathway with enzyme involved and ORFs identified in CH25 as well as the enzymes involved and ORFs for L-arabinose utilization in RIMD2210633 and FORC_008.

Conclusion

A key mechanism in the evolution of bacteria is horizontal gene transfer and the acquisition of clusters of genes that encode new phenotypes. Many studies have

shown the importance of the acquisition of virulence genes on mobile genetic elements, such as phage, plasmids and pathogenicity islands, in the emergence of new pathogens and the enhancement of established pathogens (Waldor and Mekalanos 1996, Hacker and Kaper 2000, Dobrindt, Hentschel et al. 2002, Dobrindt, Hochhut et al. 2004, Kaper, Nataro et al. 2004, Boyd, Almagro-Moreno et al. 2009, Boyd 2012, Boyd, Carpenter et al. 2012, Pilla and Tang 2018). In *V. parahaemolyticus*, its emergence as a human pathogen resulted from the acquisition of T3SS-2 that is absent from non-pathogenic strains (Makino, Oshima et al. 2003, Hurley, Quirke et al. 2006). Several studies have demonstrated differences in genome content among *V. parahaemolyticus* strains with an emphasis on potential virulence factors and pathogen markers (Hurley, Quirke et al. 2006, Boyd, Cohen et al. 2008, Chen, Stine et al. 2011, Loyola, Navarro et al. 2015, Espejo, Garcia et al. 2017).

However, other factors can play a key role in the emergence of new strains in new environments such as metabolic traits. Metabolic capacity is an important fitness factor both within the natural environment and within host species allowing bacteria to use new nutrient sources and/or acquire and use nutrients more efficiently. Little is known about how metabolic differences emerge or how widespread these differences are within and between species. This study examined both fine scale differences (transporter differences, multiple copies of the same protein) and absolute differences (present and absent of metabolic cluster) among *V. parahaemolyticus* isolates. We show that differences in the types of transporters present within a catabolism gene cluster can lead to differences in utilization efficiencies. The significance of possessing high and low affinity transporters will depend to a large extend on the niche the strain is present in. Several EDA homologues were present in *V.*

parahaemolyticus and was a feature of the Campbellii clade in general, suggesting that sugar acids are an important nutrient source for this group.

We identified multiple metabolism islands carrying different catabolic gene clusters (citrate, L-rhamnose, L-arabinose, L-arabinogalactan, D-galactitol, and OGA) present in a range of *V. parahaemolyticus* strains and diverse *Vibrio* species. Some metabolic regions showed a correlation with the emergence of *V. parahaemolyticus* pathogenic clones such as the presence of citrate fermentation in strains from the 2009 Peru outbreak. On the other hand, many of the metabolic differences did not correlate with pathogenic strains, indicating that environmental fitness is an essential driver of metabolic diversity. Many of the metabolic traits examined had a limited distribution outside of the Campbellii clade, and others, for example L-arabinose utilization, had a limited distribution outside of *V. parahaemolyticus,* suggesting a defining feature of the species. Metabolic traits such as citrate fermentation had a limited distribution within *V. parahaemolyticus* and the Campbellii clade, in general but were prevalent among members of the Cholerae clade, a key biochemical test for *V. cholerae*.
The identification of transposase genes associated with the metabolic systems present on islands suggested a mechanism by which these regions are lost and gained. For the 135-kb metabolic island that contained multiple metabolism clusters inserted at VPA0712 relative to RIMD2210633, we identified a number of variant islands (**Figure 15**). The most evolutionary parsimonious scenario for how these variants arose is that a progenitor island contained all three metabolic gene clusters, the type IV pilus and the type I-F CRISPR-Cas system (**Figure 15**). We speculate that many of the variant islands arose from single deletion events from this progenitor island whereas other variants arose by two deletion events (**Figure 15**). The presence of transposase

genes and direct repeats flanking the functional modules suggests a mechanism of how the progenitor island arose and how variant islands can be formed from deletion events. Of course, it is also possible that these variants could arise by accretions of systems within an island mediated by transposons as well but would require many more steps. The presence of these regions in both clinical and environmental strains indicates that the gene pool or reservoir for novel phenotypes is large. The presence of these regions in both clinical and environmental strains indicates that the gene pool or reservoir for novel phenotypes is large.

The identification of a possible Tn7-like transposon-CRISPR-Cas system interaction involved in the mobilization of T3SS-2 systems within and between *V. parahaemolyticus* strains is curious and noteworthy. The T3SS-2 system genes are all within the Tn7-like transposon R end and L end attachment sites strongly suggesting the 80-kb region is carried on the transposon. The transposon genes *tnsABC* are adjacent to the insertion site *yciA*, which is typical of Tn7-like systems, but lack the *tnsD* or *tnsE* genes involved in target integration. A *tnsD* homologue, *tniQ* is present within a CRISPR-Cas operon encoding *tniQcas5cas7cas6f*, which lacks the *cas1cas2* genes required for spacer acquisition. The mini CRISPR-Cas system always co-occurs with *tnsABC* not only in *V. parahaemolyticus* but also in other species (Peters, Makarova et al. 2017, McDonald, Regmi et al. 2019). One could speculate that the mini CRISPR-Cas system allows the transposon to target and insert into novel sites permitting its propagation. Whether the two systems have co-opted each other to form a functional unit involved in the mobilization of large bacterial regions remains to be

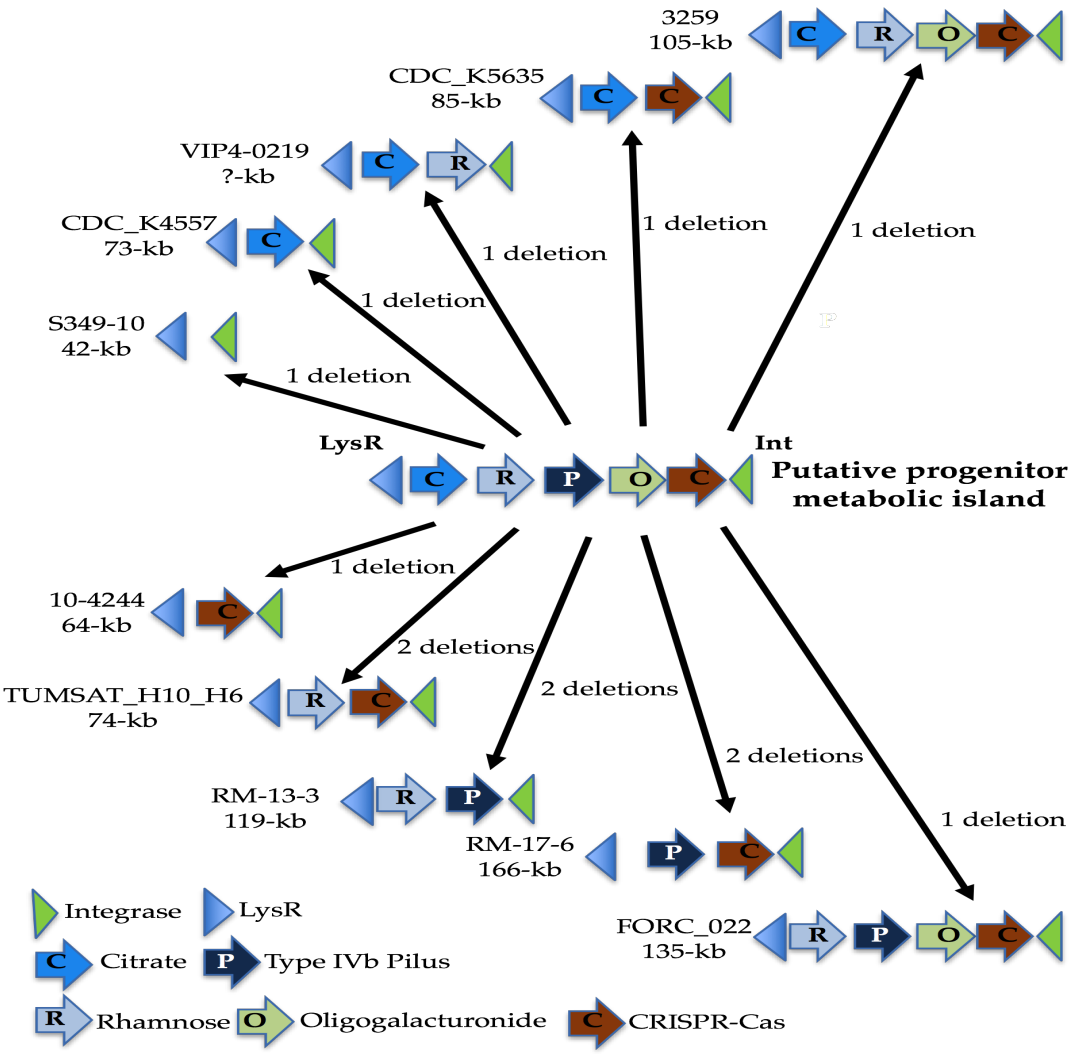


Figure 15 **Predicted model of the 135-kb Metabolic island emergence.** A model predicting how variants of the metabolic island could have arisen in *V. parahaemolyticus* is shown depicting single and double deletion events. The putative progenitor metabolic island was not identified in any strain in the genome databases. Shown are the most evolutionary parsimonious steps required to explain how these variants arose. In all, we identified 10 variants and for the sake of simplicity, we only show five genes clusters from these islands, which also contain restriction modification systems amongst others.

determined. Indeed, it will be an exciting prospect to determine whether this proposed non-canonical function of a CRISPR-Cas system plays a role in the acquisition of T3SSs.

Lastly, our phenotypic microarray analysis of growth of RIMD2210633 and UCM-V493 on different carbon sources demonstrate that UCM-V493 could use a wider array of carbon sources and use many substrates significantly better than RIMD2210633. Also, UCM-V493 could grow on tween 20, glycyl-L-glutamic acid, propionic acid and chondroitin sulfate c, whereas RIMD2210633 could not. Interestingly both strains grew significantly better on α and γ-cyclodextrins than D-glucose. Whether these growth differences are strain specific or a reflection of wider patterns between clinical and environmental isolates within the species will need to be examined further.

Chapter 3

TWO ADENYLATE CYCLASES CyaA AND CyaB ARE REQUIRED FOR ACTIVATION OF cAMP RECEPTOR PROTEIN (CRP) IN *VIBRIO PARAHAEMOLYTICUS*

This work will be submitted to the journal *mBio*

Regmi, A. and E.F. Boyd. 2019. Two adenylate cyclases CyaA and CyaB activate cAMP receptor protein (CRP) in *Vibrio parahaemolyticus*. mBio

Introduction

Many bacterial species repress alternative carbon catabolism genes expression when glucose, the preferred carbon source, is present, a mechanism known as carbon catabolite repression (CCR) (Botsford and Harman 1992). Two major players are involved in CCR, adenylate cyclase (AC) which produces cyclic AMP (cAMP) from ATP and cAMP receptor protein (CRP), which is activated by cAMP. The cyclic nucleotide cAMP is a second messenger that is widely used in all domains of life. In *E. coli* and many other *Enterobacteria*, in the presence of glucose, cAMP levels are low and in the absence of glucose, cAMP levels are high and activate the CRP transcription factor (Green, Stapleton et al. 2014). There are six classes of AC based on primary amino acid sequence, with different catalytic mechanisms and structures indicating different evolutionary origins (Barzu and Danchin 1994, Ramakrishnan,

Jain et al. 2016). *Escherichia coli* and other *Proteobacteria* contain the *cyaA* gene, which encodes a class I AC (AC-I), an 810 amino acid protein with two functionally distinct domains; the N-terminal catalytic cyclase domain and the C-terminal regulatory domain. In the absence of glucose (preferred carbon source), the cAMP-CRP complex activates alternative catabolic gene expression (Botsford and Harman 1992, Gorke and Stulke 2008). Recent studies have shown an expanded physiological role of AC-I in bacteria, specifically in virulence, biofilm formation, motility and type III secretion (Skorupski and Taylor 1997, Silva and Benitez 2004, Smith, Wolfgang et al. 2004, Liang, Pascual-Montano et al. 2007, Fong and Yildiz 2008, Kim, Lee et al. 2013). In addition, the regulatory role of AC-I is amplified by its regulation of other global regulators such as the stress response sigma factor RpoS and quorum sensing global regulators (Hengge-Aronis 2002, Liang, Pascual-Montano et al. 2007, Lee, Park et al. 2008, McDonough and Rodriguez 2011, Kim, Kim et al. 2012). Class II ACs are comprised of the bacterial exotoxins such as the calmodulin-activated exotoxin from *Bordetella pertussis*, which is a 1,706 amino acid protein with an N-terminal cyclase catalytic domain and a C-terminal hemolytic domain. This toxin is translocated into target host eukaryotic cells, where it is activated by calmodulin to produce toxic levels of cAMP (Glaser, Ladant et al. 1988, Ladant and Ullmann 1999). The class III ACs are the most diverse and widespread and encompass adenylate cyclases and guanylate cyclases (GCs) from eukaryotic and prokaryotic species (Sinha and Sprang 2006). A well characterized class III AC enzyme from bacteria is CyaC from *Pseudomonas aeruginosa* that has receptor binding activity as well as cyclase activity. This species also contains a class I and a class II AC (Smith, Wolfgang et al. 2004). A class IV AC was first described in *Aeromonas hydrophila* named CyaB, a

small 181 amino acid protein with unusual alkalinophilic and thermophilic activity (Sismeiro, Trotot et al. 1998, Iyer and Aravind 2002). A class IV AC was also described in *Yersinia pestis* and its structure determined (Gallagher, Smith et al. 2006, Smith, Kim et al. 2006, Gallagher, Kim et al. 2011). A structurally diverse AC was identified in *Prevotella ruminicola* and formed the sole representative of a class V AC (Cotta, Whitehead et al. 1998). A class VI AC was described in *Rhizobium elti* and was shown to have cyclase activity by complementation of an *E. coli cyaA* mutant (Tellez-Sosa, Soberon et al. 2002).

The role of CRP as a transcription factor has been extensively studied in *E. coli* where it is required for the expression of multiple carbon metabolism genes (Gosset, Zhang et al. 2004). Phenotypic growth analyses of Δ*crp* in carbon and nitrogen sources showed growth defects in most of the substrates tested demonstrating the essential nature of CRP (Gosset, Zhang et al. 2004). A global transcriptome analysis identified a minimum of 378 operons directly under the control of CRP further establishing CRP as a global regulator in *E. coli* (Shimada, Fujita et al. 2011). Similar to *E. coli*, *V. cholerae*, a human gastrointestinal pathogen, has a single AC, CyaA that binds a homologue of CRP to activate transcription at multiple sites (Liang, Pascual-Montano et al. 2007). CRP also plays a crucial role in controlling virulence gene expression in *V. cholerae* mainly through its regulation of *hapR*, the quorum sensing master regulator LuxR homologue (Skorupski and Taylor 1997, Liang, Pascual-Montano et al. 2007). Transcriptomic analysis demonstrated that 176 genes were differentially regulated in a Δ*crp* mutant strain compared to wild type. This study demonstrated that CRP regulated quorum sensing, motility and multiple genes associated with intestinal colonization (Liang, Pascual-Montano et al. 2007). A recent

study using ChIP-Seq demonstrated 119 CRP binding sites in the genome of *V. cholerae* most overlapping with the previous identified CRP BS (Manneh-Roussel, Haycocks et al. 2018).

Vibrio parahaemolyticus is an emerging human pathogen naturally inhabiting the marine environment, where it exists as free living or associated with shellfish (Kaneko and Colwell 1973, Joseph, Colwell et al. 1982, Nordstrom and DePaola 2003, Zimmerman, DePaola et al. 2007). Consumption of raw or undercooked shellfish infected with *V. parahaemolyticus* causes gastroenteritis in human and *V. parahaemolyticus* is the leading cause of bacterial seafood borne gastroenteritis worldwide (Nair, Ramamurthy et al. 2007, Su and Liu 2007). Previous studies have shown that metabolism is important for *in vivo* fitness and survival of *V. parahaemolyticus* (Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017). However, the role of CRP in *V. parahaemolyticus* physiology has not been previously examined.

In this study, the function and activation of CRP in *V. parahaemolyticus* was investigated using mutational and phenotypic analysis. We constructed deletions of *crp* (VP2793) and *cyaA* (VP2987) in *V. parahaemolyticus* RIMD2201633 and examined growth patterns in various carbon sources. The Δ*crp* mutant grew on minimal medium M9 3% NaCl (M9S) supplemented with glucose (M9SG) but was unable to grow on L-arabinose, N-acetyl-D-glucosamine, D-glucosamine, D-ribose, or D-gluconate as sole carbon sources. The Δ*crp* mutant also showed defects in motility, capsule polysaccharide (CPS) production and biofilm formation. Interestingly, the Δ*cyaA* mutant grew on all carbon sources similar to wild type and showed no motility, biofilm or CPS defects. This suggested the presence of an additional *cya* gene in the *V.*

parahaemolyticus genome. Bioinformatics analysis identified a gene VP1760 that encodes a putative class IV AC *cyaB* that is present in all *V. parahaemolyticus* strains. In motility assays the single Δ*cyaA* and Δ*cyaB* mutants behaved similar to wild type whereas a double Δ*cyaB*/Δ*cyaA* mutant showed growth and motility defects similar to the Δ*crp* mutant. However, the Δ*cyaB* mutant showed a similar defect as the Δ*crp* mutant in CPS production and biofilm formation. Both *cyaA* and *cyaB* complemented an *E. coli* Δ*cyaA* mutant indicating that both genes encode proteins that produce cAMP. These data indicate that *V. parahaemolyticus* produces two ACs but CyaB is essential for CPS production and biofilm formation. Using *A. franciscana* as a model of infection, cAMP and CRP were essential for *V. parahaemolyticus* virulence.

Materials and Methods

Bacterial Strains, Media, and Culture Conditions

All strains and plasmids used in this study are listed in **Table 10**. A streptomycin-resistant clinical isolate *V. parahaemolyticus* RIMD2210633 was used throughout the study. Unless stated otherwise, all *V. parahaemolyticus* strains were grown in Lysogeny Broth (LB) medium (Fischer Scientific, Pittsburgh, PA) containing 3% NaCl (LBS) at 37°C with aeration or M9 medium media (Sigma Aldrich, St. Louis, MO) supplemented with 3% NaCl supplement (M9S) with various carbon sources. Antibiotics were added to growth media at the following concentrations: Ampicillin (Amp), 100 μg/ml, streptomycin (Sm), 200 μg/ml and chloramphenicol (Cm), 10 μg/ml, Kanamycin (Km), 50 μg/ml when required.

Table 10 Strains and plasmids used in this study

Bacterial strains	**Genotype or description**	**References or sources**
Vibrio parahaemolyticus		
RIMD2210633	O3:K6 clinical isolate, Sm^r	
Δ*crp*	RIMD2210633 Δ*crp* (VP2793), Sm^r	This study
Δ*cyaA*	RIMD2210633 Δ*cyaA* (VP2987), Sm^r	This study
Δ*cyaB*	RIMD2210633 Δ*cyaB* (VP1760), Sm^r	This study
Δ*cyaB*/Δ*cyaA*	RIMD2210633 Δ*cyaB* Δ*cyaA* (VP2987), Sm^r	This study
Escherichia coli		
BW25113	Wild-Type *E. coli* K12 strain	
Δ*cyaA*	BW25113, F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, ΔcyaA751::kan, Δ(rhaD-rhaB)568, hsdR514*	
DH5αλpir		Laboratory collection
B2155λpir	*ΔdapA::emr pir*, for bacterial conjugation	Laboratory collection
B2155pDS*crp*	B2155 strain harboring pDS*crp*	This study
B2155pDS*cyaA*	B2155 strain harboring pDS*cyaA*	This study
B2155pDS*cyaB*	B2155 strain harboring pDS*cyaB*	This study
Complement strains		This study
E. coli Δ*cyaA*pCyaA	*E. coli* Δ*cyaA* harboring pBBR*cyaA*	This study
E. coli Δ*cyaA*pCyaB	*E. coli* Δ*cyaA* harboring pBBRcyaA	This study
Plasmids		
pDS132	Suicide plasmid, Cm^r, SacB	
pDSΔ*crp*	pDS132 harboring truncated *crp*, Cm^r	This study
pDSΔ*cyaA*	pDS132 harboring truncated *cyaA*, Cm^r	This study
pDSΔ*cyaB*	pDS132 harboring truncated *cyaB*, Cm^r	This study
pBBR1MCS	Expression vector, lacZ promoter, Cm^r	This study
pBBR*cyaA*	pBBR1MCS harboring full-length *cyaA* (VP2987), Cm^r	This study
pBBR*cyaB*	pBBR1MCS harboring full-length *cyaB*(VP1760), Cm^r	This study

Construction of Mutant Strains in *V. parahaemolyticus* RIMD2210633

Splice by overlapping extension (SOE) PCR and an allelic exchange method (Ho, Hunt et al. 1989) was used to construct an in-frame, non-polar deletion mutant of *crp* (VP2793), *cyaA* (VP2987), and *cyaB* (VP1760) from *V. parahaemolyticus* RIMD2210633. Briefly, primers were designed to the VP2793, VP2987 and VP1760 separately using *V. parahaemolyticus* RIMD2210633 genomic DNA as a template. All primers used in this study are listed in **Table 11**. SOE PCR was conducted to obtain truncated versions of VP2793 (81-bps of 633-bp), VP2987 (153-bp of 2,529-bp) and VP1760 (84-bp of 546-bp). The truncated PCR fragments were then cloned into the suicide vector pDS132 and named pDS Δ*crp,* pDS Δ*cyaA,* and pDS Δ*cyaB*. pDS Δ*crp,* pDS Δ*cyaA,* and pDS Δ*cyaB* were then transformed into *E. coli* strain β2155 λ*pir*, which were then conjugated separately into *V. parahaemolyticus* RIMD2210633. Conjugation was conducted by cross streaking both strains onto LBS plate containing 0.3 mM DAP. Bacterial growth lawn from this plate was scrapped, suspended in LB, serially diluted and plated on LBS plate containing streptomycin and chloramphenicol. Next day, the colonies were verified for single crossover via PCR. The colonies with single crossover were grown over night in LBS with no antibiotic added and were plated onto LBS plate containing 10% sucrose to select for double crossover deletion mutant. Double deletion mutant was constructed by deleting *cyaA* from Δ*cyaB* background following the above-mentioned protocol. The gene deletions were confirmed by PCR

Table 11 Primers used in this study.

Primer name and use	Sequence (5'-3')[a]	Melting temp (°C)	Size (bp)
SOE PCR			
VP1760 SOEA	GGATCTAGACAGCTCAACTGGGTTTGGTT	62	408
VP1760 SOEB	TTCAAACTGCCCTTGGAAATG	54	
VP1760 SOEC	CATTTCCAAGGGCAGTTTGAAGAAGCGGATAGAGAGCATCG	66	513
VP1760 SOED	CACGAGCTCGCTGCTTATCTTTGCGCTTT	64	
VP1760 SOEFF	ACGGTGAAAGCCATCAAAAC	54	1,708
VP1760 SOEFR	CCTTCGGTCTTACAATAACCAATC	54	
VP2793 SOEA	TCTAGATTGTGGTGATTCGTCTGCGT	60	458
VP2793 SOEB	TGTTGGGTCGGTTTGAGGTT	57	
VP2793 SOEC	AACCTCAAACCGACCCAACACTAATCTCTGCGCACGGTAA	68	451
VP2793 SOED	GCATGCCCATCTTCGTCAAGCAGGGT	65	
VP2793 SOEFF	CTGCACACCTTTGACGAAGC	57	2,408
VP2793 SOEFR	GAGTCGTTATGACGCCCGAT	57	
VP2987 SOEA	TCTAGACAGGATGAGCGGCTTGAAGA	61	513
VP2987 SOEB	GCGATCGATACGCTGCTGGTT	60	
VP2987 SOEC	**AACCAGCAGCGTATCGATCGC**CCACCCTGAAGAGGGTAACGC	71	503
VP2987 SOED	GAGCTCGCACCACACTCAGGAAAAGC	64	
VP2987 SOEFF	CGCACCTTGAGGGAGTTCAT	57	4,042
VP2987 SOEFR	GAGGATGTTGACCCACAGCA	57	
Complementation primers			
cyaA com F	CAAGGATCCTGTTCAACCAAGGAATTACCC	61	2,563
cyaA com R	GCATCTAGAGGATATGTTTACGTTACGCG	59	
cyaB com F	CTTGGATCCAGTTTATGTAGTAGAGGCGA	59	576
cyaB com R	CACTCTAGAGAAATAGCGACTATGCTGAT	57	

[a]Underlined base pairs indicate restriction sites

Complementation Assay

Primer pairs were designed to amplify *cyaA* (VP2987) and *cyaB* (VP1760) using the genomic DNA of RIMD2210633 as template (**Table 11**). Purified PCR products of *cyaA* and *cyaB*, and the vector pBBR1MCS were digested using BamHI

and XbaI. Both the insert and the vector were gel purified. The genes were each cloned into the vector pBBR1MCS creating pBB*cyaA* or pBB*cyaB*, which were then transformed into *E. coli* B2155 using calcium chloride transformation method. *E. coli* B2155 containing pBB*cyaA* and pBB*cyaB* were conjugated into *E. coli* Δ*cyaA* (Baba, Ara et al. 2006). Conjugation was conducted as described above and the selection for the complement strain was done using LB agar supplemented with Km and Cm. Complement strains were verified using PCR. IPTG (0.5mM) was used to induce the transcription of the respective genes from the vector.

Growth Pattern Analysis

Bacterial cultures were grown overnight in LBS or M9S supplemented with 10mM glucose at 37°C with aeration. This overnight growth culture was spun down, suspended and washed with PBS and a 1: 40 dilution was then used to conduct growth pattern analysis in LBS or M9S supplemented with 10 mM of each individual carbon source. Growth curves were conducted in a 96 well plate with shaking for 1000 sec in between cycles and optical density (OD595) was measured each hour for 24 hours. The data was plotted using GraphPad Prism 7.00, GraphPad software, La Jolla California USA, www.graphpad.com.

Swimming, Swarming, CPS production, and Biofilm Assay

Swimming assays were performed in semi solid media, LB 2% NaCl with 0.3 % agar and swarming assays were performed in heart infusion (HI) broth containing 2% NaCl and 1.5 % agar. To study swimming behavior, a single colony of the test bacterium was picked and stabbed in the center of the swimming media, and the plates were incubated at 37°C for 24 hrs. For the swarming assay a single colony was picked

and gently inoculated on the surface of the swarming media, and the plates were incubated at 30°C for 48 hrs. To examine CPS production, HI media containing 1.5% agar, 2.5mM $CaCl_2$, and 0.25% Congo red dye was inoculated on the surface of the plate and incubated at 30°C for 48 hrs. Biofilm formation assays were performed using crystal violet staining. Cultures were grown overnight in LBS and 1:40 dilution was then inoculated in LBS in a 96 well plate, static at 37°C for 24 hrs. After 24 hrs, the wells are washed with PBS, stained with crystal violet for 30 min and biofilm formation was then documented. The biofilms are then dissolved in DMSO and the OD was measured to quantify the biofilm.

Polymyxin B Sensitivity Assays

Overnight grown bacterial cultures were diluted (1:50) into LBS and grown for 2 hrs (early exponential stage), 1 ml of culture was spun down and suspended in 1 ml of 1X PBS. 200 µl PBS suspended cultures were added in a 96 well plates. Polymyxin B sulphate (8 µg/ml) (Sigma-Aldrich) was then added to each of the wells containing the bacterial cultures. Killing assay was conducted by obtaining the OD every 5 min for 2 hrs. The data was plotted using GraphPad Prism 7.00, GraphPad software, La Jolla California USA, www.graphpad.com. The experiment was conducted in triplicate using two biological replicates.

Bioinformatics Analysis to Identify CRP Binding Sites

The regulatory region upstream of the targeted operons of *V. parahaemolyticus* RIMD210633 were obtained using the NCBI nucleotide database. An online tool virtual footprint was used to identify the CRP binding sites using *E. coli* CRP consensus binding sequence (Münch, Hiller et al. 2003).

Artemis franciscana Infection Assay

The *A. franciscana* infection assay was conducted using previously described method with modification as required (Ha, Kim et al. 2014, Lee, Kim et al. 2014, Kim, Jang et al. 2018, McDonald, DeMeester et al. 2018). Briefly, cysts of *A. franciscana* (San Francisco Bay Brand ASF65031) were hatched using artificial seawater (Instant Ocean SS15-10 08/08) at 28°C with light and aeration for 48 hrs. After 48 hrs, 20 nauplii were transferred into 5 ml of artificial sea water into a 25 mm petri dish. *V. parahaemolyticus* strains were grown for 2 hrs and bacterial culture were added to the 25 mm petri dish to a final concentration of $1X10^7$ cells/ml. The petri dish was incubated at 30^o C, and at 24 and 48 hrs, the nauplii were inspected visually and scored as dead and alive. Two biological replicates with three technical replicate plate of nauplii per biological replicate were used to conduct the infection assay.

Results

CRP is an Essential Global Regulator in *Vibrio parahaemolyticus*

cAMP-CRP is a major regulatory mechanism for carbon metabolism and virulence in many bacterial species (Gosset, Zhang et al. 2004, Liang, Pascual-Montano et al. 2007, Shimada, Fujita et al. 2011, Kim, Lee et al. 2013, Manneh-Roussel, Haycocks et al. 2018). We identified homologues of CRP and adenylate cyclase CyaA in *V. parahaemolyticus*, VP2793 and VP2987, respectively. The CRP protein of *V. parahaemolyticus* shared 98% amino acid identity to CRP in *V. cholerae* (**Figure 16A**).

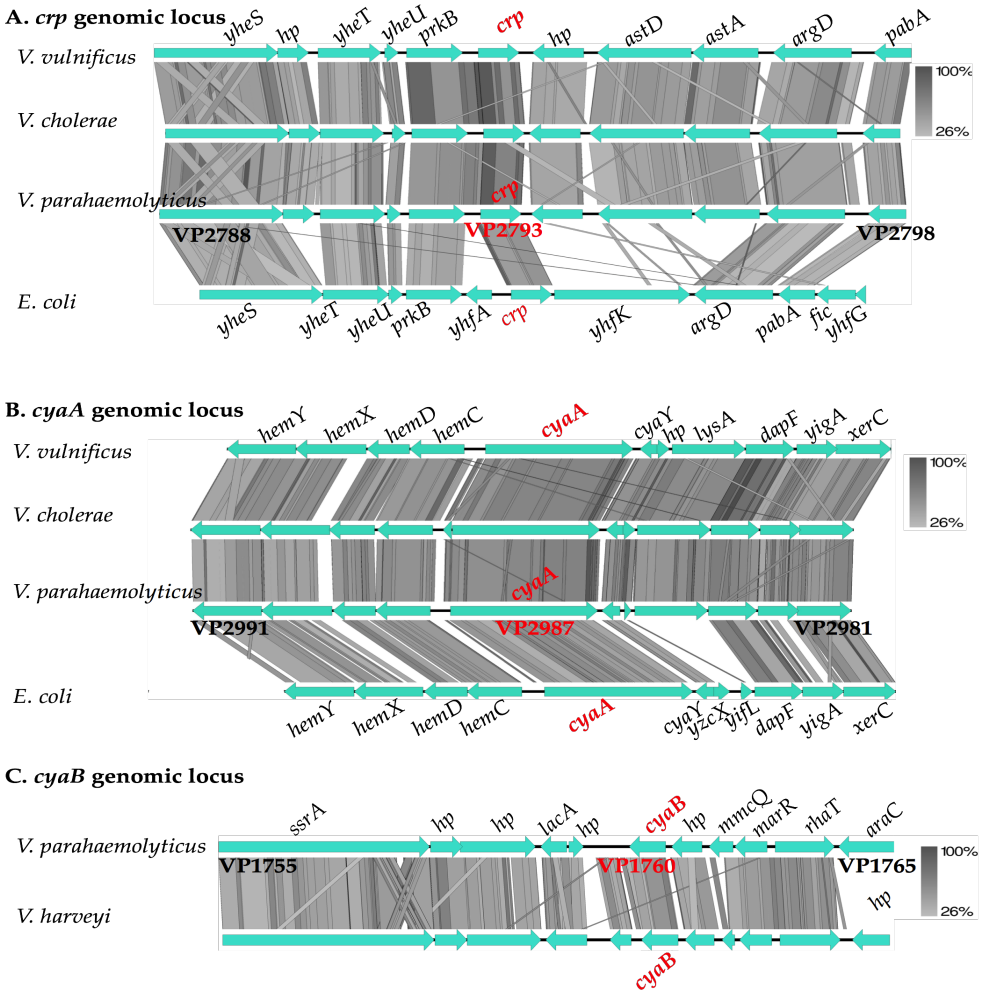


Figure 16 **Genomic context comparison of *crp*, *cyaA* and *cyaB* loci.** The genomic region upstream and downstream of **(A)** *crp* and **(B)** *cyaA* **(C)** *cyaB* from *V. parahaemolyticus* was compared with that of *crp* and *cyaA* from *V. cholerae*, *V. vulnificus* and *E. coli* and *cyaB* from *V. harveyi* using tBLASTx, and figures were constructed using Easyfig.

Similarly, the 842 amino acid sequence of CyaA shared 82% amino acid identity with that of *V. cholerae* (**Figure 16B**). The genome context of both genes was conserved across *V. parahaemolyticus*, *V. cholerae* and *E. coli* (**Figure 16A and 16B**). To begin to investigate the physiological role of CRP in *V. parahaemolyticus,* we first

constructed an in-frame deletion of *crp* (VP2793) and *cyaA* (VP2987) in *V. parahaemolyticus* RIMD2210633. The Δ*crp* mutant displayed small and translucent colony morphology on LBS plates **(Figure 17A)**.

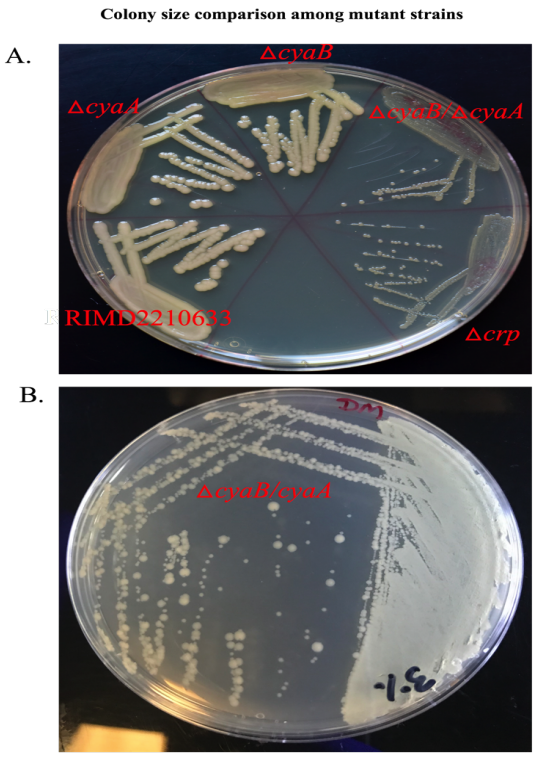


Figure 17 **Colony morphology of *Vibrio parahaemolyticus* strains. (A)** Single colony of *V. parahaemolyticus* RIMD2210633 and its isogenic mutant strains Δ*cyaA*, Δ*cyaB*, Δ*cyaB*/Δ*cyaA,* and Δ*crp* were plated on LBS agar plate and grown overnight at 37^{o}C before the picture was taken. **(B)** An inoculum from the glycerol stock of Δ*cyaB*/Δ*cyaA* was plated on LBS agar and grown overnight at 37^{o}C before the picture was taken.

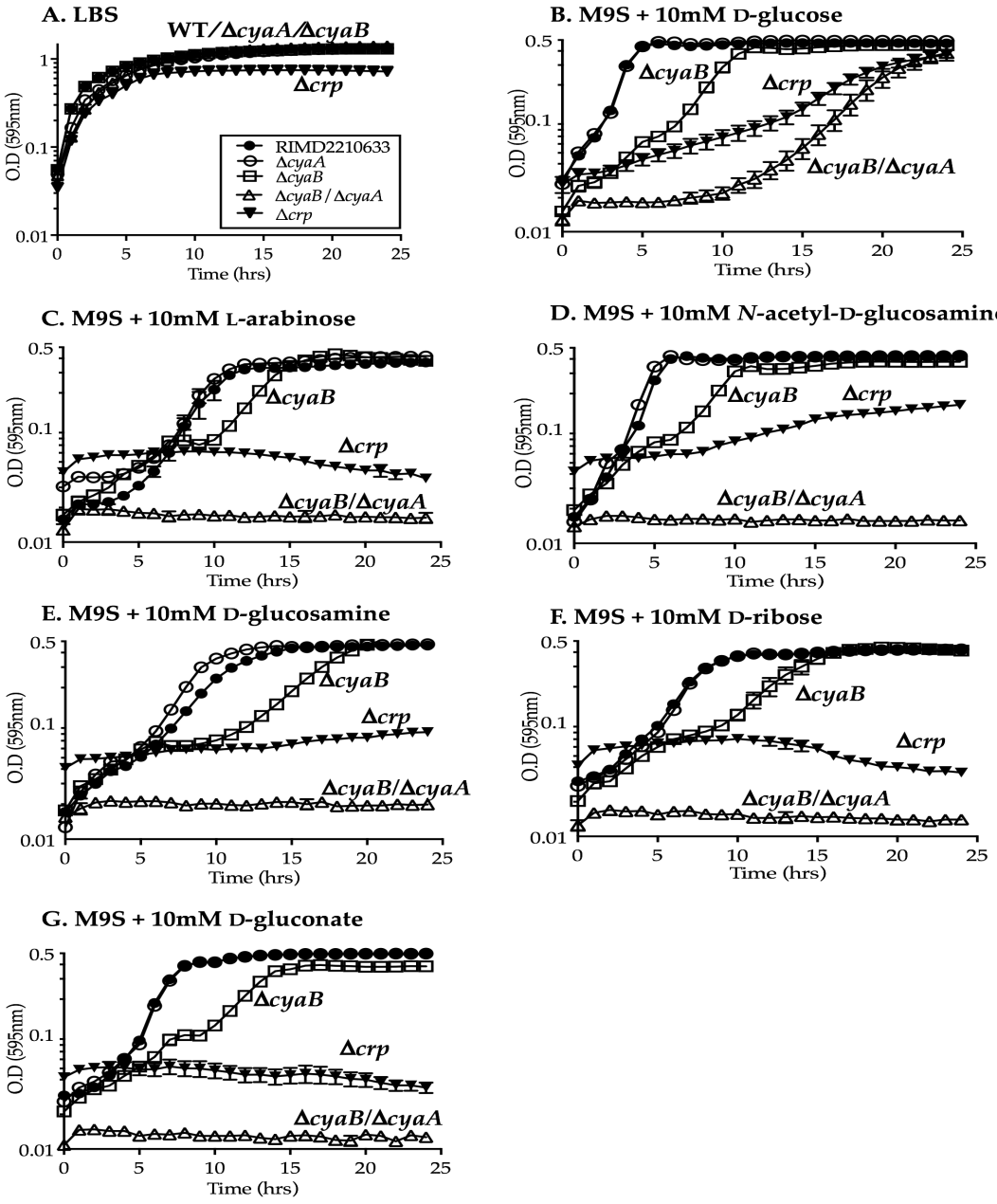


Figure 18 **Growth curve analysis.** Growth pattern analysis of RIMD2210633 (closed circle), Δ*cyaA* (open circle), Δ*cyaB* (open square), Δ*cyaB/*Δ*cyaA* (open triangle) and Δ*crp* (closed downward facing triangle) in **(A)** LBS **(B)** M9S + 10mM D-glucose **(C)** M9S + 10mM L-arabinose **(D)** M9S + 10mM NAG **(E)** M9S + 10mM D-glucosamine **(F)** M9S + D-ribose **(G)** M9S + 10mM D-gluconate. For growth in minimal media, the strains were grown overnight in M9S + 10mM glucose.

Growth curve analysis demonstrated that the Δ*crp* strain had a growth defect in LBS broth and in M9S D-glucose **(Figure 18A and 18B**). Whereas in M9S supplemented with L-arabinose, *N*-acetyl-D-glucosamine, D-ribose, or D-gluconate, the Δ*crp* strain was unable to grow (**Figure 18C-Figure 18G**). These data demonstrate that CRP is required for the catabolism on alternative carbon sources similar to many other species. In addition, bioinformatics analysis identified strong putative CRP binding sites in the regulatory region of L-arabinose, *N*-acetyl-D-glucosamine, D-glucosamine, D-ribose, and D-gluconate catabolism operons suggesting that CRP directly regulates their transcription (**Figure 19A-Figure 19E**).

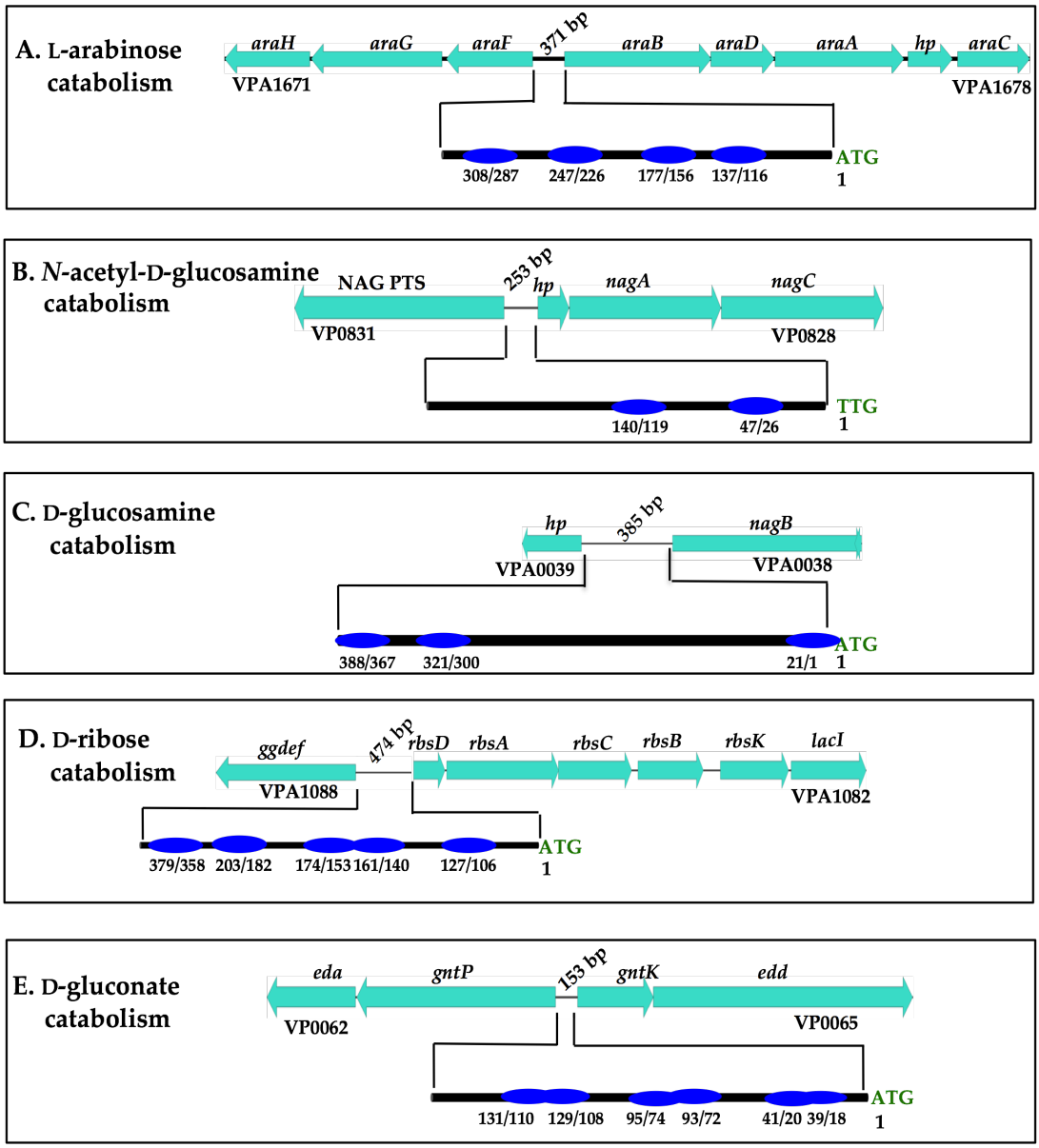


Figure 19 ***In silico* analysis of CRP binding sites in metabolism gene clusters.** Putative CRP binding sites were identified in the intragenic region of catabolism genes of **(A)** L-arabinose **(B)** NAG **(C)** D-glucosamine **(D)** D-ribose **(E)** D-gluconate. Virtual footprint program and the CRP binding consensus sequence of *E. coli* was used. Blue ovals represent the predicted CRP binding sites. The numbers indicate the start and the end of the CRP binding sites respective to the "A" of the ATG.

CRP requires cAMP for activation, which is produced by adenylate cyclase encoded by *cyaA* (VP2987) in *V. parahaemolyticus* RIMD2210633. Thus, deletion of *cyaA* (Δ*cyaA*) should give similar growth phenotypes as the Δ*crp* deletion mutant. Growth of Δ*cyaA* on LBS showed identical colony morphology to wild type (**Figure 17A**). Similarly, we examined the growth pattern in LBS broth and M9S supplemented with D-glucose, L-arabinose, *N*-acetyl-D-glucosamine, D-glucosamine, D-ribose, or D-gluconate, and found that the mutant grew similar to wild type (**Figure 18A-Figure 18G**). These data suggested that *V. parahaemolyticus* RIMD2210633 contained an addition uncharacterized adenylate cyclase.

BLAST analysis using CyaA as a seed against the RIMD2210633 genome, gave only one hit to itself indicating no other homologue of the class I adenylate cyclase is present. We then searched the *V. parahaemolyticus* genome for class II, III and class IV adenylate cyclases and identified a homologue of CyaB, a class IV adenylate cyclase, previously identified in *Yersinia pestis*. The CyaB protein of *V. parahaemolyticus* shared 45 % amino acid identity with CyaB of *Y. pestis* and 56% identity with CyaB of *Aeromonas hydrophila*. The CyaB in both these species has been structurally characterized and shown to have AC activity generating cAMP (Sismeiro, Trotot et al. 1998, Smith, Kim et al. 2006, Bettendorff and Wins 2013). The X-ray structure of this putative adenylate cyclase Q87NV8 from *V. parahaemolyticus* has been resolved at the 2.25 A resolution (pdb2ACA). Similar to *Y. pestis* class IV adenylate cyclase, CyaB (VP1760) is predicted to cyclizes both ATP and GTP. CyaB lacks transmembrane domains and signal peptide cleavage sites indicating this is a cytosolic protein. No additional adenylate cyclases were identified in *V. parahaemolyticus*. To determine whether CyaB produces cAMP, we complemented an

E. coli Δ*cyaA* mutant with either *V. parahaemolyticus cyaA* or *cyaB* and examined growth in several carbon sources. Both genes complemented the *E. coli* Δ*cyaA* mutant indicating both have adenylate cyclase activity (**Figure 20)**.

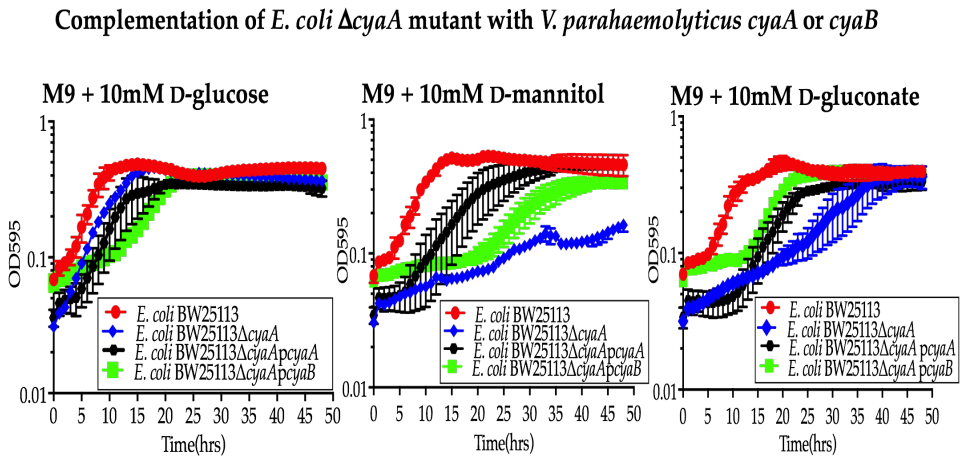


Figure 20 **Complementation analysis of *E. coli* Δ*cyaA*.** The *E. coli* Δ*cyaA* strain was complemented separately with a functional copy of CyaA and CyaB from *V. parahaemolyticus* and growth curves were conducted in **(A)** M9 + 10 mM glucose, M9 + 10mM mannitol, and M9 + 10mM gluconate for 48 hours.

We reasoned that our lack of a Δ*cyaA* mutant phenotype is due to CyaB complementing the Δ*cyaA* mutant. To investigate this further, we constructed a Δ*cyaB* strain and examined growth. The Δ*cyaB* mutant had similar colony morphology to wild type and showed identical growth patterns in LBS broth **(Figure 17A and Figure 18A)**. The Δ*cyaB* mutant showed longer doubling times in M9S supplemented with D-glucose, L-arabinose, *N*-acetyl-D-glucosamine, D-glucosamine, D-ribose, or D-gluconate but reached a similar biomass to wild type **(Figure 18B-Figure 18G)**.

Next, we constructed a Δ*cyaB*/Δ*cyaA* double mutant strain by deleting *cyaA* from the Δ*cyaB* background. The double mutant demonstrated small translucent colony morphology on LBS plates similar to the colonies observed for Δ*crp* **(Figure 17A)**. Interestingly, inoculum obtained from glycerol stocks of the double mutant on a LBS gave small translucent colonies and normal opaque colonies suggesting suppressor mutations were occurring. We did not see this phenotype in either of the *cya* single mutants (**Figure 17B**). In *V. cholerae,* which contains one adenylate cyclase*,* the Δ*cyaA* mutant reverting to wild-type phenotypes was observed (Skorupski and Taylor 1997). In *E. coli* it was demonstrated that suppression of Δ*cyaA* may lead to compensatory mutation in *crp* leading to CRP function being independent of cAMP to some degree (Harman and Dobrogosz 1983). For our analysis small colonies were always used for characterization studies. The double mutant showed a growth defect on M9SG and showed no growth in M9S supplemented with L-arabinose, *N*-acetyl-D-glucosamine, D-glucosamine, D-ribose, or D-gluconate **(Figure 18B-Figure 18G)**. Overall the growth assays suggest that CyaB is the predominant adenylate cyclase used under the conditions examined and CyaA and CyaB can compensate for each other.

Swimming and Swarming Motility

Vibrio parahaemolyticus produces both polar and lateral flagella required for swimming and swarming motility, respectively. To investigate the role of CRP in the motility of *V. parahaemolyticus*, we conducted swimming and swarming assays and found that the Δ*crp* mutant was defective in swimming compared to wild type (**Figure 21B**).

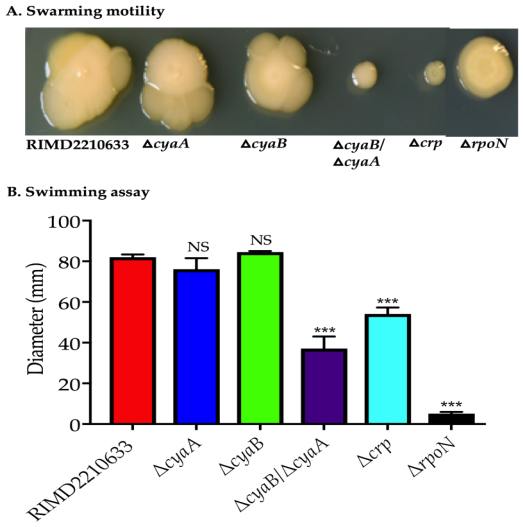


Figure 21 **Motility characterization of *V. parahaemolyticus* mutants. (A)** Colonies of *V. parahaemolyticus* RIMD2210633 and mutant strains Δ*crp,* Δ*cyaA,* Δ*cyaB*, and Δ*cyaB*/Δ*cyaA* were accessed for swarming motility in semi solid media over 48 hrs. **(B**) Swimming assays were conducted using *V. parahaemolyticus* RIMD210633 and mutant strains Δ*crp,* Δ*cyaA,* Δ*cyaB*, and Δ*cyaB*/Δ*cyaA* and the diameter of the region covered by the bacterial strains were measured and represented as bar graph. For both assays Δ*rpoN* is also used as a control as the strain was previously shown to be defective in both swimming and swarming. Error bars indicate standard error of mean (SEM). Unpaired Student *t* test was conducted to obtain the *P* values. The significant difference in swimming is denoted by asterisks. *, P<0.05, **, P<0.01, ***, P<0.001. NS, not significant.

Similarly, the Δ*crp* strain also demonstrated a defect in swarming as indicated by the absence of the budding phenotype compared to wild type (**Figure 21A**). Neither of the single Δ*cyaA* nor Δ*cyaB* mutants showed a motility defect but the double mutant did in both the swimming and swarming assays similar to Δ*crp* (**Figure 21B and 21A**). The motility assays demonstrated that the cAMP-CRP is important for the motility of *V. parahaemolyticus*. We examined the regulatory region of the operons containing *fliA* (VP2230) and *lafS* (VPA1555), genes encoding sigma factors for polar and lateral

flagella biosynthesis respectively and identified putative CRP binding sites in both regulatory regions indicating that CRP maybe a direct regulator of polar and lateral biosynthesis in *V. parahaemolyticus* (**Figure 22A and 22B**).

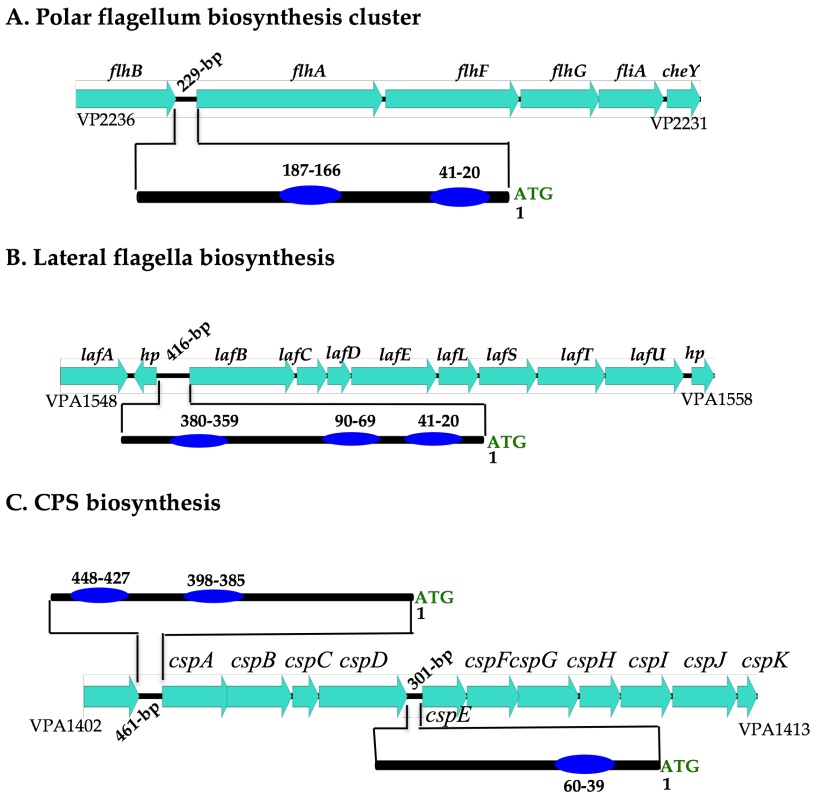


Figure 22 ***In silico* analysis of CRP binding sites. (A)** Putative CRP binding sites were identified in the intragenic region of polar flagellum biosynthesis genes **(B)** lateral flagella biosynthesis genes and (C) CPS biosynthesis genes. Virtual footprint program and the CRP binding consensus sequence of *E. coli* was used for the *in silico* analysis. Blue ovals represent the predicted CRP binding sites. The numbers indicate the start and the end of the CRP binding sites respective to the "A" of the ATG.

Capsule Polysaccharide (CPS) Production and Biofilm Formation

Next, we examined capsule polysaccharide (CPS) production as it was previously shown in *V. vulnificus* that the *crp* mutant produced less capsule (Kim, Lee et al. 2013). In our CPS assays, the Δ*crp* mutant had a smooth and translucent colony morphology compared to the opaque and rugose colony morphology of wild type (**Figure 23A**).

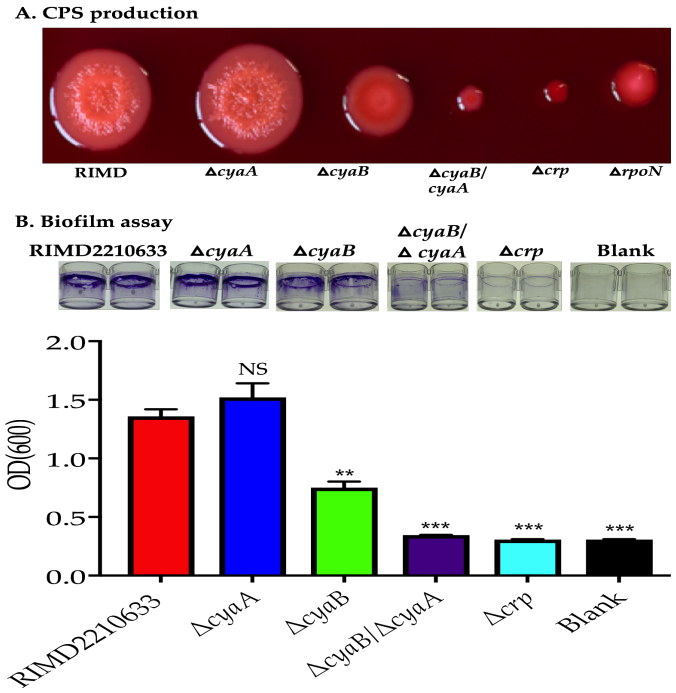


Figure 23 **Capsule polysaccharide production and biofilm formation of *V. parahaemolyticus* mutants. (A)** Colonies of *V. parahaemolyticus* RIMD2210633 and mutant strains were inoculated on agar plate containing Congo red dye. **(B)** The biofilm produced by *V. parahaemolyticus* strains was accessed using crystal violet dye after 24 hours after the start of the assay. The biofilms were dissolved using DMSO, and the optical density of the biofilms were measured at OD (600) and represented as bar graphs. Error bars indicate standard error of mean (SEM). Unpaired Student *t* test was conducted to obtain the *P* values. The significant difference in biofilm production is denoted by asterisks. *, P<0.05, **, P<0.01, ***, P<0.001. NS, not significant.

This indicates that in the mutant less CPS is produced similar to what was shown in *V. vulnificus*. The Δ*cyaA* mutant had an opaque and rugose colony morphology similar to wild type, but the Δ*cyaB* mutant produced a smooth and translucent colony morphology similar to the Δ*crp* mutant. This indicates that Δ*cyaB* mutant cannot be compensated by Δ*cyaA* in the production of CPS, and has a requirement for CyaB. The Δ*cyaB*/Δ*cyaA* double mutant gave a smooth and translucent colony morphology demonstrating less CPS is produced.

Defects in CPS production prompted us to investigate the role of CRP in biofilm production as previous work in this species has shown that CPS production results in robust biofilm formation (McCarter 1999, Enos-Berlage, Guvener et al. 2005). *Vibrio parahaemolyticus* biofilm production was visualized using a crystal violet stain and quantified by dissolving the biofilm in DMSO and measuring the optical density. The biofilm assay showed that the Δ*crp* mutant was significantly defective in biofilm production as indicated by the absence of a dense biofilm ring compared to the wild type strain (**Figure 23B**). Quantification of the biofilm also demonstrated the biofilm produced by Δ*crp* strain was significantly less compared to the wild type strain (**Figure 23B**). Analysis of the Δc*yaA* mutant showed similar levels of biofilm to wild type, however the Δ*cyaB* mutant showed significant defects. As expected, the Δ*cyaB*/Δ*cyaA* double mutant also had defects in biofilm formation.

Antimicrobial Peptide Resistance

To investigate the role of CRP in antimicrobial peptide resistance, a killing assay was conducted by measuring the OD of the bacterial strains after treating them with polymyxin B (8 µg/ml). The decreasing OD indicates cell lysis due to polymyxin B. The wild type strain demonstrated intermediate killing as indicated by the gradual

decrease in the OD over time, however, a polymyxin B sensitive strain Δ*rpoE* (Haines-Menges, Whitaker et al. 2014) demonstrated enhanced killing (**Figure 24**).

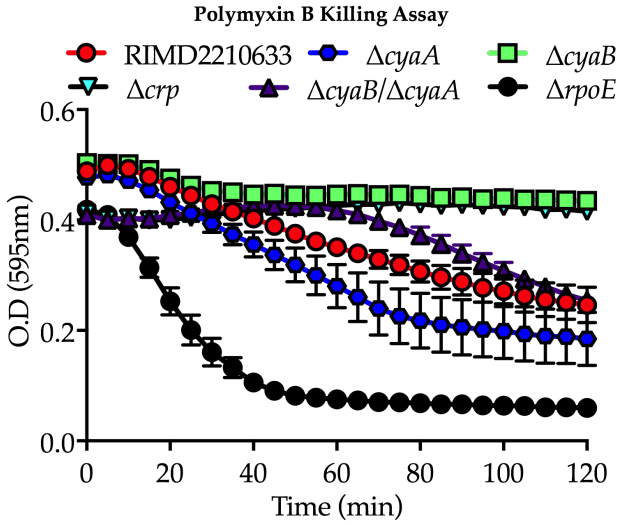


Figure 24 **Polymyxin B killing assay.** *V. parahaemolyticus* wild type and mutant strains Δ*crp,* Δ*cyaA,* Δ*cyaB*, and Δ*cyaB*/Δ*cyaA* were exposed to 8 µg/ml of polymyxin B and the OD at 595nm was measured every five minutes.

The Δ*crp* mutant demonstrated resistance towards polymyxin B, with no cell lysis occurring (**Figure 24**). Similarly, the *cyaB* mutant showed resistance to polymyxin B. A common feature of both these mutants is the absence of a CPS, which could suggest that CPS increases sensitivity to polymyxin. Further investigation is required to determine the molecular mechanism for polymyxin B resistance.

The Double Mutant (Δ*cyaB/ΔcyaA*) and the Δ*crp* Mutant Have Attenuated Virulence Towards *Artemia franciscana*

To determine whether CyaA, CyaB, and CRP are essential for virulence, we conducted *Artemia franciscana* killing assays. We used RIMD2210633 and heat killed RIMD2210633 as our controls. After 24 hours exposure to the shrimps with RIMD2210633 and heat killed RIMD2210633, no death of the shrimp was observed in both conditions, however, after 48 hours of exposure, only 69% of the shrimp exposed to RIMD2210633 survived and 92% of the shrimp exposed to heat killed RIMD2210633 survived demonstrating the virulence of RIMD2210633 in this model (**Figure 25**). Next, we conducted similar a killing assay using Δ*cyaA,* Δ*cyaB,* Δ*cyaB/*Δ*cyaA, and* Δ*crp* mutants. Almost 100% of the shrimps exposed to these mutants survived after 24 hours of exposure, however, only 70.5% and 71.5%, of the shrimp survived after 48 hours of exposure to Δ*cyaA* and Δ*cyaB,* whereas 92% and 95% of the Δ*cyaB/*Δ*cyaA* and Δ*crp* mutants, respectively (**Figure 25**). The percent survival of shrimp exposed to Δ*cyaA* and Δ*cyaB* at 48 hours were almost identical with that of RIMD2210633. The percent survivals of shrimp exposed to Δ*cyaB/*Δ*cyaA* and Δ*crp* were significantly higher compared to RIMD2210633 (**Figure 25**). The data from the killing assay demonstrated that cAMP-CRP is an essential virulence factor for *V. parahaemolyticus*.

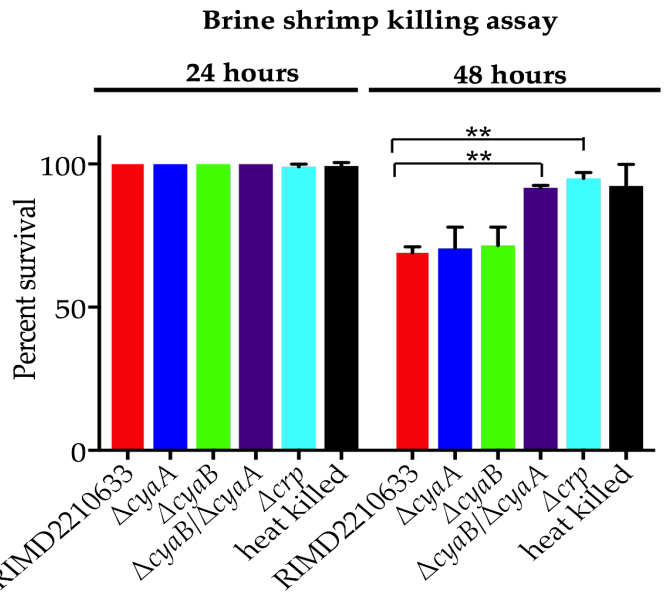


Figure 25 **. *Artemia franciscana* killing assay.** *Artemia* nauplii were exposed to $1X10^7$ CFU of either RIMD2210633, Δ*cyaA,* Δ*cyaB*, Δ*cyaB*/Δ*cyaA,* Δ*crp and* heat killed (HK) RIMD. The number of nauplii survived were then counted at 24 and 48 hours post exposure. The killing assay was conducted using at least two biological replicates and three technical replicates. Bar graph indicates mean, and error bars represents SEM. One-way ANOVA was conducted to identify the P-values. *, P>0.05; **, P>0.01, ***, P>0,001.

Distribution and Phylogenetic Analysis of *cyaB*.

CyaB was first discovered in the late 1990s and found in only a handful of species, notably *A. hydrophila, Y. pestis* and a number of hyperthermophilic archaebacteria (Sismeiro, Trotot et al. 1998). We used *cyaB* (VP1760) as a seed and first examined the distribution among members of the family *Vibrionaceae* in the NCBI genome database. The *cyaB* gene was present in all *V. parahaemolyticus* strains and was highly conserved. CyaB was absent from most *V. alginolyticus* strains, a closely related species but was present in other members of the Campbellii clade to

which *V. parahaemolyticus* belongs. The *cyaB* gene was present in *V. harveyi, V. campbellii, V. jasicida, V. owensii, V. jasicida, and V. rotiferanus* from the Campbellii clade (**Figure 26**).

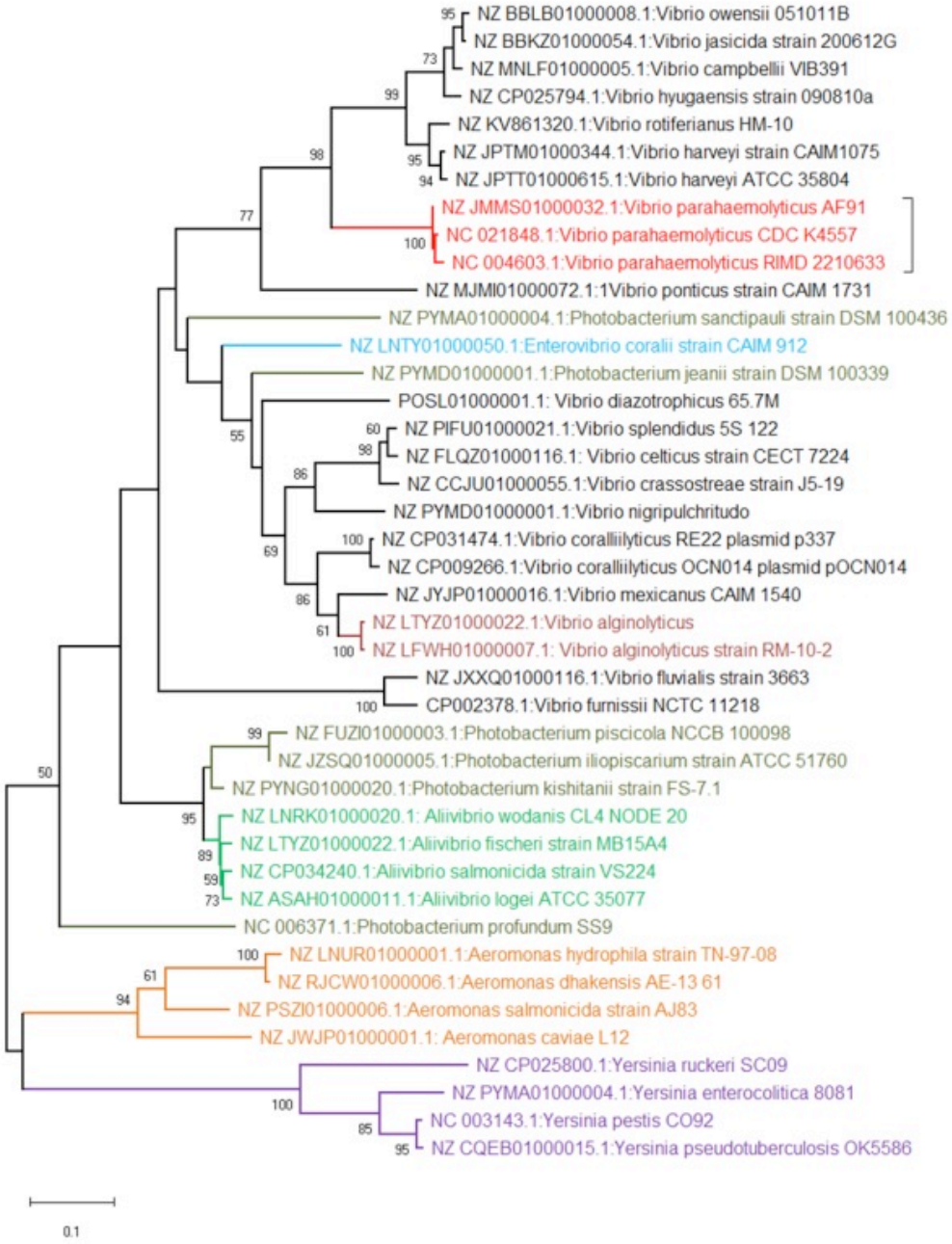


Figure 26 **Evolutionary analysis by Maximum Likelihood method of *cyaB*.** The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model (Kimura, 1980). The tree with the highest log likelihood (-5729.11) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.4899)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 42 nucleotide sequences. Codon positions included were 1st+2nd. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 360 positions in the final dataset. Evolutionary analyses were conducted in MEGA X ((Kumar, Stecher et al. 2018).

On the *cyaB* gene tree these species formed highly divergent branches and only show ~70% amino acid identity with CyaB from *V. parahaemolyticus* where as CyaA shows ~95 amino acid identity among these species. Most of the *cyaB* genes from *Aliivibrio* and *Photobacterium* cluster closely together and divergent from these are *cyaB* genes from several *Vibrio* species. Of note is *cyaB* from a strain of *V. alginolyticus*, a sister species of *V. parahaemolyticus* on a divergent branch and clustering with *cyaB* from *V. coralliilyticus,* which is present on a plasmid. Forming separate divergent cluster are *cyaB* from *Aeromonas* species and *Yersinia* species. Overall the phylogeny of *cyaB* suggest very different evolutionary histories in different species.

Discussion

CRP is an important global regulator in many bacterial species and is activated after binding to its allosteric effector, cAMP, a second messenger, which is synthesized by the enzymatic action of adenylate cyclase encoded by *cyaA*. For the first time, we demonstrate that in *V. parahaemolyticus*, CRP is an important global regulator of metabolism and also regulates motility, CPS production and biofilm formation. In addition, deletion of *crp* results in virulence attenuation in a *A. franciscana* model of infection. Unlike in *V. cholerae* and *V. vulnificus,* where deletion of *cyaA* gave phenotypes identical to Δ*crp*, the *V. parahaemolyticus cyaA* mutant gave phenotypes similar to wild type. This led us to the identification and characterization of a class IV AC present in all *V. parahaemolyticus* strains but is absent in its sister species *V. alginolyticus*. Homologues of CyaB are present in other members of the Campbellii clade, however they shared <70% amino acid identity with *V. parahaemolyticus* whereas CyaA shares >95% amino acid identity among this group. Our study shows for the first time a role for a class IV AC as a second messenger and that the class IV AC is the predominant type use in *V. parahaemolyticus*.

In all non-glucose carbon sources examined, the *V. parahaemolyticus* Δ*crp* mutant strain was unable to grow, similar to what was observed for the *crp* mutants of *E. coli* and *V. cholerae* (Gosset, Zhang et al. 2004, ChenBaoli, LiangWeili et al. 2013). In *V. cholerae,* using *in silico* analysis, putative CRP binding sites were identified in the regulatory region of the carbon catabolism genes and confirmed by EMSA (ChenBaoli, LiangWeili et al. 2013). Using bioinformatics analysis, we also identified multiple putative CRP binding sites in the regulatory region of multiple carbon sources suggesting direct regulation.

Deleting *crp* in *V. parahaemolyticus* resulted in diminished CPS production. In *V. cholerae,* CRP negatively regulates *vpsR*, an activator of *vps* genes expression that synthesize the capsule (Fong and Yildiz 2008) . In *V. parahaemolyticus,* a homologue of *vpsR* is present *cpsR* (VP0514) along with *cspQ* (VPA1446) an additional CPS regulator both of which are dispersed from the CPS gene cluster (VPA1403-1413) (Guvener and McCarter 2003, Ferreira, Chodur et al. 2012). We identified CRP binding sites in the CPS biosynthesis gene clusters and in the regulatory region of both regulators (**Figure 22 and Figure 27**). This binding needs to be confirmed but does suggest direct regulation of CPS production by CRP. Previous work in *V. parahaemolyticus* has shown that the quorum sensing master regulator OpaR (LuxR homologue) is a direct regulator of CPS biosynthesis. Indeed it has been shown that deletion of *opaR* results in a phenotype identical to the Δ*crp* mutant (Kalburge, Carpenter et al. 2017). We identified CRP binding sites in the *opaR* regulatory region suggesting that the Δ*crp* CPS phenotype could be through regulation of OpaR (**Figure 27**).

Regulatory region of *opaR* (VP2516) in *Vibrio parahaemolyticus*

BS1
ACTGTGCTCAATTTAGTTTGTGGGTTGAGGTAGGTCGCACGCAAACGATCACCTAATCTT
BS2 BS3
ACGGCGTGTAGTGTATCACCACGAAGAGTGGGTTGAAAGTCACATCCTTTTGCGGAAAAT
BS4
TCTTGGGAGCGCTGCCAAATTTTCGAAAGTTTACAGCAAAACTTATCAATAATAATGACA
BS5
TTACTGTCTATAAATTATCCATCAGATGTTGACCTTGTGCATATGCACCATTACACTCAT

CACTGCTTAAAGCAATTATTAAAATAATCAATTAGGGGGTTACCCCCAACAACAACTCAA

ATGGCAAGGAAAATGGAT**ATG**

Regulatory region of *cspR* (VP0514) in *Vibrio parahaemolyticus*

BS1
TTCTTTGCTGAACAATGACAAAAAAACCACCTTCGGGTGGTTTTTTTGTATCTTTAATTTCCTA

AAATTCAATAGGTTGCATTTTGATTTTGATGGCATATGCAAATAAGTATTCGCCAAATCTCAAT
BS2
CTATTGCATAGTCTTAATATTGAGACAGACTTCGAGCGTCGTAGTCTCGATAATGTTACCCGTT

CGCAAAGATGGGTTTGTGGCGCGAAATTTCTATCAACAAAAGAATAAGGGTTAGAGA**ATG**

Figure 27 **Regulatory region of *opaR* and *cspR* in *V. parahaemolyticus.*** The nucleotide sequence from the regulatory region of *opaR* and *cspR* were obtained and used as an input to identify the putative CRP binding sites using Virtual footprint database.

Decreased CPS production in the Δ*crp* mutant could, in part, explain, the significantly reduced biofilm produced by this mutant. Similar decreased biofilm production was observed for the Δ*crp* mutant in *E. coli*, however, in *V. cholerae,* Δ*crp* mutant resulted in increased biofilm production. It was shown that in quorum sensing positive *V. cholerae* strains (*hapR*, the *luxR* homologue is functional), CRP positively regulates *hapR*, which inhibits biofilm formation. HapR regulation of biofilm is indirect acting through repression of *vspR*, an activator of CPS production. However, in *V. cholerae* strains with a non-functional HapR, CRP represses expression of CdgA, a cyclic diguanylyl cyclase that produces c-di-GMP that positively regulates biofilm formation (Fong and Yildiz 2008). We speculate that CRP positively regulates both *cps* and biofilm matrix production genes in *V. parahaemolyticus*.

Previous studies in *E. coli*, *V. cholerae* and *V. vulnificus* have demonstrated that cAMP-CRP is a positive regulator of major flagella biosynthesis genes (Bartlett, Frantz et al. 1988, Liang, Pascual-Montano et al. 2007, Kim, Lee et al. 2013). In our study, deletion of *crp* resulted in motility defects, and we identified CRP binding sites in both polar and lateral flagella regulatory regions suggesting direct regulation by CRP.

The *crp* mutant in *V. cholerae* demonstrated attenuated virulence in an infant mouse model (Skorupski and Taylor 1997). More recently, CRP was demonstrated to be required for the efficient colonization of a zebrafish larva and was shown to be dependent upon the CRP induction of *rtxBDE* genes encoding RTX toxin export machinery (Manneh-Roussel, Haycocks et al. 2018). In our study, in *V. parahaemolyticus* the *crp* mutant demonstrated attenuated *A. franciscana* killing after 48 hours of exposure.

Multiple studies in *E. coli* and *Enterobacteria* in general have demonstrated that deleting *cyaA* impairs the cell from cAMP production thus deactivating CRP, which led to *cyaA* and *crp* mutants demonstrating identical phenotypes. The unexpected finding that the *cyaA* mutant behaved similar to wild type prompted us to search proteins with AC function and led us to the identification of *cyaB* (VP1760). CyaB belongs to CYTH superfamily of proteins, a class IV AC, which contains one domain. The canonical AC CyaA, which is a class I type, contains two domains, a cyclase domain and a regulatory domain. CyaB has been identified only in *A. hydrophila*, and *Y. pestis* and hyperthermophilic archaebacteria, previously (Sismeiro, Trotot et al. 1998). In *A. hydrophila*, CyaB was only active in high temperature and alkaline PH condition and not active under normal laboratory condition. In that study,

it was demonstrated that deleting *cyaB* did not show any cAMP-CRP phenotype defect. However, when expressed from a plasmid in a *cyaA* and *cyaA/cyaB* mutant strains, it was able to restore the phenotypes back to wild type. In our study, deleting *cyaB* demonstrated *crp* mutant specific phenotypes but with less severity suggesting that CyaA could be compensating for the loss of CyaB. Double mutant Δ*cyaB*/Δ*cyaA* demonstrated phenotypes identical to Δ*crp* indicating only after inactivating both CyaA and CyaB, the cells become devoid of cAMP. We speculate that CyaB is the major AC in *V. parahaemolyticus* under the conditions we examined.

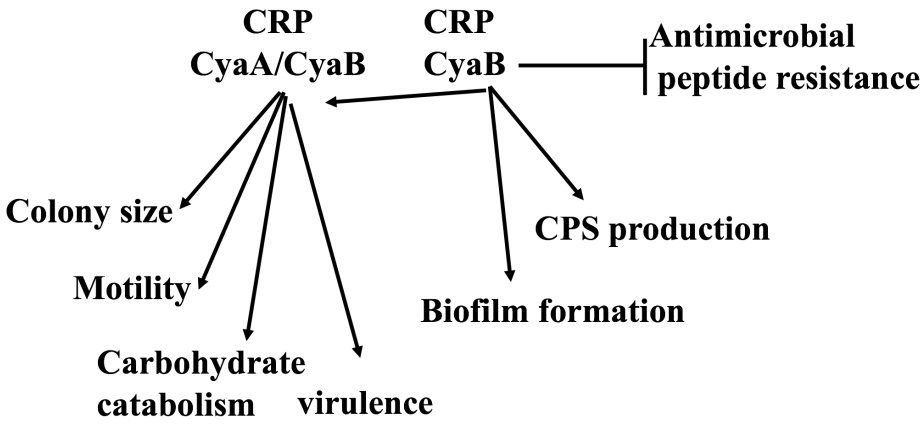


Figure 28 **A predicted model of cAMP-CRP function in *V. parahaemolyticus*.** cAMP in *V. parahaemolyticus* can be synthesized via the enzymatic action of either CyaA or CyaB, which in turn activates CRP and CRP represses *cyaA*. CRP in *V. parahaemolyticus* is an activator of genes involved in carbon catabolism, motility, CPS production, biofilm formation, virulence and negative regulator genes involved in antimicrobial peptide resistance. The requirement for CyaB is specific for CPS production and biofilm formation and increases sensitivity to the antimicrobial peptide polymyxin B.

Based on the proposed model (**Figure 28**), in normal laboratory condition, deleting *cyaA* will not have any impact in the cAMP biosynthesis and CRP function, thus we do not see any phenotypic changes, however, deleting *cyaB* will impair the cell from cAMP resulting in loss of cAMP-CRP function, which in turn releases repression of *cyaA* and restores cAMP production and CRP function. To investigate whether our hypothesis is correct, we conducted bioinformatics analysis to identify the putative CRP binding sites in the regulatory region of *cyaA* and *cyaB.* We identified

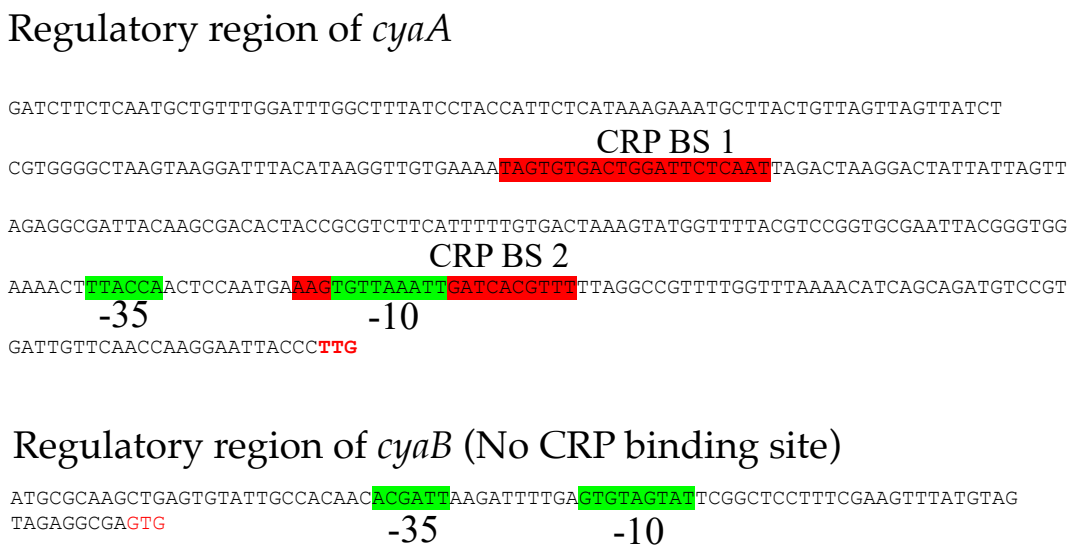


Figure 29 **Regulatory region of *cyaA* and *cyaB* in *V. parahaemolyticus.*** The nucleotide sequence from the regulatory region of *cyaA* and *cyaB* were obtained and used as an input to identify the putative CRP binding sites using Virtual footprint database and RNA polymerase binding site using SoftBerry-BPROM. Highlighted in red are the predicted CRP binding sites and highlighted in green are the predicted -10 and -35 site.

two putative CRP binding sites in the regulatory region of *cyaA*, one of which overlaps the RNAP binding site, characteristic of a repressor (**Figure 29**). No CRP binding sites were identified in the regulatory region of *cyaB* suggesting this gene is regulated very differently than *cyaA*.

Chapter 4

FACTOR FOR INVERSION STIMULATION, FIS, IS REQUIRED FOR OPTIMAL CARBOHYDRATE UTILIZATION AND IN VIVO SURVIVAL IN *VIBRIO PARAHAEMOLYTICUS*

This work will be submitted for publication

Regmi, A. and E.F. Boyd. 2019. Factor for inversion stimulation (FIS) in *Vibrio parahaemolyticus* is required for optimal carbohydrate utilization and *in vivo* survival

Introduction

Vibrio parahaemolyticus is the leading cause of bacterial seafood borne gastroenteritis worldwide and in the United States alone, 45,000 cases of *V. parahaemolyticus* infections are reported each year (Nair, Ramamurthy et al. 2007, Su and Liu 2007, @CDCgov 2018). A pandemic clone of *V. parahaemolyticus* belonging to O3:K6 serogroup emerged in 1995 (Nair, Ramamurthy et al. 2007). The type strain for the pandemic clone RIMD2210633 was sequenced and shown to contain two type three secretion systems (T3SS), one on each chromosome, designated T3SS-1 and T3SS-2 (Makino, Oshima et al. 2003). The T3SS is a contact based secretion system present in a wide range of enteric species that directly injects effector toxins into eukaryotic host cells (Galán and Waksman 2018). T3SS-1 is present in all isolates and is ancestral to the species whereas T3SS-2 is present within a pathogenicity island and is only present in pathogenic strains (Makino, Oshima et al. 2003, Hurley, Quirke et

al. 2006, Boyd, Cohen et al. 2008). T3SS-1 effector proteins play a role in cytotoxicity, whereas T3SS-2 effectors are required for enterotoxicity (Park, Ono et al. 2004, Hiyoshi, Kodama et al. 2010, Pineyro, Zhou et al. 2010, Ritchie, Rui et al. 2012).

In addition to the T3SSs and their effectors, other factors have also been shown to be important for virulence. Previously, it was shown that the regulator ToxR is required for acid and bile salt stresses that the bacterium encounters *in vivo*. The data suggest that in a *ΔtoxR* mutant that the defect in stress tolerance is due to the ToxRS regulation of *ompU*, which encodes a major outer membrane protein and this is required for *in vivo* survival (Whitaker, Parent et al. 2012). Furthermore deletion of the alternative sigma factor *rpoE* enders the mutant more sensitive to cell envelope stresses such as polymyxin B (a cationic antimicrobial peptide), ethanol, and high temperature and was defective in intestinal colonization(Haines-Menges, Whitaker et al. 2014). Whitaker and colleagues demonstrated the importance of bacterial metabolism to host colonization using a streptomycin pretreated mouse model (Whitaker, Richards et al. 2014). Their study showed that the sigma factor RpoN, had superior intestinal colonization compared to the wild type due to a faster doubling time on mouse intestinal mucus and mucus sugars L-arabinose, D-gluconate and D-ribose, indicating that RpoN is an important regulator of metabolism (Whitaker, Richards et al. 2014). A more recent study has also suggested that metabolic regulation is important *in vivo* (Kalburge, Carpenter et al. 2017). In this study, it was demonstrated that bacterial metabolism and transporter genes were down regulated in a *luxO* mutant (LuxO is the quorum sensing (QS) response regulator) and was attenuated for intestinal colonization (Kalburge, Carpenter et al. 2017). These studies demonstrated

that carbon metabolism plays a key role in colonization in this species (Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017).

The factor for inversion stimulation (FIS) belongs to a family of nucleoid associated protein (NAP) and like other nucleoid-associated proteins has two major functions, chromosome organization and gene regulation (Kahramanoglou, Seshasayee et al. 2011). FIS is a highly conserved protein with conserved functions among many enteric bacteria (Beach and Osuna 1998). FIS along with other NAPs is an important positive regulator of ribosome synthesis, which is finely tuned to the nutritional status of the cell (Schneider, Ross et al. 2003, Dennis, Ehrenberg et al. 2004). As a transcription factor, it can act both as a repressor or as an activator based on the position of its binding relative to the RNA polymerase. FIS has been shown to be a negative regulator of gyrase A and gyrase B (encoded by *gyrA* and *gyrB*), required for DNA negative supercoiling (Schneider, Travers et al. 1999). In many enteric pathogens, FIS has also been demonstrated to regulate genes required for motility and virulence, making FIS a versatile global regulator (Goldberg, Johnson et al. 2001, Kelly, Goldberg et al. 2004, Bradley, Beach et al. 2007, Lautier and Nasser 2007, Steen, Harrison et al. 2010, Zhang, Ni et al. 2012, Wang, Liu et al. 2013).

We characterized the role of FIS in the gastrointestinal pathogen, *V. parahaemolyticus* RIMD2210633 for the first time. We constructed a Δ*fis* mutant strain and examined various phenotypes previously shown in other species to required FIS. The Δ*fis* mutant was defective in both swimming and swarming motility and we show this effect was direct by demonstrating binding of purified FIS protein to the regulatory region of flagellum biosynthesis genes. In in vitro growth competition assays we found that Δ*fis* had a defect in several carbon sources. Putative FIS binding

sites were identified within the regulatory region of these carbon sources and using EMSAs showed direct binding. In a pretreated streptomycin adult mouse model of colonization, we show that the Δ*fis* mutant had a fitness defect.

Materials and Methods

Bacterial Strains, Media, and Culture Conditions

All strains and plasmids used in this study are listed in **Table 12**. A streptomycin-resistant clinical isolate *V. parahaemolyticus* RIMD2210633 was used throughout the study. Unless stated otherwise, all *V. parahaemolyticus* strains were grown in Lysogeny Broth (LB) medium (Fischer Scientific, Pittsburgh, PA) containing 3% NaCl (LBS) at 37°C with aeration or M9 medium media (Sigma Aldrich, St. Louis, MO) supplemented with 3% NaCl (M9S). Antibiotics were added to growth media at the following concentrations: Ampicillin (Amp), 100 μg/ml, streptomycin (Sm), 200 μg/ml and chloramphenicol (Cm), 10 μg/ml when required.

Table 12 Bacterial strains and plasmids used in this study

Strain	**Genotype or description**	**Reference or sources**
Vibrio parahaemolyticus		
RIMD2210633	O3:K6 clinical isolate, Str^r	(Makino, Oshima et al. 2003)
Δ*fis*	RIMD2210633, Str, Δ*fis*	This study
WBWlacZ	RIMD2210633, Str, lacZ	(Whitaker, Parent et al. 2012)
Δ*fis*pBBR*fis*	Δ*fis* harboring pBBR*fis*, Cm^r	This study

Escherichia coli		
DH5αλpir		Laboratory collection
B2155λpir	*ΔdapA::emr pir*, for bacterial conjugation	Laboratory collection
BL21(DE3)	Expression strain	Laboratory collection
Plasmids		
pDS132	Suicide plasmid, Cm^r, SacB	
pDSΔ*fis*	pDS132 harboring truncated *fis*, Cm^r	This study
pBBR1MCS	Expression vector, lacZ promoter, Cm^r	
pBBR*fis*	pBBR1MCS harboring full-length *sglt*, Cm^r	This study

Construction of Δ*fis* Mutant in *V. parahaemolyticus* RIMD2210633.

Splice by overlapping extension (SOE) PCR and an allelic exchange method (Ho, Hunt et al. 1989) was used to construct an in-frame, non-polar deletion mutant of *fis* (VP2885) in *V. parahaemolyticus* RIMD2210633. Briefly, primers were designed to VP2885 using *V. parahaemolyticus* RIMD2210633 genomic DNA as a template. All primers used in this study are listed in **Table 13**. SOE PCR was conducted to obtain an 18 bp-truncated version of VP2885 (297-bp). The Δ*fis* PCR fragments were cloned into the suicide vector pDS132 and named pDSΔ*fis*. pDS Δ*fis* was then transformed into *E. coli* strain β2155 λ*pir*, and was conjugated into *V. parahaemolyticus* RIMD2210633. Conjugation was conducted by cross streaking both strains onto LB plate containing 0.3 mM DAP. Bacterial growth lawn from this plate was scraped, suspended in LB, serially diluted and plated on LB plate containing streptomycin and chloramphenicol. Next day, the colonies were verified for single crossover via PCR. The colonies with single crossover were grown over night in LBS

with no antibiotic added and were plated onto LBS plate containing 10% Sucrose to select for double crossover deletion mutant. The gene deletion was confirmed by PCR and sequencing.

Table 13 Primers used in this study

Primer name	**Sequences (5'-3')**	**Tm(ºC)**	**Product size (bp)**
SOE PCR			
VP2885 SOEA	TCTAGAGCATGGCAGAACAAAAGCGT	58	459
VP2885 SOEB	GAACATATTCGGTCTAGCTCT	51	
VP2885 SOEC	AGAGCTAGACCGAATATGTTCGGCATGAACTAAGACCGGA	66	634
VP2885 SOED	GAGCTCCTCGAGCTTCTCATCGCCTT	63	
VP2885 SOEFF	GATGTCCGCCAACCCAAAAC	55	2,982
VP2885 SOEFR	GCTGGTGATGCAAAAGGTGG	54	
Complementation			
FisComp F	CGAGGATCCAGAGAAGAGCTAGACCGAAT	62	305
FisComp R	CACTCTAGATCCGGTCTTAGTTCATGCCGTA	62	
EMSA			
VPA1424F	ACACCGGACTGGACTTTTGT	57	130
VPA1424R	CCATTTGAAAAGGGATTGAGTG	52	
VPA1424F	GAGCAAAGCCACATTTCACT	54	156
VPA1424R	TGATGAAAAGACTGTAGTGAAGAAAAA	53	
VPA1674F	CCATCCACTGGTAGAGGTGAG	57	138
VPA1674R	TTGAGAGCTGCTTCATAGCC	55	
VP0063F	CATAGGTTCACCTTTAAGAA	47	138
VP0063R	AATTTGCGCGTGATCTCAAT	53	
VPA0038F	GCGTAAAGCAAACGTAAAACTG	53	154

Primer name	Sequences (5'-3')	Tm(°C)	Product size (bp)
SOE PCR			
VP2885 SOEA	TCTAGAGCATGGCAGAACAAAAGC GT	58	459
VPA0038R	TTTGTGCCCCTAGTTATTGGA	54	
VP2235F	CAATGGATTGTGCTGTTATTTAGC	53	161
VP2235R	TCACGCGAATACGCACTTTA	55	
VPA1550F	AGCCTGACGTTGGTAAGGTG	57	244
VPA1550R	TATGTGTTCCGCCTTCCTCT	56	
VP1932F	CTTTTCTCACAATCGTTACAGGA	53	229
VP1932R	GCCATTATCCCTCTATTTTCTGA	52	
FIS purification primers			
FisFWDpMAL	TGCCCATGGGAATGTTCGAACAAAA TCTGACTTC	63	308
FisREVpMAL	CTCGGATCCGAAATCCGGTCTTAGT TCATGCCGTA	66	

Growth Pattern Analysis

Bacterial cultures were grown overnight in 5 ml LBS or 5 ml of M9S supplemented with 10mM glucose. One ml of overnight culture was pelleted, suspended and washed with PBS. A 1: 40 dilution of the culture was then used to conduct growth curves in LBS or M9S supplemented with 10 mM individual carbon sources. Growth analysis was performed in a 96 well plate with shaking for 1000 sec in between cycles and optical density (OD595) was measured each hour for 24 hours. The data was plotted using GraphPad Prism 7.00, GraphPad software, La jolla California USA.

In Vitro Growth Competition Assays

In vitro growth competition assays were performed by diluting the inoculum 1:50 into LBS broth, and separately in M9S supplemented with 10mm of individual carbon sources, D-glucose, L-arabinose, D-mannose, D-ribose, D-gluconate, D-glucosamine, and *N*-acetyl-D-glucosamine (NAG). In all cases, the culture was incubated at 37°C for 24 hours and then serial diluted and plated on LBS plus Streptomycin and X-gal. The competitive Index (CI) was determined using the following equation: CI=ratio out (Δ*fis* /WBWlacZ) / ratio in (Δ*fis*/WBWlacZ). A CI of < 1 indicate WBWlacZ outcompetes the Δ*fis*, a CI of >1 indicates that Δ*fis* outcompetes WBWlacZ. The ratio of Δ*fis* to WBWlacZ in the inoculum mixture is termed as "Ratio in" and the ratio of Δ*fis* to WBWlacZ colonies recovered from the mouse intestine is referred as "Ratio out."

In Vivo Colonization Competition Assays

All mice experiments were approved by the University of Delaware Institutional Animal Care and Use Committee. A β-galactosidase knock in *V. parahaemolyticus* RIMD2210633 strain, WBWlacZ, which was previously shown to grow similar to wild type *in vitro* and *in vivo* was used for all the competition assays (Whitaker, Parent et al. 2012, Haines-Menges, Whitaker et al. 2014, Whitaker, Richards et al. 2014). Inocula for competition assays were prepared using overnight grown cultures of WBWlacZ and Δ*fis* diluted into fresh LBS media and grown for four hours. These exponential phase cultures were then pelleted by centrifugation at 4,000 × g, washed and suspended in PBS. One ml of WBWlacZ and one ml of Δ*fis* were prepared corresponding to 1×10^{10} CFU of each strain based on the previously determined OD and CFU ratio. A 500 µl aliquot of Δ*fis* was combined with 500 µl of

the WBWlacZ, yielding a total bacterial concentration of 1×10^{10} CFU/ml. The inoculum was serially diluted and plated on LBS agar plate supplemented with 200 µg/ml streptomycin and 8 µg /ml of 5-bromo-4-chloro-3-indolyl-B-D-galactoside (X-gal) to determine the exact ratio in the inocula. Male C57BL/6 mice, aged 6 to 10 weeks were housed under specific-pathogen-free conditions in standard cages in groups (5 per group) and provided standard mouse feed and water *ad libitum*. Pre-treatment of mice with streptomycin were performed as previously described (Whitaker, Parent et al. 2012, Boyd, Carpenter et al. 2015). Mice were inoculated with 100 µl of the bacterial suspension and 24 hours post-infection, mice were sacrificed and the entire gastrointestinal tract was harvested. Samples were placed in 8 ml of sterile PBS, mechanically homogenized and serially diluted in PBS. Diluted samples were plated for CFU's on LBS, streptomycin and X-gal for a blue (WBWLacZ) versus white (Δ*fis*) screen of colonies after incubation at 37°C overnight. The competitive Index (CI) was determined as described above.

Swimming, Swarming, CPS production, and Biofilm Assays

Swimming assays were conducted in swimming media containing LB 2% NaCl and 0.3 % agar, whereas swarming assays were conducted in heart infusion (HI) media with 2% NaCl and 1.5 % agar. To study swimming behavior, a single colony of the test bacterium was picked and stabbed on the center of the plate, and the plates were incubated at 37°C for 24 hrs. For the swarming assays, plates were incubated at 30°C for 48 hrs. To observe capsule polysaccharide formation (CPS), HI media containing 1.5 % agar, 2.5mM $CaCl_2$, and 0.25% Congo red dye was used and incubated at 30°C. Biofilm assays were conducted using crystal violet staining. Cultures were grown overnight in LBS and the overnight grown culture (1:40 dilution)

was then inoculated in LBS in a 96 well plate, static at 37°C. After 24 hours, the wells are washed with PBS, stained with crystal violet for 30 min and then accessed for biofilm formation. The biofilms are then dissolved in DMSO and OD595 was measured.

Polymyxin B Sensitivity Assays

Overnight grown bacterial cultures were diluted (1:50) into fresh LBS and grown for 2 hours until they reached early exponential phase. Then cultures were spun down and suspended in 5 ml LBS. Polymyxin B sulphate (Sigma-Aldrich) was added to the cultures (final concentration of 40 μg/ml) and were incubated at 37°C for 1 hr. For zero min time point, an aliquot was obtained before adding the polymyxin B, and then aliquots were taken at 30 min and 60 min time points. The aliquots were serially diluted and plated to determine the CFU at each specific time points. Percent survival was calculated by dividing the CFUs at 30 min and 60 min with that of 0 min and multiplied by 100. For disk assay, similar growth condition as used as described for the survival assay were used, and after 2 hour of growth a sterile cotton swap was used to obtain sample and using the same swap cultures were spread throughout the plate for 4 times. Polymyxin B disks (100 μg of polymyxin in each disk) were placed on the plates with the culture, and the plates were incubated for 24 hours at 37°C before the zone of inhibitions were measured.

Bioinformatics Analysis to Identify Putative FIS Binding Sites

The regulatory region of gene clusters of interest of *V. parahaemolyticus* RIMD210633 were obtained using NCBI nucleotide database. Virtual footprint, an online database, was used to identify the FIS binding sites using the *E. coli* FIS

consensus binding sequence (Münch, Hiller et al. 2003). The 229 bp, 416 bp, 371 bp, 385 bp, 153 bp and 385 bp DNA region upstream of *flhA* (VP2235), *lafB* (VPA1550), *araB* (VPA1674), *nagB* (VPA0038), *gntK* (VP064), and *fru pts IIA/B/C* (VPA1424) were used as an input for FIS binding DNA region and default settings were used to obtain the output.

FIS Protein Purification

FIS was purified using a method previously described with modifications as necessary (Carpenter, Rozovsky et al. 2016, Kalburge, Carpenter et al. 2017). Briefly, *fis* was cloned into the pMAL-c5x expression vector in which maltose binding protein (MBP) and an N-terminal 6X His tag is fused to *fis* separated by a tobacco etch virus (TEV) protease cleavase site. PCR was conducted using Accura HiFidelity Polymerase, the primer pairs FisFWDpMAL and FisREVpMAL (**Table 13**) and *V. parahaemolyticus* RIMD2210633 genomic DNA to amplify *fis* (VP2885). The *fis* PCR product was gel cut purified, digested with NcoL and BamHI to create sticky ends, and purified using NucleoSpin Gel and PCR cleanup kit (Macherey-Nagel). Similarly, the pMAL-c5x vector was digested with NcoL and BamHI and gel cut purified using NucleoSpin Gel and PCR cleanup kit (Macherey-Nagel). The double digested pMAL-c5x and the *fis* DNA fragment were ligated and transformed into DH5α using CaCl2 transformation procedure. The plasmid DNA (pMAL-c5x*fis*) was extracted and verified by PCR and sequencing. The vector pMAL-c5x*fis* was then transformed into *E. coli* BL21 (DE3). *E. coli* BL21 harboring pMAL-c5x*fis* was grown overnight in 100 ml LB supplemented with 100 µg/ml ampicillin and 0.2% glucose at 37°C. The overnight grown culture was inoculated into 1 L of fresh LB supplemented with 100 µg/ml ampicillin and 0.2% glucose and was grown at 37°C

until the OD reached 0.4, at which point, the culture was moved to 18°C untill OD reached 0.6. Expression of the plasmid was induced by adding 0.5 mM IPTG at OD of 0.6 and the cells were allowed to grow for 18 hours at 18°C. Cells were pelleted at 5000 rpm and resuspended in 15 ml of 50 mM sodium phosphate, 200 mM NaCl, 0.5 mM Benzamidine and 1mM phenyalmethanesulfonyl fluoride (PMSF) at pH 7.5. Bacterial cells were lysed using high pressure homogenizer, spun down at 15000 RPM for 60 min, and the supernatant was collected. The supernatant was passed through a 20 ml amylose column equilibrated with 5 column volumn (CV) of column buffer (50 mM sodium phosphate, 200mM NaCl, 2mM EDTA, 5mM BME, pH 7.5) and the column was washed with 12 CVs of column buffer. FIS fused with MBP-TEV and 6X His was then eluted with three CV column buffer supplemented with 20 mM maltose. Using TEV protease (1:10, TEV: protein in 50mM sodium phosphate, 200mM NaCl, 10mM imidazole, 5mM BME, pH 7.5) the fused protein was cleaved at TEV cleavage site. The cleaved fused protein was run through IMAC column equilibrated with 50mM sodium phosphate, 2000-500 mM NaCl, 10mM imidazole, 5mM BEM, pH 7.5 and the flow through was collected, Which contained the FIS protein. Mass spectrometry was conducted to identify the protein and SDS-PAGE was conducted to determine its purity.

Electrophoretic Mobility Shift Assay (EMSA)

A 138 bp fragment of *araB* (VPA1674, arabinose catabolism), a 154 bp fragment of *nagB* (VPA0038, glucosamine catabolism), 138 bp fragment of *gntK* (VP0064, gluconate catabolism), a 156 bp DNA fragment of *pts* (VPA1424, mannose transport), a 161 bp fragment of *flhA* (VP2235), and a 244 bp fragment of *lafB* (VPA1550) regulatory regions were used as probes in electrophoretic mobility shift

assays (EMSA). The fragments were PCR amplified using Accura Hifidelity Polymerase in 50 µl reaction mixture using respective primers sets listed in **Table 13** using *V. parahaemolyticus* RIMD2210633 genomic DNA as template. The PCR amplified product was run in 1.5% agarose gel and was purified after cutting the band using a NucleoSpin Gel and PCR cleanup kit (Macherey-Nagel). The concentration of the purified DNA probe was determined using NanoDrop spectrophotometer. Various concentrations of purified FIS 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, pH7.4) for 20 min at room temperature. A native acrylamide 6% gel was prepared and preran for 2 hours (200 V at 4°C), which was then used to load 10 µl of the target DNA probe which was incubated for 20 min in binding buffer. The gel was run at 200V for 2 hours in 1X Tris-acetate-EDTA (TAE) buffer at 4°C, which was then stained in an ethidium bromide bath (0.5 µg/ml) for 20 min, washed with water and imaged.

Results

FIS is Conserved Among Bacterial Species

In *V. parahaemolyticus* FIS is encoded by an ORF VP2885 and is in an operon with VP2884, which encodes for tRNA-dihydrouridine synthase B (DusB) (**Figure 30**).

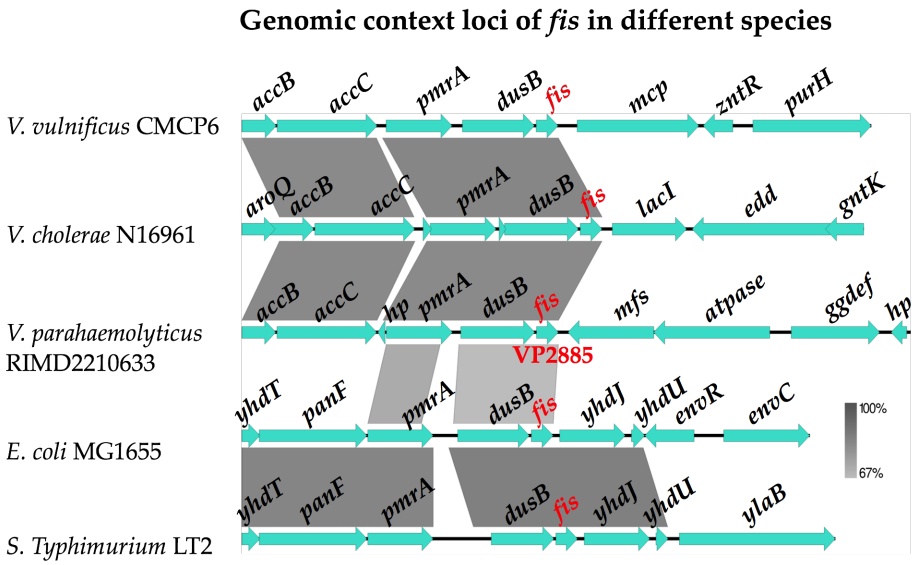


Figure 30 **Genomic context comparison of *fis* loci.** The genomic region upstream and downstream of *fis* (VP2885) from *V. parahaemolyticus* RIMD2210633 was compared with that of the region upstream and downstream of *fis* from *V. cholerae*, *V. vulnificus*, *E. coli* and *S. typhimurium* using blastn, and figures were constructed using Easyfig. Grey shade indicates nucleotide homology and cyan arrow indicate ORFs.

Comparative analysis of FIS and its genome context among *Vibrio cholerae*, *Vibrio vulnificus*, *Escherichia coli* and *Salmonella enterica* was performed. Analysis identified that FIS from *V. parahaemolyticus* shared 100%, 100%, 82.65% and 82.65% amino acid identity with that of *V. cholerae*, *V. vulnificus, E. coli* and *S. enterica* respectively, indicating that this is a highly conserved protein (**Figure 30**). Examination of the genomic regions immediately upstream and downstream of *fis* also showed highly conserved genome context *(***Figure 30***)*. Based on the conserved homology of FIS among the five species, we speculated that the function of FIS in *V. parahaemolyticus* could be similar to other species.

FIS is Required for Motility in *V. parahaemolyticus*

To characterize the role of FIS in *V. parahaemolyticus*, a mutant strain, Δ*fis*, was constructed by deleting 179-bps from VP2885, the gene coding for FIS. The mutant strain demonstrated small colony size compared to wild type (WT) on LBS plates, however, the mutant strain grew identical to WT in LBS broth (**Figure 31B and 31A**).

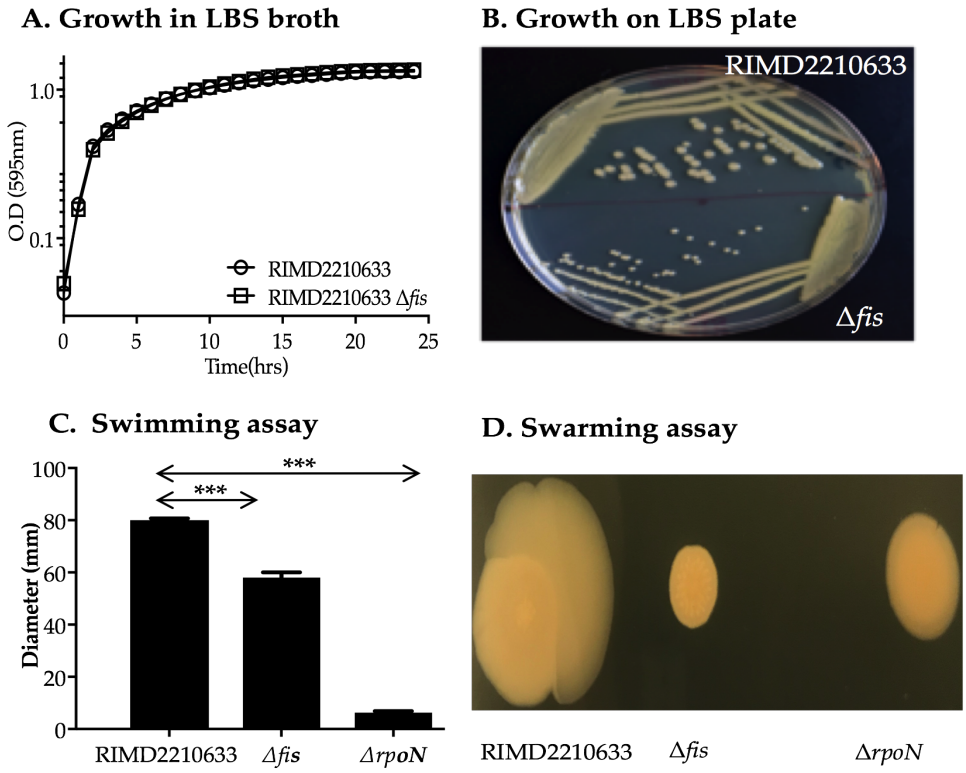


Figure 31 **Growth and motility assay (A)** Growth curve of RIMD2210633 and Δ*fis* in LBS for 24 hrs. **(B)** Colony morphology of RIMD2210633 and Δ*fis* in LBS agar. **(C)** Diameters of the bacterial growth in swimming media were measured to quantify the swimming behavior. **(D)** Single colonies of RIMD 2210633, Δ*fis, and* Δ*rpoN* were incubated on the

surface of the swarming plate and Growth assay was conducted in triplicate using two biological replicates. Swimming assay was repeated twice using three biological replicates each time. Swarming assays were conducted using at least two biological replicates. Error bars indicate standard error of mean (SEM). Unpaired Student *t* test was conducted to obtain the *P* values. The significant difference in swimming is denoted by asterisks. *, P<0.05, **, P<0.01, ***, P<0.001.

Swimming and swarming assays were conducted to investigate the role of FIS in the motility of *V. parahaemolyticus*. The Δ*fis* mutant demonstrated a significant swimming and swarming defect compared to WT suggesting that FIS is a positive regulator of motility (**Figure 31C, 31D and Figure 32A**). To investigate this further, the regulatory region of the operons containing the sigma factor *fliAP* (VP2232 *fliA* for polar flagellum synthesis) and *fliAL (lafS VPA1555* for lateral flagella synthesis) required were analyzed for the presence of putative FIS binding sites using the consensus sequence from *E. coli* (**Figure 33C**). Previously, deletions of *fliAF* and *fliAL* were demonstrated to impair swimming and swarming in *V. parahaemolyticus* RIMD2210633 (Whitaker, Richards et al. 2014). Our bioinformatics analysis identified putative FIS binding sites in the regulatory region of the gene cluster containing *fliAL* **(Figure 33A)** and the regulatory region of the cluster containing *lafS* (**Figure 33B**). EMSA was conducted using a purified FIS protein from RIMD2210633. We used a fragment of DNA with no putative FIS binding site as our non-binding control (**Figure 33D**) and the regulatory region of *gyrA* as a FIS binding positive control (**Figure 33E**) for the EMSA. FIS has been demonstrated to bind to the regulatory region of *gyrA* in *E. coli* and *S. enterica* previously (Schneider, Travers et al. 1999, Keane and Dorman 2003). EMSA results demonstrated that FIS binds to the DNA probes from the regulatory region of *fliAP* and *fliAL* and the *gyrA* promoter

*(***Figure 33A, 33B and 33E***)*. The data suggest that FIS is a positive regulator of motility genes and the regulation is direct.

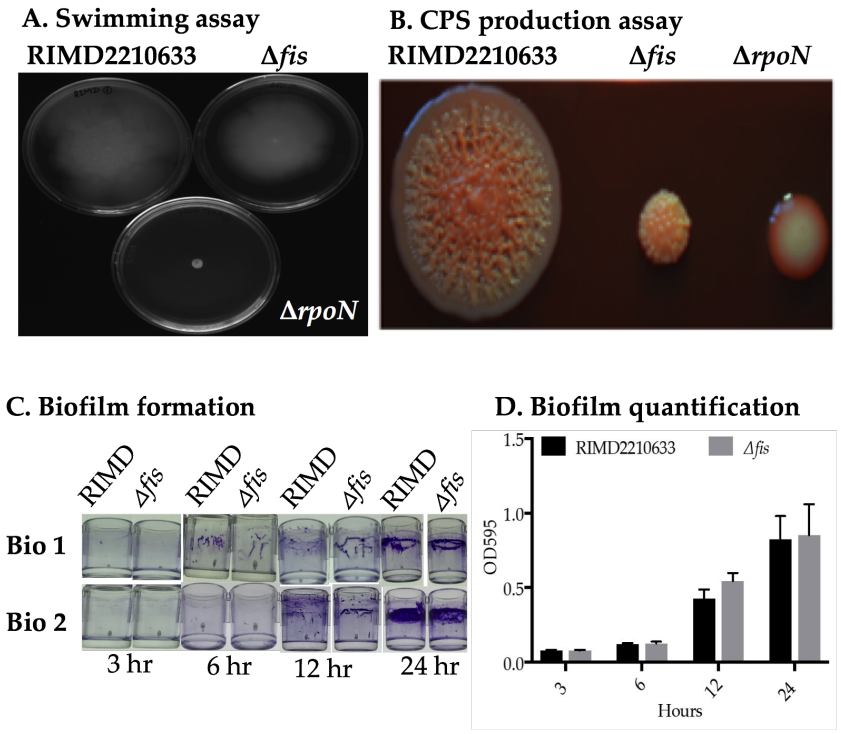


Figure 32 **Motility, CPS production and Biofilm assay. (A)** Strains RIMD2210633, Δ*fis, and* Δ*rpoN* on swimming media. **(B)** congo red plate incubated for 48 hrs at 30 degrees to investigate the CPS production. **(C)** Crystal violet staining was used to conduct the biofilm assay using RIMD2210633 and Δ*fis.* Pictures were taken before dissolving the biofilm with DMSO. The biofilm production was quantified by dissolving the stained biofilm in DMSO and measuring the OD595 after diluting 1:10 in PBS. The biofilm assays were performed in triplicates using two biological replicates. Error bar indicates SEM. Unpaired Student *t* test was conducted to obtain the *P* values.

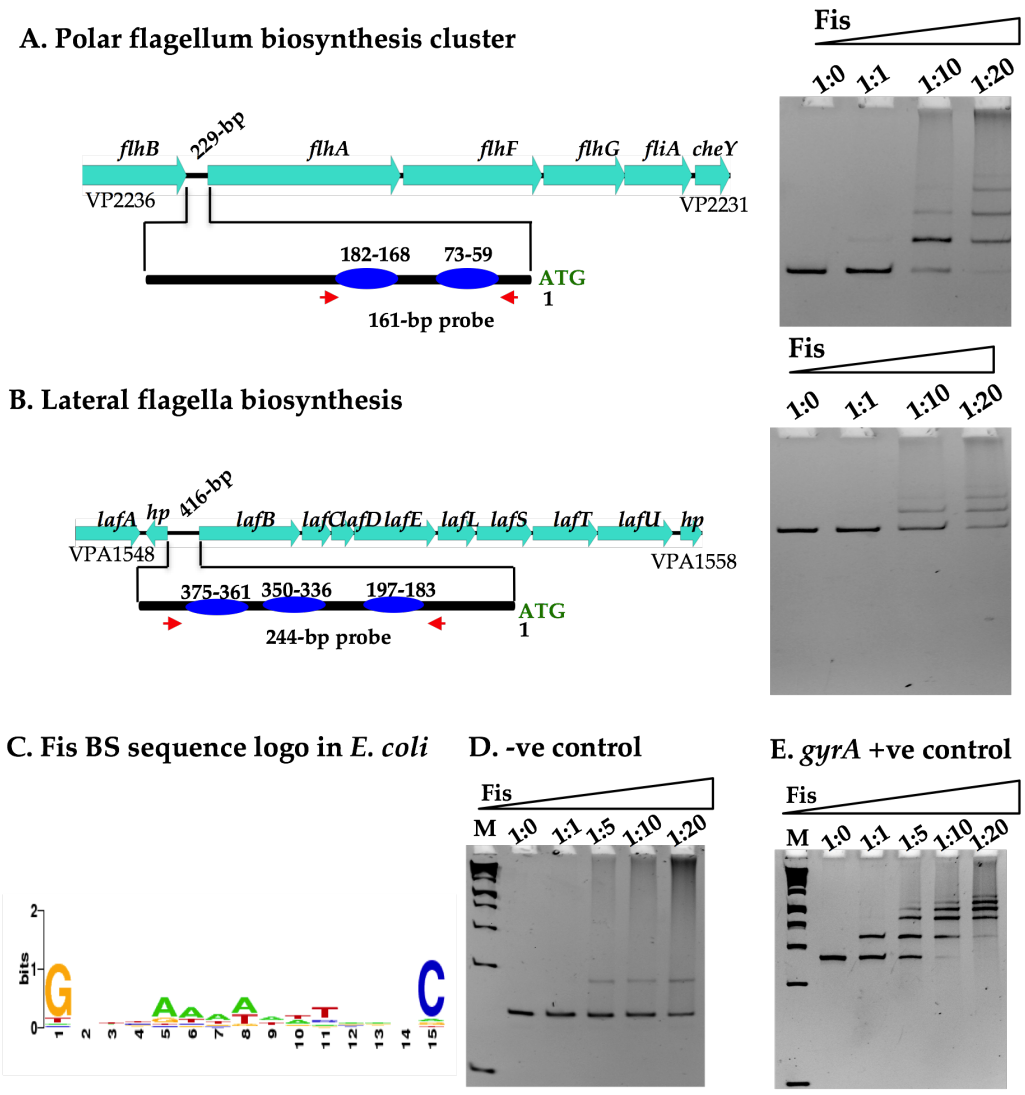


Figure 33 **Identification of FIS binding sites and FIS binding assay. (A)**The regulatory region of polar flagellum biosynthesis (upstream of VP2235) was investigated for the presence FIS binding sites and EMSA was conducted to demonstrate FIS binding in the predicted region. **(B)** The regulatory region of lateral flagella biosynthesis (upstream of VPA1500) cluster of *V. parahaemolyticus* RIMD2210633 were investigated for the

presence FIS binding sites and EMSA was conducted to demonstrate FIS binding in the predicted region. **C)** Consensus FIS binding sequence logo from *E. coli* **(F)** Consensus FIS binding sequence from the predicted binding sites identified in *V. parahaemolyticus* was constructed using WebLogo. **(D)** EMSA demonstrating the inability of FIS to bind in a region with no predicted FIS binding sites. **(E)** EMSA demonstrating FIS binding in the regulatory region of *gyrA*. The putative FIS binding sites (blue ovals) were identified relative to the translational start of the above-mentioned genes. The ratio above each EMSA gel indicates DNA to protein concentration.

FIS Plays a Role in the Carbon Catabolism Regulation in *V. parahaemolyticus*

To investigate whether FIS plays a role in growth of *V. parahaemolyticus*, we conducted growth competition assays between WT and Δ*fis* in LBS or M9S supplemented with 10 mM D-glucose, mouse intestinal mucin, L-arabinose, D-glucosamine, D-gluconate or D-mannose (**Figure 34).** In these competition assays, Δ*fis* was significantly outcompeted by the WT in mouse intestinal mucus (CI 0.61), L-arabinose (CI 0.4), D-glucosamine (CI 0.68) and D-gluconate (CI 0.78), whereas the Δ*fis* significantly outcompeted the WT in D-mannose (CI 1.64). These data suggested that FIS plays a positive role in the growth of this pathogen in intestinal mucus, and L-arabinose, D-glucosamine, D-gluconate and a negative role in growth in D-mannose. Bioinformatics analysis was conducted using the virtual footprint program (Münch, Hiller et al. 2003) and putative FIS binding sites were identified in the regulatory regions of VPA1674 (L-arabinose catabolism), VPA0038 (D-glucosamine catabolism), VP0064 (D-gluconate catabolism), and VPA1424 (D-mannose catabolism) (**Figure 35A-35D**).

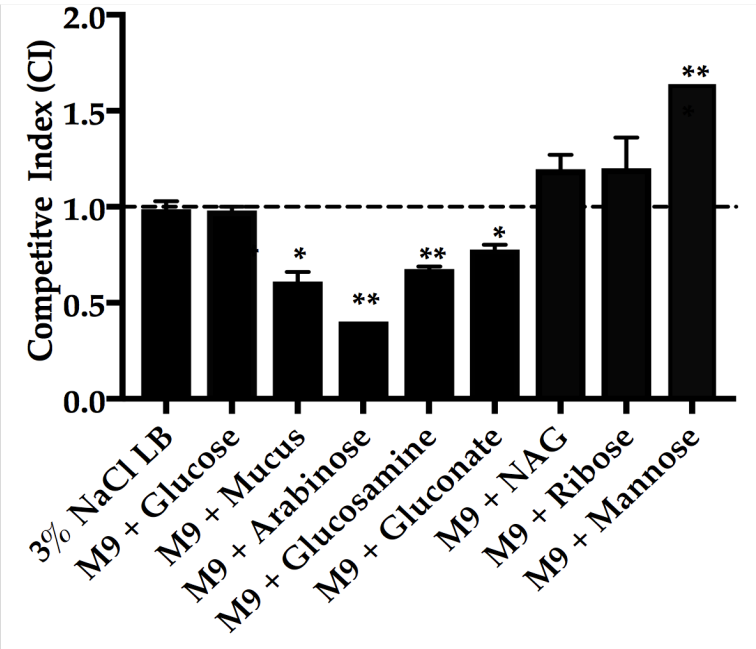


Figure 34 ***In vitro* competition assay.** WBWlacZ and Δ*fis* were grown in co-culture (1:1 ratio) for 24 hours using 3% NaCl LB and M9 supplemented with 100 μg/ml mouse intestinal mucin, 10 mM of individual carbon sources, D-glucose, L-arabinose, D-glucosamine, D-gluconate, *N*-acetyl-D-glucosamine (NAG), D-ribose, and D-mannose. Competitive indices were calculated for LBS and for each of the carbon sources mentioned above to access competition between the two strains. The assays were conducted in triplicate using two biological replicates. Error bar indicates SEM. Unpaired Student *t* test was conducted to obtain the *P* values. The significant difference in CI is denoted by asterisks. *, P<0.05, **, P<0.01, ***, P<0.001.

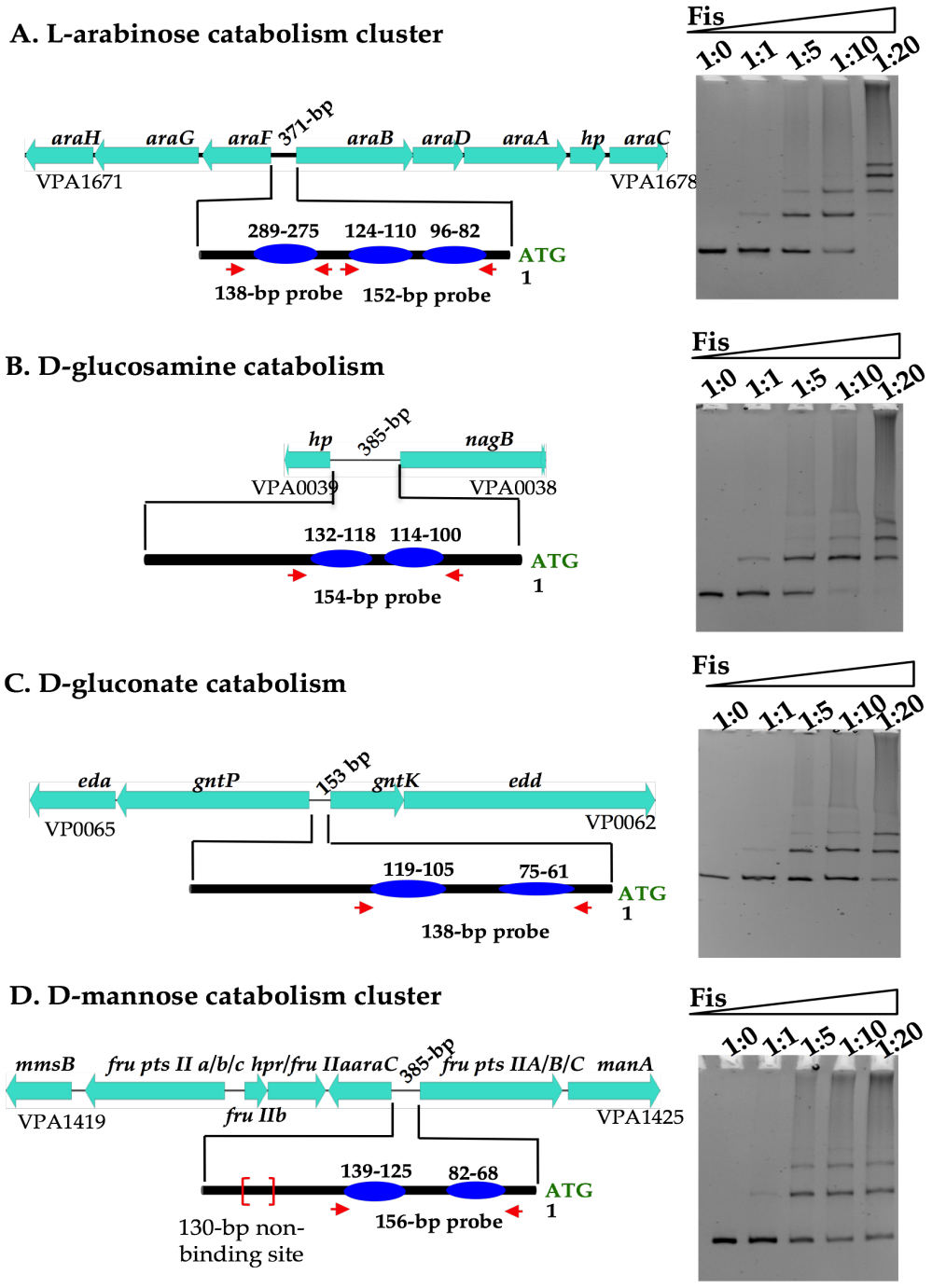


Figure 35 **Identification of FIS binding sites and FIS binding assay. (A)** Putative FIS binding sites in the regulatory region of L-arabinose catabolism

(upstream of VPA1674) cluster and EMSA to demonstrate binding in the predicted region. Similarly, putative FIS binding sites and EMSA in the regulatory region of **(B)** D-glucosamine catabolism (upstream of VPA0038) cluster **(C)** D-gluconate catabolism (upstream of VP0063) cluster **(D)** D-mannose catabolism (upstream of VPA1424) cluster. The putative FIS binding sites (blue ovals) are numbered relative to the first nucleotide of the start codon. The red arrows indicate the DNA region amplified for EMSA probe. The red brackets in the mannose regulatory region indicates region devoid of FIS binding site as predicted by Virtual footprint.

To validate that FIS binds to the predicted binding sites in the above-mentioned regulatory regions, DNA probes containing the FIS binding sites were constructed (**Table 13**) and EMSA was conducted using FIS protein that was purified to homogeneity. EMSA demonstrated that FIS binds to the 152 bp DNA probe of the L-arabinose regulatory region, and the binding of FIS was concentration dependent (**Figure 35A**). Similarly, probes for the promoter regions of D-glucosamine (154-bp), D-gluconate (138-bp) and D-mannose (130-bp) were also bound by FIS in a concentration dependent manner (**Figure 35B-35D**). The EMSA results demonstrated that FIS is a direct regulator of catabolism genes in *V. parahaemolyticus*.

FIS Mutant is Sensitive to the Antimicrobial Peptide Polymyxin B

To investigate whether FIS has any role in antimicrobial peptide resistance, the Δ*fis* strain was examined for polymyxin B sensitivity using survival and disk diffusion assays. For the disk diffusion assay, a disk containing 100 μg total polymyxin B was used and the Δ*fis* mutant had significantly larger zone of inhibition compared to WT (**Fig. 36A)** indicating the mutant is more sensitive towards polymyxin B compared to the WT. To verify this further, we conducted survival assay using 200 μg of total polymyxin B in 5ml of LBS media. CFUs were counted at 0, 30 and 60 min post exposure. Significantly lower CFUs for the Δ*fis* were recovered at 30 and 60 minute

compared to WT (**Figure 36B**). We also used Δ*rpoE* as a control, which was previously

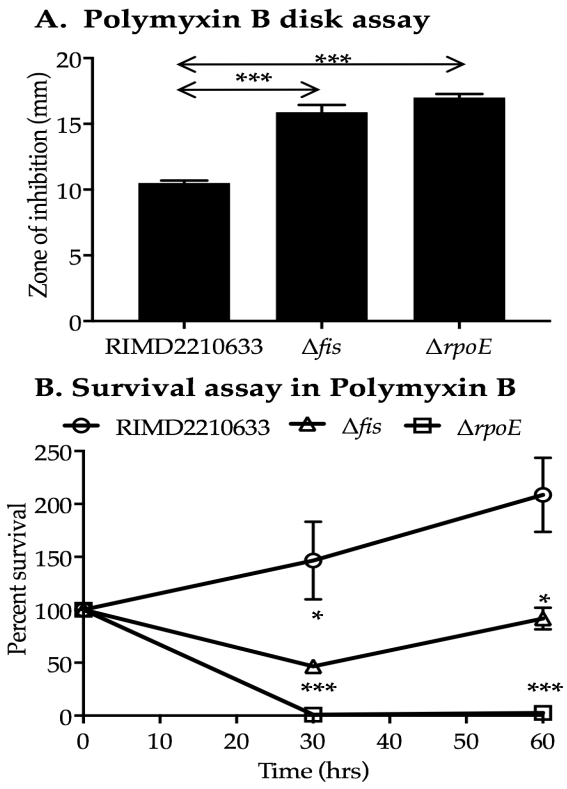


Figure 36 **Polymyxin B sensitivity assay. (A)** Disk containing 100 µg of total polymyxin B was used to identify the zone of inhibition for RIMD2210633, Δ*fis* and Δ*rpoE* and was quantified by measuring the diameter of the zone of inhibition. **(B)** Survival assays were conducted using RIMD 2210633, Δ *fis, and* Δ *rpoE* after treating the bacterial cultures with polymyxin B (final concentration of 40 µg/ml) and calculating percent survival at 30 min and 60 min. The disk assay was performed in duplicate using two biological replicates. The survival assay was performed in triplicate using two biological replicates. Unpaired t-test was conducted to determine the p-value. *, P<0.05, **, P<0.01, ***, P<0.001.

shown to be significant sensitivity to polymyxin B (Haines-Menges, Whitaker et al. 2014). The data demonstrated that FIS is required for polymyxin B resistance in *V. parahaemolyticus*, although the mechanism remains unknown.

One mechanism by which bacteria are more resistant towards antimicrobial peptide is by producing a robust capsule polysaccharide (CPS) (Llobet, Tomas et al. 2008, Olaitan, Morand et al. 2014). To investigate if the polymyxin B sensitivity was due to less CPS production in the mutant, we conducted a CPS production assay and identified that Δ*fis* strain had opaque colony morphology with rugose structure, which suggested that the mutant strain is proficient in producing CPS (**Figure 32B**). Similarly, our data demonstrated no significant difference in the biofilm production between Δ*fis* and the wild type at various time-points tested (**Figure 32C and 32D**).

FIS Contributes to Fitness in the *V. parahaemolyticus* Colonization of Adult Mouse

In vitro competition assays determined that the Δ*fis* strain was significantly outcompeted in mouse intestinal mucin and mucus sugars like arabinose, glucosamine and gluconate. To elucidate if this defect translates *in vivo*, we conducted *in vivo* competition assay using streptomycin treated adult mouse model of colonization. For the competition assays, we used a β-galactosidase knock-in strain of RIMD2210633, WBWlacZ, which was previously demonstrated to behave like wild type *in vitro* and *in vivo* (Whitaker, Parent et al. 2012, Boyd, Carpenter et al. 2015). Adult C57BL/6 mice were pretreated with streptomycin (20 mg/mouse) 24 hours prior to orogastric co-inoculation with a mixture of 5X 10^8 CFU of Δ*fis* and 5X 10^8 CFU of WBWlacZ. Twenty-four hours post infection mice were scarified and the entire gastrointestinal

tract of each mouse was removed, homogenized, and plated for CFUs using a blue/white screen to assay for the presence of both strains.

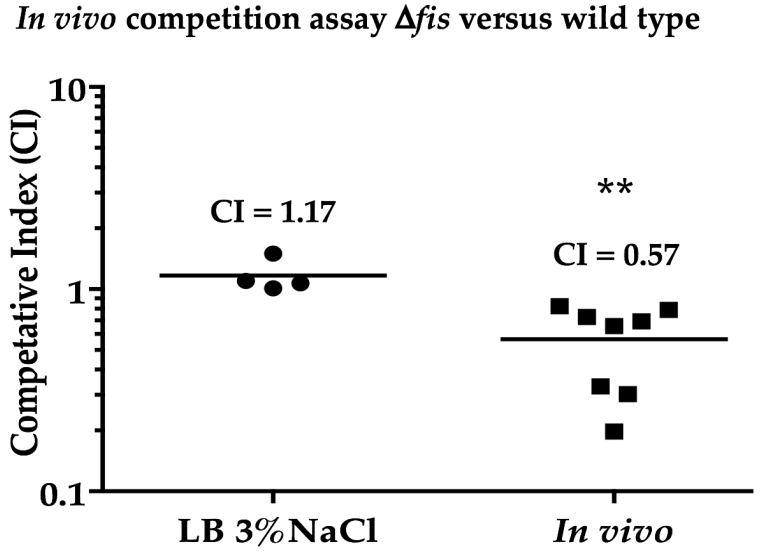


Figure 37 ***In vivo* competition assay.** Streptomycin treated adult mice were orogastrically infected with a 1:1 mixed culture of the WBWlacZ strain and Δ*fis* strain. Gastrointestinal tracts were harvested, homogenized and plated on X-gal plates. CFU was determined using a blue/white colony screen. Using same or similar bacterial mix culture, *in vitro* competition assays between WBWlacZ and Δ*fis* strain were conducted in LBS in test tubes. Competitive index (CI) was calculated as CI = ratio out (Δ*fis*/WBWlacZ)/ratio in (Δ *fis*/WBWlacZ). One sample t-test was conducted to determine the P-value. *, $P < 0.05$; **, $P < 0.01$.

The *in vivo* competition assay between Δ*fis* and WBWlacZ determined the competitive index to be 0.57 (**Figure 37**), indicating that Δ*fis* was significantly outcompeted by the WT. We also conducted *in vitro* competition assays using LBS

and determined that the competition index to be 1.17 (**Figure 37**). These data suggested that FIS is essential for *in vivo* fitness of *V. parahaemolyticus*.

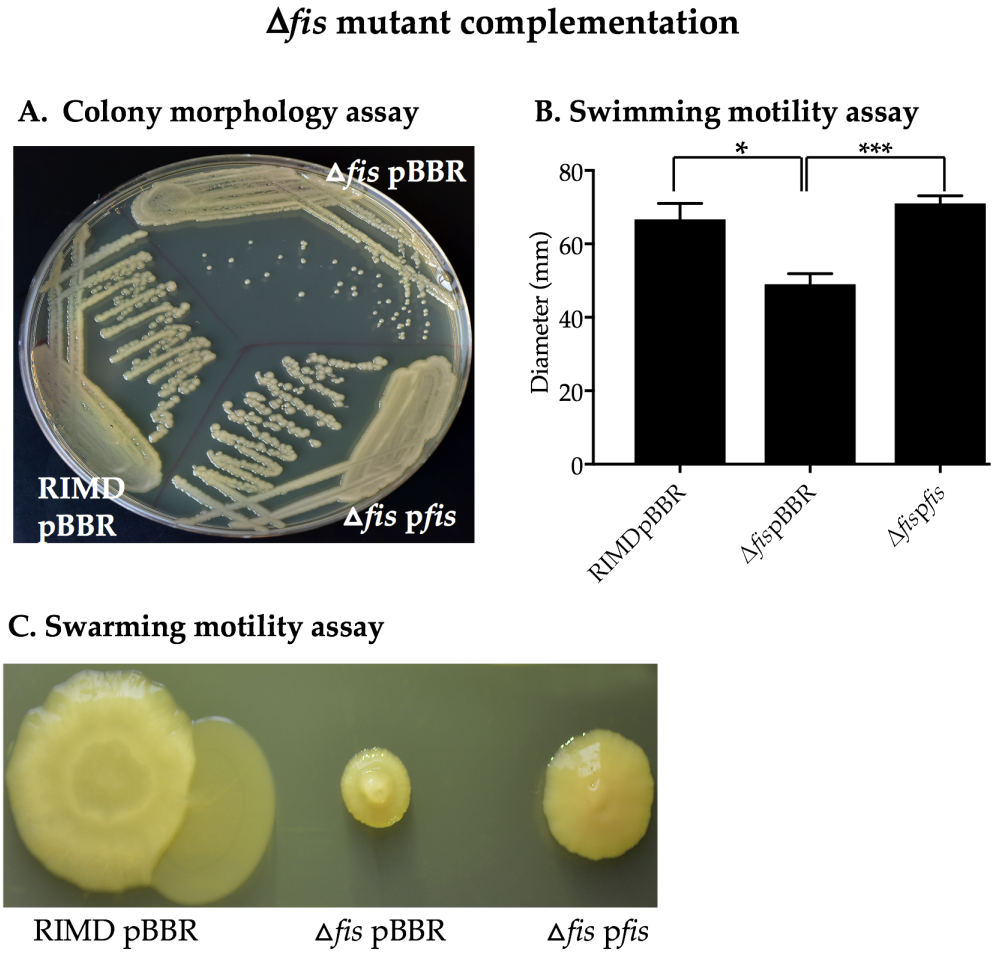


Figure 38 **Complementation analysis of Δ*fis* strain.** *Vibrio parahaemolyticus* RIMD2210633 Δ*fis* strain was complemented with a functional copy of FIS and **(A)** the colony morphology **(B)** swimming motility and **(C)** swarming motility was assessed. Error bar indicates SEM. Unpaired Student *t* test was conducted to obtain the *P* values. The significant difference in swimming is denoted by asterisks. *, P<0.05, **, P<0.01, ***, P<0.001.

Discussion

FIS is a global regulator ubiquitous among Proteobacteria, and in this study, we investigated its role in the biology of the gastrointestinal pathogen *V. parahaemolyticus*. Our work identified FIS as an important regulator of both swimming and swarming motility and is required for efficient use of several carbohydrates. We found that the *fis* mutant was sensitive to the antimicrobial peptide polymyxin B but not as sensitive as the *rpoE* mutant. In *in vivo* competition assays using a streptomycin pretreated adult mouse colonization model, the *fis* mutant was out competed by wild type indicating a fitness defect. Expressing FIS from an expression vector in Δ*fis* strain was able to restore colony size, and motility of the mutant strain (**Figure 38A-38C**) demonstrating that the phenotypes observed in this study are specific to the lack of FIS protein in Δ*fis* strain.

Our analysis identified that FIS is a positive regulator of motility, both swimming and swarming, and the regulation is direct. Similar results were shown in a study in *Salmonella,* in which it was demonstrated that when *fis* was deleted, the mutant strain was defective in swimming. The study identified that FIS directly regulates motility in *Salmonella* (Kelly, Goldberg et al. 2004). The present study identified that, under the condition tested, FIS has no role in biofilm production in *V. parahaemolyticus*.

Our results also highlighted the role of FIS in resistance against polymyxin B, an antimicrobial peptide, however, ruled out the possibility that the sensitivity of the Δ*fis* strain observed was due to defect in CPS production. Other mechanisms by which bacteria become resistant to polymyxin B are by modifying the LPS, upregulating efflux pumps and outer membrane proteins, which needs to be examined further in *V. parahaemolyticus.*

Various studies have demonstrated the role of FIS in bacterial metabolism (Gonzalez-Gil, Bringmann et al. 1996, Kelly, Goldberg et al. 2004), and in this study we identified that FIS is required for growth in mouse intestinal mucus and mucus sugars L-arabinose, D-glucosamine and D-gluconate. Our study also demonstrated that FIS binds to the regulatory region of the L-arabinose, D-glucosamine and D-gluconate catabolism cluster indicating that FIS is a direct positive regulator. However, Δ*fis* strain demonstrated enhanced growth in D-mannose indicating that FIS act as a direct transcription repressor for D-mannose catabolism. Overall, the data indicate that FIS is a major regulator of carbon catabolism in *V. parahaemolyticus*, however, its function as an activator or repressor depends on the specific carbon source.

The *in vivo* competition assay in streptomycin treated mouse model using Δ*fis* strain and WBWLacZ demonstrated that deleting *fis* resulted in attenuated colonization by the Δ*fis* strain demonstrating that FIS is required for the *in vivo* colonization. We speculate that the defect in colonization is due to its inability to efficiently utilize mucus sugars L-arabinose, D-glucosamine and D-gluconate and significant lack of resistance towards polymyxin B. These phenotypes were previously implicated in attenuated colonization of *V. parahaemolyticus* strains in streptomycin treated adult mouse model of colonization (Haines-Menges, Whitaker et al. 2014, Whitaker, Richards et al. 2014, Kalburge, Carpenter et al. 2017).

Chapter 5

PERSPECTIVES AND FUTURE DIRECTIONS

The findings in this dissertation demonstrate the importance of diverse and efficient carbon source utilization in the survival and fitness of *V. parahaemolyticus*. First, we demonstrated that efficiency in carbon source utilization can be gained by having efficient substrate transporters and diversity can be obtained by acquisition or loss of carbon source utilization clusters. A large-scale comparative genomic analysis identified multiple metabolic islands carrying complete sets of different carbon utilization genes. These metabolic islands were demonstrated to be acquired via HGT and suggested to help *V. parahaemolyticus* in various niche adaptation. Then, the dissertation work explored the importance of two major global regulators, CRP and FIS in *V. parahaemolyticus* that in other species are regulators of carbon metabolism genes. This work, for the first time, demonstrated that cAMP receptor protein, CRP, is essential for different physiological functions in *V. parahaemolyticus*, especially it is required for carbon utilization. CRP was also identified to act as a virulence factor in *V. parahaemolyticus.* We also discovered that *V. parahaemolyticus* harbors two functional adenylate cyclases, CyaA and CyaB. Although CyaB was the major adenylate cyclase in *V. parahaemolyticus*, both adenylate cyclases were involved in activating CRP. Another global regulator, FIS, was shown to be essential for optimal carbohydrate catabolism and *in vivo* colonization in *V. parahaemolyticus.*

Acquisition and Loss of Metabolic Islands

In the second chapter of this dissertation, we identified that *V. parahaemolyticus* strains have gained and lost multiple carbohydrate catabolism systems carried on metabolism islands of different size. Bioinformatics and phylogenetic analysis identified that the metabolic islands were horizontally acquired. The catabolic clusters were flanked by transposases and direct repeats indicating acquisition and loss as a modular unit. Our analysis identified that the newly identified catabolic clusters of citrate, L-rhamnose, oligogalacturonides, L-arabinogalactan, and L-arabinose utilization contain complete set of genes for the transport, catabolism and regulation of the respective substrate. It would be of interest in the future to identify if the genotype translates into phenotype and determine whether CRP and/or FIS is required for their regulation. Expression analysis of the representative genes can be conducted to identify if the transport and catabolic genes are expressed when exposed to the respective carbon sources. Also, growth analysis can be conducted to identify if the *V. parahaemolyticus* strain carrying the catabolic cluster can utilize the carbon sources.

A novel metabolism island ranging in size from 64-kb to 165-kb was identified, and the variation in the size of the island was due to acquisition and/or loss of modular units from the island. Further bioinformatics analysis is required to identify if there exist a pattern for the excision or integration of the modules from the island. To accomplish this, the updated *V. parahaemolyticus* genomes from the NCBI genome database should be screened for the presence of the metabolism island variants, and the variant islands should be compared to come up with a pattern. Also, bioinformatics analysis in this study identified that these clusters may act as modular units and can excise and integrate in the genome. Further work is needed to demonstrate this

phenomenon of the movement of the modular unit and to identify what triggers such movement.

Role of cAMP-CRP in *V. parahaemolyticus* metabolism

cAMP receptor protein (CRP) is an essential global regulator in many bacterial species regulating operons of multiple functional categories including carbon and nitrogen metabolism. To investigate the role of cAMP-CRP in *V. parahaemolyticus,* first we constructed Δ*crp* strain and characterized this mutant strain. We identified that cAMP-CRP is required for carbohydrate catabolism in *V. parahaemolyticus* by demonstrating the requirement of CRP for L-arabinose, D-gluconate, *N*-acetyl-D-glucosamine, D-glucosamine, and D-ribose catabolism. We can expand this study by using phenotypic microarray plates and conduct high throughput growth screening to identify a list of carbon sources that has a strict requirement of CRP for their catabolism. Via bioinformatics analysis, we identified putative CRP binding sites in the regulatory region of the operon for L-arabinose, D-gluconate, *N*-acetyl-D-glucosamine, D-glucosamine, and D-ribose catabolism indicating that CRP is a direct regulator. We can expand this analysis to identify putative binding sites in the regulatory region of the carbon metabolism operons identified from the high throughput growth screening. Similarly, cAMP-CRP was essential for motility, CPS production and biofilm production, however, was not essential for polymyxin B resistance. We also identified putative CRP binding sites in the regulatory region of polar and lateral flagella and CPS biosynthesis operons indicating CRP directly regulates these operons. These data indicated that cAMP-CRP is a global regulator in *V. parahaemolyticus*, which directly regulates the targeted operons. Expression analysis of the catabolic, motility and CPS biosynthesis genes should be conducted to

further confirm that cAMP-CRP is a positive regulator of these operons. Also, the cAMP-CRP targets identified in this dissertation are essential targets, however, do not represent a complete set of targets under the control of CRP in *V. parahaemolyticus*. A global approach using ChipSeq should be conducted to identify the genome wide targets of cAMP-CRP in *V. parahaemolyticus*. Also, using the data from Chip-Seq, a consensus cAMP-CRP binding sequence for *V. parahaemolyticus* should be determined. To validate the Chip-Seq results CRP protein should be purified and Electrophoretic mobility shift assay (EMSA) should be conducted to demonstrate binding of CRP in the targeted regions.

The Δ*crp* and Δ*cyaB*/Δ*cyaA* strains demonstrated attenuated virulence in the brine shrimp. It would be of interest in the future to identify what factors were involved in the killing of the brine shrimp. Transcriptomic analysis utilizing RNAseq will help answer this question. RNA extracted from WT strain from the infected brine shrimp, RNA from WT strain from Seawater and RNA from Δ*crp* strain isolated from the brine shrimp should be compared to identify the factor for this virulence potential of *V. parahaemolyticus*.

One interesting pattern that was observed was the similar targets of CRP and quorum sensing regulators. For example, like cAMP-CRP, quorum sensing regulates CPS production, biofilm production and motility in *V. parahaemolyticus*. It would be of interest to investigate if there is a link between CRP and quorum sensing in *V. parahaemolyticus,* or CRP and quorum sensing act independently on the targeted operons.

Finally, we identified that *V. parahaemolyticus* contained two functional adenylate cyclases, CyaA and CyaB. Under the current laboratory condition, CyaB

was the major adenylate cyclase, and CyaA and CyaB compensated for each other and both are involved in activating CRP. In closely related species like *V. cholerae* and *V. vulnificus*, CyaA is the sole adenylate cyclase. It will be interesting to identify, under what conditions is *cyaA* active and acts as the major adenylate cyclase.

Our data also identified that there is a specific requirement of CyaB in the CPS biosynthesis and biofilm production in *V. parahaemolyticus*. We speculate that this requirement is independent of AC function of CyaB and could be due to a second function of the CyaB. In *Yersinia pestis* CyaB can act as adenylate cyclase and guanylate cyclase (Gallagher, Kim et al. 2011). It would be of interest in the future to identify what other function does CyaB in *V. parahaemolyticus* have.

FIS is required for optimal catabolism in *V. parahaemolyticus*

In chapter 3, we demonstrated that a nucleotide associated protein, FIS is involved in multiple physiological behaviors of *V. parahaemolyticus* including efficient utilization of mouse intestinal mucus, L-arabinose, D-glucosamine and D-gluconate. Using bioinformatics analysis, we identified potential FIS binding sites in the regulatory regions of L-arabinose, D-glucosamine and D-gluconate catabolism gene cluster and using EMSA, we demonstrated binding of FIS indicating that FIS is a positive regulator of L-arabinose, D-glucosamine and D-gluconate catabolism and the regulation is direct. Expression analysis of L-arabinose, D-glucosamine and D-gluconate catabolism genes is further required to fully support the notion that FIS is a positive regulator of these carbohydrate sources. Also, we tested limited carbon sources and focused on the carbohydrates found in the intestinal mucus, however a global RNAseq analysis should be conducted to identify other nutrient sources under

the regulation of FIS. Growth analysis using high throughput screens and *in vitro* competition assays can follow to support the findings of the RNAseq analysis.

REFERENCES

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): 3807-3816.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner and H. Mori (2006). "Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection." Mol Syst Biol **2**: 2006.0008.

Bartlett, D. H., B. B. Frantz and P. Matsumura (1988). "Flagellar transcriptional activators FlbB and FlaI: gene sequences and 5' consensus sequences of operons under FlbB and FlaI control." J Bacteriol **170**(4): 1575-1581.

Barzu, O. and A. Danchin (1994). "Adenylyl cyclases: a heterogeneous class of ATP-utilizing enzymes." Prog Nucleic Acid Res Mol Biol **49**: 241-283.

Beach, M. B. and R. Osuna (1998). "Identification and characterization of the *fis* operon in enteric bacteria." J Bacteriol **180**(22): 5932-5946.

Bettendorff, L. and P. Wins (2013). "Thiamine triphosphatase and the CYTH superfamily of proteins." Febs j **280**(24): 6443-6455.

Biswas, A., R. H. Staals, S. E. Morales, P. C. Fineran and C. M. Brown (2016). CRISPRDetect: A flexible algorithm to define CRISPR arrays. BMC Genomics. **17**:356.

Botsford, J. L. and J. G. Harman (1992). "Cyclic AMP in prokaryotes." Microbiol Rev **56**(1): 100-122.

Boyd, E. F. (2012). "Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions." Adv Virus Res **82**: 91-118.

Boyd, E. F., S. Almagro-Moreno and M. A. Parent (2009). "Genomic islands are dynamic, ancient integrative elements in bacterial evolution." Trends Microbiol **17**(2): 47-53.

Boyd, E. F., M. R. Carpenter and N. Chowdhury (2012). Mobile effector proteins on phage genomes. Bacteriophage. **2:** 139-148.

Boyd, E. F., M. R. Carpenter, N. Chowdhury, A. L. Cohen, B. L. Haines-Menges, S. S. Kalburge, J. J. Kingston, J. B. Lubin, S. Y. Ongagna-Yhombi and W. B. Whitaker (2015). "Post-Genomic Analysis of Members of the Family *Vibrionaceae*." Microbiol Spectr **3**(5).

Boyd, E. F., A. L. Cohen, L. M. Naughton, D. W. Ussery, T. T. Binnewies, O. C. Stine and M. A. Parent (2008). "Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*." BMC Microbiol **8**: 110.

Bradley, M. D., M. B. Beach, A. P. de Koning, T. S. Pratt and R. Osuna (2007). "Effects of Fis on *Escherichia coli* gene expression during different growth stages." Microbiology **153**(Pt 9): 2922-2940.

Carpenter, M. R., S. Rozovsky and E. F. Boyd (2016). "Pathogenicity Island Cross Talk Mediated by Recombination Directionality Factors Facilitates Excision from the Chromosome." J Bacteriol. **198**(5):766-76.

Chen, Y., O. C. Stine, J. H. Badger, A. I. Gil, G. B. Nair, M. Nishibuchi and D. E. Fouts (2011). "Comparative genomic analysis of *Vibrio parahaemolyticus*: serotype conversion and virulence." BMC Genomics **12**: 294.

ChenBaoli, LiangWeili, WuRui, LiangPu and KanBiao (2013). "Phenotype microarray screening of carbon sources used by *Vibrio cholerae* identifies genes regulated by the cAMP receptor protein." http://dx.doi.org/10.1139/cjm-2013-0084.

Conway, T. (1992). "The Entner-Doudoroff pathway: history, physiology and molecular biology." FEMS Microbiol Rev **9**(1): 1-27.

Conway, T. and P. S. Cohen (2015). "Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut." Microbiol Spectr **3**(3).

Cotta, M. A., T. R. Whitehead and M. B. Wheeler (1998). "Identification of a novel adenylate cyclase in the ruminal anaerobe, *Prevotella ruminicola* D31d." FEMS Microbiol Lett **164**(2): 257-260.

Daniels, N. A., L. MacKinnon, R. Bishop, S. Altekruse, B. Ray, R. M. Hammond, S. Thompson, S. Wilson, N. H. Bean, P. M. Griffin and L. Slutsker (2000). "*Vibrio parahaemolyticus* infections in the United States, 1973-1998." J Infect Dis **181**(5): 1661-1666.

Dennis, P. P., M. Ehrenberg and H. Bremer (2004). "Control of rRNA synthesis in *Escherichia coli*: a systems biology approach." Microbiol Mol Biol Rev **68**(4): 639-668.

Dobrindt, U., U. Hentschel, J. B. Kaper and J. Hacker (2002). "Genome plasticity in pathogenic and nonpathogenic enterobacteria." Curr Top Microbiol Immunol **264**(1): 157-175.

Dobrindt, U., B. Hochhut, U. Hentschel and J. Hacker (2004). "Genomic islands in pathogenic and environmental microorganisms." Nature Reviews Microbiology **2**(5): 414.

Enos-Berlage, J. L., Z. T. Guvener, C. E. Keenan and L. L. McCarter (2005). "Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*." Mol Microbiol **55**(4): 1160-1182.

Espejo, R. T., K. Garcia and N. Plaza (2017). "Insight Into the Origin and Evolution of the *Vibrio parahaemolyticus* Pandemic Strain." Front Microbiol **8**: 1397.

Felsenstein, J. (1985). "Phylogenies and the Comparative Method." The American Naturalist **125**(1): 1-15.

Ferreira, R. B., D. M. Chodur, L. C. Antunes, M. J. Trimble and L. L. McCarter (2012). "Output targets and transcriptional regulation by a cyclic dimeric GMP-responsive circuit in the *Vibrio parahaemolyticus* Scr network." J Bacteriol **194**(5): 914-924.

Fong, J. C. N. and F. H. Yildiz (2008). Interplay between Cyclic AMP-Cyclic AMP Receptor Protein and Cyclic di-GMP Signaling in *Vibrio cholerae* Biofilm Formation. J Bacteriol. **190:** 6646-6659.

Franzin, F. M. and M. P. Sircili (2015). "Locus of Enterocyte Effacement: A Pathogenicity Island Involved in the Virulence of Enteropathogenic and Enterohemorragic *Escherichia coli* Subjected to a Complex Network of Gene Regulation." Biomed Res Int **2015**:534738.

Freter, R. (1988). Mechanisms of bacterial colonization of the mucosal surfaces of the gut. Virulence mechanisms of bacterial pathogens. Washington DC, American Society of Microbiology**:** 45-60.

Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005). "Mobile genetic elements: the agents of open source evolution." Nat Rev Microbiol **3**(9): 722-732.

Fujino, T., Y. Okuno, D. Nakada, A. Aoyama, K. Fukai, T. Mukai and T. Ueho (1953). "On the bacteriological examination of Shirasu-food poisoning." Med Jour Osaka Univ **4**(2/3): 299-304.

Gallagher, D. T., S. K. Kim, H. Robinson and P. T. Reddy (2011). "Active-site structure of class IV adenylyl cyclase and transphyletic mechanism." J Mol Biol **405**(3): 787-803.

Gallagher, D. T., N. N. Smith, S. K. Kim, A. Heroux, H. Robinson and P. T. Reddy (2006). "Structure of the class IV adenylyl cyclase reveals a novel fold." J Mol Biol **362**(1): 114-122.

Galán, J. E. and G. Waksman (2018). "Protein Injection Machines in Bacteria." Cell **172**(6): 1306-1318.

Gavilan, R. G., M. L. Zamudio and J. Martinez-Urtaza (2013). "Molecular epidemiology and genetic variation of pathogenic *Vibrio parahaemolyticus* in Peru." PLoS Negl Trop Dis **7**(5): e2210.

Giraud, M. F. and J. H. Naismith (2000). "The rhamnose pathway." Curr Opin Struct Biol **10**(6): 687-696.

Glaser, P., D. Ladant, O. Sezer, F. Pichot, A. Ullmann and A. Danchin (1988). "The calmodulin-sensitive adenylate cyclase of Bordetella pertussis: cloning and expression in *Escherichia coli*." Mol Microbiol **2**(1): 19-30.

Goldberg, M. D., M. Johnson, J. C. Hinton and P. H. Williams (2001). "Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli*." Mol Microbiol **41**(3): 549-559.

Gonzalez-Escalona, N., R. G. Gavilan, M. Toro, M. L. Zamudio and J. Martinez-Urtaza (2016). "Outbreak of *Vibrio parahaemolyticus* Sequence Type 120, Peru, 2009." Emerging Infectious Diseases **22**(7): 1235-1237.

Gonzalez-Gil, G., P. Bringmann and R. Kahmann (1996). "FIS is a regulator of metabolism in *Escherichia coli*." Mol Microbiol **22**(1): 21-29.

Gorke, B. and J. Stulke (2008). "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." Nat Rev Microbiol **6**(8): 613-624.

Gosset, G., Z. Zhang, S. Nayyar, W. A. Cuevas and M. H. Saier (2004). Transcriptome Analysis of Crp-Dependent Catabolite Control of Gene Expression in *Escherichia coli*. J Bacteriol. **186:** 3516-3524.

Green, J., M. R. Stapleton, L. J. Smith, P. J. Artymiuk, C. Kahramanoglou, D. M. Hunt and R. S. Buxton (2014). "Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches." Curr Opin Microbiol **18**: 1-7.

Grenier, F., D. Matteau, V. Baby and S. Rodrigue (2014). Complete Genome Sequence of *Escherichia coli* BW25113. Genome Announc. **2**.

Grissa, I., G. Vergnaud and C. Pourcel (2007). "CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats." Nucleic Acids Res **35**(Web Server issue): W52-57.

Grissa, I., G. Vergnaud and C. Pourcel (2008). "CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats." Nucleic Acids Res **36**(Web Server issue): W145-148.

Guvener, Z. T. and L. L. McCarter (2003). "Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*." J Bacteriol **185**(18): 5431-5441.

Ha, C., S. K. Kim, M. N. Lee and J. H. Lee (2014). "Quorum sensing-dependent metalloprotease VvpE is important in the virulence of *Vibrio vulnificus* to invertebrates." Microb Pathog **71-72**: 8-14.

Hacker, J., G. Blum-Oehler, I. Muhldorfer and H. Tschape (1997). "Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution." Mol Microbiol **23**(6): 1089-1097.

Hacker, J. and E. Carniel (2001). "Ecological fitness, genomic islands and bacterial pathogenicity." EMBO Rep. **2**(5):376-81.

Hacker, J. and J. B. Kaper (2000). "Pathogenicity islands and the evolution of microbes." Annu Rev Microbiol **54**: 641-679.

Haines-Menges, B., W. B. Whitaker and E. F. Boyd (2014). "Alternative sigma factor RpoE is important for *Vibrio parahaemolyticus* cell envelope stress response and intestinal colonization." Infect Immun **82**(9): 3667-3677.

Harman, J. G. and W. J. Dobrogosz (1983). "Mechanism of CRP-mediated cya suppression in *Escherichia coli*." J Bacteriol **153**(1): 191-199.

Hehemann, J. H., L. V. Truong, F. Unfried, N. Welsch, J. Kabisch, S. E. Heiden, S. Junker, D. Becher, A. Thurmer, R. Daniel, R. Amann and T. Schweder (2017). "Aquatic adaptation of a laterally acquired pectin degradation pathway in marine gammaproteobacteria." Environmental Microbiology **19**(6): 2320-2333.

Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter and C. M. Fraser (2000). "DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*." Nature **406**(6795): 477-483.

Hengge-Aronis, R. (2002). "Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase." Microbiol Mol Biol Rev **66**(3): 373-395.

Herrington, D. A., R. H. Hall, G. Losonsky, J. J. Mekalanos, R. K. Taylor and M. M. Levine (1988). "Toxin, toxin-coregulated pili, and the toxR regulon are essential for *Vibrio cholerae* pathogenesis in humans." J Exp Med **168**(4): 1487-1492.

Hiyoshi, H., T. Kodama, T. Iida and T. Honda (2010). "Contribution of *Vibrio parahaemolyticus* virulence factors to cytotoxicity, enterotoxicity, and lethality in mice." Infect Immun **78**(4): 1772-1780.

Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen and L. R. Pease (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene **77**(1): 51-59.

Hochhut, B., K. Jahreis, J. W. Lengeler and K. Schmid (1997). "CTnscr94, a conjugative transposon found in enterobacteria." J Bacteriol **179**(7): 2097-2102.

Hondo, S., I. Goto, I. Minematsu, N. Ikeda, N. Asano, M. Ishibashi, Y. Kinoshita, N. Nishibuchi, T. Honda and T. Miwatani (1987). "Gastroenteritis due to Kanagawa negative *Vibrio parahaemolyticus*." Lancet **1**(8528): 331-332.

Hurley, C. C., A. Quirke, F. J. Reen and E. F. Boyd (2006). "Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates." BMC Genomics **7**: 104.

Iyer, L. M. and L. Aravind (2002). "The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates." BMC Genomics **3**(1): 33.

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-Sgm **148**: 3681-3693.

Joseph, S. W., R. R. Colwell and J. B. Kaper (1982). "*Vibrio parahaemolyticus* and related halophilic *Vibrios*." Crit Rev Microbiol **10**(1): 77-124.

Joshi, J., J. Srisala, V. H. Truong, I. T. Chen, B. Nuangsaeng, O. Suthienkul, C. F. Lo, T. W. Flegel, K. Sritunyalucksana and S. Thitamadee (2014). "Variation in *Vibrio parahaemolyticus* isolates from a single Thai shrimp farm experiencing an outbreak of acute hepatopancreatic necrosis disease (AHPND)." Aquaculture **428**: 297-302.

Kahramanoglou, C., A. S. N. Seshasayee, A. I. Prieto, D. Ibberson, S. Schmidt, J. Zimmermann, V. Benes, G. M. Fraser and N. M. Luscombe (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. Nucleic Acids Res. **39:** 2073-2091.

Kalburge, S. S., M. R. Carpenter, S. Rozovsky and E. F. Boyd (2017). "Quorum Sensing Regulators Are Required for Metabolic Fitness in *Vibrio parahaemolyticus*." Infection and Immunity **85**(3).

Kaneko, T. and R. R. Colwell (1973). "Ecology of *Vibrio parahaemolyticus* in Chesapeake Bay." J Bacteriol **113**(1): 24-32.

Kaper, J. B., J. P. Nataro and H. L. Mobley (2004). "Pathogenic *Escherichia coli*." Nat Rev Microbiol **2**(2): 123-140.

Karaolis, D. K., J. A. Johnson, C. C. Bailey, E. C. Boedeker, J. B. Kaper and P. R. Reeves (1998). "A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains." Proc Natl Acad Sci U S A **95**(6): 3134-3139.

Keane, O. M. and C. J. Dorman (2003). "The gyr genes of *Salmonella enterica* serovar Typhimurium are repressed by the factor for inversion stimulation, Fis." Mol Genet Genomics **270**(1): 56-65.

Kelly, A., M. D. Goldberg, R. K. Carroll, V. Danino, J. C. D. Hinton and C. J. Dorman (2004). "A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar Typhimurium. Microbiology **150**(pt7):2037-53.

Kim, B. S., S. Y. Jang, Y. J. Bang, J. Hwang, Y. Koo, K. K. Jang, D. Lim, M. H. Kim and S. H. Choi (2018). "QStatin, a Selective Inhibitor of Quorum Sensing in *Vibrio* Species." MBio **9**(1).

Kim, S. P., C. M. Kim and S. H. Shin (2012). "Cyclic AMP and cyclic AMP-receptor protein modulate the autoinducer-2-mediated quorum sensing system in *Vibrio vulnificus*." Curr Microbiol **65**(6): 701-710.

Kim, Y. R., S. E. Lee, B. Kim, H. Choy and J. H. Rhee (2013). "A dual regulatory role of cyclic adenosine monophosphate receptor protein in various virulence traits of *Vibrio vulnificus*." Microbiol Immunol **57**(4): 273-280.

Kimura, M. (1980). "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences." J Mol Evol **16**(2): 111-120.

Kornberg, H. L. and C. Riordan (1976). "Uptake of galactose into *Escherichia coli* by facilitated diffusion." J Gen Microbiol **94**(1): 75-89.

Kovach, M. E., R. W. Phillips, P. H. Elzer, R. M. Roop, 2nd and K. M. Peterson (1994). "pBBR1MCS: a broad-host-range cloning vector." Biotechniques **16**(5): 800-802.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura (2018). "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." Mol Biol Evol **35**(6): 1547-1549.

Ladant, D. and A. Ullmann (1999). "*Bordatella pertussis* adenylate cyclase: a toxin with multiple talents." Trends Microbiol **7**(4): 172-176.

Lange, S. J., O. S. Alkhnbashi, D. Rose, S. Will and R. Backofen (2013). "CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems." Nucleic Acids Res **41**(17): 8034-8044.

Lautier, T. and W. Nasser (2007). "The DNA nucleoid-associated protein Fis co-ordinates the expression of the main virulence genes in the phytopathogenic bacterium *Erwinia chrysanthemi*." Mol Microbiol **66**(6): 1474-1490.

Le, S. Q. and O. Gascuel (2008). "An improved general amino acid replacement matrix." Mol Biol Evol **25**(7): 1307-1320.

Lee, C. T., I. T. Chen, Y. T. Yang, T. P. Ko, Y. T. Huang, J. Y. Huang, M. F. Huang, S. J. Lin, C. Y. Chen, S. S. Lin, D. V. Lightner, H. C. Wang, A. H. J. Wang, L. I. Hor and C. F. Lo (2015). "The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin (vol 112, pg 10798, 2015)." Proceedings of the National Academy of Sciences of the United States of America **112**(39): E5445-E5445.

Lee, H. J., S. J. Park, S. H. Choi and K. H. Lee (2008). "*Vibrio vulnificus* rpoS expression is repressed by direct binding of cAMP-cAMP receptor protein complex to its two promoter regions." J Biol Chem **283**(45): 30438-30450.

Lee, M. N., S. K. Kim, X. H. Li and J. H. Lee (2014). "Bacterial virulence analysis using brine shrimp as an infection model in relation to the importance of quorum sensing and proteases." J Gen Appl Microbiol **60**(5): 169-174.

Liang, W., A. Pascual-Montano, A. J. Silva and J. A. Benitez (2007). "The cyclic AMP receptor protein modulates quorum sensing, motility and multiple genes that affect intestinal colonization in *Vibrio cholerae*. Microbiology **153**(pt9):2964-75.

Llobet, E., J. M. Tomas and J. A. Bengoechea (2008). "Capsule polysaccharide is a bacterial decoy for antimicrobial peptides." Microbiology **154**(Pt 12): 3877-3886.

Lombard, V., H. G. Ramulu, E. Drula, P. M. Coutinho and B. Henrissat (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic Acids Research **42**(D1): D490-D495.

Loyola, D. E., C. Navarro, P. Uribe, K. Garcia, C. Mella, D. Diaz, N. Valdes, J. Martinez-Urtaza and R. T. Espejo (2015). "Genome diversification within a clonal population of pandemic *Vibrio parahaemolyticus* seems to depend on the life circumstances of each individual bacteria." BMC Genomics **16**: 176.

Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori and T. Iida (2003). "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of V cholerae." Lancet **361**(9359): 743-749.

Manneh-Roussel, J., J. R. J. Haycocks, A. Magán, N. Perez-Soto, K. Voelz, A. Camilli, A. M. Krachler and D. C. Grainger (2018). cAMP Receptor Protein Controls *Vibrio cholerae* Gene Expression in Response to Host Colonization. mBio. **9**(4).

Manning, P. A. (1997). "The tcp gene cluster of *Vibrio cholerae*." Gene **192**(1): 63-70.

Martinez-Urtaza, J., A. Lozano-Leon, A. DePaola, M. Ishibashi, K. Shimada, M. Nishibuchi and E. Liebana (2004). "Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates." J Clin Microbiol **42**(10): 4672-4678.

McCarter, L. (1999). "The multiple identities of *Vibrio parahaemolyticus*." J Mol Microbiol Biotechnol **1**(1): 51-57.

McDonald, N. D., K. E. DeMeester, A. L. Lewis, C. L. Grimes and E. F. Boyd (2018). "Structural and functional characterization of a modified legionaminic acid involved in glycosylation of a bacterial lipopolysaccharide." J Biol Chem **293**(49): 19113-19126.

McDonald, N. D., J. B. Lubin, N. Chowdhury and E. F. Boyd (2016). "Host-Derived Sialic Acids Are an Important Nutrient Source Required for Optimal Bacterial Fitness In Vivo." MBio **7**(2).

McDonald, N. D., A. Regmi, D. P. Morreale, J. D. Borowski and E. Fidelma Boyd (2019). "CRISPR-Cas systems are present predominantly on mobile genetic elements in *Vibrio* species." BMC Genomics **20**(1): 105.

McDonough, K. A. and A. Rodriguez (2011). "The myriad roles of cyclic AMP in microbial pathogens: from signal to sword." Nat Rev Microbiol **10**(1): 27-38.

Miyamoto, Y., T. Kato, Y. Obara, S. Akiyama, K. Takizawa and S. Yamai (1969). "In vitro hemolytic characteristic of *Vibrio parahaemolyticus*: its close correlation with human pathogenicity." J Bacteriol **100**(2): 1147-1149.

Münch, R., K. Hiller, H. Barg, D. Heldt, S. Linz, E. Wingender and D. Jahn (2003). PRODORIC: prokaryotic database of gene regulation. Nucleic Acids Res. **31:** 266-269.

Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda and D. A. Sack (2007). "Global dissemination of *Vibrio parahaemolyticus* serotype O3 : K6 and its serovariants." Clinical Microbiology Reviews **20**(1): 39.

Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda and D. A. Sack (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 and E. F. Boyd (2011). "Dichotomy in the evolution of pathogenicity island and bacteriophage encoded integrases from pathogenic *Escherichia coli* strains." Infect Genet Evol **11**(2): 423-436.

Nishibuchi, M., A. Fasano, R. G. Russell and J. B. Kaper (1992). "Enterotoxigenicity of *Vibrio parahaemolyticus* with and without genes encoding thermostable direct hemolysin." Infect Immun **60**(9): 3539-3545.

Nordstrom, J. L. and A. DePaola (2003). "Improved recovery of pathogenic *Vibrio parahaemolyticus* from oysters using colony hybridization following enrichment." J Microbiol Methods **52**(2): 273-277.

Noriea, N. F., 3rd, C. N. Johnson, K. J. Griffitt and D. J. Grimes (2010). "Distribution of type III secretion systems in *Vibrio parahaemolyticus* from the northern Gulf of Mexico." J Appl Microbiol **109**(3): 953-962.

Okada, N., T. Iida, K. S. Park, N. Goto, T. Yasunaga, H. Hiyoshi, S. Matsuda, T. Kodama and T. Honda (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." Infection and Immunity **77**(2): 904-913.

Okuda, J., M. Ishibashi, E. Hayakawa, T. Nishino, Y. Takeda, A. K. Mukhopadhyay, S. Garg, S. K. Bhattacharya, G. B. Nair and M. Nishibuchi (1997). "Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan." J Clin Microbiol **35**(12): 3150-3155.

Olaitan, A. O., S. Morand and J. M. Rolain (2014). "Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria." Front Microbiol **5**:643.

Park, K. S., T. Ono, M. Rokuda, M. H. Jang, K. Okada, T. Iida and T. Honda (2004). "Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*." Infect Immun **72**(11): 6659-6665.

Patra, T., H. Koley, T. Ramamurthy, A. C. Ghose and R. K. Nandy (2012). "The Entner-Doudoroff pathway is obligatory for gluconate utilization and contributes to the pathogenicity of *Vibrio cholerae*." J Bacteriol **194**(13): 3377-3385.

Peters, J. E., A. D. Fricker, B. J. Kapili and M. T. Petassi (2014). "Heteromeric transposase elements: generators of genomic islands across diverse bacteria." Mol Microbiol **93**(6): 1084-1092.

Peters, J. E., K. S. Makarova, S. Shmakov and E. V. Koonin (2017). "Recruitment of CRISPR-Cas systems by Tn7-like transposons." Proc Natl Acad Sci U S A **114**(35): E7358-e7366.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.

Philippe, N., J. P. Alcaraz, E. Coursange, J. Geiselmann and D. Schneider (2004). "Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria." Plasmid **51**(3): 246-255.

Pilla, G. and C. M. Tang (2018). "Going around in circles: virulence plasmids in enteric pathogens." Nature Reviews Microbiology **16**(8): 484.

Pineyro, P., X. Zhou, L. H. Orfe, P. J. Friel, K. Lahmers and D. R. Call (2010). "Development of two animal models to study the function of *Vibrio parahaemolyticus* type III secretion systems." Infect Immun **78**(11): 4551-4559.

Ramakrishnan, G., A. Jain, N. Chandra and N. Srinivasan (2016). "Computational recognition and analysis of hitherto uncharacterized nucleotide cyclase-like proteins in bacteria." Biol Direct **11**: 27.

Ritchie, J. M., H. Rui, X. Zhou, T. Iida, T. Kodoma, S. Ito, B. M. Davis, R. T. Bronson and M. K. Waldor (2012). "Inflammation and disintegration of intestinal villi in an experimental model for *Vibrio parahaemolyticus*-induced diarrhea." PLoS Pathog **8**(3): e1002593.

Rodionov, D. A., M. S. Gelfand and N. Hugouvieux-Cotte-Pattat (2004). "Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria." Microbiology-Sgm **150**: 3571-3590.

Rohmer, L., D. Hocquet and S. I. Miller (2011). "Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis." Trends Microbiol **19**(7): 341-348.

Rotman, B., A. K. Ganesan and R. Guzman (1968). "Transport systems for galactose and galactosides in *Escherichia coli*: II. Substrate and inducer specificities." Journal of Molecular Biology **36**(2): 247-260.

Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Mol Biol Evol **4**(4): 406-425.

Schmidt, H. and M. Hensel (2004). Pathogenicity Islands in Bacterial Pathogenesis. Clin Microbiol Rev. **17:** 14-56.

Schneider, D. A., W. Ross and R. L. Gourse (2003). "Control of rRNA expression in *Escherichia coli*." Curr Opin Microbiol **6**(2): 151-156.

Schneider, R., A. Travers, T. Kutateladze and G. Muskhelishvili (1999). "A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*." Mol Microbiol **34**(5): 953-964.

Schwarz, R. and Dayhoff, M. (1979) Matrices for Detecting Distant Relationships. In Dayhoff, M., Ed., Atlas of Protein Sequences, National Biomedical Research Foundation, 353-358.

Shimada, T., N. Fujita, K. Yamamoto and A. Ishihama (2011). "Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources." PLoS One **6**(6): e20081.

Silva, A. J. and J. A. Benitez (2004). "Transcriptional regulation of *Vibrio cholerae* hemagglutinin/protease by the cyclic AMP receptor protein and RpoS." J Bacteriol **186**(19): 6374-6382.

Sinha, S. C. and S. R. Sprang (2006). "Structures, mechanism, regulation and evolution of class III nucleotidyl cyclases." Rev Physiol Biochem Pharmacol **157**: 105-140.

Sismeiro, O., P. Trotot, F. Biville, C. Vivares and A. Danchin (1998). "*Aeromonas hydrophila* adenylyl cyclase 2: a new class of adenylyl cyclases with thermophilic properties and sequence similarities to proteins from hyperthermophilic archaebacteria." J Bacteriol **180**(13): 3339-3344.

Skorupski, K. and R. K. Taylor (1997). "Cyclic AMP and its receptor protein negatively regulate the coordinate expression of cholera toxin and toxin-coregulated pilus in *Vibrio cholerae*." Proc Natl Acad Sci U S A **94**(1): 265-270.

Smith, N., S. K. Kim, P. T. Reddy and D. T. Gallagher (2006). "Crystallization of the class IV adenylyl cyclase from *Yersinia pestis*." Acta Crystallogr Sect F Struct Biol Cryst Commun **62**(Pt 3): 200-204.

Smith, R. S., M. C. Wolfgang and S. Lory (2004). "An adenylate cyclase-controlled signaling network regulates *Pseudomonas aeruginosa* virulence in a mouse model of acute pneumonia." Infect Immun **72**(3): 1677-1684.

Soding, J., A. Biegert and A. N. Lupas (2005). "The HHpred interactive server for protein homology detection and structure prediction." Nucleic Acids Research **33**: W244-W248.

Song, Y., P. Yu, B. Li, Y. Pan, X. Zhang, J. Cong, Y. Zhao, H. Wang and L. Chen (2013). "The mosaic accessory gene structures of the SXT/R391-like integrative and conjugative elements derived from *Vibrio* spp. isolated from aquatic products and environment in the Yangtze River Estuary, China." BMC Microbiol **13**: 214.

Soto-Rodriguez, S. A., B. Gomez-Gil, R. Lozano-Olvera, M. Betancourt-Lozano and M. S. Morales-Covarrubias (2015). "Field and Experimental Evidence of *Vibrio parahaemolyticus* as the Causative Agent of Acute Hepatopancreatic Necrosis Disease of Cultured Shrimp (Litopenaeus vannamei) in Northwestern Mexico." Applied and Environmental Microbiology **81**(5): 1689-1699.

Steen, J. A., P. Harrison, T. Seemann, I. Wilkie, M. Harper, B. Adler and J. D. Boyce (2010). "Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors." PLoS Pathog **6**(2): e1000750.

Su, Y. C. and C. Liu (2007). "*Vibrio parahaemolyticus*: a concern of seafood safety." Food Microbiol **24**(6): 549-558.

Sullivan, M. J., N. K. Petty and S. A. Beatson (2011). "Easyfig: a genome comparison visualizer." Bioinformatics **27**(7): 1009-1010.

Suvorova, I. A., M. N. Tutukina, D. A. Ravcheev, D. A. Rodionov, O. N. Ozoline and M. S. Gelfand (2011). "Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria." J Bacteriol **193**(15): 3956-3963.

Taylor, R. K., V. L. Miller, D. B. Furlong and J. J. Mekalanos (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): 2833-2837.

Tellez-Sosa, J., N. Soberon, A. Vega-Segura, M. E. Torres-Marquez and M. A. Cevallos (2002). "The *Rhizobium etli* cyaC product: characterization of a novel adenylate cyclase class." J Bacteriol **184**(13): 3560-3568.

Thompson, J. D., D. G. Higgins and T. J. Gibson (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): 4673-4680.

Turk, E., O. Kim, J. le Coutre, J. P. Whitelegge, S. Eskandari, J. T. Lam, M. Kreman, G. Zampighi, K. F. Faull and E. M. Wright (2000). "Molecular characterization of *Vibrio parahaemolyticus* vSGLT: a model for sodium-coupled sugar cotransporters." J Biol Chem **275**(33): 25711-25716.

Vibrio Species Causing Vibriosis | *Vibrio* Illness (Vibriosis) | CDC." from https://www.cdc.gov/vibrio/. Accessed 16 Aug 2018.

Waldor, M. K. and J. J. Mekalanos (1996). "Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin." **272**(5270).

Wang, H., B. Liu, Q. Wang and L. Wang (2013). "Genome-wide analysis of the salmonella Fis regulon and its regulatory mechanism on pathogenicity islands." PLoS One **8**(5): e64688.

Whitaker, W. B., M. A. Parent, A. Boyd, G. P. Richards and E. F. Boyd (2012). "The *Vibrio parahaemolyticus* ToxRS regulator is required for stress tolerance and colonization in a novel orogastric streptomycin-induced adult murine model." Infect Immun **80**(5): 1834-1845.

Whitaker, W. B., G. P. Richards and E. F. Boyd (2014). "Loss of sigma factor RpoN increases intestinal colonization of *Vibrio parahaemolyticus* in an adult mouse model." Infect Immun **82**(2): 544-556.

Wozniak, R. A., D. E. Fouts, M. Spagnoletti, M. M. Colombo, D. Ceccarelli, G. Garriss, C. Dery, V. Burrus and M. K. Waldor (2009). "Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs." PLoS Genet **5**(12): e1000786.

Xu, F., N. Gonzalez-Escalona, K. P. Drees, R. P. Sebra, V. S. Cooper, S. H. Jones and C. A. Whistler (2017). "Parallel evolution of two clades of a major Atlantic endemic *Vibrio parahaemolyticus* pathogen lineage by independent acquisition of related pathogenicity islands." Appl Environ Microbiol. **83** (18).

Xu, F., N. Gonzalez-Escalona, J. Haendiges, R. A. Myers, J. Ferguson, T. Stiles, E. Hickey, M. Moore, J. M. Hickey, C. Schillaci, L. Mank, K. DeRosia-Banick, N. Matluk, A. Robbins, R. P. Sebra, V. S. Cooper, S. H. Jones and C. A. Whistler (2017). Sequence Type 631 *Vibrio parahaemolyticus*, an Emerging Foodborne Pathogen in North America. J Clin Microbiol. **55:** 645-648.

Xu, F., S. Ilyas, J. A. Hall, S. H. Jones, V. S. Cooper and C. A. Whistler (2015). "Genetic characterization of clinical and environmental *Vibrio parahaemolyticus* from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages." Front Microbiol. **6**:272.

Zhang, H., B. Ni, X. Zhao, I. Dadzie, H. Du, Q. Wang, H. Xu and X. Huang (2012). "Fis is essential for the stability of linear plasmid pBSSB1 and affects the motility of *Salmonella enterica* serovar Typhi." PLoS One **7**(7): e37462.

Zimmerman, A. M., A. DePaola, J. C. Bowers, J. A. Krantz, J. L. Nordstrom, C. N. Johnson and D. J. Grimes (2007). "Variability of total and pathogenic *Vibrio parahaemolyticus* densities in northern Gulf of Mexico water and oysters." Appl Environ Microbiol **73**(23): 7589-7596.

Appendix A

CHARACTERIZATION OF AN ALTERNATIVE SIGMA FACTOR VP2210 IN *V. PARAHAEMOLYTICUS*

Sigma factors are bacterial proteins that binds the regulatory DNA region and interacts with RNA polymerase to initiate transcription. Bacteria harbors two major family of sigma factors, sigma 70 and sigma 54. In *V. parahaemolyticus*, the sigma 54 encoded by *rpoN* was demonstrated to be involved in motility and carbon and nitrogen metabolism. In a streptomycin treated adult mouse model, Δ*rpoN* strain demonstrated superior colonization and was suggested due to its efficient carbon utilization potential (Whitaker, Richards et al. 2014). In another study, one of the sigma 70 sigma factor RpoE was demonstrated to be involved in cell envelop stress response and a mutant strain devoid of RpoE had attenuated colonization of a streptomycin treated adult mouse and this phenotype was due to the inability of the mutant to overcome cell envelop stress (Haines-Menges, Whitaker et al. 2014) The genome of *V. parahaemolyticus* contains 11 total sigma factors and the functions of all the sigma factors are not characterized. In this work, we investigated the potential role of a sigma factor with a locus tag VP2210. To investigate the role of VP2210 in *V. parahaemolyticus,* we constructed an in-frame deletion mutant of VP2210 and characterized the mutant under various stress conditions. Our analysis found that under our laboratory condition, VP2210 was not involved in any of the stress condition we tested, and VP2210 is not essential for *in vivo* colonization of a streptomycin treated adult mouse model.

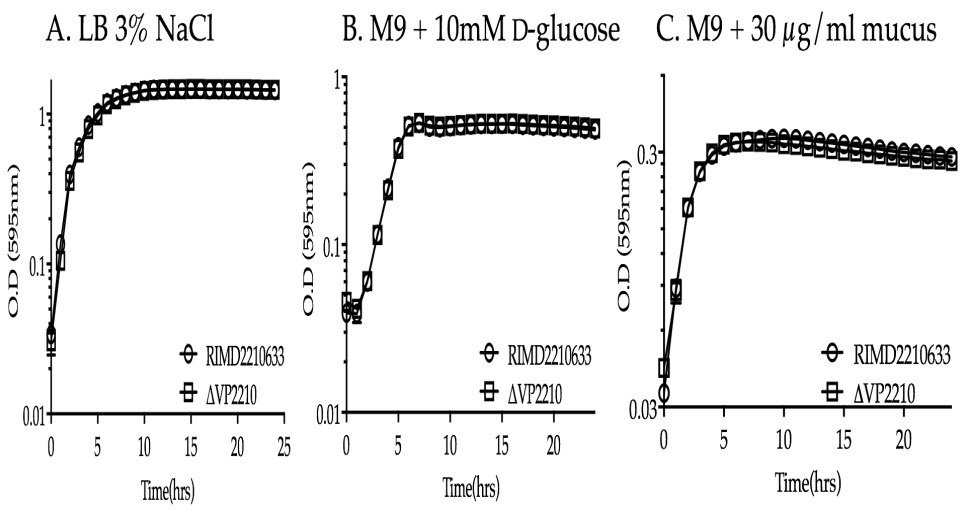


Figure 39 **Growth pattern analysis of Δ*VP2210*.** Bacterial culture of RIMD2201633 and ΔVP2210 were grown in **(A)** LB 3% NaCl **(B)** M9 + 10 mM glucose and **(C)** M9 + 30 μg/ml mouse intestinal mucin for 24 hours. The Δ*VP2210* mutant strain grew identical to the WT in all the growth media.

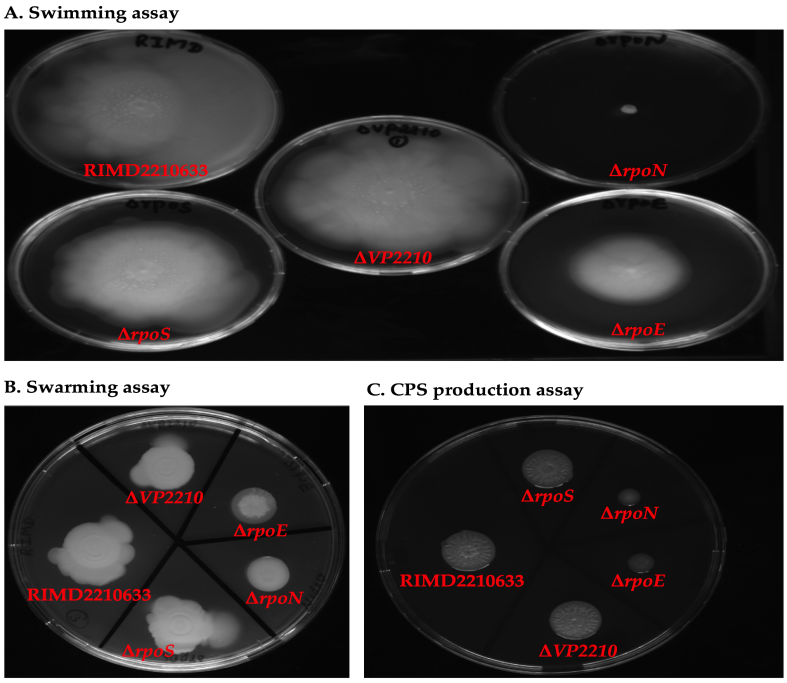


Figure 40 **Swimming, swarming and CPS production assay.** *Vibrio parahaemolyticus* strains RIMD2210633, Δ*VP2210*, Δ*rpoE*, Δ*rpoN* and Δ*rpoS* were assessed for their ability to **(A)** swim **(B)** swarm, and **(C)** CPS production. RIMD2210633 is efficient in swimming, swarming and CPS production and Δ*rpoN* is defective in all these three behaviors. The Δ*rpoS* and Δ*VP2210* behaved like RIMD2210633 and only Δ*rpoE* demonstrated swimming, swarming and CPS production defect.

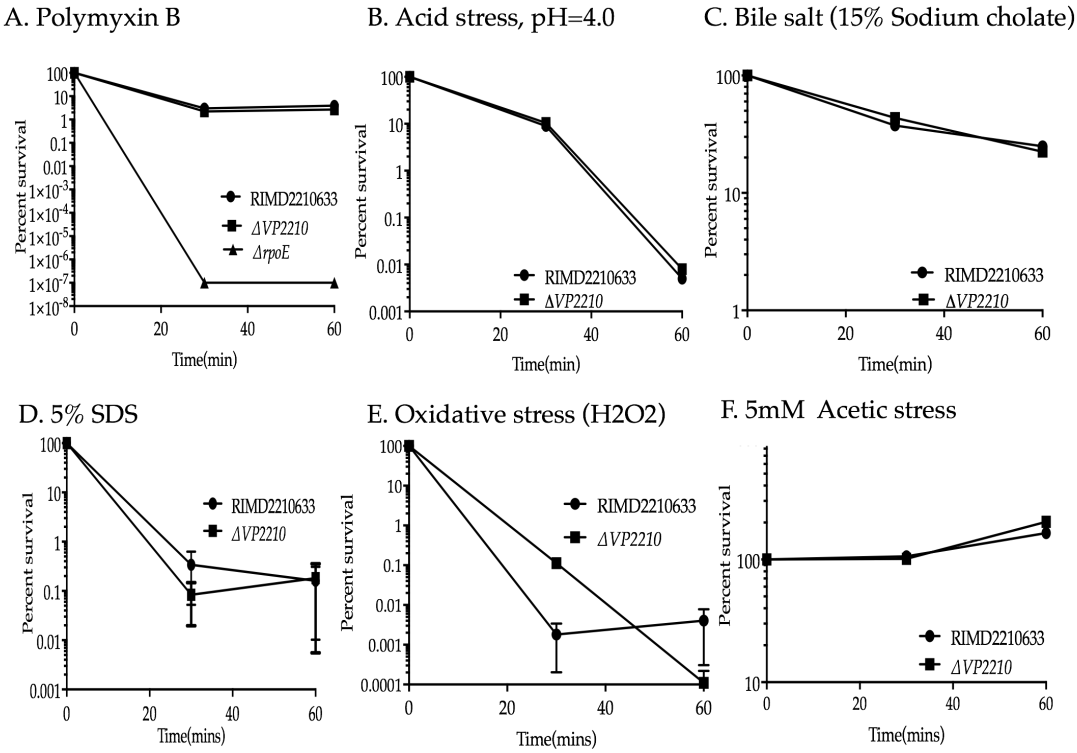


Figure 41 **Characterization of Δ*VP2210* strain in various stress conditions.** The overnight culture of RIMD2210633 and Δ*rpoE* were grown for 2 hrs (early log phase) and treated with either **(A)** 40 µg/ml polymyxin B **(B)** pH stress (LB 3% NaCl pH 4.0) **(C)** 15% sodium cholate **(D)** 5% SDS **(E)** 5 mM H202 or **(F)** 5mM acetic acid. At 0 min, 30 min and 60 min, aliquot of the bacterial culture was obtained, serially diluted and plated to obtain CFUs. Percent survival was then calculated to difference in survival between the RIMD2210633 and ΔVP2210 strain. For polymyxin B stress assay Δ*rpoE* was used as a control. The Δ*VP2210* strain behaved identical to RIMD2210633 in all the tested conditions indicating VP2210 is not involved in any of these stress tolerance mechanism.

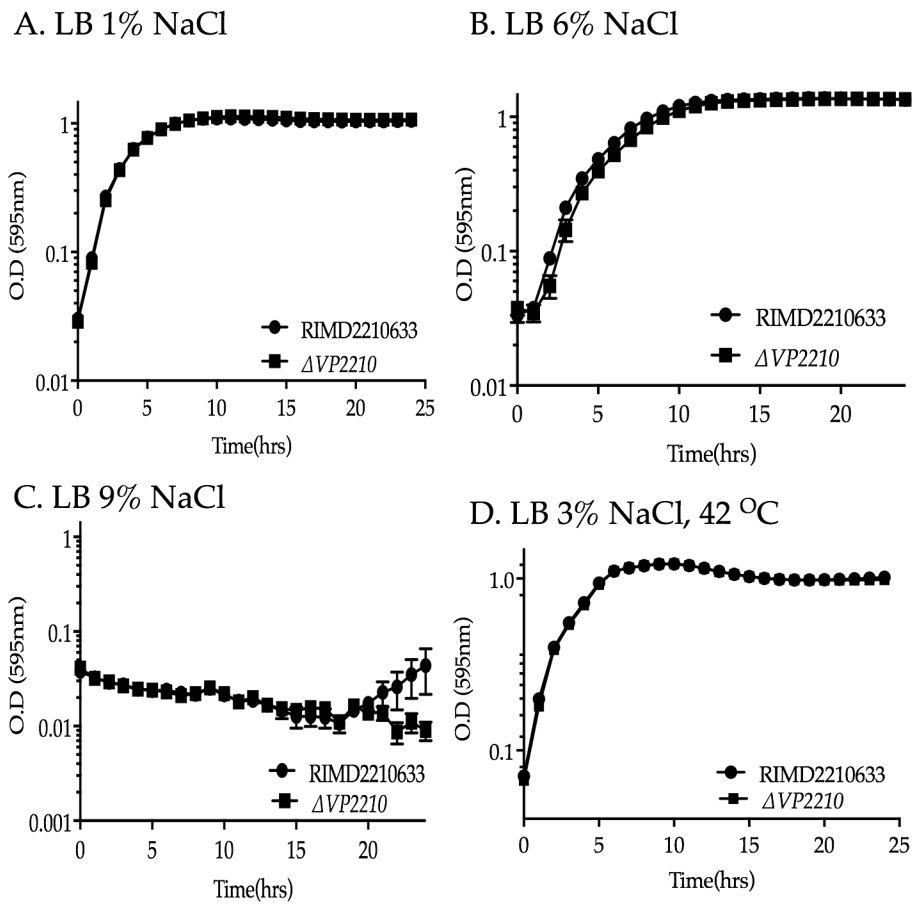


Figure 42 **Characterization of Δ*VP2210* strain in salt and temperature stress**. *Vibrio parahaemolyticus* strains RIMD2210633 and Δ*VP2210* were grown in **(A)** LB 1%NaCl **(B)** LB 6%NaCl **(C)** LB 9% NaCl at 37 degree Celsius to investigate the role of VP2210 in salt stress and **(D)** LB 3% NaCl at 42 degree Celsius to investigate the role of VP2210 in temperature stress. In all the conditions tested the Δ*VP2210* strain behaved like RIMD2210633 indicating that VP2210 is not involved in salt or temperature stress.

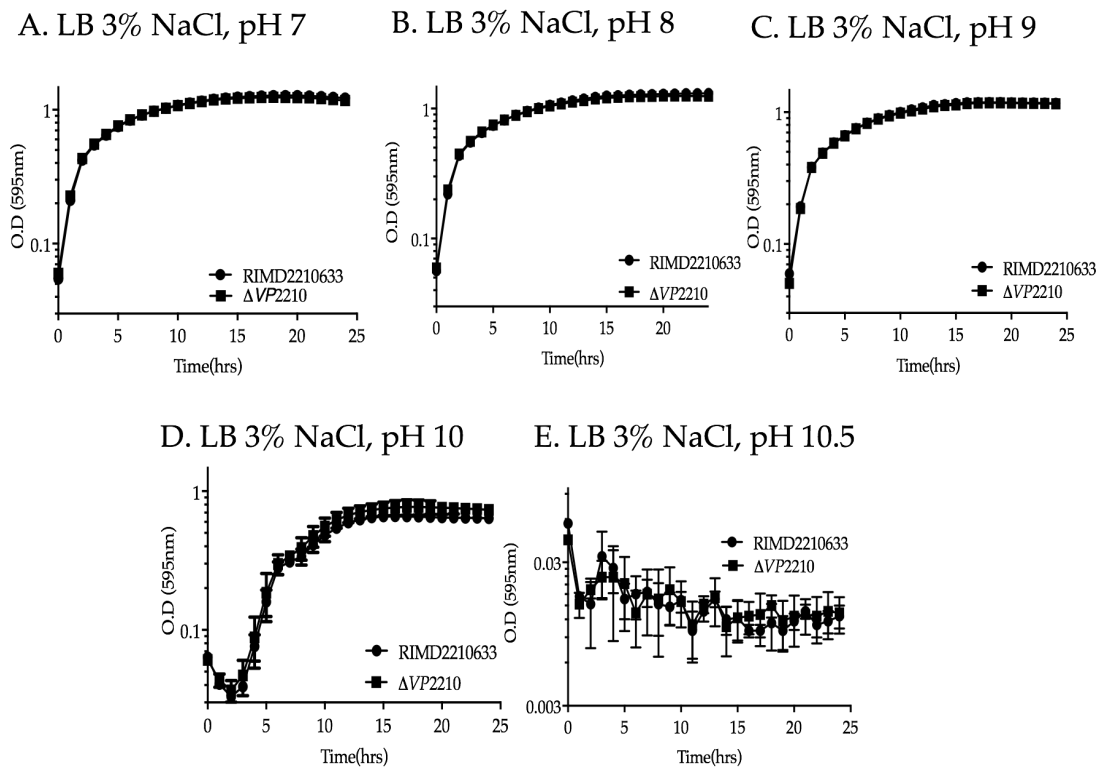


Figure 43 **Characterization of Δ*VP2210* strain in alkaline stress.** *Vibrio parahaemolyticus* strains RIMD2210633 and Δ*VP2210* were grown in **(A)** LB 3% pH 7 **(B)** LB 3% pH 8 **(C)** LB 3% pH 9 **(D)** LB 3% pH 10 **(E)** LB 3% pH 10.5 for 24 hrs to investigate the potential role of VP2210 in alkaline stress tolerance. The Δ*VP2210* strain grew identical to RIMD2210633 in all the condition tested indicating that VP2210 doesn't play any role in alkaline stress tolerance in *V. parahaemolyticus*.

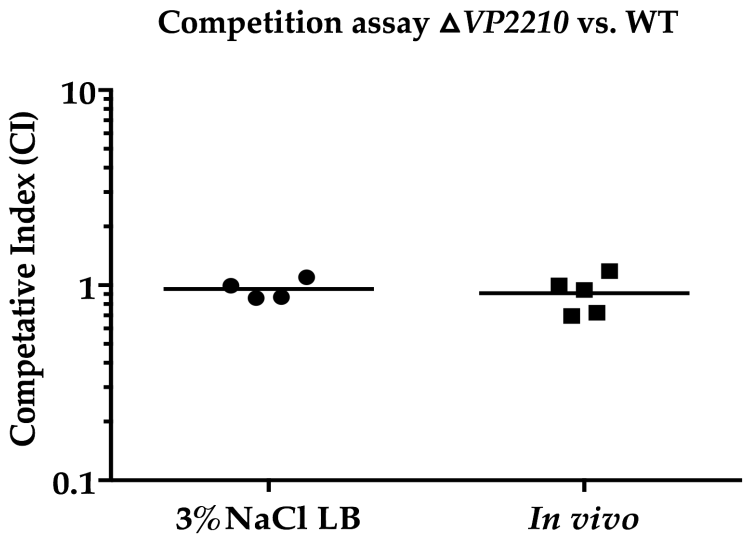
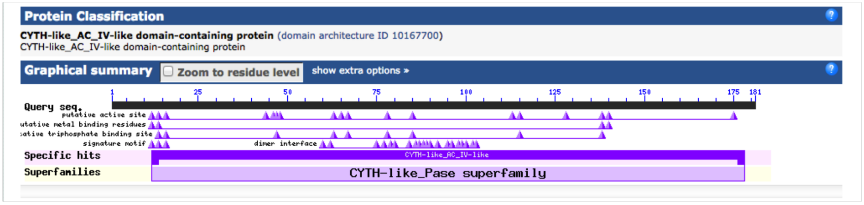


Figure 44 ***In vivo* competition assay between Δ*VP2210 and* WT strain**. A streptomycin pretreated adult mouse model of colonization was used to investigate the colonization role of VP2210. The competitive index of the assay was identified to be almost one indicating that there is no competition between the WT and the mutant strain demonstrating that VP2210 doesn't play any role in *in vivo* colonization of the streptomycin pretreated adult mouse.

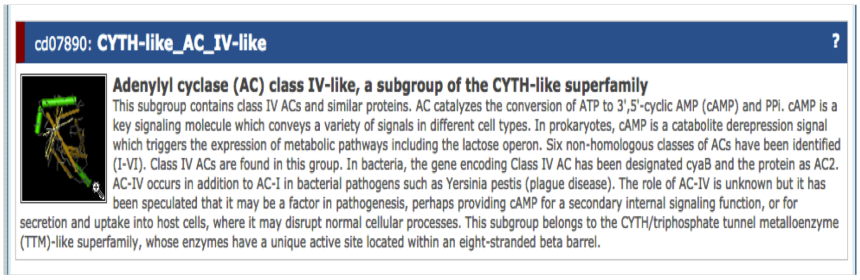
Appendix B

STRUCTURAL ANALYSIS OF CyaB

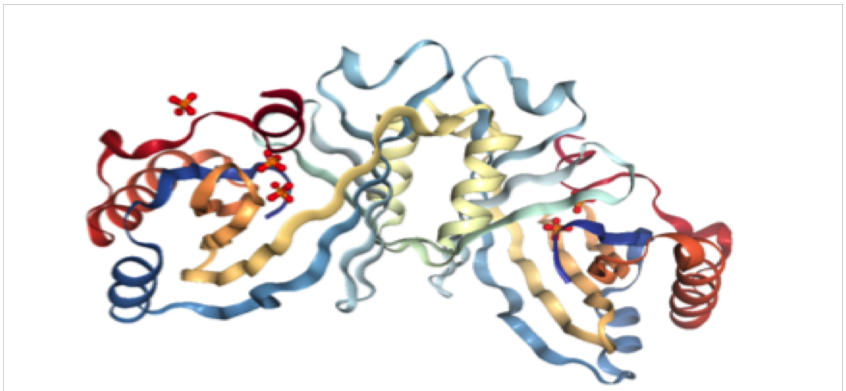
A.



B.



C.



X-ray structure of a putative adenylate cyclase Q87NV8 from Vibrio parahaemolyticus (VP1760) at the 2.25 A resolution. Northeast Structural Genomics Target VpR19. 10.2210/pdb2ACA/pdb

Figure 45 **Structural analysis of CyaB. (A)** BLAST analysis of *V. parahaemolyticus* VP1760. **(B)** NCBI Description of VP1760 (CyaB). **(C)** X-ray structure of a putative adenylate cyclase Q87NV8 from *Vibrio parahaemolyticus* (VP1760) at the 2.25 A resolution. Northeast Structural Genomics Target VpR19. 10.2210/pdb2ACA/pdb.

Appendix C

DISTRIBUTION AND PHYLOGENETIC ANALYSIS OF CYAB

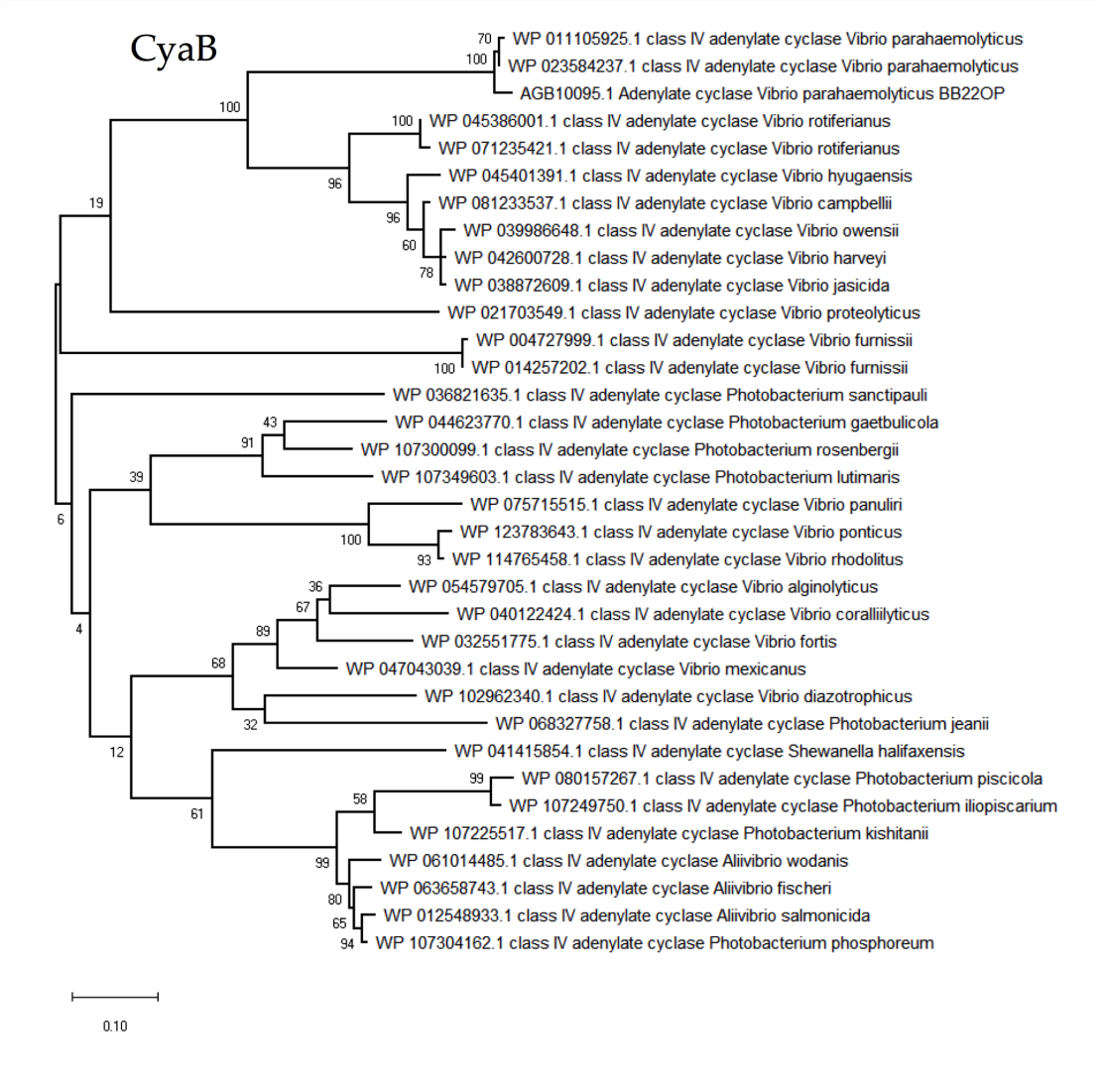


Figure 46 **Distribution and Phylogenetic analysis of CyaB.**

Appendix D

SXT/R391 LIKE ICE IN *V. PARAHAEMOLYTICUS* STRAINS

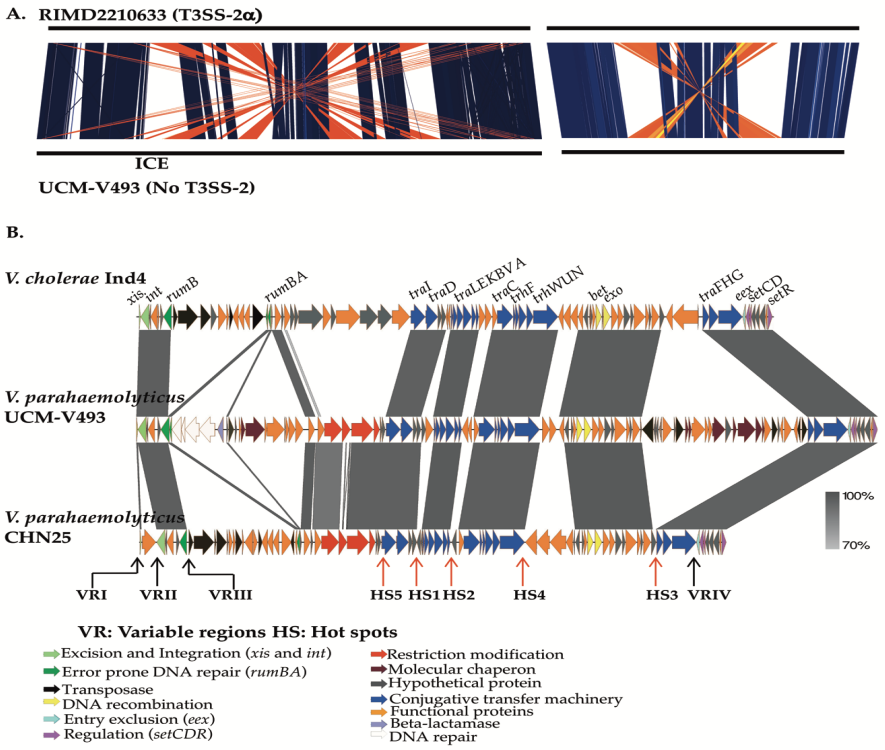
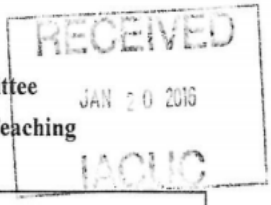


Figure 47 **SXT/R391 like ICE in *V. parahaemolyticus* strains.** (**A**). Whole genome comparison between RIMD2210633 and UCM-V493. Black bars indicate chromosomes. Blue and red blocks represent regions of nucleotide homology between strains. Red indicates inversion of the region. White blocks indicate regions present in one strain and absent in the other. (**B**). Comparative analysis of ICE present in strains UCM-V493 and CHN25 with that of *V. cholerae* Ind4, which contains SXT/R391 family ICE. Gray shade, homologous region. Color arrows, genes encoding functional and hypothetical proteins. Black arrow, transposase. Hot spots and variable regions within the ICE were assigned as previously described (Wozniak, Fouts et al. 2009, Song, Yu et al. 2013).

Appendix E

ANIMAL WORK APPROVAL



University of Delaware
Institutional Animal Care and Use Committee
Application to Use Animals in Research and Teaching

Title of Protocol: Studies of *Vibrio species* pathogenesis	
AUP Number: 1169-2016-0	← (4 digits only — if new, leave blank)
Principal Investigator: Dr. E. Fidelma Boyd	
Common Name: Mouse **Genus Species:** *Mus musculus*	

Pain Category: *(please mark one)*

USDA PAIN CATEGORY: *(Note change of categories from previous form)*	
Category	**Description**
☐ B	Breeding or holding where NO research is conducted
☐ C	Procedure involving momentary or no pain or distress
☐ D	Procedure where pain or distress is alleviated by appropriate means (analgesics, tranquilizers, euthanasia etc.)
☒ E	Procedure where pain or distress cannot be alleviated, as this would adversely affect the procedures, results or interpretation

Official Use Only

IACUC Approval Signature: _____

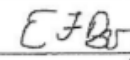
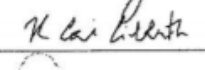
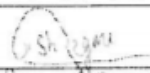
Date of Approval: 3-1-16

Principal Investigator Assurance

1. I agree to abide by all applicable federal, state, and local laws and regulations, and UD policies and procedures.
2. I understand that deviations from an approved protocol or violations of applicable policies, guidelines, or laws could result in immediate suspension of the protocol and may be reportable to the Office of Laboratory Animal Welfare (OLAW).
3. I understand that the Attending Veterinarian or his/her designee must be consulted in the planning of any research or procedural changes that may cause more than momentary or slight pain or distress to the animals.
4. I declare that all experiments involving live animals will be performed under my supervision or that of another qualified scientist. All listed personnel will be trained and certified in the proper humane methods of animal care and use prior to conducting experimentation.
5. I understand that emergency veterinary care will be administered to animals showing evidence of discomfort, ailment, or illness.
6. I declare that the information provided in this application is accurate to the best of my knowledge. If this project is funded by an extramural source, I certify that this application accurately reflects all currently planned procedures involving animals described in the proposal to the funding agency.
7. I assure that any modifications to the protocol will be submitted to by the UD-IACUC and I understand that they must be approved by the IACUC prior to initiation of such changes.
8. I understand that the approval of this project is for a maximum of one year from the date of UD-IACUC approval and that I must re-apply to continue the project beyond that period.
9. I understand that any unanticipated adverse events, morbidity, or mortality must be reported to the UD-IACUC immediately.
10. I assure that the experimental design has been developed with consideration of the three Rs: reduction, refinement, and replacement, to reduce animal pain and/or distress and the number of animals used in the laboratory.
11. I assure that the proposed research does not unnecessarily duplicate previous experiments. ***(Teaching Protocols Exempt)***
12. I understand that by signing, I agree to these assurances. E Fidelma Boyd — Signature of Principal Investigator 1/20/2016 — Date

NAMES OF ALL PERSONS WORKING ON THIS PROTOCOL

I certify that I have read this protocol, accept my responsibility and will perform only those procedures that have been approved by the IACUC.

Name	Signature
1. Dr. E. Fidelma Boyd	
2. Sai Kalburge	
3. Abish Regmi	
4. Nathan McDonald	
5.	
6. Click here to enter text.	
7. Click here to enter text.	
8. Click here to enter text.	
9. Click here to enter text.	
10. Click here to enter text.	
11. Click here to enter text.	
12. Click here to enter text.	
13. Click here to enter text.	
14. Click here to enter text.	