

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Discovery and description of a novel toxin within the pan- genome of the waterborne pathogen *Vibrio cholerae*

Permalink

<https://escholarship.org/uc/item/9bs7f43g>

Author

Purdy, Alexandra Elizabeth

Publication Date

2007

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Discovery and Description of a Novel Toxin Within the Pan-Genome of the
Waterborne Pathogen *Vibrio cholerae***

A Dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Marine Biology

by

Alexandra Elizabeth Purdy

Committee in charge:

Professor Douglas H. Bartlett, Chair
Professor Farooq Azam
Professor Bianca Brahamsha
Professor Joshua Fierer
Professor Kit Pogliano

2007

Copyright

Alexandra Elizabeth Purdy, 2007

All rights reserved.

The Dissertation of Alexandra Elizabeth Purdy is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2007

DEDICATION

Dedicated to my family
Mom, Dad, Chris, and Pippa

EPIGRAPH

If I could do it all over again ... I would be a microbial ecologist.

Ten billion bacteria live in a gram of ordinary soil, a mere pinch held between thumb and forefinger. They represent thousands of species, almost none of which are known to science. Into that world I would go with the aid of modern microscopy and molecular analysis.

I would cut my way through clonal forests sprawled across grains of sand, travel in an imagined submarine through drops of water proportionately the size of lakes, and track predators and prey in order to discover new life ways and alien food webs.

All this, and I need venture no farther than ten paces outside my laboratory building.

Naturalist

Edward O. Wilson

TABLE OF CONTENTS

Signature Page.....	iii
Dedication.....	iv
Epigraph.....	v
Table of Contents.....	vi
List of Abbreviations.....	vii
List of Figures.....	viii
List of Tables.....	ix
Acknowledgements.....	x
Vita.....	xiv
Abstract.....	xvi
Chapter 1. Perspectives on <i>Vibrio cholerae</i> : Microbial ecology, genomic diversity, and virulence mechanisms.....	1
Chapter 2. A glimpse into the expanded genome content of the <i>Vibrio cholerae</i> species through the identification of genes present in environmental strains.....	27
Chapter 3. Cholix toxin, a novel ADP-ribosylating factor from <i>Vibrio cholerae</i>	74
Chapter 4. Diversity and distribution of cholix toxin among <i>Vibrio</i> <i>cholerae</i> strains from around the world.....	108
Chapter 5. Ecological and pathogenic implications for cholix toxin in <i>Vibrio cholerae</i> : future research directions.....	151

LIST OF ABBREVIATIONS

ADP	adenosine diphosphate
ADPRT	ADP-ribosyl transferase
bp	basepair
dN/dS	non-synonymous vs. synonymous amino acid changes
DNA	deoxyribonucleic acid
eEF2	eukaryotic elongation factor 2
ER	endoplasmic reticulum
ETA	exotoxin A from <i>Pseudomonas aeruginosa</i>
FEL	Fixed Effects Likelihood analysis
GARD	Genetics Algorithms for Recombination Detection program
indel	insertion-deletion
ISR	intergenic spacer region
KDEL	Lys-Asp-Glu-Leu endoplasmic reticulum retention protein sequence
LB	Luria-Bertani broth
LPS	lipopolysaccharide
LRP1	low density lipoprotein receptor-like protein/alpha-2 macroglobulin receptor
Mb	mega basepairs
MEF1	mouse embryonic fibroblast cell line
N16961	<i>Vibrio cholerae</i> strain O1 El Tor N16961
NAD ⁺	nicotinamide adenine dinucleotide
PCR	polymerase chain reaction
PEA13	mouse embryonic fibroblast cell line lacking LRP1
REL	Random Effects Likelihood analysis
RNA	ribonucleic acid
RTX	repeats-in-structure toxin
SIO	<i>Vibrio cholerae</i> strain SIO
SLAC	Single Likelihood Ancestor Counting method
SSH	suppressive subtractive hybridization
TP	<i>Vibrio cholerae</i> strain TP

LIST OF FIGURES

Figure 2.1:	Phylogeny of <i>Vibrio cholerae</i> strains SIO and TP based upon (A) 16S rDNA sequence and (B) <i>xerC</i> sequence.....	54
Figure 2.2:	Structure of plasmid pSIO1 (4906 bp).....	55
Figure 2.3:	Distribution of selected environmental genes among individual <i>Vibrio</i> isolates.....	56
Figure 2.4:	Percentage distribution of selected environmental genes among <i>Vibrio</i> strains.....	57
Figure 3.1:	Cholix toxin is cytotoxic towards mouse fibroblasts.....	94
Figure 3.2:	Biological activity of cholix toxin.....	95
Figure 3.3:	The structure of cholix toxin.....	96
Figure 3.4:	The structure of cholix toxin.....	97
Figure 3.S1.	Supplementary Figure 1. Electron density maps of the full-length cholix and cholix _c structures.....	107
Figure 4.1.	<i>chxA</i> sequences fall into three clades.....	138
Figure 4.2:	Neighbor joining trees automatically generated by GARD.....	139
Figure 4.3:	Annotated cholix toxin protein sequence alignment.....	141
Figure 4.4:	Surface view of cholix toxin structure: regions of interest within the receptor binding domain.....	144
Figure 4.5:	Cholix toxin protein structure: regions of interest within the catalytic domain.....	145

LIST OF TABLES

Table 2.1:	Percent nucleotide sequence identity of phylogenetic marker genes to published N16961 sequences among <i>Vibrio cholerae</i> strains used in this study.....	50
Table 2.2:	Functional categories of genes that are significantly similar to SIO and TP sequence fragments using the BlastX algorithm...	51
Table 2.3:	Top BlastX matches of genes of interest.....	53
Table 2.S1:	Supplementary Materials Table 1.....	70
Table 2.S2:	Supplementary Materials Table 2.....	72
Table 3.1:	Comparison of the ADPRT kinetic parameters for ETA _c and cholix _c toxins.....	93
Table 3.S1:	Supplementary Materials Table 1. Data collection statistics, phasing and refinement of cholix structures.....	106
Table 4.1:	Primers used in this study.....	130
Table 4.2:	Presence of <i>chxA</i> among <i>Vibrio cholerae</i> strains isolated in Bangladesh.....	131
Table 4.3:	Presence of <i>chxA</i> among <i>Vibrio cholerae</i> strains isolated in Mexico.....	133
Table 4.4:	Presence of <i>chxA</i> among <i>Vibrio cholerae</i> strains from Senegal and Peru.....	135
Table 4.5:	<i>V. cholerae</i> strains from southern California: strains carrying <i>chxA</i>	136
Table 4.7:	Additional strains which were tested for <i>chxA</i>	137

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my dissertation advisor, Doug Bartlett, for his unbounded optimism, enthusiasm, and support. With his door always open, Doug welcomed discussions, and he was seemingly ready with new ideas at all times. He encouraged me to be fearless and pursue this project unconditionally, which vastly expanded the array of techniques and collaborative interactions that I developed. He did all that he could to help me along the way, whether it was to provide contacts who could supply equipment, lab space or reagents, or to find ways to keep me supported during the last months of my project. He is the consummate scientist, continuously enthralled by his subject of study, and truly dedicated. At the same time, he was an inspirational mentor, and I can only hope to have learned from his example.

Of course I would also like to extend my sincere thanks to the members of my thesis committee: Farooq Azam, whose discussions always bestowed me with new hope and ideas for further research directions; Kit Pogliano, Bianca Brahmsha and Joshua Fierer whose expertise, optimism and encouragement were truly appreciated, and who each contributed valuable ideas. I am grateful that I was guided by a group of such superb scientists.

I would like to thank the members of the Bartlett lab, past and present, fixtures and visitors alike, who helped me throughout the project. I would like to especially to thank my officemates, Leena Palekar and Simran Saini, for their eternal patience and

support. The folks who helped with this project also deserve a great deal of thanks, especially Debbie Balch and Collin Chang. I would also like to thank all of the members of the Azam lab for their advice, patience, and borrowed equipment! After six years as a Hubbs Hall resident, there is hardly a person in the building who did not help me at some point along the way, and for this I am extremely grateful. These folks include (but are certainly not limited to) Eddie Kisfaludy, Roger Chastain, Mark Hildebrand, Hope Johnson, Brian Hillier, Sheila Podell, and Eric Allen. Folks outside of Hubbs Hall also helped a great deal, especially Sara Kelly, whose guidance with tissue culture work was tremendously generous, and Art Poon whose knowledge and patience with my naïve questions regarding phylogenetics and evolution were phenomenal. Also, numerous folks around the country helped us out by testing our *V. cholerae* strains for virulence phenotypes, and I am grateful for their help as well. I would also like to thank Xavier Lauth and Jason Stannard at Kent SeaTech, a small aquaculture firm in San Diego, who let me work in their laboratories for several months, and graciously helped me with virulence assays.

I would especially like to thank Rod Merrill at the University of Guelph in Ontario, Canada for being an absolutely fantastic collaborator. He graciously hosted me in his laboratory for two weeks, and patiently shared his expertise regarding enzyme kinetics, protein purification, and structural biochemistry. I would also like to thank the folks in the Merrill lab who were so welcoming during my stay in Guelph, and who have continued to help me by providing purified proteins, molecular biology advice, and support throughout the project, including Gerry Prentice, Dawn White,

Rob Fieldhouse, and Rene Jorgensen. Working with all them has been an absolute pleasure, and I thank each of them for their patience as this aspiring microbial geneticist was taught to become a protein biochemist and a cell biologist!

Of course I would also like to thank the friends who have supported me along the way: Leena & Demian, Christine, Robert, Michele, and Rita, as well as many other friends who made my years in San Diego no less than epic.

Lastly, I would like to thank my family: Mom, Dad, Christopher, and Pippa. Their unwavering faith in me is both a mystery and a tremendous support.

Chapter 2, in full, is a reprint of the material as it appears in Purdy, A., F. Rohwer, R. Edwards, F. Azam, and D. H. Bartlett. A glimpse into the expanded genome content of the *Vibrio cholerae* species through the identification of genes present in environmental strains. 2005. J. Bacteriol. 187: 2992-3001. The dissertation author was the primary investigator and author of this paper, with the co-authors listed above.

Chapter 3, in full, has been submitted for publication. Jorgensen, R., A. E. Purdy, R. J. Fieldhouse, M. S. Kimber, D. H. Bartlett, and A. R. Merrill. Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. Submitted. The dissertation author was the co-first author of this paper, and was the primary investigator and author of sections relating to the biological role of cholix toxin.

Chapter 4, in part, will be submitted for publication. Purdy, A. E., D. Balch, and D. H. Bartlett. Diversity and distribution of cholix toxin among *Vibrio cholerae* strains from around the world. The dissertation author was the primary investigator and author of this work.

VITA

- 1996 High school diploma, Falmouth High School, Falmouth, MA
- 2000 Bachelor of Science, The College of William and Mary
- 2000 - 2001 Research Assistant, The Jackson Laboratory
- 2001 - 2006 Howard Hughes Medical Institute Predoctoral Fellow
Scripps Institution of Oceanography
University of California, San Diego
- 2003 Advanced Bacterial Genetics, Cold Spring Harbor Laboratory
- 2004, 2006 Teaching Assistant, BIMM120 Bacteriology
Division of Biological Sciences, University of California, San Diego
- 2007 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Case, L. K., L. Petell, L. Yurkovestkiy, A. Purdy, K. J. Savage, and T. V. Golovkina. 2007. Replication of beta- and gamma-retroviruses is restricted in I/LnJ mice via the same genetic mechanism. *J Virol.* In press.

Case, L. K., A. Purdy, and T. V. Golovkina. 2005. Molecular and cellular basis of the retrovirus Resistance in I/LnJ mice. *J Immunol.* **175**:7543-7549.

Jorgensen, R., A. E. Purdy, R. J. Fieldhouse, M. S. Kimber, D. H. Bartlett, and A. R. Merrill. 2007. Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. Submitted.

Makri, S., A. E. Purdy, D. Bartlett, and J. Fierer. 2007. Pathogenicity of environmental isolates of *V. cholerae* in mice. *Microb Infect.* **9**:1351-1358.

Purdy, A., L. Case, M. Duvall, M. Overstrom-Coleman, N. Monnier, A. Chervonsky, and T. Golovkina. 2003. Unique resistance of I/LnJ mice to a retrovirus is due to sustained interferon gamma-dependent production of virus-neutralizing antibodies. *J. Exp. Med.* **197**:233-243.

Purdy, A., F. Rohwer, R. Edwards, F. Azam, and D. H. Bartlett. 2005. A glimpse into the expanded genome content of the *Vibrio cholerae* species through the identification of genes present in environmental strains. *J. Bacteriol.* **187**:2992-3001.

Valentine, D. L., M. Kastner, G. D. Wardlaw, X. Wang, A. Purdy, and D. H. Bartlett. 2005. Biogeochemical investigations of marine methane seeps, Hydrate Ridge, Oregon. *J. Geophys. Res. Biogeosciences*:G02005, doi:10.1029/2005JG000025.

ABSTRACT OF THE DISSERTATION

Discovery and Description of a Novel Toxin within the Pan-Genome of the
Waterborne Pathogen *Vibrio cholerae*

by

Alexandra Elizabeth Purdy

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2007

Professor Douglas H. Bartlett, Chair

The waterborne pathogen *Vibrio cholerae* continues to be responsible for devastating global pandemics of cholera disease. Strains of the O1 or O139 serogroups that carry the cholera toxin can colonize the small intestine of humans and

cause massive efflux of water from the victim, resulting in dehydration and death if treatment with saline fluids is not initiated quickly. However, the majority of strains isolated from environmental settings do not carry cholera toxin and are serogroups other than O1 or O139 (hence, their non-O1, non-O139 strain designation). Although these strains are often lumped together when discussing causative agents of cholera disease, they are in fact quite genotypically and phenotypically diverse. In Chapter 2 of this thesis, a first look at the types of genes that account for the genotypic diversity is presented, and this work demonstrated that the *V. cholerae* pan-species genome is quite extensive. Chapter 3 then follows up on the discovery of a gene encoding a toxin similar to exotoxin A of *Pseudomonas aeruginosa*, a potent secreted virulence factor of this opportunistic pathogen. In a joint publication, our work describing the biological effects of the new toxin, now designated cholix toxin, complements the work of our collaborators at the University of Guelph, who solved the crystal structure of both full length cholix toxin and its catalytic domain. Chapter 4 then takes a phylogenetic approach to understanding the role the toxin may be playing in *V. cholerae* virulence or environmental survival, by assessing the presence or absence of the gene encoding cholix toxin, *chxA*, in a collection of *V. cholerae* strains amassed from around the world. The *chxA* gene was then sequenced from a wide variety of clinical and environmental strains, revealing the presence of specific regions with extensive sequence variation. A final chapter describes further efforts to uncover the purpose of this toxin, as well as ideas for future research directions to pursue this goal.

Cholix toxin provides a model for understanding selective pressures and the evolution of toxin genes in environmental settings.

Chapter 1.

Perspectives on *Vibrio cholerae*: Microbial ecology, genomic diversity, and virulence mechanisms

Abstract

Vibrio cholerae, the bacterial species responsible for causing cholera disease, can be isolated from coastal marine, estuarine, and riverine systems throughout the world, suggesting that the reservoir for this infectious disease agent is the aquatic environment. Cholera pandemics have been caused by only two of the more than 200 *V. cholerae* serogroups thus far characterized, but the vast majority of strains isolated from environmental settings are genetically diverse and usually do not carry the genes required for causing cholera. The sharp phenotype difference among strains within this single species provides a system in which to investigate genome dynamics governing bacterial speciation (such as recombination) under conditions that might be expected to promote diversity and subsequent divergence. This introductory chapter will provide a background of *V. cholerae* biology in order to place the dissertation's discoveries of specific genes present in environmental strains, and the subsequent characterization of a novel toxin in *V. cholerae*, into a larger context.

Introduction. Around the globe, infectious diseases account for millions of deaths annually (62). As parts of the world become both politically and ecologically unstable, waterborne diseases become a greater concern because the best prevention mechanism -- supplying the most vulnerable populations with clean drinking water -- becomes increasingly difficult (12, 95). Cholera disease is transmitted by water carrying high concentrations of *Vibrio cholerae*, a gram-negative bacterium capable of colonizing the human small intestine and secreting cholera toxin, which results in rapid and massive diarrhea and fluid loss from the victim. Estimates have suggested that cholera can affect as many five to seven million people per year worldwide, causing greater than >100,00 deaths (80), although the reported number of cases is much lower, with 101,000 cases reported in 2004, resulting in 2345 deaths (36). It is often easily treatable if adequate fluids are administered within a short time of the onset of symptoms (81), and vaccines against the disease are available, but they are not fully protective, and their effectiveness is limited to time scales of months to a few years (36). It is likely that the disease will never be completely eradicated, however, because *Vibrio cholerae* is an autochthonous member of the aquatic microbial community, and its extensive genetic diversity and propensity for homologous recombination ensure that new variants will emerge to evade immune defenses. Therefore understanding the nature of *V. cholerae* genetic diversity and the means by which it survives in the environment is critical to both defining parameters that allow

new pathogenic strains emerge and predicting the onset of conditions favoring *V. cholerae* proliferation, thereby preventing or limiting the extent of future pandemics.

Historical context. Diseases with symptoms similar to cholera have been described in written records dating from more than 2000 years ago, suggesting that cholera disease was present among the global population prior to the early 1800s, when the first pandemics were officially documented (12, 66). In 1849, during the second pandemic, Dr. John Snow hypothesized the waterborne nature of cholera disease transmission based upon the co-localization of cholera cases with particular public drinking water sources (12). The *Vibrio cholerae* bacterium was later discovered independently by both Filippo Pacini in Italy and the German bacteriologist Robert Koch. Since 1817, at least seven global pandemics have been recorded, and it is thought that an eighth pandemic is developing concurrently with the present seventh pandemic due to the emergence of a new *V. cholerae* serogroup (20).

***Vibrio cholerae* virulence.** *Vibrio cholerae* strains are frequently described in terms of their serogroups, which are defined by cell surface properties. Of the more than 200 serogroups designated so far, only O1 and O139 strains have proliferated into pandemic clones. The O1 serogroup, which has been responsible for the first seven pandemics, is divided into two major serotypes, Ogawa and Inaba, which are in turn divided into two biotypes, classical and El Tor, which are determined by biochemical properties and bacteriophage susceptibility (20). The classical biotype of the O1 serogroup was the predominant clinical strain until El Tor was identified in 1961, and since then El Tor strains have largely replaced classical strains throughout the world

(14, 20). In 1991, the O139 serogroup emerged as a second cholera-causing serogroup in the Bay of Bengal (and thus was given the synonym ‘Bengal’) from an ancestral O1 El Tor strain that underwent significant changes within the *wb** region in the genome that is responsible for encoding the O-antigen (25).

The pandemic O1 and O139 strains carry cholera toxin, a secreted two domain (AB₅) enterotoxin, encoded by the *ctxAB* genes within the CTX prophage (92). The B subunit of the toxin is a lectin that binds to the GM₁ ganglioside, a membrane glycolipid receptor present in lipid rafts on the cell membrane (54). The toxin enters the cell and proceeds via retrograde transport to the endoplasmic reticulum where the A₁ subunit undergoes retro-translocation into the cytosol (54). The A₁ subunit then ADP-ribosylates the G α heterotrimeric G protein, which activates adenyl cyclase, resulting in an increase in intracellular cAMP, which in turn misregulates the K⁺ and Cl⁻ ion channels. This causes osmosis of H₂O out of the epithelial cells lining the intestine, resulting in the loss of massive volumes of fluid referred to as ‘rice stool’, due to the sloughed epithelial cells and mucus that give it a white, flaky appearance. Concentrations of *V. cholerae* in this fluid can reach as high as 10⁸ cells ml⁻¹ (44). Cholera toxin may play an additional role during infection as an immunomodulatory agent, perhaps by inhibiting production of TNF α , a cytokine produced by macrophages that stimulates the immune response (26, 28).

Prior to secreting cholera toxin, *V. cholerae* must first colonize the mammalian gut. The primary virulence factor required for colonization is the toxin co-regulated type IV pilus (TCP). The TCP is encoded within the *Vibrio* pathogenicity island

(VPI), a genomic region that has some prophage-like characteristics (45). This pilus is coordinately regulated with CTX, which are both controlled by a regulatory cascade involving ToxR, ToxT and TcpP (53, 89). Virulence genes in *V. cholerae* are also regulated via quorum sensing (100), which consists of at least three converging pathways that sense concentrations of quorum molecules secreted by nearby bacteria (68).

Other virulence factors in pandemic O1, O139 *V. cholerae* strains may contribute to disease. In addition to CTX and TCP, other virulence factors are present in the *V. cholerae* genome. In order to develop an O1 El Tor strain safe to use as a vaccine, numerous factors aside from CTX must be deleted because they can cause residual reactogenicity in volunteers and vaccine recipients (55). In contrast, deletion of CTX from an O1 classical strain resulted in a safe, live attenuated vaccine strain (75) suggesting that these additional factors are not universally present or functional in all *V. cholerae* strains.

The *V. cholerae* RTX (repeats-in-toxin) toxin, a 4,546 amino acid protein encoded by the *rtxA* gene, causes cell rounding and covalent actin cross-linking in eukaryotic cells (27, 57). The RTX family of toxins consists of pore-forming toxins with a consensus Ca^{2+} -binding repeat sequence, of which the best-studied member is *E. coli* hemolysin (96). An RTX toxin gene has been implicated in environmental survival of *Legionella pneumophila* in amoeba (11). The RTX toxin gene is widespread among *V. cholerae* strains as determined by a PCR assay, but expression

of this toxin is prevented in O1 classical biotype strains due to a significant deletion within the *rtxACBD* operon (10, 57).

Pandemic *V. cholerae* strains also carry several hemolysins, the best studied of which is the HlyA or El Tor hemolysin, encoded by the *hlyA* gene. This hemolysin lyses erythrocytes and also exhibits enterotoxicity in rabbit ileal loop studies (38, 99). Other hemolysins carried by *V. cholerae* include lecithinase LEC (of the thermolabile hemolysin family), HLX, and Vc-delta thermostable hemolysin (reviewed in (99)). Interestingly, a number of other hemolysins and hemolysin-like genes are annotated in the *V. cholerae* O1 El Tor N16961 genome, but they may not be true hemolysins, and instead may play a role in biofilm formation (70). Furthermore, the contributions of hemolysins to virulence of *V. cholerae* in a mouse infection model are not clearly understood (18, 26).

The hemagglutinin/protease (HA/P), a Zn-metalloprotease (mucinase), is capable of a number of different functions, including the inactivation of CTX ϕ (51), but perhaps most importantly, it has been associated with the detachment of *V. cholerae* from cultured intestinal cells (2), and therefore may act as a ‘detachase’. However, its contribution to virulence in a mouse model is also not well understood (26).

Pandemic isolates of *Vibrio cholerae* can also be virulent towards non-mammalian hosts. A *Drosophila* model of *Vibrio cholerae* pathogenesis revealed that cholera toxin plays a role in virulence, but the mechanism of infection must be significantly different; in contrast to its virulent effects in mammalian systems,

administration of CTX alone has no effect on flies (5). Pandemic isolates also showed toxic effects towards grazing protozoa, although the cholera toxin appears to be uninvolved (63), and instead the effect seems to be mediated by a secreted protease that also causes toxicity towards *Caenorhabditis elegans* (91).

Virulence in non-O1, non-O139 *V. cholerae* isolates. Non-O1, non-O139 isolates from both environmental and clinical sources do not cause cholera, but they can be responsible for limited gastrointestinal disease outbreaks (71). Functional virulence genes are found in these strains, including HlyA, RTX toxin, and HA/P. Purified HA/P from a non-O1, non-O139 isolate causes enterotoxicity in *in vivo* models, such as the rabbit ileal loop assay (30). Furthermore, a type III secretion system (17), and a type VI secretion system (79) have recently been discovered in non-O1, non-O139 strains. A heat-stable enterotoxin in non-O1, non-O139 *V. cholerae* strains (NAG-ST), encoded by the *stn/sto* gene, has also been detected (33, 85). Genes encoding TCP and CTX are also found in non-O1, non-O139 strains of both clinical and environmental origin, perhaps indicating a source of genetic diversity among these strains (22, 31, 56, 74, 82, 84). The presence of additional virulence factors is also suggested by the observation that non-O1, non-O139 strains that do not carry CTX or TCP are capable of colonizing mice or causing fluid accumulation in the rabbit ileal loop assay, and virulence does not correlate with the presence or absence of certain virulence genes (22).

Genetic diversity of *Vibrio cholerae* of both environmental and clinical origin. *V. cholerae* is a member of the Vibrionaceae, a family of gamma-

proteobacteria. This species encompasses large variations in genome architecture, sequence types, and biological phenotypes. The extensive genomic variation among non-O1, non-O139 strains had been documented via several PCR methods, including amplified fragment length polymorphism (AFLP) (42, 43), Box elements PCR (BOX-PCR) and enterobacterial repetitive intergenic consensus sequence PCR (ERIC-PCR) (87). Furthermore, multi-locus enzyme electrophoresis (MLEE) (19) and then multi-locus sequence typing (MLST), documented significant variation among specific protein and gene sequences, respectively. In comparing phylogenies of O1/O139 and non-O1, non-O139 strains of both clinical and environmental origin, there appeared to be no correlation with the origin of the strain; that is, strains of clinical and environmental origin were not phylogenetically distinct (19, 87). Furthermore, non-O1, non-O139 strains can be very closely related to the pandemic O1/O139 strains suggesting that some cholera disease could be caused by the local emergence of a toxigenic strain (3, 43, 87). Constructing phylogenies with each gene sequence also revealed the prevalence of homologous recombination occurring both inter- and intra-genically (7, 52). Recombination is presumed to be responsible for the unexpected polymorphism in *tcpA* between VPI genes in strains from the sixth pandemic (Classical 395) and the seventh pandemic (El Tor N16961) (46). Extensive genomic variation in other *Vibrios* in environmental settings has been discovered, with individual genotypes present at very low abundance (90). The frequency of both homologous and illegitimate (non-homologous) recombination affects the maintenance or purging of sequence diversity, and therefore bacterial speciation, but

this is dependent upon both the strength of selection, and species-specific recombination mechanisms (78). The interplay of selective pressure and genome architecture remains at the heart of understanding the ecological significance of the observed extensive genomic diversity within the aquatic *Vibrios*.

Novel gene sequence can be introduced by several mechanisms, the most significant of which is an integron system (sometimes referred to as a ‘super-integron’) present on the second chromosome of *V. cholerae* (as well as other *Vibrios*) (64). Integrons consist of the *intI* gene, which encodes the IntI site-specific recombinase, an *attI* site where gene ‘cassettes’ are inserted, and Pc, the promoter that expresses them. Successive gene cassettes can be inserted into the *attI* site, such that tens or hundreds of genes can be brought into the chromosome. As a result, the second chromosome possesses many more unknown and hypothetical genes than the first chromosome (34). However, in one study of both toxigenic and non-toxigenic O1 and O139 *V. cholerae* strains, chromosome 2 was found to be quite stable outside of the superintegron region (76). DNA can be brought into the bacterial cell via transduction, conjugation, or transformation. *V. cholerae* can take part in conjugation; the 62-kB SXT transposon-like element that confers resistance to different antibiotics is transferred via conjugation (93). The *ctxAB* genes are carried on a prophage that can infect other *V. cholerae* strains carrying the TCP (92) and other prophages have been shown to transfer DNA between *V. cholerae* strains (see for example, (8)). Also, the capability to become competent to take up naked DNA in the presence of chitin

(65) may be a widespread phenomenon in the environment (69). This could have significant implications for the frequency of gene transfer among *V. cholerae* strains.

Via comparative genomic hybridization microarray analysis, genes missing in O1 El Tor and classical and O139 *V. cholerae* strains relative to the O1 El Tor strain N16961 were described, and as a group, they lacked only ~1% of N16961 genes (16). In contrast, in a study of environmental strains from southern California, roughly ~81% of N16961 genes were present in all of the other environmental strains (69). Of the genes missing from the environmental strains, ~4% were missing from all strains, 29% of which were virulence-related genes (69). Another comparative genomic hybridization study used a set of O1 and O139 *Vibrio cholerae* strains that had been isolated in China over a period of decades, most of which were of clinical origin, although a number of the strains lacked *ctxAB* or both *ctxAB* and *tcpA*. When these were compared to the N16961 genome on a microarray, ~76% of genes surveyed were present in all strains, suggesting that these genes represented the functional core of the *V. cholerae* genome (76). Interestingly, within the O1 or O139 serogroup strains, non-toxigenic strains exhibited greater sequence heterogeneity based upon analysis of fluorescence ratios (76).

***Vibrio cholerae* as an aquatic bacterium.** *V. cholerae* can be detected in coastal environments throughout the world as part of the aquatic microbial community (1, 4, 47-49, 60). In aquatic environments and laboratory microcosm experiments, its propensity to attach to surfaces has been documented (6, 9, 15, 37, 39-41, 86) although this phenomenon has not been universally observed (98) and colonization is often

much greater on surfaces of dead organisms than live organisms (73, 88). In the environment, it presumably participates in the formation of biofilms, which could serve a number of functions, including protection from stressful chemicals (94), or protozoan grazing (63). By forming aggregates, its ability to survive in the environment is enhanced, and each aggregate could constitute a dose adequate for infecting humans (21), although other factors such as O₂ limitation could affect the growth advantage. Whether the formation of aggregates explains the ‘hyperinfectious’ state of the *V. cholerae* bacterium following an initial passage through the human host (67) is unknown. In regions where the disease is endemic, it was found that filtering water for drinking through several layers of sari cloth (thus creating an effective pore size of ~20 µm) decreases cholera cases by ~50% (13). The ability of *V. cholerae* to survive in the planktonic form may depend on its ability to utilize available nutrients, perhaps in the form of dissolved organic carbon following blooms of dinoflagellates (72) in order to grow at rates exceeding the rate of grazing by protists (98).

In the aquatic environment, *V. cholerae* faces numerous dangers. *Vibrio cholerae* is subject to lysis by phage, and the prevalence of lytic phage and susceptible *Vibrio cholerae* strains has been shown to be inversely correlated (24), and this also correlates with cholera incidence as well (23). Different strains of *V. cholerae* appear to be differentially grazed by protists (63, 79, 83). Whether this differential grazing can be attributed to cell surface characteristics, as has been proposed for *Salmonella* (97), or whether specific virulence mechanisms are involved (79) has yet to be determined (83). Furthermore, *V. cholerae* is susceptible to secondary metabolites

secreted by other bacteria, suggesting that bacteria-bacteria interactions also affect *V. cholerae* proliferation and ecology (59).

Large-scale changes in climate and temperature also affect cholera incidence, and these observations may help to develop predictive models of cholera outbreaks (58, 77). Data from remote sensing satellites revealed that fluctuations in both sea surface temperature (which is related to phytoplankton concentrations) and sea surface height (a proxy for human-plankton contact) exhibited a bimodal pattern similar to that observed for cholera cases in Bangladesh. A time delay existed such that the outbreak periods of spring and fall were correlated with higher sea surface temperatures observed approximately one month to several months previously (58). A much longer time series of cholera cases and sea surface temperature anomalies in Bangladesh revealed interannual variability in cholera cases which suggested a link to the El Niño-Southern Oscillation (ENSO), again where higher temperatures correlated with increased cholera cases (77). Mechanistic explanations of these observations have not been provided, despite their importance to developing more accurate predictive models. However, it was recently suggested that an increase in sea surface height, which would transport marine waters into freshwater regions, might promote the dispersal of biofilms resulting in an increase in planktonic *V. cholerae*, and therefore affecting *V. cholerae* dynamics (50). Freshwater has a low concentration of Ca^{2+} , and Ca^{2+} was found to be required for biofilm formation for O139 strains in seawater (50).

There is a dearth of information regarding concentrations of *V. cholerae* in the environment. This is a challenging problem, because *V. cholerae* concentrations are low. A study performed in the Chesapeake Bay using fluorescence *in situ* hybridization (FISH) was successful in enumerating the *V. mimicus-V. cholerae* group (35), but the recent introduction of new technologies should spur the field to investigate this in more detail. For example, a new microarray method for detecting *in situ* levels of *Vibrios* in the environment was developed, and this method is sensitive enough to distinguish between the species (61). Also, progress towards developing PCR-based methods to detect *V. cholerae in situ* has been made but detection limits are still quite high (29, 32). This will help to better correlate actual *V. cholerae* concentrations with environmental parameters and cholera cases in areas where cholera disease is endemic.

Although progress has been made towards understanding *Vibrio cholerae* survival in the environment, a holistic model has not integrated information on multiple scales. Hypotheses developed in laboratory investigations need to be tested both in microcosm assays and in the environment. Also, genome-enabled studies are necessary in order to describe *V. cholerae* genome dynamics, adaptation, and recombination in the face of a multiplicity of selective pressures.

This thesis provides a first ‘glimpse’ into the *V. cholerae* pan-species genome and describes additional genes involved in eukaryotic interactions. A novel toxin is

characterized and shown to inhibit protein synthesis in eukaryotic cells. Finally, phylogenetic studies reveal the presence of three clades of this toxin gene in *V. cholerae* strains from around the world. Further work to characterize the role for this toxin in the ecology and virulence of the *V. cholerae* is needed in order to understand selective forces governing its evolution.

References

1. **Alam, M., N. A. Hasan, A. Sadique, N. A. Bhuiyan, K. U. Ahmed, S. Nusrin, G. B. Nair, A. K. Siddique, R. B. Sack, D. A. Sack, A. Huq, and R. R. Colwell.** 2006. Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl. Environ. Microbiol.* **72**:4096-4104.
2. **Benitez, J. A., A. J. Silva, and R. A. Finkelstein.** 2001. Environmental signals controlling production of hemagglutinin/protease in *Vibrio cholerae*. *Infect. Immun.* **69**:6549-6553.
3. **Bik, E. M., R. D. Gouw, and F. R. Mooi.** 1996. DNA fingerprinting of *Vibrio cholerae* strains with a novel insertion sequence element: a tool to identify epidemic strains. *J. Clin. Microbiol.* **34**:1453-1461.
4. **Binsztein, N., M. C. Costagliola, M. Pichel, V. Jurquiza, F. C. Ramirez, R. Akselman, M. Vacchino, A. Huq, and R. Colwell.** 2004. Viable but nonculturable *Vibrio cholerae* O1 in the aquatic environment of Argentina. *Appl. Environ. Microbiol.* **70**:7481-7486.
5. **Blow, N. S., R. N. Salomon, K. Garrity, I. Reveillaud, A. Kopin, F. R. Jackson, and P. I. Watnick.** 2005. *Vibrio cholerae* infection of *Drosophila melanogaster* mimics the human disease cholera. *PLoS Pathog.* **1**:e8.
6. **Broza, M., H. Gancz, M. Halpern, and Y. Kashi.** 2005. Adult non-biting midges: possible windborne carriers of *Vibrio cholerae* non-O1 non-O139. *Environ. Microbiol.* **7**:576-585.
7. **Byun, R., L. D. H. Elbourne, R. Lan, and P. R. Reeves.** 1999. Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infect. Immun.* **67**:1116-1124.
8. **Campos, J., E. Martinez, K. Marrero, Y. Silva, B. L. Rodriguez, E. Suzarte, T. Ledon, and R. Fando.** 2003. Novel type of specialized transduction for CTX-phi or its satellite phage RS1 mediated by filamentous phage VGJ-phi in *Vibrio cholerae*. *J. Bacteriol.* **185**:7231-7240.

9. **Chiavelli, D. A., J. W. Marsh, and R. K. Taylor.** 2001. The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl. Environ. Microbiol.* **67**:3220-3225.
10. **Chow, K. H., T. K. Ng, K. Y. Yuen, and W. C. Yam.** 2001. Detection of RTX toxin gene in *Vibrio cholerae* by PCR. *J. Clin. Microbiol.* **39**:2594-2597.
11. **Cirillo, S. L. G., L. Yan, M. Littman, M. M. Samrakandi, and J. D. Cirillo.** 2002. Role of the *Legionella pneumophila rtxA* gene in amoebae. *Microbiology* **148**:1667-1677.
12. **Colwell, R. R.** 1996. Global climate and infectious disease: the cholera paradigm. *Science* **274**:2025-2031.
13. **Colwell, R. R., A. Huq, M. S. Islam, K. M. A. Aziz, M. Yunus, N. H. Khan, A. Mahmud, R. B. Sack, G. B. Nair, J. Chakraborty, D. A. Sack, and E. Russek-Cohen.** 2003. Reduction of cholera in Bangladeshi villages by simple filtration. *Proc. Natl. Acad. Sci. U. S. A.* **100**:1051-1055.
14. **Cottingham, K. L., D. A. Chiavelli, and R. K. Taylor.** 2003. Environmental microbe and human pathogen: the ecology and microbiology of *Vibrio cholerae*. *Front. Ecol. Environ.* **1**:80-86.
15. **DePaula, A., M. W. Presnell, M. L. Motes, R. M. McPhearson, R. M. Twedt, R. E. Becker, and S. R. Zywno.** 1983. Non-O1 *Vibrio cholerae* in shellfish, sediment and waters of the U. S. Gulf Coast. *J. Food Prot.* **46**:802-806.
16. **Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos.** 2002. Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. U. S. A.* **99**:1556-1561.
17. **Dziejman, M., D. Serruto, V. C. Tam, D. Sturtevant, P. Diraphat, S. M. Faruque, M. H. Rahman, J. F. Heidelberg, J. Decker, L. Li, K. T. Montgomery, G. Grills, R. Kucherlapati, and J. J. Mekalanos.** 2005. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proc. Nat. Acad. Sci. U. S. A.* **102**:3465-3470.

18. **Fallarino, A., S. R. Attridge, P. A. Manning, and T. Focareta.** 2002. Cloning and characterization of a novel haemolysin in *Vibrio cholerae* O1 that does not directly contribute to the virulence of the organism. *Microbiology* **148**:2181-2189.
19. **Farfan, M., D. Minana, M. C. Fuste, and J. G. Loren.** 2000. Genetic relationships between clinical and environmental *Vibrio cholerae* isolates based on multilocus enzyme electrophoresis. *Microbiology* **146**:2613-2626.
20. **Faruque, S. M., M. J. Albert, and J. J. Mekalanos.** 1998. Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* **62**:1301-1314.
21. **Faruque, S. M., K. Biswas, S. M. N. Udden, Q. S. Ahmad, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2006. Transmissibility of cholera: *In vivo*-formed biofilms and their relationship to infectivity and persistence in the environment. *Proc. Natl. Acad. Sci. U. S. A.* **103**:6350-6355.
22. **Faruque, S. M., N. Chowdhury, M. Kamruzzaman, M. Dziejman, M. H. Rahman, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2004. Genetic diversity and virulence potential of environmental *Vibrio cholerae* population in a cholera-endemic area. *Proc. Natl. Acad. Sci. U. S. A.* **101**:2123-2128.
23. **Faruque, S. M., M. J. Islam, Q. S. Ahmad, A. S. G. Faruque, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2005. Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc. Natl. Acad. Sci. U. S. A.* **102**:6119-6124.
24. **Faruque, S. M., I. B. Naser, M. J. Islam, A. S. Faruque, A. N. Ghosh, G. B. Nair, D. A. Sack, and J. J. Mekalanos.** 2005. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc. Natl. Acad. Sci. U. S. A.* **102**:1702-1707.
25. **Faruque, S. M., D. A. Sack, R. B. Sack, R. R. Colwell, Y. Takeda, and G. B. Nair.** 2003. Emergence and evolution of *Vibrio cholerae* O139. *Proc. Natl. Acad. Sci. U. S. A.* **100**:1304-1309.

26. **Fullner, K. J., J. C. Boucher, M. A. Hanes, G. K. Haines, B. M. Meehan, C. Walchle, P. J. Sansonetti, and J. J. Mekalanos.** 2002. The contribution of accessory toxins of *Vibrio cholerae* O1 El Tor to the proinflammatory response in a murine pulmonary cholera model. *J. Exp. Med.* **195**:1455-1462.
27. **Fullner, K. J., and J. J. Mekalanos.** 2000. *In vivo* cross-linking of cellular actin by the *Vibrio cholerae* RTX toxin. *EMBO J.* **19**:5315-5323.
28. **Fullner Satchell, K. J.** 2003. Activation and suppression of the proinflammatory immune response by *Vibrio cholerae* toxins. *Microbes Infect.* **5**:1241-1247.
29. **Fykse, E. M., G. Skogan, W. Davies, J. S. Olsen, and J. M. Blatny.** 2007. Detection of *Vibrio cholerae* by real-time nucleic acid sequence-based amplification. *Appl. Environ. Microbiol.* **73**:1457-1466.
30. **Ghosh, A., D. R. Saha, K. M. Hoque, M. Asakuna, S. Yamasaki, H. Koley, S. S. Das, M. K. Chakrabarti, and A. Pal.** 2006. Enterotoxigenicity of mature 45-kilodalton and processed 35-kilodalton forms of hemagglutinin protease purified from a cholera toxin gene-negative *Vibrio cholerae* non-O1, non-O139 strain. *Infect. Immun.* **74**:2937-2946.
31. **Ghosh, C., R. K. Nandy, S. K. Dasgupta, G. B. Nair, R. H. Hall, and A. C. Ghose.** 1997. A search for cholera toxin (CT), toxin coregulated pilus (TCP), the regulatory element ToxR and other virulence factors in non-O1/non-O139 *Vibrio cholerae*. *Microb. Pathogen.* **22**:199-208.
32. **Gubala, A. J., and D. F. Proll.** 2006. Molecular-beacon multiplex real-time PCR assay for detection of *Vibrio cholerae*. *Appl. Environ. Microbiol.* **72**:6424-6428.
33. **Guglielmetti, P., L. Bravo, A. Zanchi, R. Monte, G. Lombardi, and G. M. Rossolini.** 1994. Detection of the *Vibrio cholerae* heat-stable enterotoxin gene by polymerase chain reaction. *Molecular and Cellular Probes* **8**:39-44.
34. **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**:477-483.
35. **Heidelberg, J. F., K. B. Heidelberg, and R. R. Colwell.** 2002. Bacteria of the gamma-subclass *Proteobacteria* associated with zooplankton in Chesapeake Bay. *Appl. Environ. Microbiol.* **68**:5498-5507.
 36. **Hill, D. R., L. Ford, and D. G. Lalloo.** 2006. Oral cholera vaccines: use in clinical practice. *Lancet Infect. Dis.* **6**:361-373.
 37. **Huq, A., E. B. Small, P. A. West, M. I. Huq, R. Rahman, and R. R. Colwell.** 1983. Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl. Environ. Microbiol.* **45**:275-283.
 38. **Ichinose, Y., K. Yamamoto, N. Nakasone, M. J. Tanabe, T. Takeda, T. Miwatani, and M. Iwanaga.** 1987. Enterotoxigenicity of El Tor-like hemolysin of non-O1 *Vibrio cholerae*. *Infect. Immun.* **55**:1090-1093.
 39. **Islam, M. S., B. S. Drasar, and D. J. Bradley.** 1990. Long-term persistence of toxigenic *Vibrio cholerae* O1 in the mucilaginous sheath of a blue-green alga, *Anabaena variabilis*. *J. Trop. Med. Hyg.* **93**:133-139.
 40. **Islam, M. S., B. S. Drasar, and D. J. Bradley.** 1989. Attachment of toxigenic *Vibrio cholerae* O1 to various freshwater plants and survival with a filamentous green alga, *Rhizoclonium fontanum*. *J. Trop. Med. Hyg.* **92**:396-401.
 41. **Islam, M. S., M. A. Miah, M. K. Hasan, R. B. Sack, and M. J. Albert.** 1994. Detection of non-culturable *Vibrio cholerae* O1 associated with a cyanobacterium from aquatic environments in Bangladesh. *Trans. Royal Soc. Trop. Med. Hyg.* **88**:298-299.
 42. **Jiang, S. C., V. Louis, N. Choopun, A. Sharma, A. Huq, and R. R. Colwell.** 2000. Genetic diversity of *Vibrio cholerae* in Chesapeake Bay determined by

- amplified fragment length polymorphism fingerprinting. Appl. Environ. Microbiol. **66**:140-147.
43. **Jiang, S. C., M. Matte, G. Matte, A. Huq, and R. R. Colwell.** 2000. Genetic diversity of clinical and environmental isolates of *Vibrio cholerae* determined by amplified fragment length polymorphism fingerprinting. Appl. Environ. Microbiol. **66**:148-153.
 44. **Kaper, J. B., J. Morris, J. G., and M. M. Levine.** 1995. Cholera. Clin. Microbiol. Rev. **8**:48-86.
 45. **Karaolis, D. K. R., 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**:3134-3139.
 46. **Karaolis, D. K. R., R. Lan, J. B. Kaper, and P. R. Reeves.** 2001. Comparison of *Vibrio cholerae* pathogenicity islands in sixth and seventh pandemic strains. Infect. Immun. **69**:1947-1952.
 47. **Kaysner, C. A., C. Abeyta Jr., M. M. Wekell, A. DePaula Jr., R. F. Stott, and J. M. Leitch.** 1987. Incidence of *Vibrio cholerae* from estuaries of the United States west coast. Appl. Environ. Microbiol. **53**:1344-1348.
 48. **Kenyon, J. E., D. C. Gillies, D. R. Piexoto, and B. Austin.** 1983. *Vibrio cholerae* (non-O1) isolated from California coastal waters. Appl. Environ. Microbiol. **46**:1232-1233.
 49. **Keymer, D. P., M. C. Miller, G. K. Schoolnik, and A. B. Boehm.** 2007. Genomic and phenotypic diversity of coastal *Vibrio cholerae* strains is linked to environmental factors. Appl. Environ. Microbiol. **73**:3705-3714.
 50. **Kierek, K., and P. I. Watnick.** 2003. The *Vibrio cholerae* O139 O-antigen polysaccharide is essential for Ca²⁺-dependent biofilm development in sea water. Proc. Natl. Acad. Sci. U. S. A. **100**:14357-14362.
 51. **Kimsey, H. H., and M. K. Waldor.** 1998. *Vibrio cholerae* hemagglutinin/protease inactivates CTXphi. Infect. Immun. **66**:4025-4029.

52. **Kotetishvili, M., O. C. Stine, Y. Chen, A. Kreger, A. Sulakvelidze, S. Sozhamannan, and J. G. Morris, Jr.** 2003. Multilocus sequence typing has better discriminatory ability for typing *Vibrio cholerae* than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness. *J. Clin. Microbiol.* **41**:2191-2196.
53. **Krukonis, E. S., and V. J. DiRita.** 2003. From motility to virulence: sensing and responding to environmental signals in *Vibrio cholerae*. *Curr. Opin. Microbiol.* **6**:186-190.
54. **Lencer, W. I., and B. Tsai.** 2003. The intracellular voyage of cholera toxin: going retro. *Trends Biochem Sci.* **28**:639-645.
55. **Levine, M. M., J. B. Kaper, D. Herrington, G. Losonsky, J. G. Morris, M. L. Clements, R. E. Black, B. Tall, and R. Hall.** 1988. Volunteer studies of deletion mutants of *Vibrio cholerae* O1 prepared by recombinant techniques. *Infect. Immun.* **56**:161-167.
56. **Li, M., M. Kotetishvili, Y. Chen, and S. Sozhamannan.** 2003. Comparative genomic analyses of the *Vibrio* pathogenicity island and cholera toxin prophage regions in nonepidemic serogroup strains of *Vibrio cholerae*. *Appl. Environ. Microbiol.* **69**:1728-1738.
57. **Lin, W., K. J. Fullner, R. Clayton, J. A. Sexton, M. B. Rogers, K. E. Calia, S. B. Calderwood, C. Fraser, and J. J. Mekalanos.** 1999. Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc. Natl. Acad. Sci. U. S. A.* **96**:1071-1076.
58. **Lobitz, B., L. Beck, A. Huq, B. Wood, G. Fuchs, A. S. Faruque, and R. Colwell.** 2000. Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc. Natl. Acad. Sci. U. S. A.* **97**:1438-1443.
59. **Long, R. A., D. C. Rowley, E. Zamora, J. Liu, D. H. Bartlett, and F. Azam.** 2005. Antagonistic interactions among marine bacteria impede the proliferation of *Vibrio cholerae*. *Appl. Environ. Microbiol.* **71**:8531-8536.

60. **Louis, V. R., E. Russek-Cohen, N. Choopun, I. N. G. Rivera, B. Gangle, S. C. Jiang, A. Rubin, J. A. Patz, A. Huq, and R. R. Colwell.** 2003. Predictability of *Vibrio cholerae* in Chesapeake Bay. *Appl. Environ. Microbiol.* **69**:2273-2285.
61. **Marcelino, L. A., V. Backman, A. Donaldson, C. Steadman, J. R. Thompson, S. P. Preheim, C. Lien, E. Lim, D. Veneziano, and M. F. Polz.** 2006. Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. *Proc. Natl. Acad. Sci. U. S. A.* **103**:13629-13634.
62. **Mathers, C. D., and D. Loncar.** 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine* **3**:e442.
63. **Matz, C., D. McDougald, A. M. Moreno, P. Y. Yung, F. H. Yildiz, and S. Kjelleberg.** 2005. Biofilm formation and phenotypic variation enhance predation-driven persistence of *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* **102**:16819-16824.
64. **Mazel, D., B. Dychinco, V. A. Webb, and J. Davies.** 1998. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**:605-608.
65. **Meibom, K. L., M. Blokesch, N. A. Dolganov, C.-Y. Wu, and G. K. Schoolnik.** 2005. Chitin induces natural competence in *Vibrio cholerae*. *Science* **310**:1824-1827.
66. **Mekalanos, J. J., E. J. Rubin, and M. K. Waldor.** 1997. Cholera: molecular basis for emergence and pathogenesis. *FEMS Immunol. & Microbiol.* **18**:241-248.
67. **Merrell, D. S., S. M. Butler, F. Qadri, N. A. Dolganov, A. Alam, M. B. Cohen, S. B. Calderwood, G. K. Schoolnik, and A. Camilli.** 2002. Host-induced epidemic spread of the cholera bacterium. *Nature* **417**:642-645.
68. **Miller, M. B., K. Skorupski, D. H. Lenz, R. K. Taylor, and B. L. Bassler.** 2002. Parallel quorum sensing systems converge to regulate virulence in *Vibrio cholerae*. *Cell* **110**:303-314.

69. **Miller, M. C., D. P. Keymer, A. Avelar, A. B. Boehm, and G. K. Schoolnik.** 2007. Detection and transformation of genome segments that differ within a coastal population of *Vibrio cholerae* strains. *Appl. Environ. Microbiol.* **73**:3695-3704.
70. **Moorthy, S., and P. I. Watnick.** 2005. Identification of novel stage-specific genetic requirements through whole genome transcription profiling of *Vibrio cholerae* biofilm development. *Mol. Microbiol.* **57**:1623-1635.
71. **Morris Jr., J. G., T. Takeda, B. D. Tall, G. A. Losonsky, S. K. Bhattacharya, B. D. Forrest, B. A. Kay, and M. Nishibuchi.** 1990. Experimental non-O group 1 *Vibrio cholerae* gastroenteritis in humans. *J. Clin. Invest.* **85**:697-705.
72. **Mourino-Perez, R. R., A. Z. Worden, and F. Azam.** 2003. Growth of *Vibrio cholerae* O1 in red tide waters off California. *Appl. Environ. Microbiol.* **69**:6923-6931.
73. **Mueller, R. S., D. McDougald, D. Cusumano, N. Sodhi, S. Kjelleberg, F. Azam, and D. H. Bartlett.** 2007. *Vibrio cholerae* strains possess multiple strategies for abiotic and biotic surface colonization. *J. Bacteriol.* **189**:5348-5360.
74. **Mukhopadhyay, A. K., S. Chakraborty, Y. Takeda, G. B. Nair, and D. E. Berg.** 2001. Characterization of VPI pathogenicity island and CTXphi prophage in environmental strains of *Vibrio cholerae*. *J. Bacteriol.* **183**:4737-4746.
75. **Olivier, V., N. H. Salzman, and K. J. F. Satchell.** 2007. Prolonged colonization of mice by *Vibrio cholerae* El Tor O1 depends on accessory toxins. *Infect. Immun.* **75**:5043-5051.
76. **Pang, B., M. Yan, Z. Cui, X. Ye, B. Diao, Y. Ren, S. Gao, L. Zhang, and B. Kan.** 2007. Genetic diversity of toxigenic and nontoxigenic *Vibrio cholerae* serogroups O1 and O139 revealed by array-based comparative genomic hybridization. *J. Bacteriol.* **189**:4837-4849.

77. **Pascual, M., X. Rodo, S. P. Ellner, R. R. Colwell, and M. J. Bouma.** 2000. Cholera dynamics and El Niño-Southern Oscillation. *Science* **289**:1766-1769.
78. **Polz, M. F., D. E. Hunt, S. P. Preheim, and D. M. Weinreich.** 2006. Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Phil. Trans. R. Soc. B.* **361**:2009-2021.
79. **Pukatski, S., A. T. Ma, D. Sturtevant, B. Krastins, D. Sarracino, W. C. Nelson, J. F. Heidelberg, and J. J. Mekalanos.** 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* model system. *Proc. Natl. Acad. Sci. U. S. A.* **103**:1528-1533.
80. **Ryan, E. T., and S. B. Calderwood.** 2000. Cholera vaccines. *Clin Infect Dis* **31**:561-5.
81. **Ryan, E. T., U. Dhar, W. A. Khan, M. A. Salam, A. S. Faruque, G. J. Fuchs, S. B. Calderwood, and M. L. Bennish.** 2000. Mortality, morbidity, and microbiology of endemic cholera among hospitalized patients in Dhaka, Bangladesh. *Am J Trop Med Hyg* **63**:12-20.
82. **Said, B., H. R. Smith, S. M. Scotland, and B. Rowe.** 1995. Detection and differentiation of the gene for toxin-coregulated pili (*tcpA*) in *Vibrio cholerae* non-O1 using the polymerase chain reaction. *FEMS Microbiol. Lett.* **125**:205-210.
83. **Saini, S.** 2007. Differential resistance of *Vibrio cholerae* strains to protozoan grazing. M.S. Thesis, University of California, San Diego.
84. **Sarkar, A., R. K. Nandy, G. B. Nair, and A. C. Ghose.** 2002. *Vibrio* pathogenicity island and cholera toxin genetic element-associated virulence genes and their expression in non-O1 non-O139 strains of *Vibrio cholerae*. *Infect. Immun.* **70**:4735-4742.
85. **Sarkar, B., T. Bhattacharya, T. Ramamurthy, T. Shimada, Y. Takeda, and G. B. Nair.** 2002. Preferential association of the heat-stable enterotoxin gene (*stn*) with environmental strains of *Vibrio cholerae* belonging to the O14 serogroup. *Epidemiol Infect.* **129**:245-251.

86. **Shukla, B. N., D. C. Singh, and S. C. Sanyal.** 1995. Attachment of non-culturable toxigenic *Vibrio cholerae* O1 and non-O1 and *Aeromonas* spp. to the aquatic arthropod *Gerris spinolae* and plants in the River Ganga, Varanasi. FEMS Immun. Med. Microbiol. **12**:113-120.
87. **Singh, D. V., M. H. Matte, G. R. Matte, S. Jiang, F. Sabeena, B. N. Shukla, S. C. Sanyal, A. Huq, and R. R. Colwell.** 2001. Molecular analysis of *Vibrio cholerae* O1, O139, non-O1, and non-O139 strains: Clonal relationships between clinical and environmental isolates. Appl Environ Microbiol. **67**:910-921.
88. **Tamplin, M. L., A. L. Gauzens, A. Huq, D. A. Sack, and R. R. Colwell.** 1990. Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. Appl. Environ. Microbiol. **56**:1977-1980.
89. **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**:2833-2837.
90. **Thompson, J. R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Distel, and M. F. Polz.** 2005. Genotypic diversity within a natural coastal bacterioplankton population. Science **307**:1311-1313.
91. **Vaitkevicius, K., B. Lindmark, G. Ou, T. Song, C. Toma, M. Iwanaga, J. Zhu, A. Andersson, M.-L. Hammarstrom, S. Tuck, and S. N. Wai.** 2006. A *Vibrio cholerae* protease needed for killing of *Caenorhabditis elegans* has a role in protection from natural predator grazing. Proc. Natl. Acad. Sci. U. S. A. **103**:9280-9285.
92. **Waldor, M. K., and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272**:1910-1914.
93. **Waldor, M. K., H. Tschape, and J. J. Mekalanos.** 1996. A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim and streptomycin in *Vibrio cholerae* O139. J. Bacteriol. **178**:4157-4165.

94. **Watnick, P., and R. Kolter.** 2000. Biofilm, city of microbes. *J. Bacteriol.* **182**:2675-2679.
95. **Weiss, R. A., and A. J. McMichael.** 2004. Social and environmental risk factors in the emergence of infectious diseases. *Nature Medicine* **10**:S70 - S76.
96. **Welch, R. A., C. Forestier, A. Lobo, S. Pellett, W. Thomas Jr., and G. Rowe.** 1992. The synthesis and function of the *Escherichia coli* hemolysin and related RTX exotoxins. *FEMS Microbiol. Immunol.* **105**:29-36.
97. **Wildschutte, H., D. M. Wolfe, A. Tamewitz, and J. G. Lawrence.** 2004. Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *Proc. Natl. Acad. Sci. U. S. A.* **101**:10644-10649.
98. **Worden, A. Z., M. Seidel, S. Smriga, A. Wick, F. Malfatti, D. Bartlett, and F. Azam.** 2006. Trophic regulation of *Vibrio cholerae* in coastal marine waters. *Environ. Microbiol.* **8**:21-29.
99. **Zhang, X.-H., and B. Austin.** 2005. Haemolysins in *Vibrio* species: a review. *J. Appl. Microbiol.* **98**:1011-1019.
100. **Zhu, J., M. B. Miller, R. E. Vance, M. Dziejman, B. L. Bassler, and J. J. Mekalanos.** 2002. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* **99**:3129-3134.

Chapter 2.

A glimpse into the expanded genome content of the *Vibrio cholerae* species through the identification of genes present in environmental strains

Abstract

Vibrio cholerae has multiple survival strategies which are reflected both in its broad distribution in many aquatic environments and its high genotypic diversity. In order to obtain additional information regarding the content of the *V. cholerae* species genome, suppression subtractive hybridization (SSH) was used to prepare libraries of DNA sequences from two southern California coastal isolates which are divergent or absent in the clinical strain, *Vibrio cholerae* O1 El Tor N16961. More than 1,400 subtracted clones were sequenced. This revealed the presence of novel sequences encoding functions related to cell surface structures, transport, metabolism, signal transduction, luminescence, mobile elements, stress resistance and virulence. Flanking sequence information was determined for loci of interest, and the distribution of these sequences was assessed for a collection of *V. cholerae* strains obtained from southern California and Mexican environments. This led to the surprising observation that sequences related to the toxin genes *toxA*, *cnfI* and *exoY* are widespread and more common in these strains than those of the cholera toxin genes which are a hallmark of the pandemic strains of *V. cholerae*. Gene transfer among these strains could be facilitated by a 4.9kbp plasmid discovered in one isolate, which possesses similarity to plasmids from other environmental Vibrios. By investigating some of the nucleotide

sequence basis for *Vibrio cholerae* genotypic diversity, DNA fragments have been uncovered which could promote survival in coastal environments. Furthermore, a set of genes has been described which could be involved in as yet undiscovered interactions between *Vibrio cholerae* and eukaryotic organisms.

Introduction

Although it is best known as the causative agent of the human disease cholera, *Vibrio cholerae* is also an autochthonous inhabitant of many aquatic environments, including estuarine and coastal waters (13). Indeed, the great majority of its more than 200 serogroups, excluding O1 and O139, are not associated with epidemic disease. *V. cholerae* has been isolated routinely from many aquatic environments throughout the world, often in association with plankton, plants, invertebrates and fish, and there are some reports of its presence in water birds, seals and diseased farm animals (2, 26, 27, 33, 43, 56, 66). The prevalence of *V. cholerae* in the environment is influenced by temperature and salinity (reviewed in (42)) as well concentrations of dissolved organic carbon (44). A number of genes have been previously implicated in environmental survival, and this has already led to a better understanding of the genetic basis for *Vibrio cholerae* adaptations. Genes have been uncovered which are important for biofilm formation (i.e., (68, 73)), zooplankton association (9), survival with filamentous blue cyanobacteria (28), and the degradation of non-biting midge (*Chironomos* sp.) egg masses (24). These genes clearly provide *Vibrio cholerae* with

mechanisms to avoid environmental stresses and obtain nutrients in aquatic environments.

The evolution of pandemic, virulent strains of *Vibrio cholerae* has required the acquisition of large islands of genes encoding pathogenicity functions, including the *Vibrio* pathogenicity island (32) and the cholera toxin phage genome (67). Additionally, human disease potential can be affected by other loci, including prophage K139, the superintegron, the *sxt* conjugative integrating element, O-antigen genes (reviewed in (18)), and cryptic plasmid pTLC (59). While horizontal gene transfer has played an important role in the emergence of *V. cholerae* strains with altered human virulence, little is known about genome alterations (i.e., insertions or deletions) which influence this species' environmental distribution. By exploring the gene content of multiple strains within the *Vibrio cholerae* species, it will be possible to discover new metabolic capabilities and survival strategies not represented in the single *V. cholerae* strain sequenced to date. To investigate this, we have undertaken a search for sequences present in two environmental *V. cholerae* strains which are lacking in the sequenced clinical O1 El Tor strain N16961 (25). The hypothesis underlying this effort is that additional genes encoding new survival strategies are likely to be found in non-pathogenic environmental *Vibrio cholerae* strains and the identification and characterization of these genes could provide useful information on factors that influence the environmental reservoir of *V. cholerae*. Consistent with this view, previous molecular fingerprinting studies have suggested a high level of genomic diversity among non-O1/non-O139 strains, suggesting that genomic

flexibility enables environmental survival in this species (see, for example, (4, 31, 75)).

A number of methods exist to characterize genetic differences among closely related bacteria. In the case of *V. cholerae*, representational difference analysis has previously been used by Calia *et al.* to uncover genes distinguishing pathogenic serotypes and biotypes (7) and this method was useful in the identification of a RTX toxin gene locus (41). In addition, comparative microarray analysis has revealed genes shared among pathogenic *V. cholerae* or unique to the El Tor biotype (15). Genetic differences among strains can also be characterized by suppression subtractive hybridization (SSH) (14), a technique that selectively amplifies DNA segments present in one bacterial genome (the “tester”), but not in another (the “driver”). These segments have arisen through one of two processes: (1) recent acquisition in the strain of interest relative to the driver strain, or (2) loss in the driver strain while being retained in the strain of interest. This and other subtractive hybridization approaches have been used to discover virulence loci and pathogenicity islands (see, for example, (53, 74)), and to explore genomic diversity in *Thermotoga maritima* (47). Here we report the use of SSH to isolate and subsequently characterize *V. cholerae* sequences present in two environmental isolates.

Materials and Methods

Strains. The *V. mimicus* and *V. cholerae* strains used in this study are listed in supplementary Table 2.S1. Briefly, 48 environmental *Vibrio cholerae* isolates from

Newport Beach, CA were kindly provided by Dr. Sunny Jiang, UC Irvine (30). These isolates were obtained over one year from several points along the San Diego Creek with different temperature and salinity profiles. Two environmental non-O1 *Vibrio cholerae* isolates from Mission Bay, San Diego, CA were kindly provided by Dr. Barbara Hemmingsen, San Diego State University. Clinical *Vibrio cholerae* strains O1 El Tor N16961 and O1 Classical O395 were kindly provided by Dr. John Mekalanos, Harvard University. An additional five *Vibrio cholerae* strains were provided by Dr. M. Lizarraga-Partida, Centro de Investigacion Cientifica y de Educacion Superior de Ensenada (CICESE), Mexico. *Vibrio mimicus* was obtained from the ATCC (ATCC 33653).

The two environmental isolates of *V. cholerae* that were the focus of this study were obtained on July 13, 2000, following an enrichment regimen similar to that which has been used by others (for example, (10, 30, 65)) coupled with a diagnostic polymerase chain reaction (PCR) screen. In particular, strain SIO was isolated from seawater collected off the Scripps Institution of Oceanography (SIO) pier during a bloom of the dinoflagellate *Lingulodinium polyedrum* and *V. cholerae* strain TP was obtained from plankton collected in the Torrey Pines Beach State Preserve estuary. Additional information concerning the isolation of these strains is presented in the supplementary data.

Additional characterization of strains. The *V. cholerae* isolates were examined to determine if they were members of the O1 serogroup using a direct fluorescent-antibody assay (Cholera DFA; New Horizons Diagnostics Corporation,

Columbia, Md.) and an Olympus BX51 fluorescence microscope. Substrate utilization preferences were scored using the Microstation ML3 system from Biolog (Hayward, CA).

Comparison of *Vibrio cholerae* strains SIO and TP by locus sequencing.

The *V. cholerae* strains were compared by sequencing rDNA (16S, 23S, and *Vibrio cholerae*-*Vibrio mimicus* specific 16S-23S intergenic spacer region (ISR) (12)) and protein markers (deoxyribodipyrimidine photolyase-1 (*phrB-1*), deoxyribodipyrimidine photolyase-2 (*phrB-2*), DNA helicase II (*uvrD*), DNA helicase IV (*helD*), integrase/recombinase (*xerC*) and site-specific recombinase (*int14*; Table 2.1 and supplementary Table 2.S2)). Primers and detailed methods used to amplify and sequence these genes are presented in the supplementary data. Furthermore, specific amplification of the *ompW* gene (46) and analysis of seven simple sequence repeat (SSR) loci was performed in order to further establish *Vibrio cholerae* identity (Y. Danin-Poleg, L. A. Cohen, H. Gancz, E. Malul, Y. Y. Broza, H. Goldschmidt, M. Halpern, M. Broza, and Y. Kashi, manuscript in preparation).

Subtractive hybridizations. Two subtracted libraries were constructed using the PCR-Select Bacterial Genome Subtraction Kit (Clontech, San Diego, CA). In both libraries, *V. cholerae* strain N16961 was used as the driver DNA in order to remove sequences shared between strain N16961 and the two environmental *V. cholerae* strains SIO and TP. The manufacturer's protocol was followed except that the final products were cloned into pCR4.0-Topo (Invitrogen, Carlsbad, CA). Clones were

picked into 96-well microtiter plates and sequenced using M13F and M13R primers at the San Diego State University Microchemical Core Facility.

Computer analyses. The ABI traces were analyzed using the PHRED/PHRAP package (<http://www.phrap.org/>) to remove vector sequence, mask poor quality sequence as well as sequence that was 85% identical to N16961 sequence over 20 bp or more. The sequence was masked and trimmed based on BLASTN searches using custom PERL scripts (available at <http://bartlettlab.ucsd.edu>). Masked sequence present at the ends of fragments was removed. SIO- and TP-specific masked and trimmed sequences were compared against the GenBank non-redundant database using BLASTX. The automated comparisons were performed using the BLASTCL3 client available from NCBI (<http://www.ncbi.nlm.nih.gov/>). The data were parsed using scripts available at <http://bartlettlab.ucsd.edu> to provide a best hit to *Vibrio* sequence, and a best hit to the database. The BLASTX results of matches to the entire database were used as the basis of the annotation described herein.

Contig assemblies of libraries of unique sequences in SIO and TP were performed using DNASTAR (Madison, WI). Contigs were generated using sequences that underwent more stringent screening for adaptor sequences and other short repeated sequences that are artifacts of the SSH procedure. Contig assembly was dependent upon finding a match of 20 bp with 90% identity, and all other parameters of the DNASTAR program were maintained on default settings. The resulting contig spectra were used as the basis to approximate the completeness of the libraries and compute novel genome sizes for SIO and TP. The single sequences used to generate the

contigs, as well as the assembled contig sequences, are available at <http://bartlettlab.ucsd.edu>.

In order to describe the DNA segments shared between SIO and TP, the sets of sequences used for contig assembly (as described above) were compared to each other using the TBLASTX algorithm, in two separate analyses (SIO vs. TP and TP vs. SIO). Sequences with an e-value less than 0.001 were considered to be significantly similar to one another. The categories of similar sequences were determined by matching each sequence to its top BlastX match, and comparing the results between SIO and TP. These sequences were not masked for N16961-like sequence (as described above), therefore DNA segments shared between SIO, TP, and N16961 were found from this analysis. The proportion of similar sequences found to be unique to and shared between SIO and TP was determined by manually filtering out sequences with top BLASTX hits to the N16961 genome.

Panhandle PCR and sequencing of selected genes. Panhandle PCR was performed in order to obtain additional genomic sequence data upstream and downstream of short reads of genes of interest in the SIO and TP environmental strains (16, 64). Detailed methods are presented in the supplementary data.

Colony hybridization. In order to analyze the distribution of genes among other environmental strains of *Vibrio cholerae*, we obtained strains isolated from southern California and Mexico, from a number of laboratories (see supplementary Table 2.S1). These strains, in addition to SIO, TP, and N16961, were streaked and patched onto L-agar plates and incubated overnight at 30°C. Colonies were blotted

onto Immobilon-NY+ 82mm nylon filters (Millipore, Bedford, MA) according to Sambrook *et al.* (61). DNA was immobilized on the filters by UV crosslinking (Stratalinker) and/or baking.

Hybridization probes were synthesized by PCR and purified with the UltraClean PCR Clean-up DNA Purification Kit (Mo Bio Laboratories, Solana Beach, CA). Primers were designed within or including regions of BlastX similarity to the genes of interest, with the following product sizes: *toxA* (1084 bp), *cnf* (704 bp), *cya* (427 bp), *luxA* (1194 bp), *rnr* (1258 bp), *rtx-ox* (324 bp). Probes to the SIO plasmid are as follows: pSIO1-1 includes *orf3*, with similarity to *virD4* genes, (823 bp); pSIO1-2 includes *orf2*, with similarity to a gene from the plasmid pOM1 from *Francisella tularensis* (1292 bp); pSIO1-3 includes a region similar to a 135 bp region from a plasmid from *Vibrio shiloi* (1042 bp); pSIO-4 includes *orf1* (1073 bp). Probe labeling, hybridization, washing, and detection was performed with the AlkPhosDirect labeling kit and ECF detection substrate (Amersham) according to the manufacturer's instructions, with modifications detailed in the supplementary data.

***Vibrio cholerae* strain SIO plasmid cloning and sequencing.** Plasmid pSIO1 was isolated using a QIAprep Spin MiniPrep Kit (Qiagen). Restriction digestion with *KpnI* resulted in a linearized fragment, which was cloned into the *KpnI* site of the pUC18 vector and sequenced using primers derived from pUC18 sequence or from prior rounds of pSIO1 sequencing.

Bioluminescence assay. *Vibrio cholerae* isolates demonstrating positive hybridization with the *luxA* probe were assayed for bioluminescence using a

luminometer. Individual strains were grown for 14 hours in LB at 30°C.

Measurements were made on 1 mL of culture transferred to a scintillation vial using an ATP Photometer Model 2000 (SAI Technology Company).

Results and Discussion

Isolation and characterization of strains SIO and TP. Two environmental strains of *V. cholerae* were isolated from a coastal and an estuary environment near Scripps Institution of Oceanography (SIO) following established enrichment techniques coupled with a PCR screen (see Materials and Methods). The strain obtained off of the SIO pier was designated SIO, and the strain recovered from the nearby Torrey Pines Beach State Preserve was designated TP. In order to confirm the identity of these isolates as *V. cholerae*, sequence information was obtained and analyzed for 16S rDNA, 23S rDNA, intergenic spacer region (ISR), *phrB-1*, *phrB-2*, *uvrD*, *helD*, *xerC* and *intl4*. The results clearly demonstrated that SIO and TP are closely related to *V. cholerae* (Table 2.1). Phylogenetic trees based upon 16S rDNA and *xerC* sequences placed both SIO and TP within the *V. cholerae-V. mimicus* cluster (Figure 2.1). Further comparisons of specific ISR regions from SIO, TP, assorted *Vibrio cholerae* strains, and *Vibrio mimicus* were performed in an attempt to assign SIO and TP to one or the other of these closely related species (12). Neighbor joining phylogenetic trees deduced from alignments of ISR type 1 and 2 sequences always produced distinct *Vibrio cholerae* and *Vibrio mimicus* clusters. SIO and TP never grouped within the *Vibrio mimicus* cluster but in some percentage of cases, SIO also

did not group with *Vibrio cholerae* (data not shown). Further identification of SIO and TP was accomplished by specifically amplifying the *ompW* gene using primers documented to yield a PCR product in *Vibrio cholerae* but not in *Vibrio mimicus* (46). This gene was amplified in both SIO and TP, thus confirming that both isolates are *Vibrio cholerae* strains (Y. Danin-Poleg and Y. Kashi, unpublished data). Simple sequence repeat (SSR) analysis demonstrated that SIO and TP are not related to strains of the O1 or O139 serogroups (Y. Danin-Poleg and Y. Kashi, unpublished data). These results are consistent with the direct fluorescence antibody assay and phylogenetic placement obtained using the BIOLOG GNTM substrate utilization assay. In light of these results both SIO and TP were assigned as non-O1/non-O139 strains of *Vibrio cholerae*.

Identification of genes unique to strains SIO and TP. SSH was used to isolate loci present in the environmental strains (tester DNA sources) that were missing in strain N16961 (driver DNA source). The complete genome sequence of clinical isolate N16961 is available (25), which makes it a useful reference strain for comparative analyses. 892 and 521 subtracted clones were obtained and sequenced from the SIO and TP libraries, respectively. These single reads were trimmed of vector sequence and then masked for N16961-like sequence greater than 85% identical over more than 20 bp, and compared against the non-redundant database using BLASTX (3). It was also possible to prepare contigs from some of the single reads. This

resulted in 114 SIO and 54 TP contigs with an average size of 472 bp (SIO) and 508 bp (TP). These sequences are available at <http://bartlettlab.ucsd.edu>.

When the masked SIO SSH single read fragments were compared to the non-redundant database using the BlastX algorithm, about 45% returned no match. After removing fragments with insignificant (e-value greater than 0.001) matches, and those which duplicated previous matches, roughly 24% of the total number of fragments originally obtained remained for further analysis. Of these, 25% were most similar to *V. cholerae* N16961 proteins, and 35% were similar to proteins found in other *Vibrio* species. Of the remainder, only two fragments were similar to *Archaea* and *Eukarya*, and the rest were similar to sequences in other *Bacteria*. However, roughly one third of the sequences with a significant match in the database were most similar to hypothetical proteins in the *Bacteria* (Table 2.2). Performing the same analysis with the SSH fragments obtained from the TP strain showed that roughly 57% of the masked sequences returned no match to the non-redundant database. After removing fragments with insignificant matches, and those which duplicated previous matches, there remained about 20% of the total number of fragments. Of these, only 7% were similar to N16961 proteins, 45% were similar to proteins found in other *Vibrio* species, and 46% were similar to other *Bacteria*. Only 1 match was to a protein found in the *Eukarya*, and there were no matches to archaeal proteins. Of the total number of sequences with significant matches to the database, roughly 20% were to hypothetical proteins (Table 2.2).

***Vibrio cholerae* genomic diversity.** Although SSH is not a quantitative procedure, it was possible to estimate the amount of novel sequence present in the genomes of SIO and TP. Calculating the total number of unique bp in the consensus sequences of the contigs and the remaining single reads provides one such estimate. After removing N16961-like sequence also recovered by SSH, there could be 0.20 Mb of novel sequence in SIO and 0.14 Mb in TP. In order to account for the incomplete recovery of DNA segments by SSH, a Chao2 estimator (8), typically applied to estimate species richness in ecological studies, was also used, indicating 0.8 Mb and 0.7 Mb of novel sequence in SIO and TP, respectively. One issue with these estimates is that the SSH library was not constructed from randomly-generated DNA segments, since a restriction enzyme, *RsaI*, was used in the first step to generate fragments of DNA. Thus, some sequences are more likely to be amplified in this procedure simply due to the proximity of *RsaI* restriction sites. With this caveat in mind it can be concluded that the amount of novel sequence is in the range of 3 - 20% of the N16961 genome size for these bacteria.

In order to determine the set of DNA segments shared between SIO and TP but absent from N16961, the SSH libraries from SIO and TP were compared to one another using the TBLASTX algorithm, after removing adaptor sequences but prior to masking for N16961-like sequence. A significant proportion of the shared segments were most similar to N16961 sequence, and were not considered further. Of the 214 types of unique gene segments found in SIO (Table 2.2), only 41 were shared with TP. This result suggests that a minority of gene fragments are shared, and each strain has

retained a significant number of unique genes. There was no tendency for one category of metabolic function to be represented more than any other. Preliminary microarray analyses comparing the number of SIO and TP genes shared with N16961 indicate that SIO and TP strains are missing 14% and 9% of N16961 genes, respectively (M. Miller and G. Schoolnik, unpublished data). Roughly 6% of N16961 genes are missing in SIO but are present in TP, while only 1% of N16961 genes are missing in TP but are present in SIO. Approximately 8% of N16961 genes are missing in both SIO and TP. In contrast, clinical, pandemic, pre-pandemic, and environmental strains of the O1 or O139 serogroups or the El Tor or Classical biotypes lacked only ~1% of N16961 genes (15).

Other studies have estimated the genomic variation among strains of the same species, and most have found similar percentages of strain-specific genes to the values reported above for SIO, TP, and N16961. The percentages can vary from less than 4% in *Pseudomonas aeruginosa* PAO1 when compared to clinical and environmental strains (70), to 2.6% - 16.3% among *Campylobacter jejuni* strains isolated from different sources (51), to 12-18% in *Helicobacter pylori* (60) and up to 20% in *Thermotoga maritima* (47). In general, genome variation within a single species can be as high as 26%, with many isolates showing variation between 5% and 15% (summarized in (47)).

Categories of genes present in SIO and TP, but not in N16961. The DNA segments recovered from the SSH were compared to the non-redundant database using

the BLASTX algorithm, and they were placed in 12 categories based upon the cellular function of the most similar sequence in the database (Table 2.2). Similar categories of novel sequences were present within the two environmental *V. cholerae* strains. In both cases a large percentage of the genes were associated with metabolism, mobile elements, restriction-modification systems, and the cell surface. However, strain SIO possessed a larger fraction of sequences related to transporters and stress resistance genes, whereas TP had more unique genes related to intracellular pathways and DNA mobilization. These categories of genes are typical of those that appear to vary between closely related strains of the same species in other bacteria. For example, extensive differences in O-antigen genes have previously been detected within strains of a number of species (55, 74) including within other *V. cholerae* strains (see for example, (39)). The *V. cholerae* examined in this study shared with *Aeromonas hydrophila* and *Helicobacter pylori* differences among related strains in restriction/modification systems (1, 74), with *Thermotoga maritima* in sugar transport and utilization genes (47), and with *E. coli* in phage sequences (52). Specific findings, highlighting DNA segments of particular interest, are presented below.

LPS structure. Several SIO segments were similar to genes within the lipopolysaccharide core oligosaccharide biosynthesis gene cluster (*wav* gene cluster) of two O6 serogroup strains (48). The TP strain had a sequence related to a O-antigen ligase present in a non-O1/non-O139 strain (48). Several other segments in TP were most similar to unknown or hypothetical proteins encoded within the O-antigen synthesis cluster of O139 and O22 serogroups (72).

Nutrient uptake. SIO and TP both had genes that appeared to be indicative of iron uptake, including a segment similar to a heme receptor present in *Vibrio mimicus*. SIO was also found to possess many sequences related to carbohydrate utilization including a starch utilization operon in *Klebsilla oxytoca* (19), and glycosidases in *Vibrio vulnificus*.

DNA transfer. Top BlastX matches to genes indicative of DNA transfer and rearrangement included transposases present in the genomes of other gram-negative bacteria. In the SIO strain, there were segments similar to cassettes associated with the super integrons in *Vibrionaceae* (58). In TP, a segment significantly similar to the retroelement EC67 originally described in clinical strains of *E. coli* was found, although retrons have been identified in *Vibrio cholerae* previously (63).

Plasmid pSIO1. The presence of a number of SSH sequences showing similarity to plasmid genes prompted further study of the possibility of extrachromosomal DNA in the environmental *V. cholerae* strains. High copy plasmid preparations from strain TP did not yield any DNA, but those from strain SIO resulted in the isolation of a 4.9 kbp element which was designated pSIO1. The complete nucleotide sequence of pSIO1 was obtained, revealing three open reading frames of approximately 300, 900, and 1800 bp in length. Intriguingly, ORF2 and ORF3 were similar to genes involved in DNA transfer, and there is a region very similar to a sequence present in plasmids from *Vibrio shiloi* and another environmental *Vibrio cholerae* isolate. These results are summarized in Figure 2.2 and Table 2.3.

Additional characterization of selected genes. Surprisingly, several DNA fragments were recovered from strains SIO and TP which had strong similarity to genes encoding virulence toxins in the bacterial pathogens *Pseudomonas aeruginosa*, uropathogenic *E. coli*, and others. Since these sequences had not previously been observed in *Vibrio cholerae*, additional sequence information surrounding these fragments was obtained in order to ascertain the following: (1) whether each of these sequences were part of uninterrupted open reading frames, (2) whether the genes were present in a single island of recently transferred DNA rather than scattered throughout the genome, (3) the GC content of these genes as a means of estimating whether they have been recently transferred, and (4) the conservation of amino acid residues in the encoded protein required for its putative enzymatic function. Finally, DNA fragments related to genes involved in bioluminescence were recovered, and as very few *Vibrio cholerae* strains are bioluminescent (54) additional sequence information was obtained. In all cases additional information was obtained by panhandle PCR. The results are summarized in Table 2.3.

Exotoxin A. A total of 3.7 kbp of sequence was obtained surrounding an ORF of 1968 bp encoding the putative *toxA* gene from TP. It is located between sequences homologous to VC1644 and VC1645 on chromosome I of the *V. cholerae* N16961 genome. *Pseudomonas aeruginosa toxA* encodes exotoxin A which is a member of a family of ADP-ribosylating enzymes, including cholera, diphtheria and pertussis toxins, *E. coli* heat-labile enterotoxin, and *Salmonella enterica* SpvB (34, 38, 49). The TP-deduced enzyme sequence includes most of the conserved residues involved with

receptor binding (with a number of insertions), and all of the residues important for disulfide bond formation, localization to the endoplasmic reticulum, as well as a glutamic acid residue within the catalytic domain which is absolutely conserved among members of this enzyme family (37, 69).

Cytotoxic Necrotizing Factor. 4.4 kbp of TP sequence was obtained surrounding a 2.5 kbp open reading frame encoding a protein similar to cytotoxic necrotizing factor (CNF) type-1 and type-2 found in *E. coli* (6) and to the putative *cnf* gene of *Vibrio fischeri* (40). In *E. coli*, CNF1 is encoded within a pathogenicity island and is associated with extraintestinal infections (5). The toxin causes reorganization of actin microfilaments, formation of stress fibers and membrane ruffling in eukaryotic cells (20, 62). The TP protein sequence shows strong conservation of the amino terminal receptor binding domain and membrane spanning domain, but the remainder of the protein, including the catalytic domain, has sustained significant deletions.

Adenylate cyclase. 4.4 kbp of TP DNA was obtained surrounding an ORF encoding a protein similar in its amino terminus to the deduced 126 Kd pathogenicity-associated protein encoded on the *V. cholerae* Vibrio Pathogenicity Island I and in its carboxy terminus to a secreted adenylate cyclase from *P. aeruginosa* (*exoY*, (71)), *Bacillus anthracis* (*cya*, (36)) and *Bordetella pertussis* (*cyaA*, (22)). The function of the *V. cholerae* 126Kd protein gene is unknown (32). Like other secreted adenylate cyclases, the TP protein lacks a signal peptide sequence and encodes all of the residues documented to be important for nucleotide binding (23, 71).

Repeat in structural toxin. 5.5 kbp of SIO DNA sequence was obtained as two contigs which included a region of BLASTX similarity to RTX (repeat in structural toxin) proteins present in *Aeromonas salmonicida* and *E. coli* with 55% identity over 251 amino acids along one portion of *Aeromonas* sequence. No RTX consensus could be found in the protein sequence, but repetitive sequence was evident: 18 amino acids were repeated five times within a total stretch of 160 amino acids. The RTX gene family comprises a group of diverse proteins present in a number of gram negative bacteria (35). These include some hemolysins, leukotoxins, and metalloproteases, which can insert themselves into target membranes and form pores (35). An RTX gene cluster has previously been found in *V. cholerae* O1 El Tor strains N16961 and E7946 to contribute to virulence, and to crosslink actin and depolymerize the actin cytoskeleton in a variety of mammalian cell lines (21, 41).

Ribonuclease R. 2.38 kbp was obtained from strain SIO which included a region with similarity to *rnr*, an exoribonuclease which is involved in the control of virulence gene expression in *Shigella flexneri* and enteroinvasive *E. coli* (11, 72).

LuxA. 2.0 kbp was obtained from TP which included ORFs related to the *luxA* and *luxD* products of *Vibrio harveyi*. The *luxA* and *luxD* genes encode the alpha subunit of the luciferase enzyme and an acyl transferase, or myristoyl-ACP-specific thioesterase, respectively. *luxA* sequences have previously been detected in other *V. cholerae* environmental strains (50, 54) and at least one bioluminescent *V. cholerae* strain has been described (50). TP and SIO were evaluated for bioluminescence using

a luminometer. Neither strain was visibly luminescent, but low-level light emission was detected from TP.

Distribution of selected genes among additional *V. cholerae* strains. The distribution of all of the above selected gene sequences, as well as segments of plasmid pSIO1, was determined by colony hybridization for more than fifty *V. cholerae* isolates. Most of these isolates were obtained by Jiang and Fu from coastal and river samples obtained in southern California near Newport Beach (30). Several additional San Diego County coastal isolates, several strains from Mexico, along with a classical (O395) and El Tor (N16961) O1 strain and a *Vibrio mimicus* control were also examined. The results are presented for each strain in Figure 2.3, and as a compilation of the results for each probe in Figure 2.4.

Most of the putative virulence genes identified in the SSH were present in many of the environmental isolates collected off the southern California coast. The *rnr* gene was widely distributed among the isolates, with greater than 80% yielding a positive signal. The putative *toxA* gene, encoding a protein similar to exotoxin A from *Pseudomonas aeruginosa*, was present in more than 60% of the isolates. Roughly half of the isolates contained the putative *cya* or *cnf* genes. In contrast, the RTX probe only hybridized to SIO, the strain from which it was derived, although previous results have shown that *V. cholerae rtxA* genes are widely distributed among O1, O139 and non-O1 environmental and clinical isolates (11). This suggests that the RTX-like region in SIO may be a narrowly-distributed sequence feature. The plasmid pSIO1

sequences were also present in only a few isolates. The *luxA* probe hybridized to 18% of isolates examined, and this hybridization frequency is similar to that previously reported by Palmer and Colwell (50). In their study, 15% of environmental and clinical *Vibrio cholerae* strains strongly hybridized to a *luxA* probe, and another 15% weakly hybridized to the probe. SIO and TP shared all putative virulence genes, except for the gene hybridizing to the RTX probe.

We hypothesize that the genes encoding proteins with significant similarity to toxins could contribute to a pathogenic phenotype with an as yet unknown host organism(s), since 1) the only known functions of these genes is virulence, 2) the sequences do not appear to be pseudogenes (i.e. containing frameshift mutations) and 3) they are widespread. By comparison, genes encoding the cholera toxin have been found in 17% of non-O1/non-O139 strains in southern California (29), 10% of non-O1/non-O139 strains from Brazil (57), and 21% of non-O1/non-O139 isolates from Calcutta, India (45). Thus, the intriguing possibility exists that the toxin-like genes uncovered here could actually be more widespread within environmental *V. cholerae* than are the *ctx* toxin genes. Furthermore, these genes do not appear to be recently horizontally transferred, based upon %GC content analysis, and they are not present in a single region in the chromosome (data not shown).

This observation may be especially timely, as environmental non-O1/non-O139 strains lacking *ctxA* and *tcpA* were recently shown to be virulent in mammalian models of *Vibrio cholerae* (17). Virulence was not necessarily correlated with the presence of known virulence factors, suggesting that environmental strains of *Vibrio*

cholerae may have additional undiscovered virulence genes. It remains to be determined whether the virulence-like gene sequences uncovered here contribute to mammalian virulence or virulence against aquatic organisms present in the environment from which strains TP and SIO were obtained.

In summary, the data presented here have provided information on the amount and the nature of DNA sequence differences existing between non-O1/non-O139 environmental *Vibrio cholerae* and the sequenced O1 El Tor strain. Much additional information is needed, including the nature of more subtle gene sequence variation among strains and genome variation in *V. cholerae* from other environments including those associated with cholera outbreaks. Furthermore, the effects of these chromosomal alterations on *V. cholerae* phenotypic diversity, including environmental fitness and biogeography, must be determined.

Acknowledgements

We are grateful to Dr. Sunny Jiang, Dr. Marcial Lizarraga-Partida, Dr. John Mekalanos, and Dr. Barbara Hemmingsen for generously providing *Vibrio cholerae* strains. We gratefully acknowledge Dr. Michael Miller and Dr. Gary Schoolnik for generously sharing unpublished results, and Dr. Yael Danin-Poleg and Dr. Yechezkel Kashi for performing the simple sequence repeat analysis, and sharing their technique with us prior to publication. We also thank Dr. Suhelen Egan for assistance with panhandle PCR, and Dr. Hideto Takami for assistance with plasmid annotation. Finally, we also thank three anonymous reviewers for their insightful comments.

This work was supported by grant AI46600-01 from the National Institute of Allergy and Infectious Disease to F.A. and D.H.B. A.P. is supported by a Howard Hughes Medical Institute Predoctoral Fellowship.

Chapter 2, in full, is a reprint of the material as it appears in Purdy, A., F. Rohwer, R. Edwards, F. Azam, and D. H. Bartlett. A glimpse into the expanded genome content of the *Vibrio cholerae* species through the identification of genes present in environmental strains. 2005. J. Bacteriol. 187: 2992-3001. The dissertation author was the primary investigator and author of this paper, with the co-authors listed above.

Table 2.1. Percent nucleotide sequence identity of phylogenetic marker genes to published N16961 sequences among *Vibrio cholerae* strains used in this study

	<i>Vibrio cholerae</i> strain TP	<i>Vibrio cholerae</i> strain SIO
Gene	Isolated from Torrey Pines Beach State Preserve estuary	Isolated from Scripps Inst. of Oceanogr. pier during dinoflagellate bloom
16S rDNA	99.9	99.5
23S rDNA	99	99
Intergenic spacer region type 2 ^a	99.0	99.0
<i>phrB-1</i>	96	96
<i>phrB-2</i>	98	97
<i>uvrD</i>	94	93
<i>helD</i>	99	95
<i>xerC</i>	98	97
<i>int14</i>	97	90

^a Intergenic spacer region (ISR) type 2 comparisons were made to the N16961 type 2 ISR amplified and sequenced for this study. All other comparisons were made to the sequenced N16961 genome.

Table 2.2. Functional categories of genes that are significantly similar to SIO and TP sequence fragments using the BlastX algorithm.

Strain	SIO		TP	
Functional categories of genes	Number of genes	Percent of total (SIO)	Number of genes	Percent of total (TP)
Hypothetical & unknown function	75	35.0%	20	21.1%
Intracellular functions	7	3.3%	10	10.5%
Bioluminescence	3		6	
Cell division	1		0	
Non-ribosomal peptide synthase	2		4	
Flagellar biosynthesis	1		0	
Mobile DNA	15	7.0%	10	10.5%
Phage and plasmid-encoded	6		5	
Super-integron-related elements	2		0	
Transposases and integrases	7		3	
Retroelement and addiction module	0		2	
DNA modification and structure	9	4.2%	5	5.3%
Restriction-modification systems	6		4	
Helicases, Dam methylase, binding protein	3		1	
Metabolism	24	11.2%	14	14.7%
Mannitol utilization	0		4	
Starch utilization	6		0	
DNA/RNA synthesis and processing	2		0	
Other anabolism/catabolism	16		10	
Transporters/Uptake/Secretion	20	9.3%	5	5.3%
ABC transporters	4		0	
PTS permease	1		1	
Secretion	1		1	
Iron uptake	2		1	
Other transporters	12		2	
EPS/OM/O-Antigen/Murein Biosynthesis	16	7.5%	11	11.6%
Extracellular polysaccharide synthesis/transport	3		1	
O-antigen and LPS synthesis	10		8	
Murein recycling	3		0	
Other	0		2	
Virulence	12	5.6%	5	5.3%
Toxins and hemolysins	9		3	
Virulence-associated factors	3		2	
Resistance to stress	14	6.5%	6	6.3%
Heat shock/protein folding	1		1	
Efflux of toxics out of cell	3		3	
Other	10		2	

Table 2.2. Functional categories of genes that are significantly similar to SIO and TP sequence fragments using the BlastX algorithm, continued.

Strain	SIO		TP	
Functional categories of genes	Number of genes	Percent of total (SIO)	Number of genes	Percent of total (TP)
Regulation and signal transduction	17	7.9%	8	8.4%
Two component regulatory systems	4		0	
LysR-like, GGDEF domain, and other regulators	10		7	
Chemotaxis	3		1	
Archaea	3	1.4%	1	1.1%
Eukarya	2	0.9%	0	0.0%
Total	214	100%	95	100%

Table 2.3. Top BlastX matches of genes of interest.

Putative gene name	Accession number	ORF length (aa)	Top matches after BLASTX search	E-value	Accession number of similar protein
<i>toxA</i>	AY876053	656	Exotoxin A precursor from <i>P. aeruginosa</i> PA01	2e-84	NP_249839.1
<i>cnf</i>	AY876054	846	Cytotoxic necrotizing factor type 1 from <i>V. fischeri</i>	1e-122	JE0362
			Cytotoxic necrotizing factor 1 from <i>E. coli</i>	3e-64	AF483829.1
<i>cya</i>	AY876055	1046	Probable inner membrane protein VC0822 from N16961	5e-42	D82276
			126K pathogenicity-associated protein from <i>V. cholerae</i>	5e-42	T09441
			Adenylate cyclase ExoY from <i>P. aeruginosa</i> PA01	1e-19	NP_250881.1
<i>rtx</i>	---	---	Unknown protein from <i>Synechocystis</i> sp. PCC 6803	2e-31	NP_442018.1
			Autotransporter adhesin from <i>V. vulnificus</i> CMCP6	1e-30	NP_761533.1
			RTX protein from <i>Aeromonas salmonicida</i>	3e-22	AF218037.1
<i>luxA</i>	AY876056	---	Bacterial luciferase alpha chain from <i>V. harveyi</i>	1e-149	P07740
<i>rnr</i>	---	---	Putative VacB/RNase R from <i>V. parahaemolyticus</i>	0.0	NP_798269.1
pSIO1 <i>orf1</i>	AY876057	104	No significant matches in database		
pSIO1 <i>orf2</i>	AY876057	312	Integrase (COG0582) from <i>Azotobacter vinelandii</i>	3e-72	ZP_00091084.1
			Orf3 on plasmid pOM1 from <i>Francisella tularensis</i>	3e-66	NP_052245.1
			Putative integrase from <i>Bordetella bronchiseptica</i>	7e-56	NP_887463.1
pSIO1 <i>orf3</i>	AY876057	595	VirD4-like protein encoded by gene on plasmid pMD136 from <i>Pediococcus pentosaceus</i>	1e-19	NP_037557.1
			TraG from <i>Pseudomonas aeruginosa</i>	2e-13	AAP22624

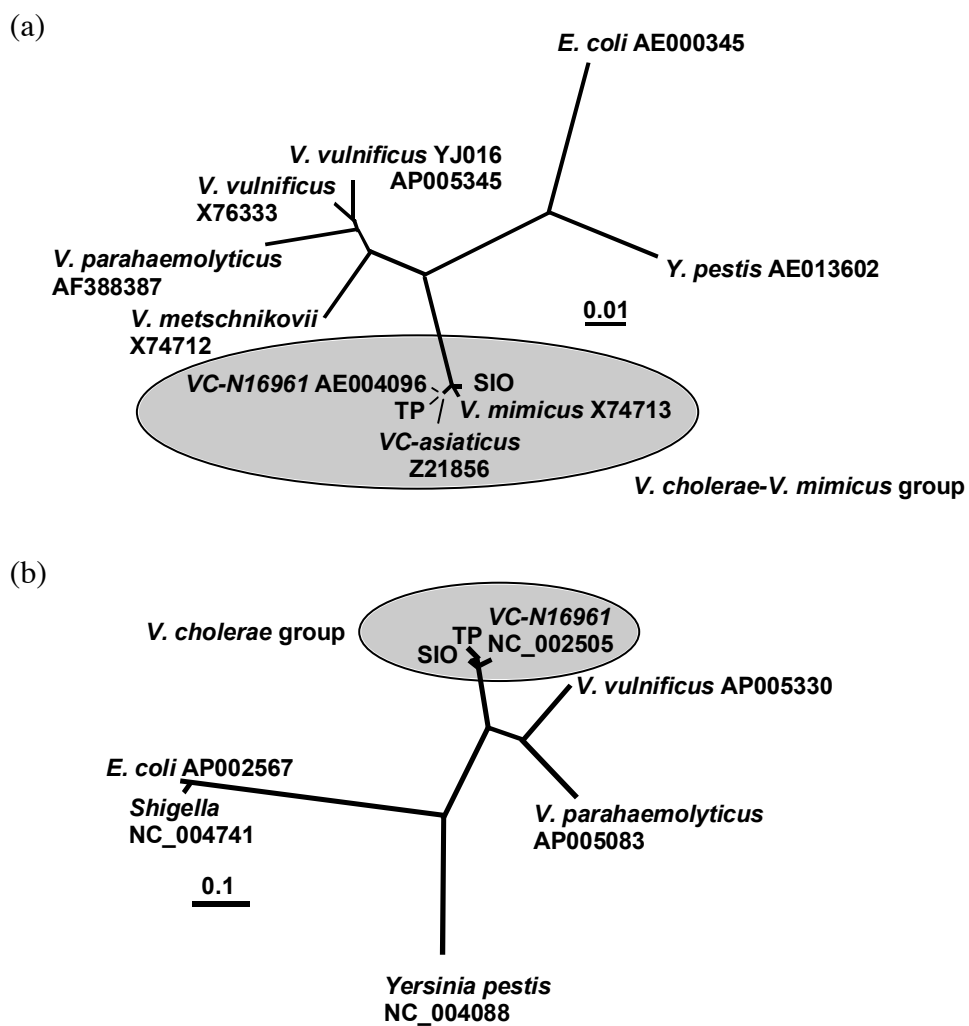


Figure 2.1. Phylogeny of *Vibrio cholerae* strains SIO and TP based upon (A) 16S rDNA sequence and (B) *xerC* sequence. Unrooted trees were generated by the neighbor-joining method. The scale bar indicates (A) 0.01 and (B) 0.1 substitutions per position.

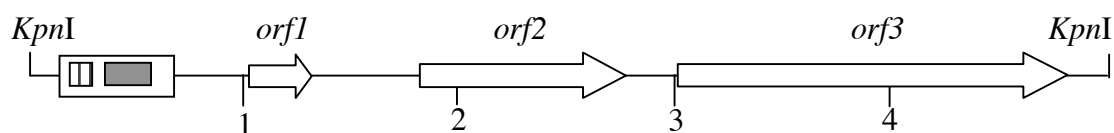


Figure 2.2. Structure of plasmid pSIO1 (4906 bp). White box indicates a 337 bp region 96% identical to a region in the cryptic plasmid pVC isolated from a *Vibrio cholerae* strain in Hong Kong. Shaded box indicates a 135 bp region 88% identical to similar region in the coral pathogen *Vibrio shiloi*. Striped box designates a 27 bp region identical to a portion of the p15A replicon. Size markings (kB) are indicated.

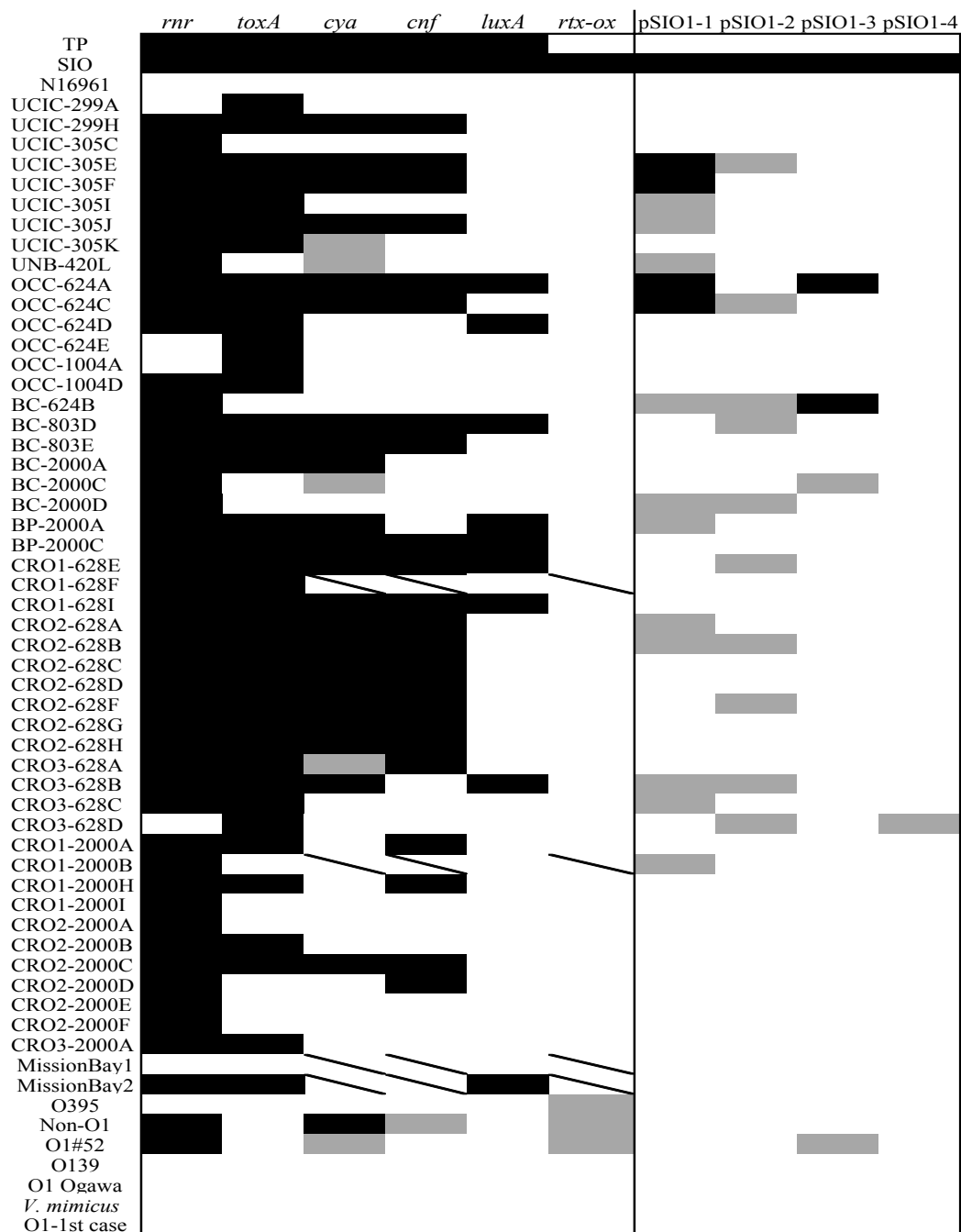


Figure 2.3. Distribution of selected environmental genes among individual *Vibrio* isolates. Colony hybridization results for a panel of environmental genes of interest (listed along the top) against a panel of *Vibrio* isolates obtained primarily from the southwestern United States (listed along the left side). Positive signal is black, no signal is white, and a faint positive signal is indicated by grey shading. A slash indicates that no data was obtained. All strains are *Vibrio cholerae*, unless otherwise designated. See Materials and Methods for additional descriptions of probes and strains used.

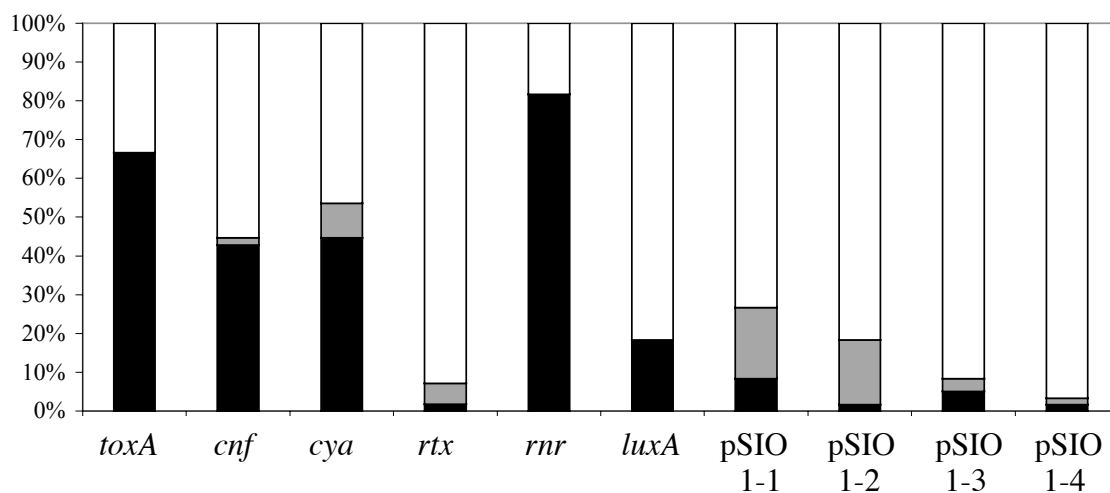


Figure 2.4. Percentage distribution of selected environmental genes among *Vibrio* strains. For the putative *cnf* and *toxA* genes, and the *rtx*-like region, a total of 55 *Vibrio cholerae* isolates and one *Vibrio mimicus* strain were included. For all other genes, a total of 59 *Vibrio cholerae* isolates and one *Vibrio mimicus* strain were included. Black shading indicates percentage of isolates which are hybridization positive for the gene of interest, grey shading indicates the percentage of isolates which exhibit weak hybridization to the gene probe.

References

1. **Akopyants, N. S., A. Fradkov, L. Diatchenko, J. E. Hill, P. D. Siebert, S. A. Lukyanov, E. D. Sverdlov, and D. E. Berg.** 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A.* **95**:13108-13113.
2. **Alexander, D., A. DePaola, S. Chirtel, and R. B. Young.** 1998. Detection of *Vibrio cholerae* in oyster (*Crassostrea virginica*) homogenate based on centrifugal removal of antimicrobial agents. *J. Microb. Methods* **33**:237-244.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
4. **Bik, E. M., R. D. Gouw, and F. R. Mooi.** 1996. DNA fingerprinting of *Vibrio cholerae* strains with a novel insertion sequence element - a tool to identify epidemic strains. *J. Clin. Microbiol.* **34**:1453-1461.
5. **Blum, G., V. Falbo, A. Caprioli, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin from the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. *FEMS Microbiol. Lett.* **126**:189-196.
6. **Boquet, P.** 2001. The cytotoxic necrotizing factor 1 (CNF1) from *Escherichia coli*. *Toxicon* **39**:1673-1680.
7. **Calia, K. E., M. K. Waldor, and S. B. Calderwood.** 1998. Use of representational difference analysis to identify genomic differences between pathogenic strains of *Vibrio cholerae*. *Infect. Immun.* **66**:849-852.
8. **Chao, A.** 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**:783 - 791.
9. **Chiavelli, D. A., J. W. Marsh, and R. K. Taylor.** 2001. The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl. Environ. Microbiol.* **67**:3220-3225.
10. **Choopun, N., V. Louis, A. Huq, and R. R. Colwell.** 2002. Simple procedure for rapid identification of *Vibrio cholerae* from the aquatic environment. *Appl. Environ. Microbiol.* **68**:995-998.
11. **Chow, K. H., T. K. Ng, K. Y. Yuen, and W. C. Yam.** 2001. Detection of RTX toxin gene in *Vibrio cholerae* by PCR. *J. Clin. Microbiol.* **39**:2594-2597.

12. **Chun, J., A. Huq, and R. R. Colwell.** 1999. Analysis of 16S-23S intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. Appl. Environ. Microbiol. **65**:2202-2208.
13. **Colwell, R. R.** 1996. Global climate and infectious disease: the cholera paradigm. Science **274**:2025-2031.
14. **Diatchenko, L., Y.-F. C. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov, and P. D. Siebert.** 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. U. S. A. **93**:6025-6030.
15. **Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos.** 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. Proc. Natl. Acad. Sci. U. S. A. **99**:1556-1561.
16. **Egan, S., S. James, C. Holmstrom, and S. Kjelleberg.** 2002. Correlation between pigmentation and antifouling compounds produced by *Pseudoalteromonas tunicata*. Environ. Microbiol. **4**:433-442.
17. **Faruque, S. M., N. Chowdhury, M. Kamruzzaman, M. Dziejman, M. H. Rahman, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2004. Genetic diversity and virulence potential of environmental *Vibrio cholerae* population in a cholera-endemic area. Proc. Nat. Acad. Sci. U. S. A. **101**:2123-2128.
18. **Faruque, S. M., and J. J. Mekalanos.** 2003. Pathogenicity islands and phages in *Vibrio cholerae* evolution. Trends in Microbiology **11**:505-510.
19. **Fiedler, G., M. Pajatsch, and A. Bock.** 1996. Genetics of a novel starch utilization pathway present in *Klebsiella oxytoca*. J. Mol. Biol. **256**:279-291.
20. **Flatau, G., E. Lemichez, M. Gauthier, P. Chardin, S. Paris, C. Fiorentini, and P. Boquet.** 1997. Toxin-induced activation of the G protein p21 Rho by deamidation of glutamine. Nature **387**:729-733.
21. **Fullner, K. J., and J. J. Mekalanos.** 2000. *In vivo* cross-linking of cellular actin by the *Vibrio cholerae* RTX toxin. EMBO J. **19**:5315-5323.
22. **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**:19-30.

23. **Glaser, P., H. Sakamoto, J. Bellalou, A. Ullman, and A. Danchin.** 1988. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. EMBO J. **7**:3997-4004.
24. **Halpern, M., H. Gancz, and Y. Kashi.** 2003. *Vibrio cholerae* hemagglutinin/protease degrades chironomid egg masses. Appl. Env. Microbiol. **69**:4200-4204.
25. **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**:477-483.
26. **Islam, M. S., B. S. Drasar, and D. J. Bradley.** 1990. Long-term persistence of toxigenic *Vibrio cholerae* O1 in the mucilaginous sheath of a blue-green alga, *Anabaena variabilis*. J. Trop. Med. Hygiene **93**:133-139.
27. **Islam, M. S., B. S. Drasar, and R. B. Sack.** 1993. The aquatic environment as a reservoir of *Vibrio cholerae*: a review. J. Diarrhoeal Dis. Res. **11**:197-206.
28. **Islam, M. S., M. M. Goldar, M. G. Morshed, M. N. H. Khan, M. R. Islam, and R. B. Sack.** 2002. Involvement of the *hap* gene (mucinase) in the survival of *Vibrio cholerae* O1 in association with the blue-green alga, *Anabaena* sp.. Can. J. Microbiol. **48**:793-800.
29. **Jiang, S., W. P. Chu, and W. X. Fu.** 2003. Prevalence of cholera toxin genes (*ctxA* and *zot*) among non-O1/O139 *Vibrio cholerae* strains from Newport Bay, California. Appl. Env. Microbiol. **69**:7541-7544.
30. **Jiang, S. C., and W. Fu.** 2001. Seasonal abundance and distribution of *Vibrio cholerae* in coastal waters quantified by a 16S-23S intergenic spacer probe. Microb. Ecol. **42**:540-548.
31. **Jiang, S. C., V. Louis, N. Choopun, A. Sharma, A. Huq, and R. R. Colwell.** 2000. Genetic diversity of *Vibrio cholerae* in Chesapeake Bay determined by amplified fragment length polymorphism fingerprinting. Appl. Environ. Microbiol. **66**:140-147.
32. **Karaolis, D. K. R., 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**:3134-3139.

33. **Kaysner, C. A., C. Abeyta Jr., M. M. Wekell, A. DePaula Jr., R. F. Stott, and J. M. Leitch.** 1987. Incidence of *Vibrio cholerae* from estuaries of the United States west coast. *Appl. Environ. Microbiol.* **53**:1344-1348.
34. **Krueger, K. M., and J. T. Barbieri.** 1995. The family of bacterial ADP-ribosylating exotoxins. *Clin. Microbiol. Rev.* **8**:34-47.
35. **Lally, E. T., R. B. Hill, I. R. Kieba, and J. Korostoff.** 1999. The interaction between RTX toxins and target cells. *Trends in Microbiology* **7**:356-361.
36. **Leppla, S. H.** 1982. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cAMP concentration in eukaryotic cells. *Proc. Natl. Acad. Sci. U. S. A.* **79**:6532-6536.
37. **Lesnick, M. L., and D. G. Guiney.** 2001. The best defense is a good offense - *Salmonella* deploys an ADP-ribosylating toxin. *Trends in Microbiology* **9**:2-5.
38. **Lesnick, M. L., N. E. Reiner, J. Fierer, and D. G. Guiney.** 2001. The *Salmonella* *spvB* virulence gene encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells. *Mol. Microbiol.* **39**:1464-1470.
39. **Li, M. R., T. Shimada, J. G. Morris, A. Sulakvelidze, and S. Sozhamannan.** 2002. Evidence for the emergence of non-O1 and non-O139 *Vibrio cholerae* strains with pathogenic potential by exchange of O-antigen biosynthesis regions. *Infect. Immun.* **70**:2441-2453.
40. **Lin, J.-W., L.-M. Chen, H.-Y. Chen, and S.-F. Weng.** 1998. Identification and analysis of the regulatory region R&R* with the *cnf1* gene encoding the cytotoxic necrotizing factor type 1 that closely links to the lux regulon of *Vibrio fischeri*. *Biochem. Biophys. Res. Commun.* **250**:462-465.
41. **Lin, W., K. J. Fullner, R. Clayton, J. A. Sexton, M. B. Rogers, K. E. Calia, S. B. Calderwood, C. Fraser, and J. J. Mekalanos.** 1999. Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc. Natl. Acad. Sci. U. S. A.* **96**:1071-1076.
42. **Lipp, E. K., A. Huq, and R. R. Colwell.** 2002. Effects of global climate on infectious disease: the cholera model. *Clin. Microbiol. Rev.* **15**:757-770.
43. **Lipp, E. K., I. N. G. Rivera, A. I. Gil, E. M. Espeland, N. Choopun, V. R. Louis, E. Russek-Cohen, A. Huq, and R. R. Colwell.** 2003. Direct detection of *Vibrio cholerae* and *ctxA* in Peruvian coastal water and plankton by PCR. *Appl. Environ. Microbiol.* **69**:3676-3680.

44. **Mourino-Perez, R. R., A. Z. Worden, and F. Azam.** 2003. Growth of *Vibrio cholerae* O1 in red tide waters off California. *Appl. Environ. Microbiol.* **69**:6923-6931.
45. **Mukhopadhyay, A. K., S. Chakraborty, Y. Takeda, G. B. Nair, and D. E. Berg.** 2001. Characterization of VPI pathogenicity island and CTXphi prophage in environmental strains of *Vibrio cholerae*. *J. Bacteriol.* **183**:4737-4746.
46. **Nandi, B., R. K. Nandy, S. Mukhopadhyay, G. B. Nair, T. Shimada, and A. C. Ghose.** 2000. Rapid method for species-specific identification of *Vibrio cholerae* using primers targeted to the gene of outer membrane protein OmpW. *J. Clin. Microbiol.* **38**:4145-4151.
47. **Nesbo, C. L., K. E. Nelson, and W. F. Doolittle.** 2002. Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*. *J. Bacteriol.* **184**:4475-4488.
48. **Nesper, J., A. Kraib, S. Schild, J. Blab, K. E. Klose, J. Bockemuhl, and J. Reidl.** 2002. Comparative and genetic analyses of the putative *Vibrio cholerae* lipopolysaccharide core oligosaccharide biosynthesis (*wav*) gene cluster. *Infect. Immun.* **70**:2419-2433.
49. **Otto, H., D. Tezcan-Merdol, R. Girisch, F. Haag, M. Rhen, and F. Koch-Nolte.** 2000. The *spvB* gene-product of the *Salmonella enterica* virulence plasmid is a mono(ADP-ribosyl)transferase. *Mol. Microbiol.* **37**:1106-1115.
50. **Palmer, L., and R. R. Colwell.** 1991. Detection of luciferase gene sequence in nonluminescent *Vibrio cholerae* by colony hybridization and polymerase chain reaction. *Appl. Environ. Microbiol.* **57**:1286-1293.
51. **Pearson, B. M., C. Pin, J. Wright, K. I'Anson, T. Humphrey, and J. M. Wells.** 2003. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Lett.* **554**:224-230.
52. **Pradel, N., S. Leroy-Setrin, B. Joly, and V. Livrelli.** 2002. Genomic subtraction to identify and characterize sequences of Shiga toxin-producing *Escherichia coli* O91 : H21. *Appl. Environ. Microbiol.* **68**:2316-2325.
53. **Radnedge, L., P. G. Agron, P. L. Worsham, and G. L. Andersen.** 2002. Genome plasticity in *Yersinia pestis*. *Microbiology* **148**:1687-1698.
54. **Ramaiah, N., J. Chun, J. Ravel, W. L. Straube, R. T. Hill, and R. R. Colwell.** 2000. Detection of luciferase gene sequences in nonluminescent bacteria from the Chesapeake Bay. *FEMS Microbiol. Ecol.* **33**:27-34.

55. **Reckseidler, S. L., D. DeShazer, P. A. Sokol, and D. E. Woods.** 2001. Detection of bacterial virulence genes by subtractive hybridization: Identification of capsular polysaccharide of *Burkholderia pseudomallei* as a major virulence determinant. *Infect. Immun.* **69**:34-44.
56. **Reddacliff, G. L., M. Hornitzky, J. Carson, R. Petersen, and R. Zelski.** 1993. Mortalities of goldfish, *Carassius-auratus* (L), associated with *Vibrio cholerae* (non01) infection. *J. Fish Dis.* **16**:517-520.
57. **Rivera, I. N. G., J. Chun, A. Huq, R. B. Sack, and R. R. Colwell.** 2001. Genotypes associated with virulence in environmental isolates of *Vibrio cholerae*. *Appl. Environ. Microbiol.* **67**:2421-2429.
58. **Rowe-Magnus, D. A., A.-M. Guerout, and D. Mazel.** 2003. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res.* **13**:428 - 442.
59. **Rubin, E. J., W. Lin, J. Mekalanos, and M. K. Waldor.** 1998. Replication and integration of a *Vibrio cholerae* cryptic plasmid linked to the CTX prophage. *Mol. Microbiol.* **28**:1247-1254.
60. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and F. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. U. S. A.* **97**:14668-14673.
61. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
62. **Schmidt, G., P. Sehr, M. Wilm, J. Selzer, M. Mann, and K. Aktories.** 1997. Gln 63 of Rho is deamidated by *Escherichia coli* cytotoxic necrotizing factor-1. *Nature* **387**:725-729.
63. **Shimamoto, T., M. Kobayashi, T. Tsuchiya, S. Shinoda, H. Kawakami, S. Inouye, and M. Inouye.** 1999. A retroelement in *Vibrio cholerae*. *Mol. Microbiol.* **34**:631-632.
64. **Siebert, P. D., A. Chenchik, D. E. Kellogg, K. A. Lukyanov, and S. A. Lukyanov.** 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* **23**:1087-1088.
65. **Staff, F. D. A.** 1998. FDA Bacteriological Analytical Manual, Eighth ed. A O A C International, Gaithersburg, MD.

66. **Visser, I. J. R., P. Vellema, H. van Dokkum, and T. Shimada.** 1999. Isolation of *Vibrio cholerae* from diseased farm animals and the surface water in the Netherlands. *Veterinary Record* **144**:451-452.
67. **Waldor, M. K., and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**:1910-1914.
68. **Watnick, P. I., and R. Kolter.** 1999. Steps in the development of a *Vibrio cholerae* El Tor biofilm. *Mol. Microbiol.* **34**:586-595.
69. **Wedekind, J. E., C. B. Trame, M. Dorywalska, P. Koehl, T. M. Raschke, M. McKee, D. FitzGerald, R. J. Collier, and D. B. McKay.** 2001. Refined crystallographic structure of *Pseudomonas aeruginosa* exotoxin A and its implications for the molecular mechanism of toxicity. *J. Mol. Biol.* **314**:823-837.
70. **Wolfgang, M. C., B. R. Kulasekara, X. Liang, D. Boyd, K. Wu, Q. Yang, C. G. Miyada, and S. Lory.** 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **100**:8484-8489.
71. **Yahr, T. L., A. J. Vallis, M. K. Hancock, J. T. Barbieri, and D. W. Frank.** 1998. ExoY, an adenylate cyclase secreted by the *Pseudomonas aeruginosa* type III system. *Proc. Natl. Acad. Sci. U. S. A.* **95**:13899-13904.
72. **Yamasaki, S., H. Shimizu, K. Hoshino, S.-T. Ho, T. Shimada, G. B. Nair, and Y. Takeda.** 1999. The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are closely related to those of *Vibrio cholerae* O22. *Gene* **237**:321-332.
73. **Yildiz, F. H., N. A. Dolganov, and G. K. Schoolnik.** 2001. VpsR, a member of the response regulators of the two-component regulatory systems, is required for expression of *vps* biosynthesis genes and EPS^{ETr}-associated phenotypes in *Vibrio cholerae* O1 El Tor. *J. Bacteriol.* **183**:1716-1726.
74. **Zhang, Y. L., C. T. Ong, and K. Y. Leung.** 2000. Molecular analysis of genetic differences between virulent and avirulent strains of *Aeromonas hydrophila* isolated from diseased fish. *Microbiology* **146**:999-1009.
75. **Zo, Y. G., I. N. G. Rivera, E. Russek-Cohen, M. S. Islam, A. K. Siddique, M. Yunus, R. B. Sack, A. Huq, and R. R. Colwell.** 2002. Genomic profiles of clinical and environmental isolates of *Vibrio cholerae* O1 in cholera endemic areas of Bangladesh. *Proc. Natl. Acad. Sci. U. S. A.* **99**:12409-12414.

Supplementary Materials and Methods

Isolation of environmental *V. cholerae* strains SIO and TP. Samples of seawater taken from nearby the Scripps Institution of Oceanography pier, as well as plankton collected with a sixty micron pore size plankton net at the Torrey Pines Estuary, were each first inoculated into alkaline peptone water (APW; 1% peptone, 1% NaCl; pH 8.6). Overnight growth in APW at 42°C was followed by transferring cells to tryptone broth (1% tryptone), culturing for an additional eight hours at 42°C and plating dilutions onto thiosulfate citrate bile salts sucrose agar (TCBS agar, Difco, Detroit, MI) agar overnight at 37°C. Yellow, flat 1-3 mm diameter colonies were restreaked onto Luria-Bertani agar (LB; 1% tryptone, 0.5% yeast extract, 1% NaCl, 1.5% agar, Difco, Detroit, MI). Presumptive *V. cholerae* were screened for sensitivity to 150 mg of the vibriostatic agent O/129 (2,4-diamino-6,7-diisopropylpteridine, Sigma Chemical Co., St. Louis, MO) and for *V. cholerae*-specific 16S rRNA gene amplification using primers VC490 and VC1261R (*Escherichia coli* 16S rRNA position numbers, supplementary Table 2.S2). These primers are closely related to the *V. cholerae*-specific primers CHO.1 and CHO.2, respectively, previously described by Nishimura *et al.* (4) and were used under standard polymerase chain reaction (PCR) conditions.

Comparison of SIO and TP strains by locus sequencing. In order to amplify genes encoding the protein markers, the *V. cholerae* strain N16961 genome was downloaded from The Institute for Genomic Research website (<http://www.tigr.org>) and MacVector (Accelrys, San Diego, CA) was used to design

primers to amplify the following genes: rDNA (16S, 23S, and *Vibrio cholerae*-*Vibrio mimicus* specific 16S-23S intergenic spacer region (ISR) (2)) and protein markers (deoxyribodipyrimidine photolyase-1 (*phrB-1*), deoxyribodipyrimidine photolyase-2 (*phrB-2*), DNA helicase II (*uvrD*), DNA helicase IV (*helD*), integrase/recombinase (*xerC*) and site-specific recombinase (*intl4*; supplementary Table 2.S2). 16S and 23S rRNA amplification was performed as described, using standard conditions and primers listed in Table 2.S2 in the supplementary data (1, 6). *Vibrio cholerae*-*Vibrio mimicus* specific ISR regions (2) from SIO, TP, *Vibrio mimicus*, and N16961 were amplified using primers listed in Table 2.S2 and standard protocols with the following cycling parameters: 94°C for 2 min, 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1.5 min, followed by a final extension step at 72°C for 10 min, and three bands were visible on an agarose gel, as expected (data not shown). The PCR products were cloned into the pCR2.1 vector using the TOPO-TA cloning kit (Invitrogen). Clones corresponding to ISR type 2 were sequenced from SIO, TP, N16961, and *Vibrio mimicus*, and a clone corresponding to type 1 was sequenced for the SIO strain. Sequencing was performed at the SDSU Microchemical Core Facility (SDSU-MCF). Sequence alignments were performed using ClustalX. Neighbor joining trees were bootstrapped and visualized in TreeView.

Panhandle PCR and sequencing of selected unique genes. Panhandle PCR was performed in order to obtain additional genomic sequence data upstream and downstream of short reads of genes of interest in the SIO and TP environmental strains (3, 5). Briefly, 1 ug of genomic DNA was added to an overnight

digestion/ligation reaction with one of several blunt-end enzymes, including *EcoRV* (GibcoBRL), *PvuII* (GibcoBRL), *RsaI* (Promega), *SspI* (New England Biolabs), and *XmnI* (Stratagene). The reactions were set up as follows: 10 pmol of Adaptors 1 and 2 (see Table 2), 10 U blunt-end restriction enzyme, 2.5 U T4 DNA ligase (Invitrogen), 2 mM ATP, 2 uL or 4 uL 10x One-Phor-All Buffer *PLUS* (Amersham), and the final volume was brought to 20 uL with sterile, distilled water. The reactions were incubated at room temperature overnight, and after 12 to 16 hours, the enzymes were heat inactivated. The DNA was precipitated and resuspended in 50 μ L of sterile distilled water. Five microliters were run on a gel to ensure that the DNA was partially digested. PCR reactions using a gene-specific primer ($T_m = 68^\circ\text{C}$) and a primer complementary to the adaptor ends (AP1 primer, see Table 2) were performed as follows: 1 μ L template DNA, 1x Taq Polymerase Reaction buffer (Invitrogen), 2.5 mM MgCl_2 , 0.5 mM dNTP mix, 10 pmol AP1 primer, 10 pmol gene-specific primer, 1 U Taq Polymerase (Invitrogen), and the final volume was brought to 20 μ L with sterile distilled water. PCR cycle parameters were as follows: 1 min at 95°C for a hot start, and then 25 cycles of 95°C for 30 sec and 68°C for 7 min. PCR reactions that yielded a single, strong band were cleaned using the UltraClean PCR Clean-up DNA Purification Kit (Mo Bio Laboratories, Solana Beach, CA) and sequenced from both directions using the gene-specific primer and the AP2 primer (see Table 2), as below. Additional PCR reactions with new specific primers were performed to walk further upstream and downstream on the chromosome. Sequencing was performed on a MegaBACE 500 sequencer (Amersham) using the DYEnamic ET Dye Terminator

Cycle Sequencing Kit (Amersham). Results were analyzed using SeqMan (DNASTAR, Madison, WI).

Colony hybridization. Probe labeling, hybridization, washing, and detection was performed with the AlkPhosDirect labeling kit and ECF detection substrate (Amersham) according to the manufacturer's instructions, with modifications detailed here. Final probe concentration was 10 ng probe per mL hybridization buffer. Stringency was adjusted to ensure that the SIO and/or the TP strain was positive for the gene of interest, and the clinical N16961 strain was clearly negative. Stringency of the hybridization was controlled by varying the concentration of NaCl in the hybridization buffer from 0.1M to 0.5M, varying the temperature of the hybridization from 55°C to 65°C, or varying the temperature of the wash steps from 55°C to 72°C. The chemifluorescent signal was detected using the Typhoon 9410 Variable Mode Imager (Molecular Dynamics, Amersham).

References for Supplementary Materials and Methods

1. **Amann, R. I., W. Ludwig, and K.-H. Schleifer.** 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143-169.
2. **Chun, J., A. Huq, and R. R. Colwell.** 1999. Analysis of 16S-23S intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. *App. Environ. Microbiol.* **65**:2202-2208.
3. **Egan, S., S. James, C. Holmstrom, and S. Kjelleberg.** 2002. Correlation between pigmentation and antifouling compounds produced by *Pseudoalteromonas tunicata*. *Environ. Microbiol.* **4**:433-442.

4. **Nishimura, M., K. Kita-Tsukamoto, K. Kogure, K. Ohwada, and U. Simidu.** 1993. A new method to detect viable bacteria in natural seawater using 16S rRNA oligonucleotide probe. *J. Oceanography* **49**:51-56.
5. **Siebert, P. D., A. Chenchik, D. E. Kellogg, K. A. Lukyanov, and S. A. Lukyanov.** 1995. An improved PCR method for walking in uncloned genomic DNA. *Nuc. Acids. Res.* **23**:1087-1088.
6. **Van Camp, G., S. Chapelle, and R. De Wachter.** 1993. Amplification and sequencing of variable regions in bacterial 23S ribosomal RNA genes with conserved primer sequences. *Curr. Microbiol.* **27**:147-151.

Supplementary Tables

Table 2.S1. Bacterial strains and isolates used in this study

Strain name ^a	Location at which strain was collected	Collection date	Strain collection ^b	Isolation source ^c
SIO	Scripps Institution of Oceanography Pier, La Jolla, CA	July 13, 2000	1	env
TP	Torrey Pines Estuary, La Jolla, CA	July 13, 2000	1	env
N16961	<i>Vibrio cholerae</i> O1 El Tor N16961, Bangladesh	1975	2	clinical
UCIC-299A	UC Irvine crew dock, Upper Newport Bay, CA	February, 1999	3	env
UCIC-299H	UC Irvine crew dock, Upper Newport Bay, CA	February, 1999	3	env
UCIC-305C	UC Irvine crew dock, Upper Newport Bay, CA	March 5, 1999	3	env
UCIC-305E	UC Irvine crew dock, Upper Newport Bay, CA	March 5, 1999	3	env
UCIC-305F	UC Irvine crew dock, Upper Newport Bay, CA	March 5, 1999	3	env
UCIC-305I	UC Irvine crew dock, Upper Newport Bay, CA	March 5, 1999	3	env
UCIC-305J	UC Irvine crew dock, Upper Newport Bay, CA	March 5, 1999	3	env
UCIC-305K	UC Irvine crew dock, Upper Newport Bay, CA	March 5, 1999	3	env
UNB-420L	Upper Newport Bay, CA	April 20, 1999	3	env
OCC-624A	Orange Coast College crew dock, Lower Newport Bay	June 24, 1999	3	env
OCC-624C	Orange Coast College crew dock, Lower Newport Bay	June 24, 1999	3	env
OCC-624D	Orange Coast College crew dock, Lower Newport Bay	June 24, 1999	3	env
OCC-624E	Orange Coast College crew dock, Lower Newport Bay	June 24, 1999	3	env
OCC-1004A	Orange Coast College crew dock, Lower Newport Bay	October 4, 1999	3	env
OCC-1004D	Orange Coast College crew dock, Lower Newport Bay	October 4, 1999	3	env
BC-624B	Big Canyon, Upper Newport Bay, CA	June 24, 1999	3	env
BC-803D	Big Canyon, Upper Newport Bay, CA	August 3, 1999	3	env
BC-803E	Big Canyon, Upper Newport Bay, CA	August 3, 1999	3	env
BC-2000A	Big Canyon, Upper Newport Bay, CA	February, 2000	3	env
BC-2000C	Big Canyon, Upper Newport Bay, CA	February, 2000	3	env
BC-2000D	Big Canyon, Upper Newport Bay, CA	February, 2000	3	env
BP-2000A	Balboa Pier, Pacific Ocean, Newport Beach, CA	February, 2000	3	env
BP-2000C	Balboa Pier, Pacific Ocean, Newport Beach, CA	February, 2000	3	env
CRO1-628E	San Diego Creek Site 1, UC Irvine, CA	June 28, 1999	3	env
CRO1-628F	San Diego Creek Site 1, UC Irvine, CA	June 28, 1999	3	env
CRO1-628I	San Diego Creek Site 1, UC Irvine, CA	June 28, 1999	3	env
CRO2-628A	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO2-628B	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO2-628C	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO2-628D	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO2-628F	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO2-628G	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO2-628H	San Diego Creek Site 2, UC Irvine, CA	June 28, 1999	3	env
CRO3-628A	San Diego Creek Site 3, UC Irvine, CA	June 28, 1999	3	env
CRO3-628B	San Diego Creek Site 3, UC Irvine, CA	June 28, 1999	3	env
CRO3-628C	San Diego Creek Site 3, UC Irvine, CA	June 28, 1999	3	env
CRO3-628D	San Diego Creek Site 3, UC Irvine, CA	June 28, 1999	3	env
CRO1-2000A	San Diego Creek Site 1, UC Irvine, CA	February, 2000	3	env
CRO1-2000A	San Diego Creek Site 1, UC Irvine, CA	February, 2000	3	env
CRO1-2000H	San Diego Creek Site 1, UC Irvine, CA	February, 2000	3	env
CRO1-2000I	San Diego Creek Site 1, UC Irvine, CA	February, 2000	3	env
CRO2-2000A	San Diego Creek Site 2, UC Irvine, CA	February, 2000	3	env
CRO2-2000B	San Diego Creek Site 2, UC Irvine, CA	February, 2000	3	env

Table 2.S1. Bacterial strains and isolates used in this study, continued.

Strain name ^a	Location at which strain was collected	Collection date	Strain collection ^b	Isolation source ^c
CRO2-2000C	San Diego Creek Site 2, UC Irvine, CA	February, 2000	3	env
CRO2-2000D	San Diego Creek Site 2, UC Irvine, CA	February, 2000	3	env
CRO2-2000E	San Diego Creek Site 2, UC Irvine, CA	February, 2000	3	env
CRO2-2000F	San Diego Creek Site 2, UC Irvine, CA	February, 2000	3	env
CRO3-2000A	San Diego Creek Site 3, UC Irvine, CA	February, 2000	3	env
MissionBay1	Mission Bay Isolate 1, San Diego, CA	November, 2002	4	env
MissionBay2	Mission Bay Isolate 2, San Diego, CA	November, 2002	4	env
O395	<i>Vibrio cholerae</i> O1 Classical O395	1964	5	clinical
Non-O1	strain collection of Dr. M. Lizarraga-Partida	N/A	6	N/A
O1#52	strain collection of Dr. M. Lizarraga-Partida	N/A	6	N/A
O139	strain collection of Dr. M. Lizarraga-Partida	N/A	6	clinical
O1 Ogawa	strain collection of Dr. M. Lizarraga-Partida	N/A	6	N/A
<i>Vibrio mimicus</i> ATCC 33653		Reference strain	7	clinical
O1-1st case	strain collection of Dr. M. Lizarraga-Partida	N/A	6	clinical

^aAll strains are *Vibrio cholerae* strains unless otherwise designated.

^bStrain collection designations are as follows: 1, Dr. Douglas H. Bartlett, SIO, UCSD; this study. 2, Dr. J. J. Mekalanos, Harvard University, (1). 3, Dr. Sunny Jiang, UC Irvine (all strains were collected as part of a study of strains around Newport Beach, CA) (2). 4, Dr. Barbara Hemmingsen, San Diego State University; unpublished. 5, Dr. J. J. Mekalanos, Harvard University, strain collection. 6, Dr. M. Lizarraga-Partida, Centro de Investigacion Cientifica y de Educacion Superior de Ensenada (CICESE), Mexico, unpublished. 7, ATCC.

^c‘Clinical’ strains were isolated in a clinical setting, and ‘env’ strains were isolated from the environment. N/A indicates that this data is unavailable.

References for Table 2.S1

1. **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**:477-483.
2. **Jiang, S. C., and W. Fu.** 2001. Seasonal Abundance and Distribution of *Vibrio cholerae* in Coastal Waters Quantified by a 16S-23S Intergenic Spacer Probe. *Microb. Ecol.* **42**:540-548.

Table 2.S2. Primers used in this study.

Loci	N16961 chromo- some	Primer set name	Primer sequence (5'-3')	Reference
16S rDNA (<i>Vibrio cholerae</i> specific)	1	VC490 VC1261R	AATACCTTAATCATTGACGTTACCT GCTCCACCTCGCGGTATC	(4) and this study
16S rDNA (eubacterial)	1	27F 1492R	AGAGTTTGATCCTGGCTCAG GGTTACCTTGTTACGACTT	(2)
23S rDNA	1	BLF 188-20 BLR 1686-20	GGAAGTGAACATCTAAGTA GGGGCCATTTTGCCGAGTTC	(6)
<i>Vibrio cholerae</i> - <i>Vibrio mimicus</i> specific Interspacer Region Types 1 and 2	1	pr16S-1541F pr23S-25R	YGGNTGGATCACCTCCTT TGACTGCCMRGGCATCCA	(1) and this study
Deoxyribodipyrimid- ine photolyase (<i>phrB-1</i>)	1	PHRB-1F PHRB-1R	GATTTACGAGTGAACGATAACCCG ATGCCGTCTAACCCATTGCG	this study
Deoxyribodipyrimid- ine photolyase (<i>phrB-2</i>)	2	PHRB-2F PHRB-2R	CCGATTTGATTTGGCGACG AAGTAGGGTGAGAGCGAACTGGTG	this study
DNA Helicase II (<i>uvrD</i>)	1	UVRDF UVRDR	CCAATAATAGCGAGCGGATGG CCAGCAGACCAGAGGTTTTTACC	this study
DNA Helicase IV (<i>hldD</i>)	2	HELDF HELDR	AGAGGCGTGTTTTGGGGAAC TGCTTGAGTTTGCTGTCATTCG	this study
Integrase/recombin- ase XerC (<i>xerC</i>)	1	XERCF XERCR	GGGATGAAAGCCAGCAGTATCG TTGTGCCTGTCCAGCAAACCAC	this study
Site-specific recombinase Intl4 (<i>intl4</i>)	2	INTL4F INTL4R	CCCATCAAGTTACTCTACGGTTTCG AGTCACGGTCTTTTCAATCCCC	this study
Panhandle PCR adaptor ligation	--	Adaptor1 Adaptor2	CTAATACGACTCACTATAGGGCTCG- AGCGGCCCGCCCGGCAGGT Phos-ACCTGCCCCG-NH ₂	(3, 5)
Panhandle PCR amplification and sequencing	--	AP1 AP2	GGATCCTAATACGACTCACTATAGGG AATAGGGCTCGAGCGGC	(3, 5)

References for Table 2.S2

1. **Chun, J., A. Huq, and R. R. Colwell.** 1999. Analysis of 16S-23S intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. *App. Environ. Microbiol.* **65**:2202-2208.
2. **Delong, E. F.** 1992. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U. S. A.* **89**:5685-5689.
3. **Egan, S., S. James, C. Holmstrom, and S. Kjelleberg.** 2002. Correlation between pigmentation and antifouling compounds produced by *Pseudoalteromonas tunicata*. *Environ. Microbiol.* **4**:433-442.
4. **Nishimura, M., K. Kita-Tsukamoto, K. Kogure, K. Ohwada, and U. Simidu.** 1993. A new method to detect viable bacteria in natural seawater using 16S rRNA oligonucleotide probe. *J. Oceanography* **49**:51-56.
5. **Siebert, P. D., A. Chenchik, D. E. Kellogg, K. A. Lukyanov, and S. A. Lukyanov.** 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucl. Acids Res.* **23**:1087-1088.
6. **Van Camp, G., S. Chapelle, and R. De Wachter.** 1993. Amplification and sequencing of variable regions in bacterial 23S ribosomal RNA genes with conserved primer sequences. *Curr. Microbiol.* **27**:147-151.

Chapter 3.

Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*

Abstract

The ADP-ribosyltransferases are a class of enzymes that display activity in a variety of bacterial pathogens responsible for causing diseases in plants and animals including those affecting mankind, such as diphtheria, cholera and whooping cough. We report the characterization of a novel toxin from *Vibrio cholerae*, which we call cholix toxin. The toxin is active against mammalian cells ($IC_{50} = 4.6 \pm 0.4$ ng/mL) and crustaceans (*Artemia* nauplii $LD_{50} = 10 \pm 2$ μ g/mL). Here we show that this toxin is the third member of the diphthamide-specific class of ADP-ribose transferases and that it possesses specific ADP-ribose transferase activity against ribosomal eukaryotic elongation factor 2. We also describe the high-resolution crystal structures of the multi-domain toxin and its catalytic domain at 2.1 and 1.25 Å resolution, respectively. The new crystal structures provide important insight into the structural basis for activation of toxin ADPRT activity. Structural data also indicate that cholix toxin possesses the necessary molecular features required for infection of eukaryotes by receptor-mediated endocytosis, translocation to the host cytoplasm, and inhibition of protein synthesis by specific modification of elongation factor 2. These results

indicate that cholera toxin may be an important virulence factor of *Vibrio cholerae* that likely plays a significant role in the survival of the organism in an aquatic environment.

Introduction

Many pathogenic bacteria utilize secreted protein toxins (exotoxins) as components of their virulence repertoire. These toxins induce cell death or alter cellular physiology by mechanisms such as proteolysis, pore formation, and covalent modification of host proteins. While some toxins are responsible for the complete pathology of a disease, others manipulate the host immune response, promote escape from the intracellular environment, release nutrients, or facilitate bacterial penetration of host barriers (3, 7). Although secreted toxins have been best characterized in terms of their virulence towards mammals, survival of these pathogens in the environment may provide additional selective pressures. For example, the secreted phospholipase of *P. aeruginosa* encoded by *plcS* (now known as *plcH*) contributes to virulence against *Candida albicans* (23), the greater wax moth, *Galleria mellonella* (26), as well as *Arabidopsis* and mice (52). Furthermore, secreted toxins may play roles in survival or colonization by non-pathogenic bacteria involved in symbioses (53).

Cholera toxin, an A-B₅ toxin that is expressed by some strains of *Vibrio cholerae*, causes cholera disease by specifically transferring an ADP-ribose group to an Arg residue of the GTP-binding protein G_s, thereby activating adenylate cyclase. Increased concentration of cAMP leads to secretion of Cl⁻, HCO₃⁻, and water from

epithelial cells at the site of colonization, resulting in dehydration and electrolyte loss from the infected patient. Production of a 'rice stool' by these patients, which carries *V. cholerae* at concentrations as high as 10^8 /mL (30), promotes dissemination of the disease among people without access to clean drinking water. Cholera cases have been linked to physical and biological conditions present in aquatic environments (2, 24, 38), and the bacterium is now known to be a constituent of the aquatic microbial community (13). Thus far, only two serogroups (O1 and O139, of greater than 200 known) are thought to have been responsible for the seven pandemics occurring since 1817 (13).

In the environment, most strains are of other serogroups (non-O1, non-O139); do not carry the genes encoding cholera toxin, and exhibit considerable genetic diversity. These strains are known to be capable of carrying many additional virulence factors, including hemolysin (18), RTX toxin (10), heat stable enterotoxin (54), hemagglutinin/protease (HA/P) (20), a type III secretion system (16), and a novel type VI secretion system associated with virulence (50), and have caused sporadic outbreaks of gastrointestinal disease distinct from cholera (14, 55) as well as extraintestinal infections (15). However, unsuccessful attempts to correlate genotypes of non-O1, non-O139 *V. cholerae* isolates with their virulence phenotypes in rabbit and mouse models suggest the presence of additional virulence factors (18). Detailed analyses of the genomes of specific non-O1, non-O139 strains of environmental origin (51) and clinical origin (8) have revealed the presence of a gene (*chxA*) encoding a novel putative secreted exotoxin (59), with similarity to exotoxin A (ExoA) of *P.*

aeruginosa. ExoA is a potent ADP-ribosylating toxin that specifically modifies eukaryotic elongation factor 2 (eEF2) leading to inhibition of protein synthesis in susceptible eukaryotic cells (29, 58).

Herein, we clearly demonstrate that the *chxA* gene encodes a second major ADP-ribosylating toxin in *V. cholerae*. This toxin is catalytically active, specific for the ribosomal eEF2 substrate, and toxic against a diverse array of eukaryotes. Cholix toxin is only the third member of the eEF2-specific ADP-ribosyltransferase toxins, in addition to ExoA and diphtheria toxin (DT). Finally, we have determined the crystal structures of the full-length cholix toxin and its catalytic C-domain (cholix_c) to 2.1 Å and 1.2 Å, respectively. Remarkably, the latter structure is the highest resolution to date for any member of the ADPRT family and its co-crystal complex with a competitive inhibitor provides new insight into the reaction mechanism of this important family of bacterial enzymes. Furthermore, its structure demonstrates striking similarities to ExoA, which consists of a tripartite domain structure, including domains I - III that function in receptor binding, membrane translocation, and enzyme catalysis, respectively. The new crystal structures reveal that inherent flexibility of Loop 1 (L1) (and perhaps also Loop 4; L4) in the toxin is a prerequisite for enzymatic activity and that disruption of specific H-bonds to domain II, either from reduction of disulfide bonds, from furin cleavage or both, is what activates the toxin upon entry into the eukaryotic host cell. In agreement with recent structural studies of the eEF2-ExoA-NAD⁺ complex (to be published) L1 of cholix also has the potential to interact

with both NAD^+ and the diphthamide target residue in eEF2 to form a solvent cover for the active site during the transferase reaction.

Materials and Methods

Cloning of *chxA* gene, expression and purification of cholix and cholix_c toxins. The 211-residue catalytic fragment of cholix (cholix_c) gene (GB AY876053) was cloned into a pET-28a(+) vector with a N-terminal 6xHis-tag and a TEV protease site. *E. coli* ER2566 cells were transformed with plasmid and were harvested by centrifugation, resuspended in 50 mM Tris-HCl, pH 7.6, 200 mM NaCl, 0.1% Tween, 1.25 mM PMSF and lysed in a French press. The cell lysate was centrifuged at 4 °C at 20,000 x g for 20 min, twice. The filtered supernatant was loaded onto a Ni-charged HiTrap™ Chelating HP column (GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 7.9 and 500 mM NaCl, washed with buffer and eluted with a 0 – 250 mM imidazole gradient. The cholix_c was dialyzed in 20 mM Tris-HCl, pH 7.6, 200 mM NaCl, and 0.1 mM PMSF and was digested with TEV (1:10 ratio) at 4 °C. The protein was then separated on a HiTrap™ Chelating HP column and the flow-through was loaded onto a Mono-Q column (Pharmacia) in 20 mM Tris-HCl pH 7.6, 10% glycerol and 25 mM NaCl and eluted with a 25-500 mM NaCl gradient.

The gene encoding the 634-residue cholix toxin for structural studies was cloned (GB AY876053) into a pET-28a(+) vector with a N-terminal 6xHis-tag. The cells were expressed and purified as for the catalytic fragment. The cholix toxin for *in vivo* studies was produced from a different construct possessing a TEV protease

digestion site between the 6xHis tag and the protein sequence. The cholix_c and cholix toxin were both concentrated to ~7 mg/mL in 100 mM NaCl, 20 mM Tris-HCl, pH 7.2 buffer.

Cytotoxicity assays. Mouse L-M fibroblasts (ATCC CCL-1.2) were maintained at 37°C (5% CO₂) in modified McCoy's 5A media supplemented with 10% fetal calf serum, 125 U/ml penicillin, 125 µg/ml streptomycin, 25 mM HEPES, and 2 mM L-glutamine. Cells were added to 24-well cell culture dishes (1 x 10⁵ cells/mL) and were incubated for 5 hours, washed with fresh media, and incubated with 0.75 mL media containing 1 µCi/mL of L-[4,5-³H]leucine (GE Healthcare Lifesciences) for 18 hours. Cells were washed twice with 0.5 mL PBS, and 0.25 mL of 0.1N NaOH was added. After 5 min at 37°C, 4 wells of each treatment were pooled together and were transferred to microcentrifuge tubes. Protein was precipitated with sodium deoxycholate and 7% trichloroacetic acid (TCA) and the precipitate was washed twice with 6% TCA, prior to dissolving in 0.2 - 0.4 mL of 0.1N NaOH for 30 minutes at 56°C. Protein concentrations were determined (Bio-Rad DC Protein Assay kit) and incorporation of ³H-leucine was measured in a Beckman LS6000TA scintillation counter (Perkin Elmer Ultima Gold cocktail). For Figure 3.1, toxin was added at the indicated dilutions in fresh media, and incubated with the cells for ~48 hours.

Cytotoxicity assays comparing MEFs and PEA13 cells. MEF-1 and PEA 13 cells were maintained in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal calf serum and incubated at 37°C under 10% CO₂. Then, 5 x 10⁵ cells/mL

were added to 24-well dishes in a 1 mL volume and experiments were performed as above, except that cells were incubated with toxin for 16 hours, then with ^3H -leucine for 3 hours. Samples were processed as above, except that ScintiSafe-30% (Fisher) scintillation cocktail was utilized.

***Artemia* toxicity assays.** *Artemia* cysts were added to filtered sea water and aerated at room temperature for ~24 hours. Within several hours of hatching, *Artemia* nauplii were placed in 24 well dishes containing 400 μL of SSW, with a total of 30 - 70 nauplii per well and incubated at 28°C for 42 - 45 hours. *Artemia* mortality was assessed using a dissecting microscope.

Detection of biotinylated ADP-ribose-eEF2. One hundred μM of Bio-NAD (Trevigen) was incubated with 5 μM toxin in the presence of 7 μL CHO cell lysate in 60 mM Tris-HCl, pH 7.6 buffer for 60 min at 25°C. The proteins were separated on by SDS-PAGE (34) and were transferred to nitrocellulose at 125 mA for 80 min. The membrane was blocked (2% BSA in PBS) for 1 h and then was incubated in 0.5% BSA in PBS with 1:5000 dilution of streptavidin-alkaline phosphate conjugate (Promega) and mixed overnight on a nutator at 4°C. The blot was then washed with 0.5 mM MgCl_2 , 40 mM NaHCO_3 , pH 9.6 buffer and developed in 10 mL of BCIP/NBT alkaline phosphate substrate for 3 min.

Crystallography. The cholix_c was co-crystallized with 5 mM PJ34 (Sigma-Aldrich) by vapor diffusion against reservoirs containing 15% PEG-8K and 20 mM KH_2PO_4 at 19 °C. Before flash freezing in liquid N_2 , the crystals were transferred to paratone-N (Hampton Research) for cryo-protection. A native 1.25 Å dataset was

collected at APS beamline 19-ID and a 1.65 Å dataset on a crystal soaked for 1 hr in 5 mM HgCl₂ was collected at beamline 17-ID. The cholix_c structure was solved using single-wavelength anomalous dispersion phases. The anomalous scattering substructure search, density modification and initial chain tracing were all performed in Phenix (1) and a single major and two minor sites were found. The resulting partial model was used for molecular replacement against the high-resolution native dataset using the CCP4 program Molrep (57), and the structure was then traced using warpNtrace (49) and was iteratively rebuilt in Coot (17) and anisotropically refined in Refmac5 (46) at 1.25 Å resolution.

Cholix toxin was crystallized by vapor diffusion against reservoirs containing 23% PEG-10K, 7.5% ethylene glycol and 0.1M HEPES pH 7.5 at 19 °C. Before flash freezing in liquid N₂ the crystals were transferred to paratone-N (Hampton Research) for cryo-protection. A native 2.1 Å dataset was collected at our in-house Enraf-Nonius FR571 diffractometer equipped with a rotating Cu anode and a Proteum Pt135 CCD detector (Bruker). The structure of cholix toxin was solved by molecular replacement with CNS 1.1 (6) and Phaser (44) using the cholix_c refined structure together with the receptor-binding and translocation domains from the structure of ExoA from *P. aeruginosa* (PDB entry 1IKQ) as search models. The solution from molecular replacement was used as input to warpNtrace, which could trace ~80% of the structure. The model was rebuilt in Coot and refined in Refmac5 using TLS at 2.1 Å resolution.

Results and Discussion

***ChxA* encodes a putative ADPRT.** Previously, we identified several DNA fragments from *V. cholerae* strains SIO and TP that have strong similarity to genes encoding virulence factors in known bacterial pathogens including a putative *toxA* gene (51) that encodes a 666-residue protein with a 32-residue leader sequence, called cholix toxin (70.7 kDa, 634-residue mature protein), and is similar to known diphthamide-specific ADPRTs (59). The cholix toxin primary structure shows 32% sequence identity with *Pseudomonas* ExoA, has a furin protease site for cellular activation (25), contains a C-terminal KDEL sequence that likely routes the toxin to the ER of the host cell (21), and possesses three classical signature regions (42, 59) peculiar to the catalytic domain of the diphthamide-specific toxins (59). Thus, all of these features within the cholix primary sequence provide a powerful indication that this protein is a new member of the eEF2-specific ADPRT group (DT group) (42, 59).

Cholix toxin is a bacterial ADPRT enzyme. The ADPRT reaction involves a nucleophilic substitution where eEF2 is the target protein (substrate) and the diphthamide residue (modified His) is the nucleophile that replaces the nicotinamide base (leaving group) of the NAD^+ substrate (59). The cellular effect of the covalent modification (ribosylation) of eEF2 is inhibition of protein synthesis leading to host cell death (48). To explore whether cholix toxin possessed ADPRT activity, we cloned both the full-length and truncated *chxA* gene into the T7-based *E. coli* pET28b

vector for expression and purification of both the whole toxin and its C-terminal catalytic domain. The proteins were purified by IMAC chromatography and were tested for both NAD-glycohydrolase (GH) and ADPRT activities using a fluorescence-based assay previously developed in our laboratory (4). The full-length recombinant toxin (634 residues, 70.7 kDa) showed only weak catalytic activity suggesting that this protein requires activation, a prerequisite for this family of enzymes/toxins (5, 11, 35). In contrast, the catalytic fragment (210 residues, 23 kDa) showed strong GH (K_M , $67 \pm 4 \mu M$; k_{cat} , $1.92 \pm 0.12 h^{-1}$) and ADPRT activities (Table 3.1). Cholix toxin is a slightly better enzyme than ExoA with a lower K_M (3-fold), similar k_{cat} , and specificity constant (2-fold higher) for the NAD^+ substrate (59). Furthermore, replacing the hallmark catalytic Glu residue within the enzyme domain of cholix toxin (Glu-581) with an Ala showed the expected result on the activity of the enzyme (~2600-fold reduction in k_{cat} , Table 3.1). Thus, cholix toxin recognizes eEF2 as the target protein substrate, possesses both GH and ADPRT enzyme activities, and this qualifies it as the third *bone fide* member of the DT group of bacterial ADPRT enzymes along with ExoA and DT (59).

Cholix toxin is active against eukaryotic cells. In order to assess whether cholix toxin possessed biological activity against eukaryotes, we dosed mouse fibroblast cells with both wild-type toxin and a catalytically inactive mutant toxin, E581A, and compared the effects on cell viability (Figure 3.1). The E581A mutant toxin had little or no effect on cell viability even at high doses (50 ng/mL) (Figure

3.1b) whereas the wild-type toxin showed considerable clearing of cell density (killing) at 1 ng/mL (Figure 3.1c) with little or no surviving fibroblast cells at 50 ng/mL (Figure 3.1f).

Using the cytotoxicity assay described in the Methods, we quantified the sensitivity of mouse fibroblast 1.2 L-M cells to cholix toxin and the results are shown in Figure 3.2a. The IC_{50} is 4.6 ± 0.4 ng/mL and this is comparable to the cytotoxicity of ExoA against this mouse fibroblast cell line (45). Furthermore, the inactive E581A mutant toxin showed little or no activity against the mouse fibroblast cells (Figure 3.2a). Figure 3.2b shows that cholix toxin recognizes the ubiquitous LRP receptor (43), which is also the specific receptor that ExoA exploits to enter the target eukaryotic cell (33) and this suggests that the cellular intoxication mechanism for cholix toxin is very similar to ExoA, but differs from that of DT (12). However, since the LRP-deficient strain also showed some sensitivity to cholix toxin, it is possible that there may be other avenues, besides the LRP receptor, by which cholix toxin can access the host cell cytoplasm.

V. cholerae is an aquatic organism that is often found attached to the exoskeletons of zooplankton (56) and this behavior may provide nutrients and protection against environmental challenges (31). Therefore, we tested the ability of purified cholix toxin to act on *Artemia nauplii* (brine shrimp) and the results are shown in Figure 3.2c. Remarkably, cholix was toxic to *Artemia nauplii* since doses near 50 μ g/mL killed all of the crustaceans, yet the E581A cholix protein had no effect on their viability.

We next examined whether cholix toxin was able to specifically and covalently modify the ribosomal eEF2 protein in a mammalian cell extract. Crude extract from Chinese Hamster Ovary (CHO) cells was incubated with toxin in the presence of biotin-NAD⁺ and the reactions were separated by SDS-polyacrylamide gel electrophoresis (PAGE), and blotted to nitrocellulose membrane and ADP-ribose-labeled bands were visualized using a streptavidin-alkaline phosphatase probe (Figure 3.2d). The eEF2 standard protein (lane 1, 0.25 µg eEF2) was labeled with biotin-ADP-ribose. Additionally, a protein corresponding to the molecular weight of mammalian eEF2 in the CHO lysate was also clearly labeled with biotin-ADP-ribose and the labelling was dependent upon the presence of cholix toxin (lanes 2 – 5). Furthermore, both the catalytic fragment of ExoA (ExoA_c, residues 399-605) and DTA labeled the eEF2 protein in the CHO cell lysate (lanes 6 and 7). A comparable experiment involving the use of ³²P-labeled NAD⁺ and cholix toxin gave similar results (data not shown). Thus, cholix toxin recognizes and labels only 1 protein in crude mammalian cell lysate that corresponds to ribosomal eEF2, which is a highly conserved and essential protein among all eukaryotes (27).

Structure of cholix toxin. The model of the 2.1 Å crystal structure for cholix toxin (no ligand or bound substrate) is shown in Figure 3.3a and it reveals that cholix toxin consists of three structural domains and is similar to the structure of *P. aeruginosa* ExoA having a 32% protein sequence identity. Cholix has a receptor-binding domain (domain Ia, 1-264, domain Ib, 387-423 (unknown function)) that

together form a 13-stranded anti-parallel β -jellyroll, a translocation domain (domain II, 265-386) consisting of a bundle of six α -helices, and catalytic domain (domain III, 424-634) with an α/β -fold topology. As seen in ExoA, the furin-cleavage site protrudes from the surface of domain II in a well-ordered loop and therefore is readily accessible for nicking by a sequence-specific endo-protease (9). Superposition of cholix toxin onto the ExoA structure (PDB entry 1IKQ) showed an overall r.m.s.d. of 2.04 Å for the C_α atoms and 1.70, 1.58, and 1.26 Å r.m.s.d. for domains I, II, and III, respectively (Figure 3.3b). Importantly, the positions of critical disulphides within cholix toxin align with those in ExoA (Figure 3.3b) and it is known that disruption of two disulphides is a prerequisite for toxin activation (5). It is clear from the cholix structure that it possesses all the necessary features for a diphthamide-specific ADPRT (59) and joins only ExoA and DT as members of this subfamily (59).

We also determined the 1.25Å crystal structure of the catalytic fragment of cholix toxin (cholix_c) with PJ34 bound (Figure 3.3c). PJ34 is a competitive inhibitor of the NAD⁺ substrate for ExoA and previously we determined its co-crystal structure with ExoA_c (60). PJ34 functions both as a ligand ($K_D = 0.51 \mu\text{M}$) and a crystallization enhancer for the cholix_c and two molecules of the inhibitor can be seen in the crystal complex. Whereas one PJ34 molecule is found in the NAD⁺ binding pocket as previously seen for ExoA ($K_D = 0.82 \mu\text{M}$) (60), a second PJ34 molecule packs between two molecules of cholix_c. The four catalytic loops surrounding the NAD⁺ binding site (L1, 477-483; L2, 547-553; L3, 574-579; L4, 511-518) corresponding to those in ExoA_c (28) are shown in yellow, cyan, orange, and red, respectively (Figure

3.3c). Figure 3.3d shows the specific binding of PJ34 within the active site of cholix_c. The inhibitor is stabilized in the NAD⁺ binding pocket through hydrophobic interactions and H-bonds. PJ34 forms two H-bonds to the backbone of Gly-461 and the phenyl moiety of Tyr-504 forms stacking interactions with the phenanthridinone ring system of PJ34. An additional H-bond is formed between His-460 and the main chain oxygen of Tyr-493 thus stabilizing these catalytic residues within the active site (Figure 3.3d), which is also observed in the βTAD and NAD⁺ complexes with toxin. The second PJ34 binds in a predominantly hydrophobic pocket formed by residues Thr-445, Leu-449, Tyr-454, Val-593 and Ile-595 in cholix_c molecule 1 and is held in place by Asn-550 in cholix_c molecule 2. Initial attempts to crystallize cholix_c without PJ34 failed suggesting that the second PJ34 has an important function in stabilizing the crystal packing.

Structural basis for cholix activation. It is well established that both ExoA and DT are activated within infected host cells by furin cleavage at an Arg-rich loop region along with reduction of critical disulphide bridges (39, 47) and that full-length toxin can be activated *in vitro* by urea and DTT (5, 19). Furthermore, PE40, a derivative of ExoA that is 20 residues longer than the furin-cleaved toxin, possesses two intact disulphide bonds and is enzymatically active (32). Figure 3.4a shows the superposition of cholix_c with the catalytic domain of full-length toxin and the r.m.s.d. is 0.67Å. The only regions where the two catalytic domains do not overlap almost perfectly are at L3 (orange) and L4 (red). L4 in cholix toxin is positioned closer to

domain II, where Thr-511 in L4 forms an H-bond to Asn-347 in α -helix D of domain II, and two water molecules mediate contact between Thr-518 and Glu-521 and between Asn-347 and Gly-352. L3 (orange) has moved slightly closer to the NAD^+ binding site in the full-length structure indicating some degree of flexibility of this loop. More importantly, the L1 residues (478-482) are not visible in the enzymatically active cholix_c structure, which most likely can be attributed to the highly flexible properties of this loop that may be important for enzymatic activation (Figure 3.4a). In the full-length structure, Asn-481 and Asn-482 in L1 form H-bonds to Arg-362 and Asn-366 in α -helix E of domain II (red) and ethylene glycol (crystallization additive) mediates contact between L1 and domain II (Figure 3.4b). This supports the idea that the flexibility of L1 (and perhaps also L4) is a prerequisite for enzymatic activity and that disruption of the H-bonds to domain II, either from reduction of disulfide bonds, from furin cleavage or both, is what activates the toxin. A similar activation mechanism has previously been suggested for ExoA (36, 37).

Although cholix_c was co-crystallized with the PJ34 inhibitor, NAD^+ is easily accommodated both electrostatically and stereochemically, within the NAD^+ binding site of cholix_c as expected for an ADPRT enzyme (Figure 3.4c). Further evidence of the structural relationship of cholix toxin with ExoA and DT can be seen from the superposition of the catalytic domains of these three toxins (Figure 3.4d) where the catalytic residues representing the ADPRT catalytic cluster (His-460, Tyr-493, Tyr-504, Glu-574, and Glu-581; cholix residue numbering) surrounding the NAD^+ ligand have overlapping positions. Residue Glu-574 in cholix, corresponding to the second

recently discovered catalytic residue in ExoA Glu-546 (to be published), has adopted a rotamer conformation, which is in agreement with the conformation of Glu-546 in ExoA structures without its natural NAD^+ substrate in the binding site. This indicates a similar catalytic function of Glu-574, however, the similarly positioned residue in DT (Phe-140) is not conserved and therefore likely makes use of a somewhat different reaction mechanism.

Superposition of cholix, cleaved at its furin cut site, onto the $\text{ExoA}_c\text{-NAD}^+$ -eEF2 complex structure (to be published) reveals that L4 of cholix fits snugly within the cleft between domain III and IV of eEF2 without any significant stereochemical difficulties and that L1 is positioned in an ideal location to form H-bonds with both the diphthamide and NAD^+ (Figure 3.4e). This suggests that the molecular basis of toxin activation involves specific structural changes within the catalytic domain that involve these loop regions. In particular, L1 likely has an active role in the stabilization of the transition state possibly by moving the diphthamide into position for the nucleophilic attack of the C1 of NAD^+ and/or by functioning as a solvent cover during the enzymatic reaction. This would be in agreement with the recent structures of the eEF2-ExoA- NAD^+ mutant complexes (to be published), where a Gln in L1 interacts with NAD^+ and an Asp interacts with the diphthamide in a conformation that also serves as a solvent cover. L1 in cholix toxin likewise contains residues (three Asn and three Glu) with the potential to interact with NAD^+ and the diphthamide upon activation of the toxin and NAD^+ binding (Figure 3.4f).

Concluding remarks

Diphtheria toxin, produced by *C. diphtheriae*, was one of the first bacterial toxins to be investigated (12). This protein was the first member of the ADPRT family to be identified and is one of the best studied and well understood bacterial toxins (12). The closest known DT relative is ExoA, produced by *P. aeruginosa*. ExoA was discovered in the 1960s and is now believed to be the most potent toxin produced by *P. aeruginosa* (40). Incredibly, DT and ExoA are only distantly related despite sharing a common ADPRT activity, target substrate protein (eEF2), and similar three-dimensional folds. Sequence similarity between the two toxins is low and is restricted to the catalytic domain. Also, the order of the domains is opposite and the receptor-binding and translocation mechanisms are different (12).

Until recently, no other members of the DT group in the ADPRT family had been identified (51). However, as bacterial genomes are being revealed, some potential diphthamide-specific toxins have recently been identified based on a search strategy that includes a limited consensus sequence pattern combined with secondary structure prediction (59). Remarkably, cholix toxin represents only the third member of the DT group and it shares much greater sequence homology with ExoA than DT. Since its original discovery in environmental isolates, *chxA* has been uncovered in other *V. cholerae* strains following genome sequencing. These include clinical isolates from the U.S. and Bangladesh (8), Japan (16) and Peru as well as an environmental isolate from Bangladesh (footnote¹). Our own analyses of strains

¹ Accession numbers: NZ_AAKI00000000 (V51), <http://www.nmpdr.org/> (NRT-36S),

collected in different continents confirm that the *chxA* gene is widely distributed (footnote²). The use of ADPRTs by *V. cholerae* is thus more extensive than previously appreciated. The specific biological target of the toxin and the nature of the symbiotic interaction associated with its activity in *V. cholerae* have yet to be determined. It could be significant that the environmental strain from which *chxA* was first discovered is capable of causing a fatal hemorrhagic pneumonia in adult mice (41). Outside of a human host, *V. cholerae* can be consumed by grazers, and interacts with both phytoplankton and zooplankton. Given that some toxins can act against multiple hosts (22), one or more of these organisms could contain the object of cholix toxin affinity and catalysis.

Data deposition. Coordinates and structure factors are deposited at the RCSB data bank, www.pdb.org as entries 2Q5T and 2Q6M.

Acknowledgements

RJ was supported by a CIHR and Carlsberg and AEP was the recipient of a pre-doctoral fellowship from Howard Hughes Medical Institute. DHB was supported by NIH and UC Marine Council and ARM was supported by CIHR and CCFF. We thank Dawn White and Gerry Prentice for excellent technical support during this research and to Sara Kelly for help with cell culture techniques. We are grateful for

NZ_AAUR00000000 (1587), NZ_AAWG00000000 (623-39), NZ_AAWF00000000 (MZO-2).

² Cholix toxin is present in *V. cholerae* strains from Mexico, Bangladesh, and Peru (Purdy *et al.*, unpublished data).

help provided by the beamline staff at 17-ID and 19-ID at the Argonne Photon Source, Chicago.

Chapter 3, in full, has been submitted for publication. Jorgensen, R., A. E. Purdy, R. J. Fieldhouse, M. S. Kimber, D. H. Bartlett, and A. R. Merrill. Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. Submitted, 2007. The dissertation author was the co-first author of this paper, and was the primary investigator and author of sections of the manuscript relating to the biological role of cholix toxin.

Table 3.1. Comparison of the ADPRT kinetic parameters for ExoA_c and cholix_c toxins

Parameter	ExoA _c	Cholix _c
K _M (NAD) (μM)	121 ± 21	45 ± 3
V _{max} (M s ⁻¹)	1.3 x 10 ⁻⁷	1.03 x 10 ⁻⁷
k _{cat} (s ⁻¹)	13 ± 2	10 ± 3
^a k _{cat} (s ⁻¹) catalytic Glu _ Ala	0.008 ± 0.0001	0.004 ± 0.003
k _{cat} /K _M (M ⁻¹ s ⁻¹)	1.1 x 10 ⁵	2.3 x 10 ⁵

The kinetic parameters were determined as described in Materials and Methods. ^aThe catalytic Glu to Ala mutations involved Glu-553 (ExoA_c) and Glu-581 (cholix_c). The values represent the mean ± SD from three independent experiments.

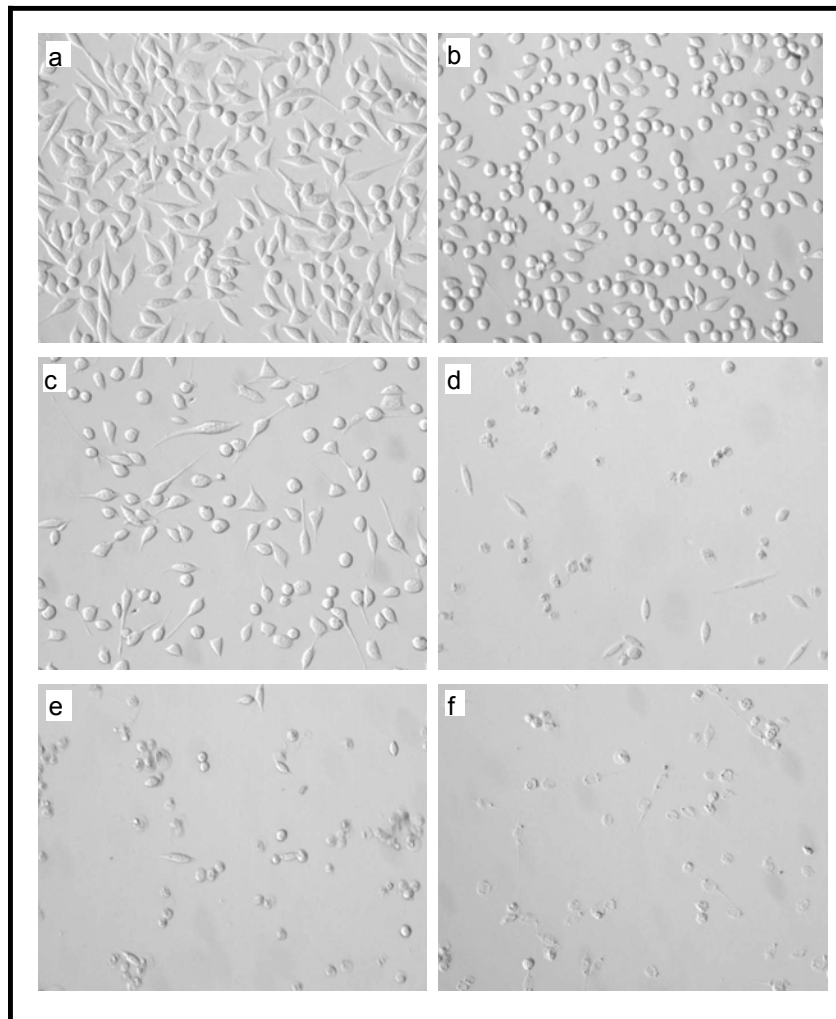


Figure 3.1. Cholix toxin is cytotoxic towards mouse fibroblasts. Cytotoxicity after ~48 hours toxin exposure. (A) Cells display several morphotypes with no toxin. (B) With 50 ng/ml E581A toxin cells still reach normal density. (C) With 1 ng/ml wild-type toxin, cells are normal. (D) 10 ng/ml, (E) 25 ng/ml, or (F) 50 ng/ml of wild-type toxin clearly results in cytotoxic effects.

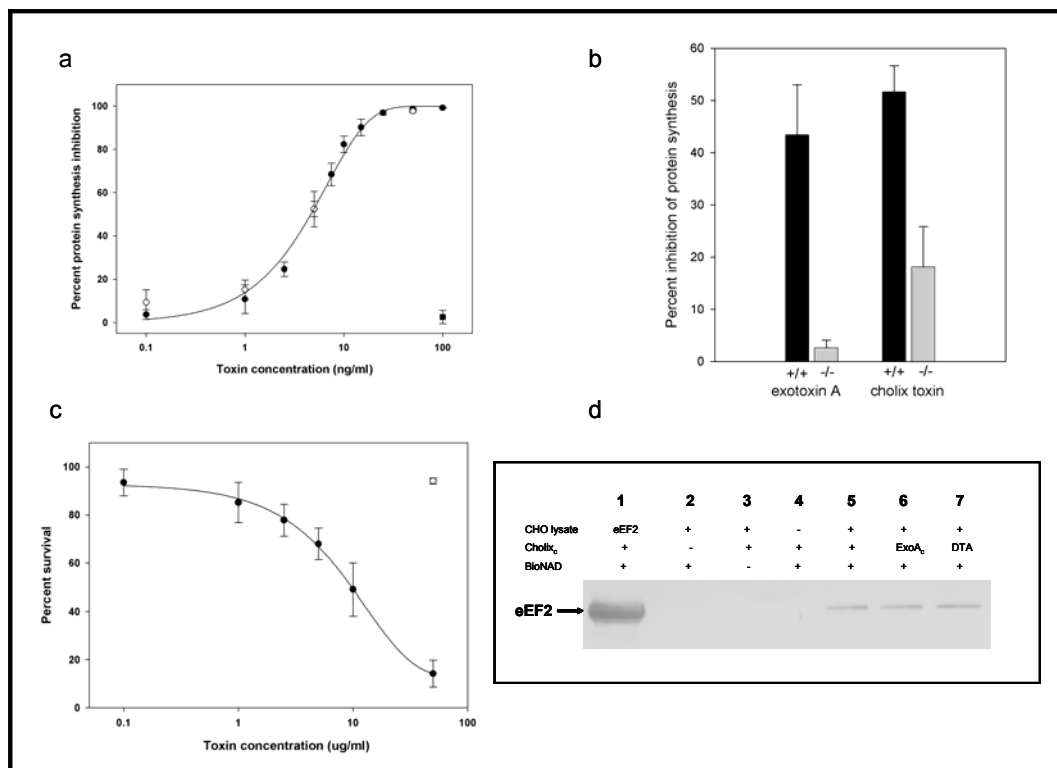


Figure 3.2. Biological activity of cholix toxin. (A) Protein synthesis inhibition in mouse fibroblasts. Cholix toxin (closed circles) inhibits protein synthesis similarly to ExoA (open circles) but E581A mutant (closed squares) does not. (B) Protein synthesis inhibition by ExoA and cholix toxin in mouse MEF1 (LRP1 +/+) and PEA13 (LRP1-/-) cells at 100 ng/mL. (C) Survival (%) of *Artemia* nauplii exposed to cholix toxin. Cholix toxin (closed circles) displays a dose-dependent effect on *Artemia* nauplii after 42 to 45 hours. Cholix toxin E581A (open circle) displays no toxic effect. (D) Western blot showing biotin-ADP-ribose labelling of eEF2 in CHO lysate. Lane 1, 0.25 μ g eEF2 incubated with cholix toxin; lanes 2 – 4, various controls; lanes 5-7, labelling of CHO lysate by cholix_c, ExoA_c, and DTA, respectively.

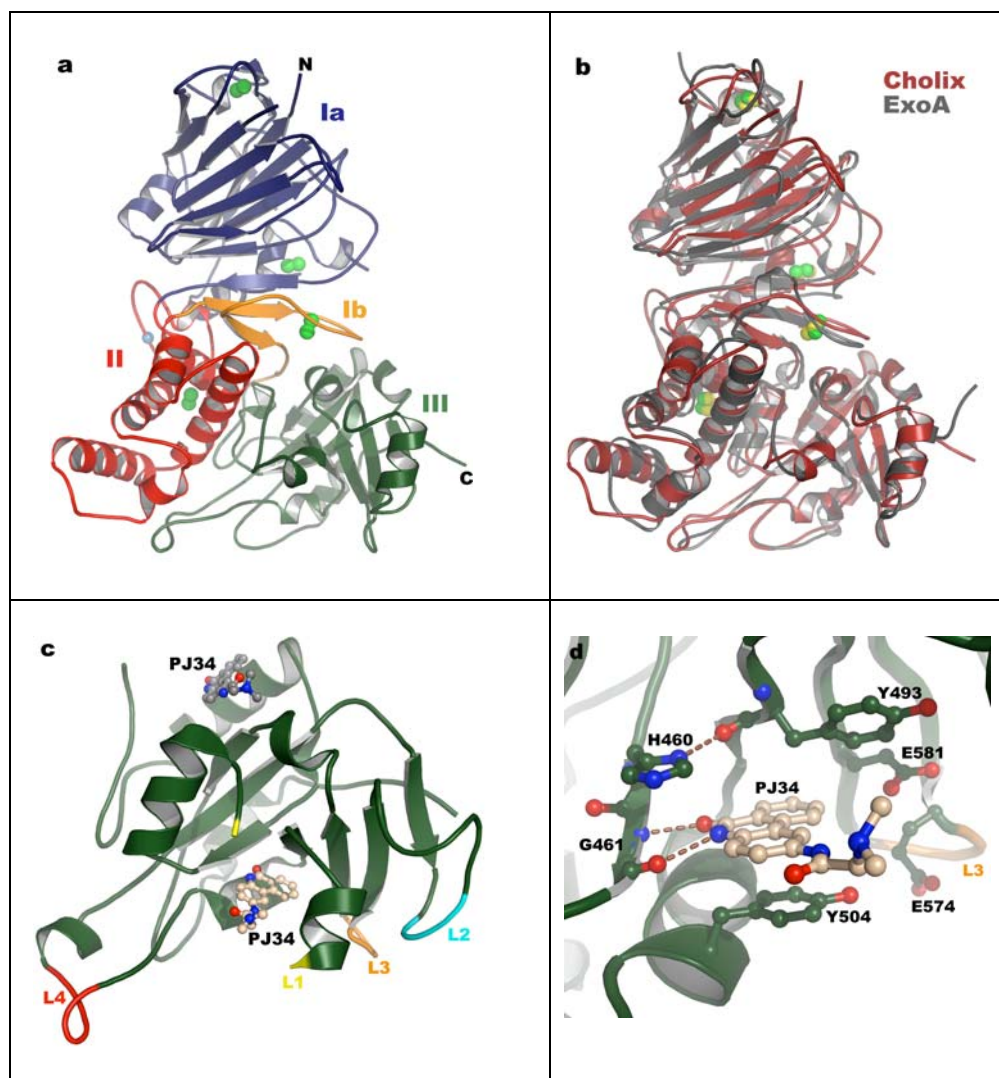
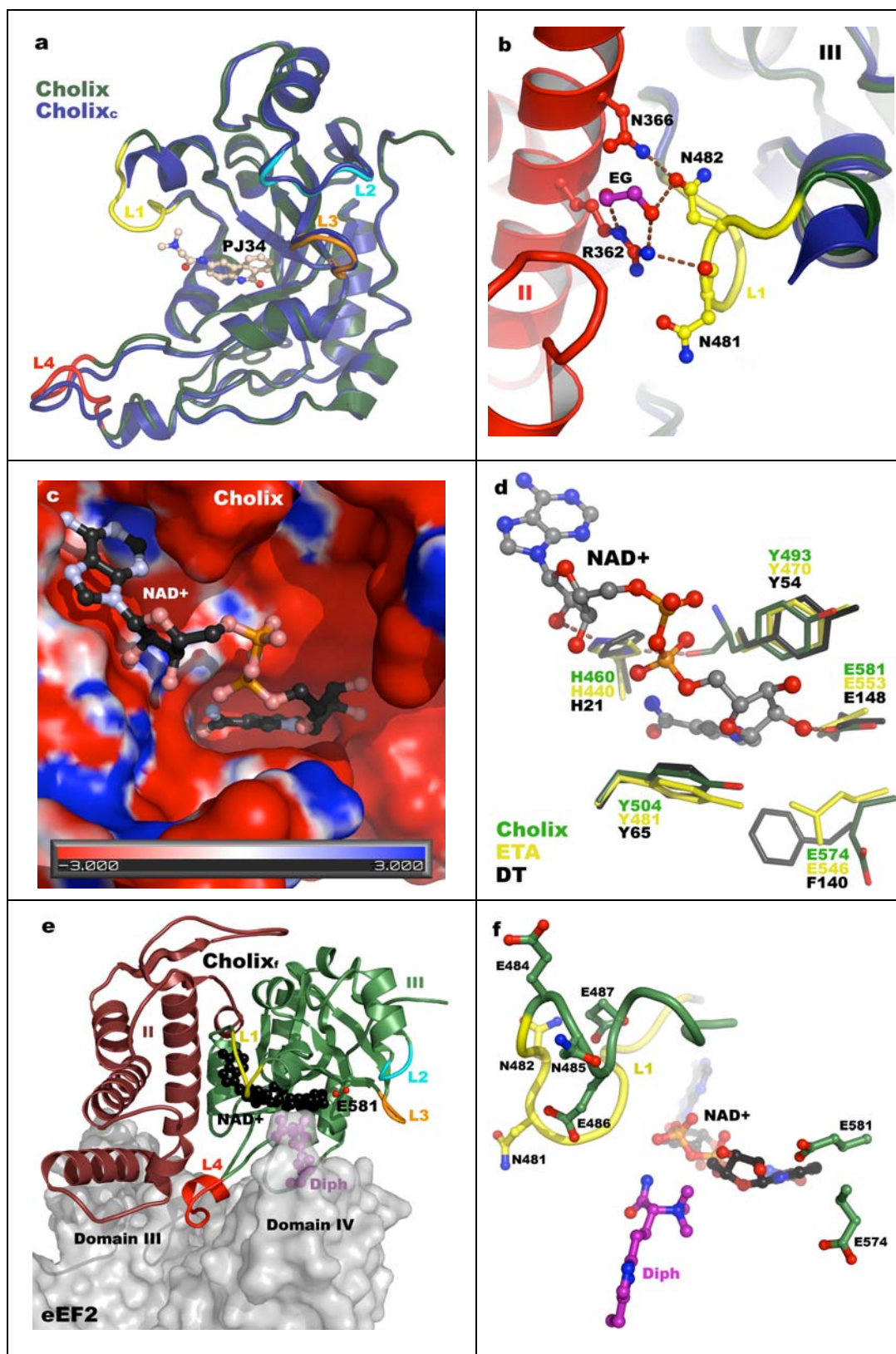


Figure 3.3. The structure of cholix toxin. (A) Ribbon drawing of the cholix structure (PDB entry 2Q5T): domain Ia, blue (1-264); domain II (265-386), red; domain Ib (387-423), orange; domain III (424-634), green. Disulfides and furin cut sites are shown as light green and blue spheres, respectively. (B) Superposition of ExoA onto the cholix structure. Disulfides are indicated by light green and yellow spheres for cholix toxin and ExoA, respectively. (C) Ribbon drawing of the catalytic domain structure (PDB entry; 2Q6M). The two PJ34 inhibitors are shown in black and grey ball and stick and the catalytic loops (L1-L4) are shown in yellow, cyan, orange and red, respectively. (D) Binding of PJ34 to the NAD⁺ binding pocket. The phenyl moiety of Tyr-504 forms stacking interactions with PJ34 (4 Å) and Tyr-493 is also adjacent to the hetero-ring system.

Figure 3.4. The structure of cholix toxin. (A) Superposition of the catalytic fragment from cholix toxin with the cholix_c structure (blue). Catalytic loops (L1-L4 in cholix) are colored as in Figure 3.3c and PJ34 (black) in the active site is shown ball and sticks. (B) L1 interactions in the cholix toxin. Asn-481 and Asn-481 (yellow) in L1 form H-bonds to Arg-362 and Asn-366 in α -helix E of domain II (red). Ethylene glycol (purple) mediates contact between L1 and domain II. (C) Binding of NAD⁺ (black) to cholix_c from superposition of cholix_c onto the eEF2-ExoA-NAD⁺ complex (to be published) The electrostatic surface potential of the cholix binding site was calculated using the APBS PyMol plugin. (D) Superposition of catalytic fragments of DT and ExoA from eEF2-ExoA-NAD⁺ onto cholix_c. Cholix_c ExoA, and DT residues are shown as green, yellow and black sticks, respectively and NAD⁺ in grey ball and stick. (E) Superposition of the furin-cleaved C-terminal fragment of full-length cholix onto ExoA of the eEF2-ExoA-NAD⁺ complex structure. Catalytic fragment (green) and domains Ib and II (brown) are shown in ribbon. L1-L4 are colored as before and NAD⁺ in black spheres. eEF2 as a white transparent surface with the diphthamide shown in purple spheres. (F) Residues in L1 of cholix with the potential to interact with the diphthamide and NAD⁺. Same superposition and coloring as in (E).



References

1. **Adams, P. D., R. W. Grosse-Kunstleve, L.-W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, and T. C. Terwilliger.** 2002. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **58**:1948-1954.
2. **Alam, M., N. A. Hasan, A. Sadique, N. A. Bhuiyan, K. U. Ahmed, S. Nusrin, G. B. Nair, A. K. Siddique, R. B. Sack, D. A. Sack, A. Huq, and R. R. Colwell.** 2006. Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl. Environ. Microbiol.* **72**:4096-4104.
3. **Alouf, J. E., and J. H. Freer (ed.).** 1999. *The Comprehensive Sourcebook of Bacterial Protein Toxins*, 2nd ed. Academic Press, San Diego.
4. **Armstrong, S., and A. R. Merrill.** 2001. Application of a fluorometric assay for characterization of the catalytic competency of a domain III fragment of *Pseudomonas aeruginosa* exotoxin A. *Anal. Biochem.* **292**:26-33.
5. **Beattie, B. K., and A. R. Merrill.** 1996. *In vitro* enzyme activation and folded stability of *Pseudomonas aeruginosa* exotoxin A and its C-terminal peptide. *Biochemistry* **35**:9042-9051.
6. **Brünger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren.** 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D, Biological crystallography* **54**:905-921.
7. **Burns, D. L., J. T. Barbieri, B. H. Iglewski, and R. Rappuoli (ed.).** 2003. *Bacterial Protein Toxins*. ASM Press, Washington, D. C.
8. **Chen, Y., J. A. Johnson, G. D. Pusch, J. G. Morris, Jr., and O. C. Stine.** 2007. The genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the

presence of pathogenic mechanisms that are distinct from those of O1 *Vibrio cholerae*. Infect. Immun. **75**:2645-2647.

9. **Chiron, M. F., C. M. Fryling, and D. J. FitzGerald.** 1994. Cleavage of pseudomonas exotoxin and diphtheria toxin by a furin-like enzyme prepared from beef liver. J. Biol. Chem. **269**:18167-18176.
10. **Chow, K. H., T. K. Ng, K. Y. Yuen, and W. C. Yam.** 2001. Detection of RTX toxin gene in *Vibrio cholerae* by PCR. J. Clin. Microbiol. **39**:2594-2597.
11. **Chung, D. W., and R. J. Collier.** 1977. Enzymatically active peptide from the adenosine diphosphate-ribosylating toxin of *Pseudomonas aeruginosa*. Infect. Immun. **16**:832-841.
12. **Collier, R. J.** 2001. Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. Toxicon **39**:1793-1803.
13. **Colwell, R. R.** 1996. Global climate and infectious disease: the cholera paradigm. Science **274**:2025-2031.
14. **Dalsgaard, A., M. J. Albert, D. N. Taylor, T. Shimada, R. Meza, O. Serichantalergs, and P. Echeverria.** 1995. Characterization of *Vibrio cholerae* non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru. J. Clin. Microbiol. **33**:2715-2722.
15. **Dalsgaard, A., A. Forslund, A. Hesselbjerg, and B. Bruun.** 2000. Clinical manifestations and characterization of extra-intestinal *Vibrio cholerae* non-O1, non-O139 infections in Denmark. Clin. Microbiol. Infect. **6**:626-628.
16. **Dziejman, M., D. Serruto, V. C. Tam, D. Sturtevant, P. Diraphat, S. M. Faruque, M. H. Rahman, J. F. Heidelberg, J. Decker, L. Li, K. T. Montgomery, G. Grills, R. Kucherlapati, and J. J. Mekalanos.** 2005. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. Proc. Nat. Acad. Sci. U. S. A. **102**:3465-3470.
17. **Emsley, P., and K. Cowtan.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. **60**:2126-2132.

18. **Faruque, S. M., N. Chowdhury, M. Kamruzzaman, M. Dziejman, M. H. Rahman, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2004. Genetic diversity and virulence potential of environmental *Vibrio cholerae* population in a cholera-endemic area. *Proc. Natl. Acad. Sci. U. S. A.* **101**:2123-2128.
19. **Galloway, D. R., R. C. Hedstrom, J. L. McGowan, S. P. Kessler, and D. J. Wozniak.** 1989. Biochemical analysis of CRM 66. A nonfunctional *Pseudomonas aeruginosa* exotoxin A. *J. Biol. Chem.* **264**:14869-14873.
20. **Halpern, M., H. Gancz, and Y. Kashi.** 2003. *Vibrio cholerae* hemagglutinin/protease degrades chironomid egg masses. *Appl. Environ. Microbiol.* **69**:4200-4204.
21. **Hessler, J. L., and R. J. Kreitman.** 1997. An early step in *Pseudomonas* exotoxin action is removal of the terminal lysine residue, which allows binding to the KDEL receptor. *Biochemistry* **36**:14577-14582.
22. **Hilbi, H., S. S. Weber, C. Ragaz, Y. Nyfeler, and S. Urwyler.** 2007. Environmental predators as models for bacterial pathogenesis. *Environmental Microbiology* **9**:563-575.
23. **Hogan, D. A., and R. Kolter.** 2002. *Pseudomonas-Candida* interactions: an ecological role for virulence factors. *Science* **296**:2229-2232.
24. **Huq, A., R. B. Sack, A. Nizam, I. M. Longini, G. B. Nair, A. Ali, J. G. Morris, Jr., M. N. H. Khan, A. K. Siddique, M. Yunus, M. J. Albert, D. A. Sack, and R. R. Colwell.** 2005. Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* **71**:4645-4654.
25. **Inocencio, N. M., J. M. Moehring, and T. J. Moehring.** 1994. Furin activates *Pseudomonas* exotoxin A by specific cleavage *in vivo* and *in vitro*. *J. Biol. Chem.* **269**:31831-31835.
26. **Jander, G., L. G. Rahme, and F. M. Ausubel.** 2000. Positive correlation between virulence of *Pseudomonas aeruginosa* mutants in mice and insects. *J. Bacteriol.* **182**:3843-3845.

27. **Jorgensen, R., A. R. Merrill, and G. R. Andersen.** 2006. The life and death of translation elongation factor 2. *Biochem. Soc. Trans.* **34**:1-6.
28. **Jorgensen, R., A. R. Merrill, S. P. Yates, V. E. Marquez, A. L. Schwan, T. Boesen, and G. R. Andersen.** 2005. Exotoxin A-eEF2 complex structure indicates ADP ribosylation by ribosome mimicry. *Nature* **436**:979-984.
29. **Jorgensen, R., S. P. Yates, D. J. Teal, J. Nilsson, G. A. Prentice, A. R. Merrill, and G. R. Andersen.** 2004. Crystal structure of ADP-ribosylated ribosomal translocase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**:45919-45925.
30. **Kaper, J. B., J. Morris, J. G., and M. M. Levine.** 1995. Cholera. *Clin. Microbiol. Rev.* **8**:48-86.
31. **Kirn, T. J., B. A. Jude, and R. K. Taylor.** 2005. A colonization factor links *Vibrio cholerae* environmental survival and human infection. *Nature* **438**:863-866.
32. **Kondo, T., D. FitzGerald, V. K. Chaudhary, S. Adhya, and I. Pastan.** 1988. Activity of immunotoxins constructed with modified *Pseudomonas* exotoxin A lacking the cell recognition domain. *J. Biol. Chem.* **263**:9470-9475.
33. **Kounnas, M. Z., R. E. Morris, M. R. Thompson, D. J. FitzGerald, D. K. Strickland, and C. B. Saelinger.** 1992. The alpha 2-macroglobulin receptor/low density lipoprotein receptor-related protein binds and internalizes *Pseudomonas* exotoxin A. *J. Biol. Chem.* **267**:12420 – 12423.
34. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685.
35. **Leppla, S. H., O. C. Martin, and L. A. Muehl.** 1978. The exotoxin of *P. aeruginosa*: A proenzyme having an unusual mode of activation. *Biochem. Biophys. Res. Commun.* **81**:532-538.
36. **Li, M., F. Dyda, I. Benhar, I. Pastan, and D. R. Davies.** 1996. Crystal structure of the catalytic domain of *Pseudomonas* exotoxin A complexed with

a nicotinamide adenine dinucleotide analog: Implications for the activation process and for ADP-ribosylation. *Proc. Natl. Acad. Sci. U. S. A.* **93**:6902-6906.

37. **Li, M., F. Dyda, I. Benhar, I. Pastan, and D. R. Davies.** 1995. The crystal structure of *Pseudomonas aeruginosa* exotoxin domain II with nicotinamide and AMP: conformational differences with the intact exotoxin. *Proc. Natl. Acad. Sci. U. S. A.* **92**:9308-9312.
38. **Lobitz, B., L. Beck, A. Huq, B. Wood, G. Fuchs, A. S. Faruque, and R. Colwell.** 2000. Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc. Natl. Acad. Sci. U. S. A.* **97**:1438-1443.
39. **Lory, S., and R. J. Collier.** 1980. Expression of enzymic activity by exotoxin A from *Pseudomonas aeruginosa*. *Infect. Immun.* **28**:494-501.
40. **Lyczak, J. B., C. L. Cannon, and G. B. Pier.** 2000. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect.* **2**:1051-1060.
41. **Makri, S., A. E. Purdy, D. Bartlett, and J. Fierer.** 2007. Pathogenicity of environmental isolates of *V. cholerae* in mice. *Microbes Infect.* **9**:1351-1358.
42. **Masignani, V., E. Balducci, D. Serruto, D. Veggi, B. Aricò, M. Comanducci, M. Pizza, and R. Rappuoli.** 2004. *In silico* identification of novel bacterial ADP-ribosyltransferases. *Int. J. Med. Microbiol.* **293**:471-478.
43. **May, P., E. Woldt, R. L. Matz, and P. Boucher.** 2007. The LDL receptor-related protein (LRP) family: An old family of proteins with new physiological functions. *Ann. Med.* **39**:219 - 228.
44. **McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni, and R. J. Read.** 2005. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D. Biol. Crystallogr.* **61**:458-464.
45. **Middlebrook, J. L., and R. B. Dorland.** 1977. Response of cultured mammalian cells to the exotoxins of *Pseudomonas aeruginosa* and

- Corynebacterium diphtheriae*: differential cytotoxicity. Can. J. Microbiol. **23**:183-189.
46. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. **53**:240-255.
 47. **Ogata, M., V. K. Chaudhary, I. Pastan, and D. J. FitzGerald.** 1990. Processing of *Pseudomonas* exotoxin by a cellular protease results in the generation of a 37,000-Da toxin fragment that is translocated to the cytosol. J. Biol. Chem. **265**:20678-20685.
 48. **Perentesis, J. P., S. P. Miller, and J. W. Bodley.** 1992. Protein toxin inhibitors of protein synthesis. BioFactors **3**:173-184.
 49. **Perrakis, A., R. Morris, and V. S. Lamzin.** 1999. Automated protein model building combined with iterative structure refinement. Nat Struct Mol Biol **6**:458-463.
 50. **Pukatski, S., A. T. Ma, D. Sturtevant, B. Krastins, D. Sarracino, W. C. Nelson, J. F. Heidelberg, and J. J. Mekalanos.** 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* model system. Proc. Natl. Acad. Sci. U. S. A. **103**:1528-1533.
 51. **Purdy, A., F. Rohwer, R. Edwards, F. Azam, and D. H. Bartlett.** 2005. A glimpse into the expanded genome content of the *Vibrio cholerae* species through the identification of genes present in environmental strains. J. Bacteriol. **187**:2992-3001.
 52. **Rahme, L. G., E. J. Stevens, S. F. Wolfort, J. Shao, R. G. Tompkins, and F. M. Ausubel.** 1995. Common virulence factors for bacterial pathogenicity in plants and animals. Science **268**:1899-1902.
 53. **Ruby, E. G., M. Urbanowski, J. Campbell, A. Dunn, M. Faini, R. Gunsalus, P. Lostroh, C. Lupp, J. McCann, D. Millikan, A. Schaefer, E. Stabb, A. Stevens, K. Visick, C. Whistler, and E. P. Greenberg.** 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. Proc. Natl. Acad. Sci. U. S. A. **102**:3004-3009.

54. **Sarkar, B., T. Bhattacharya, T. Ramamurthy, T. Shimada, Y. Takeda, and G. B. Nair.** 2002. Preferential association of the heat-stable enterotoxin gene (*stn*) with environmental strains of *Vibrio cholerae* belonging to the O14 serogroup. *Epidemiol Infect.* **129**:245-251.
55. **Sharma, C., M. Thungapathra, A. Ghosh, A. K. Mukhopadhyay, A. Basu, R. Mitra, I. Basu, S. K. Bhattacharya, T. Shimada, T. Ramamurthy, T. Takeda, S. Yamasaki, Y. Takeda, and G. B. Nair.** 1998. Molecular analysis of non-O1, non-O139 *Vibrio cholerae* associated with an unusual upsurge in the incidence of cholera-like disease in Calcutta, India. *J. Clin. Microbiol.* **36**:756-763.
56. **Tamplin, M. L., A. L. Gauzens, A. Huq, D. A. Sack, and R. R. Colwell.** 1990. Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. *Appl. Environ. Microbiol.* **56**:1977-1980.
57. **Vagin, A., and A. Teplyakov.** 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D. Biol. Crystallogr.* **56**:1622-1624.
58. **Wilson, B. A., and R. J. Collier.** 1992. Diphtheria toxin and *Pseudomonas aeruginosa* exotoxin A: active site structure and enzymic mechanism. *Curr. Top. Microbiol. Immunol.* **175**:27-41.
59. **Yates, S. P., R. Jorgensen, G. R. Andersen, and A. R. Merrill.** 2006. Stealth and mimicry by deadly bacterial toxins. *Trends Biochem Sci.* **31**:123-133.
60. **Yates, S. P., P. L. Taylor, R. Jorgensen, D. Ferraris, J. Zhang, G. R. Andersen, and A. R. Merrill.** 2005. Structure-function analysis of water-soluble inhibitors of the catalytic domain of exotoxin A from *Pseudomonas aeruginosa*. *Biochem. J.* **385**:667-675.

Supplementary Information

Table 3.S1. Data collection statistics, phasing and refinement of cholix structures.

Complex	Data collection		
	Cholix toxin	Cholix _c	
Beamline	Native	Native	HgCl ₂ derivative
	Rotating anode $\lambda = 1.5418 \text{ \AA}$	APS 19-ID $\lambda = 1.033 \text{ \AA}$	APS 17-ID $\lambda = 1.000 \text{ \AA}$
Symmetry and unit cell	P2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Resolution (Å)	a = 50.94 Å, b = 88.34 Å, c = 79.93 Å, $\beta = 95.32^\circ$	a = 40.09 Å, b = 64.76 Å, c = 78.00 Å	a = 40.06 Å, b = 64.63 Å, c = 78.14 Å
	50-2.05 (2.15-2.05)	30.0-1.25 (1.29-1.25)	50.0-1.65 (1.71-1.65)
R _{sym} ¹ (%)	9.4 (51.51)	5.5 (33.7)	6.9 (33.6)
Completeness (%)	98.1 (88.8)	97.9 (83.4)	93.6 (96.4)
Average I/⟨I⟩	8.62 (1.31)	23.5 (3.4)	28.6 (3.0)
Redundancy	3.90 (1.97)	10.7 (3.9)	3.0 (2.8)
Phasing			
Resolution			
Figure of Merit			
50.0-1.65			
0.32			
Refinement			
Resolution (Å)	50.0-2.1	25.2-1.25	
Reflections			
Work set/Test set	38767/2059	52798/2821	
Number of atoms			
Protein	4811	1578	
Ligand (PJ34)	-	44	
Waters	406	356	
Ethylene Glycol	36	-	
Chlorides	2	-	
Overall B-factor (Å ²)	16.91	17.5	
R.m.s. Deviation			
Bond lengths (Å)	0.008	0.018	
Bond angles (°)	1.031	1.797	
R-factor ² (%)	19.13	14.67	
R _{free} -factor ³ (%)	23.93	18.03	
Ramachandran plot (%)			
Most favored	90.1	92.3	
Additionally allowed	9.9	7.7	

Values in brackets are for the outer shell. $R_{\text{sym}} = (\sum_h \sum_j=1, N |I_h - I_h(j)| / \sum N I_h)$ for the intensity of a reflection measured N times and I_h is the mean intensity of the symmetry related reflections.

² R-factor = $(\sum_h ||F_o| - k|F_c|| / \sum_h |F_o|)$ where F_o and F_c are the observed and calculated structure factor respectively, and k is a scaling factor.

³ R_{free}-factor is identical to R-factor on a subset of test reflections not used in refinement.

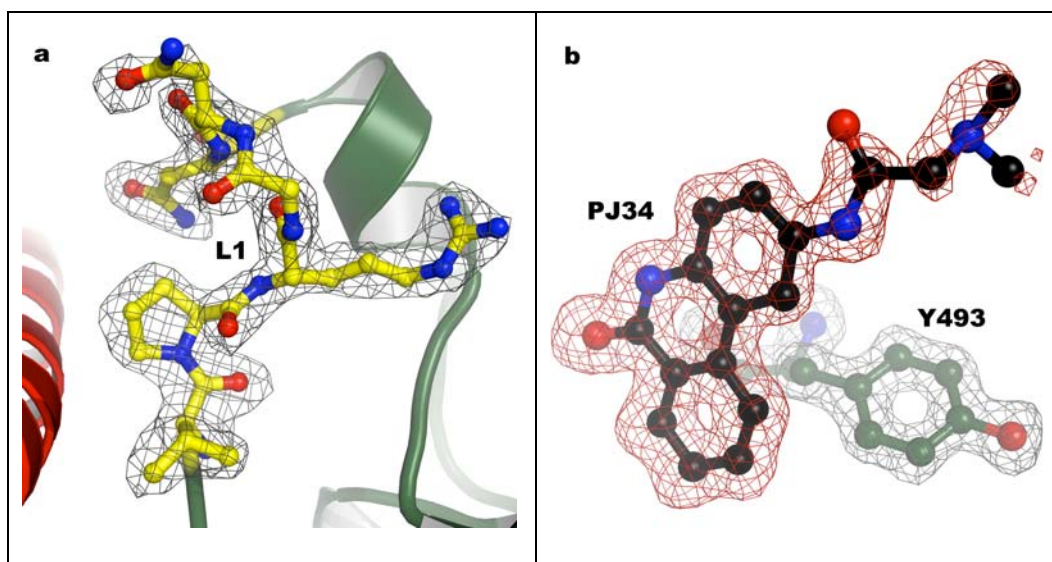


Figure 3.S1. Electron density maps of the full-length cholix and cholix_c structures. (A) Omit F_o-F_c map around residues 476-483 in L1 (yellow) of the full-length structure (grey, contoured at 2.8 σ). (B) Omit F_o-F_c map around PJ34 (black) and Tyr-493 (green) in cholix_c (red, contoured at 1.4 σ ; grey, contoured at 3.0 σ).

Chapter 4.

Diversity and distribution of cholix toxin among *Vibrio cholerae* strains from around the world

Abstract

Non-toxigenic non-O1, non-O139 *Vibrio cholerae* strains isolated from both environmental and clinical settings carry a suite of virulence factors aside from cholera toxin. We recently demonstrated that *V. cholerae* strains isolated from coastal waters of southern California carry cholix toxin, an ADP-ribosylating factor that is capable of shutting down protein synthesis in eukaryotic cells (16, 27). In order to better understand the role this toxin may play in *V. cholerae* virulence and environmental survival, DNA slot blot hybridization was utilized to assess the prevalence of the gene encoding cholix toxin, *chxA*, among a global collection of diverse *Vibrio cholerae* strains. Overall, *chxA* was present in a large proportion of non-O1, non-O139 strains, and it was rarely found in O1/O139 strains. A total of 75 *chxA* gene sequences were obtained and compared in order to decipher regions of the toxin that are under particularly strong selective pressures. Homologous recombination appears to be playing a significant role in the reassortment of toxin domains.

Introduction

Vibrio cholerae, the causative agent of cholera, possesses a number of virulence factors and secreted toxins which contribute to disease in human hosts. For survival within the environment, it appears that *V. cholerae* may deploy a distinct, but overlapping, arsenal of factors. The best-characterized virulence factors are those involved in causing cholera disease: cholera toxin (CTX) and the toxin co-regulated pilus (TCP), which are most frequently found in strains with pandemic capabilities (strains of the O1 or O139 serogroups). CTX has no known role in environmental survival, although a hypothesis involving mutualistic relationships with aquatic organisms in which CTX participates in ion regulation has been proposed (5). On the other hand, TCP has recently been shown to be involved in biofilm differentiation on chitinous surfaces (28). Other toxins carried by pandemic strains include the RTX (repeats in structure toxin), the HA/P protease, and hemolysin (11). HA/P may also have an environmentally-relevant role; it breaks down the egg masses of chironomid flies (13).

Strains not associated with cholera pandemics (the non-O1, non-O139 strains) rarely carry CTX or TCP (6, 10, 30, 31), but they often possess virulence genes such as the RTX or hemolysin (10), similarly to the pandemic strains. However, additional virulence genes and virulence-like factors may be employed for both environmental survival and disease, including NAG-ST (29), a HA/P that displays enterotoxicity (12), a type III secretion system (8), a type VI secretion system (26), a PrtV protease

that is involved in *C. elegans* killing and grazing resistance to protozoa (23, 32) as well as others that have yet to be discovered (10, 27).

We have previously shown that environmental non-O1, non-O139 strains can carry a gene encoding a toxin that acts similarly to *Pseudomonas aeruginosa* exotoxin A, the most potent toxin that this bacterium produces (27). The gene is present in more than half of all *V. cholerae* strains isolated from different aquatic settings in southern California. The protein, designated cholix toxin, presumably gains access to eukaryotic cells via receptor-mediated endocytosis, and subsequently undergoes a cleavage event in the endosome to release the catalytic domain, which proceeds to the cytosol. The toxin then catalyzes the transfer of an ADP-ribose group from NAD^+ to a post-translationally modified histidine residue (diphthamide) in eukaryotic eEF2, an essential translational elongation factor associated with the ribosome (16). This shuts down protein synthesis, and ultimately leads to cell death. Cholix toxin is the third member of the eEF2-specific ADP-ribosylating toxins, in addition to exotoxin A from *P. aeruginosa* and diphtheria toxin from *Corynebacterium diphtheriae*. The diphthamide residue is required for the catalytic reaction, and it is located in a region of eEF2 essential for its interaction with 18S rRNA that is also conserved in archaea and eukaryotes. Exotoxin A essentially mimics the 18S rRNA in its interactions with eEF2, thus ensuring that the evolution of resistance is extremely unlikely (15). The structural similarity of exotoxin A and cholix toxin suggests that cholix toxin must act in a comparable manner.

The gene encoding cholix toxin (*chxA*) gene was first discovered in a *V. cholerae* strain isolated from estuarine sediments, and a global survey of *V. cholerae* strains of both clinical and environmental origin was undertaken to describe the prevalence of *chxA*. Many strains contained the *chxA* gene, and from these, sequences of 75 individual *chxA* genes (resulting in 37 unique sequences) revealed three major clades. Analyses for positive and negative selection were performed, and the sequence differences between the clades may have functional significance, supported by the observation that the gene segments may be homologously recombined.

Materials and Methods

Strain collections. Strains were kindly provided from the culture collections of Dr. M. Leonardo Lizarraga-Partida, Departamento de Biotecnología Marina, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Mexico; Dr. Jaime Martinez-Urtaza Universidad de Santiago de Compostela, Spain; Dr. M. S. Islam, International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B); Dr. Greig Steward, University of Hawaii; and Dr. Sunny Jiang, University of California, Irvine. Strains were routinely grown in LB or 2216 marine broth at 37°C and stored in 30% glycerol-LB at -80°C.

Lateral gene transfer analysis. GC contents for the coding regions of *V. cholerae* sequenced genomes were calculated by downloading these values for the individual genes from the Joint Genome Institute's Integrated Microbial Genomes (IMG) database (22), and subsequently Microsoft Excel was used to determine the

average and standard deviation. This was then compared to the GC content of *chxA*. Codon usage bias was compared by assessing the codon usage bias for the *chxA* gene, and comparing this to the codon usage biases for other *V. cholerae* genomes. These values were obtained from the codon usage database at <http://www.kazusa.or.jp/codon/>. For each codon, the fraction of the time it was used for a particular amino acid in *chxA* was plotted against the same value obtained for N16961. Similar codon usages would result in a linear relationship that can be described by linear regression. The R value of the fitted curve indicates the statistical significance of the linear relationship.

Slot blot hybridization. Strains were inoculated into 2216 or LB/2 for overnight growth at 37°C. A slot blot apparatus (S&S Minifold I - Slot Blot) was used to filter ~100 µl of overnight culture onto a nylon 0.22 µm membrane (MagnaGraph, Osmonics). The cells were lysed and the DNA was affixed to the membrane as described (27). A DNA probe to detect *chxA* was generated by amplifying 1085 bp of the *chxA* gene from *V. cholerae* TP using primers G2TPExoAF and TPExoAF2 (Table 4.1). In order to decrease background signal, this PCR was performed using full-length *chxA* that had been cloned into a plasmid vector in *E. coli* as the DNA template. This PCR product was then labeled with alkaline phosphatase and the hybridization was performed with the AlkPhos Direct Labelling and Detection system (Amersham, GE Lifesciences). The manufacturer's instructions were followed with minor modifications: hybridization was performed overnight at 60°C, and subsequent washes were performed for 30 minutes each at 60°C. ECF was used as a substrate to

detect chemifluorescent signal, which was detected by scanning membranes on a flatbed scanner (Typhoon 9410, Molecular Dynamics, Amersham, GE Lifesciences). Each blot contained both positive and negative control cultures (TP and N16961, respectively).

Amplification and sequencing of *chxA*. Genomic DNA was isolated using the Wizard kit (Promega). The HighFidelity PCR kit (Roche) was used for amplification of genes to be sequenced. Other PCRs were performed using Taq polymerase (Invitrogen). Primers used to amplify the *chxA* genes were TPToxAF6 and L-ToxExpBamHI-LongR or ChxAR3 (Table 4.1). Based upon the TP sequence, these amplifications result in products of 1922 bp or 1861 bp, respectively, which were then sequenced. Sequencing was performed at the University of Hawaii Center for Genomics, Proteomics, and Bioinformatics.

In order to assess the chromosomal location of the *chxA* gene, sequences of open reading frames known to be adjacent to the *chxA* gene in TP (VC1645 and VC1644) were obtained from the genomes of sequenced *V. cholerae* strains in order to design conserved primers to those genes. Conserved primers for *chxA* were also designed and these were used to amplify from *chxA* to the adjacent sequences. According to the TP sequence, the primer pair ChxAF3 and VC1644_hypo_F should yield an amplicon of 2132 bp and the other primer pair, ChxAR1 and VC1645_hydro_R, should yield an amplicon of 1497 bp.

Based upon alignments of *chxA* alleles below, a set of primers conserved among all three clades of *chxA* sequences were designed. These are primers ChxAF2 and ChxAR1, which should result in a 439 bp amplicon.

Phylogenetic analysis. The *chxA* gene was amplified from *V. cholerae* strains in our collection, and a total of 1803 bp of *chxA* DNA sequence from 75 strains was used for DNA sequence alignments using MUSCLE (9). The alignments were then manually manipulated using MEGA3.1 and Se-AI (<http://tree.bio.ed.ac.uk/software/seal/>) in order to generate a DNA sequence alignment that takes into account codon spacing. From this alignment, a maximum parsimony tree was constructed in MEGA 3.1 with bootstrap values $\geq 60\%$ displayed (Figure 4.1).

To search for evidence of recombination, positive selection, and negative selection, programs freely available on the Datamonkey server (<http://www.datamonkey.org>) (25) were employed. Homologous recombination was detected using the Genetic Algorithms for Recombination Detection (GARD) program (19), which assigns recombination breakpoints when the alignment can be broken into discrete segments that have different phylogenies. To detect the presence of positive and negative selection, a series of programs capable of taking recombination into account were used for the analysis. These included the Single Likelihood Ancestor Counting (SLAC) analysis, Fixed Effects Likelihood (FEL) analysis, and Random Effects Likelihood (REL) analysis (18). These methods differ in the models used to detect dN/dS ratios and their ability to take into account differences in nucleotide

substitution rates. The consensus decision of these three programs is a reliable method for determining amino acid residues that are under particular selective pressures. Sequence features were mapped onto the structure of cholix toxin (16) using Mac PyMol.

Results

Many *V. cholerae* strains carry the *chxA* gene. *V. cholerae* strains were obtained from around the world and tested for the presence of the *chxA* gene via slot blot hybridization with a probe against *chxA*. This initial screening method was preferable to PCR because it allows the detection of diverse sequences. Overall, a total of 31 strains were positive out of 116 screened. In strains from Bangladesh, none of the 19 O1/O139 strains carried *chxA*, but 11 of the 30 non-O1, non-O139 strains were positive (Table 4.2). In a strain collection from Mexico, the majority of which are of clinical origin, a total of 12 non-O1, non-O139 strains carried *chxA* out of 20 total (Table 4.3). Interestingly, the only O1/O139 strains that carry the *chxA* gene were from this collection, in which 7 strains were positive and 14 strains were negative. The O1 designation of these strains was confirmed via DFA antibody analysis (data not shown). None of the 10 O1 strains obtained from Peru were positive for *chxA* although the only non-O1, non-O139 strain to be tested was positive (Table 4.4). From Senegal, none of the 4 O1 strains were positive (Table 4.4). Strains from southern California have been probed for this gene (*toxA*) previously (27). In this study, out of a total of 47 strains from this culture collection tested for *chxA*, 37

were positive. *chxA* sequences from these strains were included in the phylogenetic analysis (Table 4.5).

A total of 15 *V. cholerae* genomes have been sequenced and have been released to the public databases, in addition to *V. cholerae* O1 El Tor N16961. The majority of these strains are non-O1, non-O139 strains, and using BLAST, *chxA* genes were detected in the genomes of five strains (Table 4.6). The *chxA* gene has also been detected in the genomes of one strain isolated from Lake Pontchartrain (New Orleans, LA) after Hurricane Katrina, as well as several other strains from various sources. The two environmental non-O1, non-O139 strains commonly used in our laboratory, SIO and TP, both contain *chxA*.

Phylogenetics of *chxA*. A maximum parsimony tree constructed from an alignment of 75 *chxA* sequences of 1803 bp clearly delineated three separate clades of *V. cholerae chxA* gene sequences (Figure 4.1). Clade 1 contains non-O1, non-O139 strains, as well as several O1 strains from Mexico, environmental strains and clinical strains, and strains from diverse geographical locations, as well as sequenced strains of clinical origin, such as NRT36S. Thus, sequences falling in this clade appear to be representative of the majority of *chxA* alleles across all types of *V. cholerae* strains.

Strikingly, a set of *chxA* sequences from both Mexico and Bangladesh formed a virtually clonal clade very much removed from the other *chxA* sequences (clade 3, Figure 4.1). These sequences included non-O1, non-O139 strains from Mexico of clinical origin, as well as one O1 strain that carried the CTX genes (M21). Two non-

O1, non-O139 strains from Bangladesh, both of which were isolated from environmental settings, also fell into this group. Clade 2 consists of strains solely from southern California, except for strain M10, which itself is far removed from the rest of this clade.

Based upon the sequence alignments used to generate this tree, primers were designed to amplify segments from diverse *chxA* genes in order to more quickly screen for the presence of *chxA* in future work. These primers are ChxAF2 and ChxAR1 (Table 4.1). These primers will yield a PCR product of ~439 bp.

No evidence for recent lateral gene transfer from another species. To investigate whether the toxin has been recently horizontally transferred, the GC content of *chxA* was compared to the GC content of the coding regions of *V. cholerae* strains N16961 and V51. The GC content of *chxA* from strain TP is 43.7%, and the average GC content for coding regions in N16961 was $47.0 \pm 4.5\%$ and for V51 it was $48.3\% \pm 3.4\%$. The GC content of *chxA* is lower than that of the *V. cholerae* strains, but not by enough to support a recent lateral transfer event.

Codon usage patterns are an additional method by which lateral gene transfer can be detected. The codon usage for the *chxA* gene was determined, and compared to the codon usage of *V. cholerae* O1 El Tor N16961. The R^2 value of the comparison to N16961 is 0.69, but comparisons to other partial *V. cholerae* sequences yielded R^2 values ranging from 0.42 - 0.89, indicating considerable variation between different *V.*

cholerae genomes. This suggests that the codon usage signature also does not support a lateral transfer event.

In strain TP, the *chxA* gene is present on chromosome 1, between VC1644 and VC1645, on the opposite strand (27). Near this region of the *V. cholerae* genome there are no other indicators of lateral gene transfer, such as phage-like sequences or insertion sequence (IS) elements. Alignments of sequenced *V. cholerae* genomes, as well as additional sequence obtained surrounding the TP *chxA*, suggest that the gene constitutes a single-gene insertion/deletion (indel) event in chromosome 1. Aside from this indel, this region of the genome is syntenic among the *V. cholerae* genomes. In order to confirm that the gene is found in the same location in the chromosome among strains in our collection, conserved primers were designed to amplify the regions on either side of the *chxA* gene to VC1644 and to VC1645 (see Materials and Methods). A selection of strains representing the three clades of *chxA* sequences were chosen, with the reasoning that differences in *chxA* sequence are more likely to be indicative of differences in chromosomal location. A total of 15 strains were assessed (clade 1: M1, M45, B31, B55, C7, C39, C160; clade 2: C62, C74; clade 3: M6, M33, B33, B54) and in all cases, the *chxA* appears to be adjacent to VC1644, the conserved hypothetical gene (data not shown). However, variation may exist in gene sequence or regulatory regions upstream relative to the *chxA* gene in clade 3, because the amplicon spanning the distance between *chxA* and VC1645 was larger in B33, smaller in B54, and in M6 and M33, no amplicon was obtained (data not shown). All other *chxA* genes appear to be present between VC1645 and VC1644, as predicted.

Significance of sequence variation: observations of alignments and conserved residues. A cholix toxin protein sequence alignment of representatives of each of the three clades (Figure 4.3) shows that residues important to toxin cell entry and catalytic function are well conserved, and the greatest variation between the strains lies in the receptor binding domain Ia. Exotoxin A from *P. aeruginosa* is known to bind to the low density lipoprotein receptor-like protein/alpha₂-macroglobulin receptor (LRP1) which is widely distributed in many tissue types (21). For receptor binding of exotoxin A, a particular lysine residue (Lys-57) is absolutely critical (14). This Lys residue is conserved in all strains except for five (C12, C158, C32, C10, and C75) where it has been changed to Met (see C32 and C12 on Figure 4.3). All five of these strains also carry an additional conserved five amino acid insertion within the receptor-binding domain. This insertion is restricted to strains which have the Lys-Met mutation. A second insertion/deletion sequence of six amino acids is also present in domain Ia.

Domain II (the translocation domain) and domain Ib (receptor binding) are relatively well conserved among all of the sequences. In particular, the cysteine residues involved in disulfide bridge formation and toxin activation are conserved in domain II.

In domain III (the catalytic domain), of the amino acids that are critical for catalytic function (His-460, Tyr-493, Tyr-504, Glu-574, and Glu-581) (16), only Glu-574 shows variation. In clade 3, Glu-574 has been mutated to an Asp residue, which is also an acidic amino acid. Glu-581 is the conserved glutamate residue critical for

the catalytic capabilities of all ADP-ribosyl transferases (ADPRTs; Figure 4.3). Additional residues critical to the functioning of this toxin include Tyr-493 and Tyr-504, which form a hydrophobic pocket in which NAD^+ can bind (reviewed in (33)), and these residues are also conserved among all strains investigated here. A number of residues shown to be structurally important in cholix toxin (16) are conserved among all of the toxin alleles. However, regions of variability in clade 3 relative to clades 1 and 2 do appear within the NAD^+ binding pocket between the two Tyr residues, and within catalytic loops 3 and 4 (Figure 4.3).

Significance of sequence variation: homologous recombination. In order to better understand forces influencing the evolution of this gene, the *chxA* sequences were analyzed for the presence of homologous recombination utilizing the Genetic Algorithms for Recombination Detection (GARD) program (19). If GARD detects different phylogenies in discrete segments along the length of the sequence alignment, it assigns breakpoints between these regions (19). In the *chxA* DNA sequence alignment, GARD detected three individual breakpoints, and neighbor joining trees were automatically constructed for each of the four segments. These trees are shown in Figure 4.2. The locations of the putative recombination breakpoints are indicated in Figure 4.3. However, it should be noted that sufficiently divergent phylogenetic trees can arise due to differences in branch lengths, corresponding to different degrees of sequence change along the length of the alignment, and not necessarily homologous recombination.

The first segment detected by GARD includes approximately the first half of the receptor binding domain (Ia) until recombination breakpoint 1 (RecBrPt1; Figure 4.3). Here, very few differences between each of the protein sequences is observed, although sequences from clade 3 are clearly distinct from the other sequences (Figure 4.2a). The second segment corresponds to the second half of domain Ia, which contains significant sequence variation. Sequences from clade 2 cluster much more closely to clade 3 sequences than to clade 1 sequences, due the deletion of six amino acids that are retained in clades 2 and 3 (Figure 4.2b and Figure 4.3). A separate indel region exists just upstream, in which five sequences carrying a five amino acid insertion relative to the other sequences fall into both the clade 2/3 cluster (C12) and clade 1 (C32) (Figure 4.3).

The third segment, between recombination breakpoints 2 and 3 (RecBrPt2 and RecBrPt3; Figure 4.3) includes the entirety of domain II and the first half of domain Ib. Interestingly, the sequence breakpoint defining the beginning of this segment corresponds quite strikingly with the domain Ia/domain II junction (Figure 4.3). The breakpoint at the end of this segment falls in the middle of domain Ib, so that the fourth segment includes the remainder of the protein sequence (domains Ib and III). In the fourth segment, the sequences of clade 3 are considerably different than the sequences of clades 1 and 2 (Figure 4.2d, Figure 4.3).

Significance of sequence variation: selection acting on specific residues.

In order to detect possible selective forces acting on individual amino acids, analyses

to detect positive selection and negative selection were performed. The observation of positive selection, in which the ratio of non-synonymous to synonymous amino acid changes (dN/dS) is greater than 1, suggests the presence of selective pressures acting to cause the rapid evolution of the particular residue. The observation of negative selection, where dN/dS is less than 1, suggests the presence of an amino acid that must be well conserved. Using selection detection programs associated with the Datamonkey webserver as described in the Materials and Methods, SLAC analysis detected zero positively selected sites, and 10 negatively selected sites and FEL analysis detected zero positively selected sites and 60 negatively selected sites. In contrast, REL detected four positively selected sites, and four negatively selected sites. These four negatively selected sites were detected by all three methods. The fact that only one program detected positive selection suggests that it must not be particularly strong. Therefore, it appears that the toxin does not contain amino acid residues that must evolve rapidly in order to maintain its toxic effects.

In contrast, there are at least four residues that definitely appear to be under negative selection. These residues had a higher ratio of synonymous mutations within the sequence alignment, suggesting that non-synonymous mutations were detrimental to the toxin. The four residues, Glu-133, Pro-395, Tyr-439, and Glu-521, were then mapped to the protein structure of cholix toxin (Figures 4 and 5). Glu-133 falls within the receptor binding domain. Pro-386 falls in a loop region just after a beta-pleated sheet, in domain 1b, which is also responsible for receptor binding. Tyr-439 is in the catalytic domain at the end of an alpha-helix, on the protein surface, in a region that

interacts with other parts of the catalytic domain. Glu-521 falls adjacent to the catalytic loop 4.

Discussion

Vibrio cholerae strains from around the world carry cholix toxin, an eEF2-specific ADP-ribosylating toxin capable of killing eukaryotic cells. This study clearly demonstrates that the toxin is preferentially found among non-O1, non-O139 strains of either clinical or environmental origin, suggesting that *V. cholerae* might utilize this toxin in a context somewhat removed from interaction with a human host. In order to determine the diversity of toxin sequences and infer selective pressures affecting the evolution of this toxin, a comprehensive analysis of 75 *chxA* sequences was performed. The *chxA* sequences fell into three distinct clades (Figure 4.1): clade 1 contains the majority of strains investigated here and appears to be representative of *chxA* alleles. Clade 3 is a virtually clonal clade far removed from the other sequences. Clade 2 contains sequence features of both clade 1 and clade 3, and therefore this cluster falls into a distinct group between the two clades. Cholix toxin, from *V. cholerae* strain TP, is in clade 1, and its crystal structure has been solved, and therefore for these purposes serves as the canonical sequence.

Cholix toxin does not appear to have entered the *V. cholerae* genome via a recent horizontal transfer event. Both the GC content and the codon usage bias analyses indicate that the *chxA* gene and *V. cholerae* whole genome sequences have similar genomic signatures. Furthermore it appears that *chxA* localization within *V.*

cholerae chromosome 1 may be conserved among a wide diversity of strains, although there is likely to be sequence variation in the region upstream of *chxA* among the strains falling into clade 3. The diversity of the *V. cholerae* strains examined is inferred solely from the vastly different locations, times, and conditions under which the strains were collected. However, further multi-locus sequence typing approaches are needed to properly assess both the diversity of strains carrying the toxin, as well as their phylogenetic relationship to strains that are lacking *chxA*. Sequence analysis of 37 unique *chxA* gene sequences provides insights into regions of the protein that may be functionally divergent or are well conserved. Most importantly, the residue critical to the ADP-ribosyl transferase reaction, Glu-581, was conserved among all *chxA* sequences, suggesting that each of the alleles included in this study are potential ADPRTs.

The receptor binding domain, and specifically one half of domain Ia, was subject to the greatest variability between the sequences. Previous studies had determined that a particular lysine residue is crucial for exotoxin A binding to its receptor, LRP1. The corresponding lysine in cholix toxin was conserved in all strains except five; in these, the lysine had been changed to a methionine residue. Each of these five strains -- and only these strains -- also possessed an insertion of five amino acids further downstream within the receptor binding domain. Mapping these features to the cholix toxin structure revealed that the insertion falls on a loop region adjacent to the methionine residue on the surface of the protein (Figure 4.4). This suggests the intriguing possibility that the combination of the insertion and the methionine residue

indicates variation in receptor binding affinity or specificity. The functional and ecological ramifications are significant: the interaction between the receptor and the receptor binding domain determines the cellular target of the toxin. Assays to assess the functional implications of this sequence variation might lead to insights regarding the ecological purpose of the toxin. Interestingly, these features are spread between two separate clades of sequences, as C32, C10, and C75 are within a separate cluster of clade 1, and C12 and C158 fall into clade 2. The receptor binding domain contains a second indel region further downstream that, when mapped to the structure of cholix toxin, is present on an exposed region, though quite distant from the region critical to binding (Figures 3 and 4). This region has significant implications for phylogeny, however, because *chxA* alleles within clades 2 and 3 share the deletion, while those in clade 1 possess an insertion. In contrast, throughout the remainder of the *chxA* gene, clade 2 sequences are more closely related to clade 1 sequences. Analyses that detect evidence of positive or negative selection, which may indicate additional residues critical to the protein's function, found one residue within the receptor binding domain Ia that may be subject to negative selection. However, this residue, Glu-133, appears to be buried in the interior of the protein (data not shown), so its role may not be in direct interaction with the receptor, but perhaps in stabilizing the overall protein structure.

The translocation domain (domain II) and the second part of the receptor binding domain (domain Ib) were well conserved, although the sequence variations that were present once again differentiated clade 3 from clades 1 and 2. No residues in

the translocation domain were subject to negative selection. One residue in domain Ib, Pro-386, appears to be negatively selected. It falls in a loop region just after a beta-pleated sheet on the surface of the protein, but it is not near the critical lysine residue, so its role is unclear (Figure 4.4b).

Variability within the catalytic domain (domain III) once again distinguished sequences from clades 1 and 2 from clade 3. All major residues required for catalysis were conserved among all clades, aside from the Glu-574 in clade 3, which has been changed to an Asp residue. Though these amino acids are similar, a Glu-Asp mutation in the catalytic region of exotoxin A abolished catalytic activity (7). The sequence variation was particularly great in the four catalytic loops, as well as the NAD⁺ binding pocket (Figure 4.3), suggesting that their loop/pocket structures might be more permissive to sequence changes. The catalytic domain harbored two potential residues subject to negative selection, Tyr-439 and Glu-521, though they have not been previously implicated in toxin function. These residues were mapped to the structure of the catalytic domain (Figure 4.5). This analysis provides additional residues that could be subjected to site-directed mutagenesis studies in order to determine their effects on catalytic ability and structure formation.

Sequence conservation at the C-terminus of the protein could not be assessed, due to primer design constraints. However, the presence of a five amino acid insertion near the C-terminus in strain M10 (data not shown) suggests that additional sequence features may be present, and the full-length sequence for each of these toxin genes should be obtained, perhaps using the conserved primers designed above to capture the

region between the 3' end of the *chxA* sequence and VC1644. Strain TP, and most likely other strains as well, carries the canonical KDEL endoplasmic reticulum retention sequence at its C-terminus. This sequence is used by exotoxin A to hijack the eukaryotic cell's own machinery, allowing the toxin to proceed through a retrograde mechanism from the Golgi apparatus to the endoplasmic reticulum, where the catalytic domain translocates into the cytosol (reviewed in (33)). However, the sequence present at the C-terminus of exotoxin A is the non-canonical sequence REDL, raising the possibility that cholix toxin is more efficiently transported to the endoplasmic reticulum.

Homologous recombination appears to be affecting *chxA* evolution. GARD places recombination breakpoints along the alignment in locations such that the phylogenetic trees of the resulting segments are not similar. This was clearly the case for cholix toxin (Figure 4.2). However, regions where the sequence is evolving at different rates can be mistaken for homologous recombination. Examination of the sequence alignment suggests that the segments defined by GARD breakpoints contain multiple sequence types, and the sequence types are reassorted through the length of the cholix toxin alleles (Figure 4.3). This would be difficult to explain by point mutation and differential rates of evolution. Many other studies have uncovered significant sequence diversity and homologous recombination associated with the *tcpA* gene, as well as other housekeeping genes, in *V. cholerae* (2, 3, 17, 20). Functional implications for differences in *tcpA* sequences among environmental non-O1, non-O139 strains have also been discovered (2, 24). This suggests that homologous

recombination contributes to the generation of sequence diversity, as new hybrids are created. Alternatively, homologous recombination may also permit the purging of diversity at individual loci, due to selective sweeps of single genes or small genomic regions. Interestingly, it appears that the clade 3 *ctxA* sequences are very nearly clonal. One hypothesis explaining the sequence homogeneity would be a recent selective sweep through a segment of the *V. cholerae* population has recently purged other closely related sequences. Theoretically, the selective pressures resulting in this sweep could be acting on several scales. At the level of the genome, a sweep due to selective pressures on a specific gene could be occurring that is purging diversity at all other loci. This appears to be unlikely, due to the observed homologous recombination in this gene and others in the *V. cholerae* genome. Alternatively, the selective sweep could be occurring at the level of the toxin gene itself, and homologous recombination could be allowing single genes to sweep the population without towing along the rest of the genome. Lastly, the sweep could be occurring at the level of individual segments within the toxin, such that the different domains of the toxin are being swapped among the strains. The latter is supported by the fact that recombination detection analysis found regions of different phylogenies throughout the gene sequence, and the breakpoints were closely aligned with the domain delineations.

The observation that cholera toxin is preferentially found in non-O1, non-O139 strains suggests that the toxin may play an additional role in environmental survival, apart from any putative involvement in virulence. These strains have been responsible

for sporadic disease outbreaks, and they are increasingly being recognized for the wide variety of virulence and toxin mechanisms that they carry. Are all of these toxins needed for chance encounters with susceptible mammals, or do they serve another purpose in the context of selective pressures that they encounter much more frequently? This is still an open question. However, cholix toxin appears to be a powerful tool within the arsenal of *V. cholerae*, and its function in virulence and/or environmental survival in relation to the entire repertoire of factors within the pan-species genome may be important to understanding mechanisms relating to the evolution of secreted toxins in bacteria.

Acknowledgements

We would like to thank the following people for graciously providing *V. cholerae* strains for this analysis: Dr. M. S. Islam, ICDDR, Bangladesh; Dr. S. Jiang, UC Irvine; and Dr. L. Lizarraga-Partida, CICESE, Mexico. We would also like to thank Art Poon, Sergei Kosolovsky Pond, Sheila Podell and Eric Allen for suggestions regarding phylogenetic analyses. This work was supported by the NIH, CEQI, and the SIO graduate department Peterson Fellowship. A.E.P. was the recipient of a Howard Hughes predoctoral fellowship.

Chapter 4, in part, will be submitted for publication. Purdy, A. E., D. Balch, and D. H. Bartlett. Diversity and distribution of cholix toxin among *Vibrio cholerae* strains from around the world. The dissertation author was the primary investigator and author of this paper.

Table 4.1. Primers used in this study.

Amplification of <i>chxA</i> probe	G2TPExoAF TPExoAF2	CGGTTAAGCAAGGGATTGAA CCAATTCTTCGTAAGCGTTCCC
Amplification of <i>chxA</i> alleles	TPToxAF6 L-ToxExpBamHI-LongR or ChxAR3	TCATCCCACATTTGCAGTC TTATTTCAAGTTCATCTTTTCGC GCAACAGCCTCCTCATCAAT
<i>chxA</i> chromosomal location determination	VC1645_hydro_F ChxAR1 VC 1644_hypo_F ChxAF3	GCGCAGTAGTGTGAGTGAG CTTGGAGAAATGGATGCGCTG TCGACACGTTCTTTGAGTCG GAAATATCATCGAGGTGCCTAA
<i>chxA</i> conserved primers	ChxAF2 ChxAR1	TGGTGAAGATTCTCCTGCAA CTTGGAGAAATGGATGCGCTG

Table 4.2. Presence of *chxA* among *Vibrio cholerae* strains isolated in Bangladesh.

Islam lab strain ID	Bartlett lab strain ID	Serogroup	Isolation source	Date of isolation	<i>chxA</i>
WSP-220	B1	O139	Sewage	12/18/00	-
WSP-210	B2	O139	Sewage	12/10/00	-
WSP-264	B3	O139	Sewage	3/4/01	-
WSP-187	B4	O139	Sewage	12/3/00	-
WSP-27	B5	O1	Sewage	6/4/00	-
MSI-14	B6	non-O1	Water	4/4/06	-
A-1431	B7	non-O1	Phytoplankton	4/27/89	-
WSP-230	B8	O139	Sewage	1/2/01	-
MSI-12	B9	non-O1	Canal	3/23/04	+
WSP-168	B10	O139	Sewage	11/12/00	-
WSP-230	B11	O139	Sewage	12/18/00	-
WSP-224	B12	O139	Sewage	12/23/00	-
WSP-44	B13	O139	Sewage	7/9/00	-
A-216	B14	non-O1	Phytoplankton	6/14/88	+
A-532	B15	non-O1	Water	8/30/88	-
A-221	B16	non-O1	Zooplankton	6/14/88	-
MSI-3	B17	non-O1	Canal	1/21/02	-
A-132	B18	non-O1	Plankton	5/31/88	-
A-185	B19	non-O1	Water plant	6/14/88	-
WSP-189	B20	O139	Sewage	12/3/00	-
WSP-39	B21	O139	Sewage	6/18/00	-
A-427	B22	non-O1	Water	8/2/88	-
A-170	B23	non-O1	Soil sediment	6/14/88	-
A-442	B24	non-O1	Water	8/2/88	-
A-81	B25	non-O1	Plankton	4/29/88	+
A-522	B26	non-O1	Phytoplankton	8/30/88	-
A-324	B27	non-O1	Phytoplankton	6/28/88	-
A-273	B28	non-O1	Water	6/28/88	-
A-47	B29	non-O1	Water	4/29/88	+
A-304	B30	non-O1	Water	6/28/88	-
A-356	B31	non-O1	Water	7/12/88	+
A-449	B32	non-O1	Water	8/2/88	+
MSI-5	B33	non-O1	Canal	7/11/02	+
WSP-8	B34	O1	Sewage	5/14/00	-
WSP-10	B35	O1	Sewage	5/21/00	-
WSP-261	B36	O139	Sewage	3/4/01	-
WSP-252	B37	O1	Sewage	1/14/01	-
A-443	B38	non-O1	Phytoplankton	8/2/88	+
MSI-2	B39	non-O1	Pond	1/10/02	-
WSP-232	B40	O1	Sewage	1/7/01	-
WSP-64	B41	O1	Sewage	8/13/00	-
WSP-36	B42	O1	Sewage	6/11/00	-
MSI-9	B43	non-O1	Canal	3/18/03	+/-
A-1443	B44	non-O1	Water plant	4/27/89	-
A-1263	B45	non-O1	Water	3/20/89	-
A-709	B46	non-O1	Water plant	9/27/88	-
A-1275	B47	non-O1	Zooplankton	3/20/89	-

Table 4.2. Presence of *chxA* among *Vibrio cholerae* strains isolated in Bangladesh, continued.

Islam lab strain ID	Bartlett lab strain ID	Serogroup	Isolation source	Date of isolation	<i>chxA</i>
A-1204	B48	non-O1	Water plant	3/7/89	-
MSI-10	B49	non-O1	Canal	12/31/03	-
A-963	B50	non-O1	Water plant	2/23/88	-
A-618	B51	non-O1	Water	9/13/88	-
A-944	B52	non-O1	Soil sediment	12/9/88	-
MSI-6	B53	non-O1	Canal	7/29/02	-
MSI-7	B54	non-O1	Canal	7/29/02	+
MSI-1	B55	non-O1	Pond	12/5/01	+
MSI-8	B56	non-O1	Canal	2/18/03	-
MSI-15	B57	non-O1	Canal	4/26/06	-
MSI-11	B58	non-O1	Canal	3/9/04	-
MSI-4	B59	non-O1	Canal	2/5/02	+
MSI-13	B60	non-O1	Canal	4/5/04	-

Table 4.3. Presence of *chxA* among *Vibrio cholerae* strains isolated in Mexico.

L. L.-P. lab strain ID	Bartlett lab strain ID	Sero-group	Sero-type	Bio-type	clin/env origin ^a	Date of isolation	Isolation location: Mexican state or city	<i>chxA</i>	<i>ctx</i>	<i>tcp</i>
I-19	M1	O1		El Tor	env	1995	Veracruz	+	-	-
I-49	M2	O1		El Tor	env	1995	Veracruz	+	-	-
I-52	M3	O1		El Tor	env	1995	Veracruz	-	-	-
91/01	M4	O1	Inaba	El Tor	clin	1991	Tenaguillo Fd. de Mexico	-		
91/285	M5	O1	Ogawa	El Tor	clin	1991	Hidalgo	-		
91/9011	M6	non-O1			clin	1991	Veracruz	+		
91/13912-a	M7a	non-O1			clin	1991	Veracruz	-		
91/13912-b	M7b	non-O1			clin	1991	Veracruz	-		
92/14800	M8	O1	Inaba	El Tor	clin	1992	Yucatan	-	+	
92/25585	M9	O1	Hikojima	El Tor	clin	1992	Colima	-	+	
93/33227	M10	non-O1			clin	1993	Veracruz	+		
93/38850-a	M11a	O1	Ogawa		clin	1993	Veracruz	+/-	+	
93/38850-b	M11b	O1	Ogawa		clin	1993	Veracruz	+	+	
93/39020-a	M12a	non-O1			clin	1993	Baja Calif Sur	+		
93/39020-b	M12b	non-O1			clin	1993	Baja Calif Sur	+		
93/40592	M13	non-O1			clin	1993	Veracruz	ND		
93/52706	M14	non-O1			clin	1993	Sinaloa	+		
94/55977	M15	non-O1			clin	1994	Veracruz	+		
94/56826	M16	non-O1			clin	1994	Hidalgo	-		
94/58806	M17	O1	Ogawa		clin	1994	Veracruz	-	+	
94/59855	M18	non-O1			clin	1994	Tamaulipas	-		
95/63340-a	M19a	non-O1			clin	1995	Yucatan	+		
95/63340-b	M19b	non-O1			clin	1995	Yucatan	+		
95/66796	M20	non-O1			clin	1995	Tamaulipas	-		
95/69769	M21	O1	Ogawa	El Tor	clin	1995	Oaxaca	+	+	
95/75918	M22	O1	Ogawa	El Tor	clin	1995	Navarit	-		
95/76273	M23	non-O1			clin	1995	Jalisco	+		
95/82398	M24	non-O1			clin	1995	Oaxaca	-		
96/891	M25	non-O1			clin	1996	Chiapas	+		
96/912	M26a	non-O1			clin	1996	Veracruz	+		
96/1032	M27	non-O1			clin	1996	Colima	+		
96/1059	M28	non-O1			clin	1996	Tabasco	-		
96/2440	M29	O1	Ogawa	El Tor	clin	1996	Veracruz	-	+	
97/266	M30	O1	Ogawa		clin	1997	Tamaulipas	-	+	
97/600	M31	non-O1			clin	1997	Veracruz	-		
97/744	M32b	non-O1			clin	1997	Veracruz	+		
97/3168	M33	non-O1			clin	1997	Veracruz	+		
98/191	M34	non-O1			clin	1998	Veracruz	-		
98/232	M35	non-O1			clin	1998	Guerrero	+		

Table 4.3. Presence of *chxA* among *Vibrio cholerae* strains isolated in Mexico, continued.

L. L.-P. lab strain ID	Bartlett lab strain ID	Sero-group	Sero-type	Bio-type	clin/env origin ^a	Year	Isolation location: Mexican state or city	<i>chxA</i>	<i>ctx</i>	<i>tcp</i>
98/659-a	M36a	O1	Ogawa	El Tor	clin	1998	San Luis Potosi	-	-	
98/659-b	M36b	O1	Ogawa	El Tor	clin	1998	San Luis Potosi	-	-	
98/1444c	M37	non-O1			clin	1998	Campeche	+		
99/19	M38	O1	Ogawa		clin	1999	Tamaulipas	-	-	
99/1127	M39	O1	Ogawa		clin	1999	Nuevo Leon	-	-	
99/1468	M40	O1	Ogawa		clin	1999	Edo. de Mex.	-	-	
99/2110	M41	O1	Ogawa		clin	1999	Tamaulipas	-	-	
99/2118	M42	O1	Ogawa		clin	1999	Tamaulipas	-	-	
00/91	M43	O1	Ogawa		clin	2000	Chiapas	-		
00/3079-a	M44a	O1	Ogawa		clin	2000	Tamaulipas	+	-	
00/3079-b	M44b	O1	Ogawa		clin	2000	Tamaulipas	+	-	
00/3240	M45	O1	Ogawa		clin	2000	Campeche	+	-	
00/3259-a	M46a	O1	Ogawa		clin	2000	Campeche	+	-	
00/3259-b	M46b	O1	Ogawa		clin	2000	Campeche	+	-	
00/3260-a	M47a	O1	Ogawa		clin	2000	Campeche	+	-	
00/3260-b	M47b	O1	Ogawa		clin	2000	Campeche	+	-	

Table 4.4. Presence of *chxA* among *Vibrio cholerae* strains from Senegal and Peru.

Lab strain ID	Bartlett Lab Strain ID	Serogroup	Isolation location	Date of isolation	<i>chxA</i>
14VC	SP1	O1	Senegal	2004	-
36VC	SP2	O1	Senegal	2004	-
81VC	SP3	O1	Senegal	2004	-
82VC	SP4	O1	Senegal	2004	-
1/1635	SP5	O1 Ogawa	Perú	1998	-
2/1650	SP6	O1 Ogawa	Perú	1998	-
3/2331	SP7	O1 Ogawa	Perú	1998	-
4/2328	SP8	O1 Ogawa	Perú	1998	-
5/3341	SP9	non-O1, non-O139	Perú	1998	+
6/2334	SP10	O1 Ogawa	Perú	1998	-
7/1657	SP11	O1 Ogawa	Perú	1998	-
8/1689	SP12	O1 Ogawa	Perú	1998	-
9/1046	SP13	O1 Ogawa	Perú	1998	-
12/1175	SP14	O1	Perú	1997	-
13/1137	SP15	O1	Perú	1997	-
14/1178	SP16	O1	Perú	1997	-

Table 4.5. *V. cholerae* strains from southern California: strains carrying *chxA* (27).

Bartlett lab strain designation	Jiang lab designation		
C1	ucic	299	A
C2	ucic	299	H
C7	ucic	305	E
C8	ucic	305	F
C10	ucic	305	I
C11	ucic	305	J
C12	ucic	305	K
C30	occ	624	A
C31	occ	624	C
C32	occ	624	D
C33	occ	624	E
C37	occ	1004	A
C39	occ	1004	D
C42	bc	803	D
C43	bc	803	E
C52	bc	2000	A
C57	bp	2000	A
C58	bp	2000	C
C62	CRO1	628	E
C63	CRO1	628	F
C64	CRO1	628	I
C65	CRO2	628	A
C66	CRO2	628	B
C67	CRO2	628	C
C68	CRO2	628	D
C70	CRO2	628	F
C71	CRO2	628	G
C72	CRO2	628	H
C73	CRO3	628	A
C74	CRO3	628	B
C75	CRO3	628	C
C76	CRO3	628	D
C158	CRO1	2000	A
C160	CRO1	2000	H
C163	CRO2	2000	B
C164	CRO2	2000	C
C170	CRO3	2000	A

Table 4.6. Additional strains which were tested for *chxA*.

Bartlett lab ID	Serogroup	Date of isolation	Isolation location	Reference	<i>chxA</i>
TP	non-O1, non-O139	2000	Torrey Pines Estuary, San Diego, CA	(27)	+
SIO	non-O1, non-O139	2000	Scripps Institution of Oceanography pier, San Diego, CA	(27)	+
MB2	non-O1, non-O139	2002	Mission Bay, San Diego, CA	(27)	+
V1A4C	non-O1, non-O139	2005	Lake Pontchartrain, New Orleans, LA		+
NRT36S*	O31	prior to 1986	Clinical, Japan, from southeast Asia	(1, 4)	+
MZO-2*	O14	2001	Clinical, Bangladesh	(8)	+
1587*	O12	1995	Peru	(6)	+
623-39	NA	NA	NA		+
N16961*	O1 El Tor	1975	Clinical	Bangladesh (8)	(incomplete sequence) -
V51*	O141	1987	Clinical, USA		+
					(incomplete sequence)

* Denotes strains with genomes that have been fully or partially sequenced which are readily available in the public databases. NA, not available.

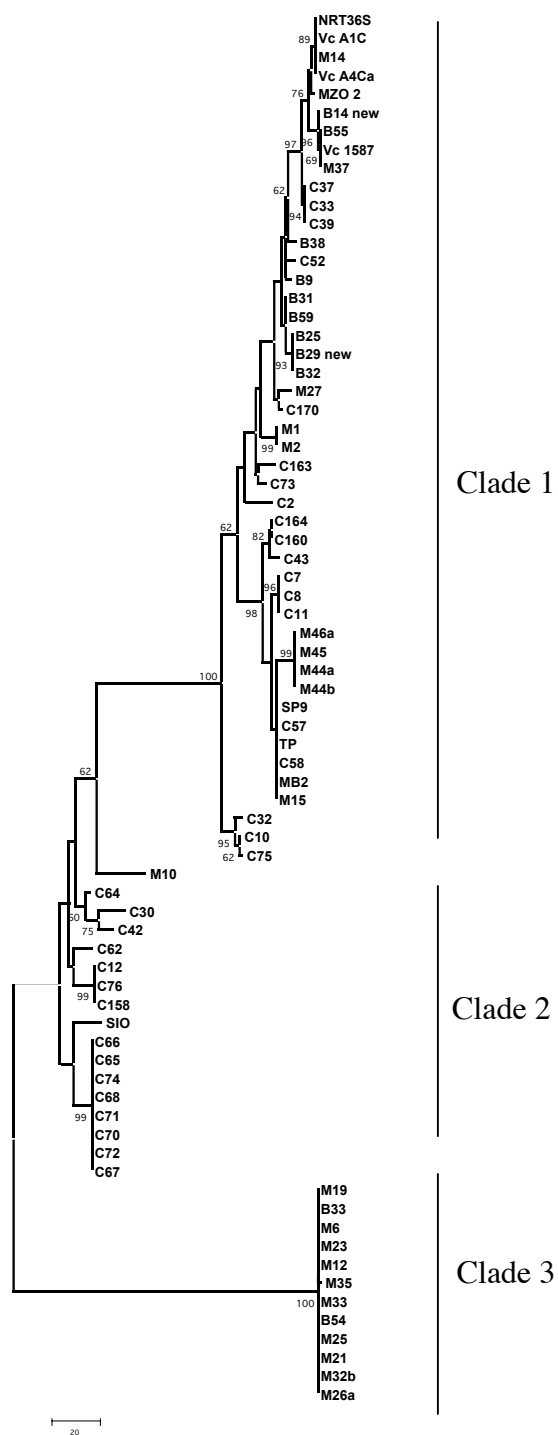
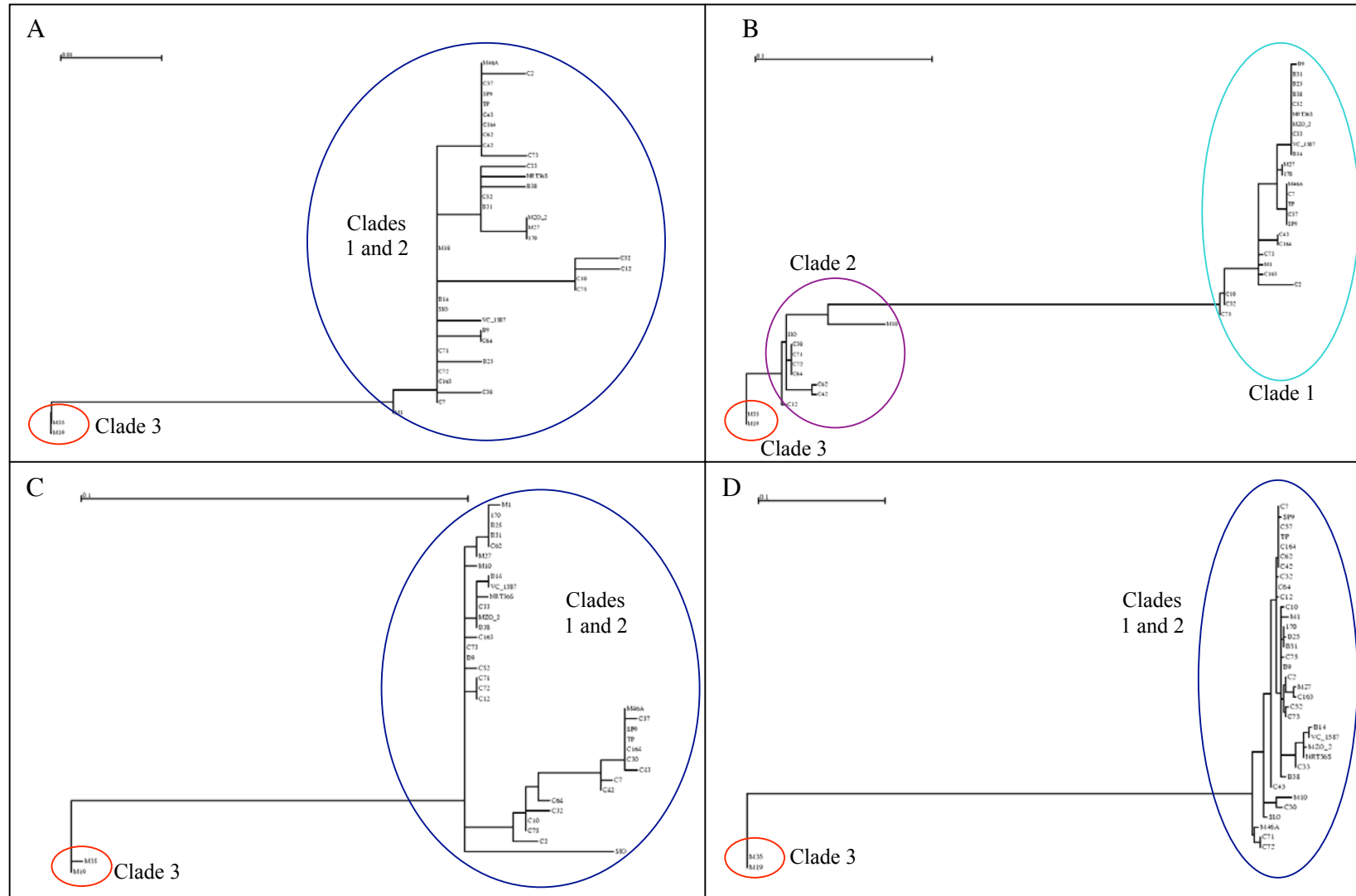


Figure 4.1. *chxA* DNA sequences fall into three clades. A maximum parsimony analysis of 1803 bp of aligned *chxA* sequence was performed, and bootstrap values $\geq 60\%$ are displayed.

Figure 4.2. Neighbor joining trees automatically generated by GARD for each of the four segments resulting from the three assigned recombination breakpoints. (A) Segment 1, first half of domain Ia. (B) Segment 2, second half of domain Ia. (C) Segment 3, domain II and half of domain Ib. (D) Segment 4, Second half of domain Ib and all of domain III. Clade designations are based upon results from Figure 4.1.



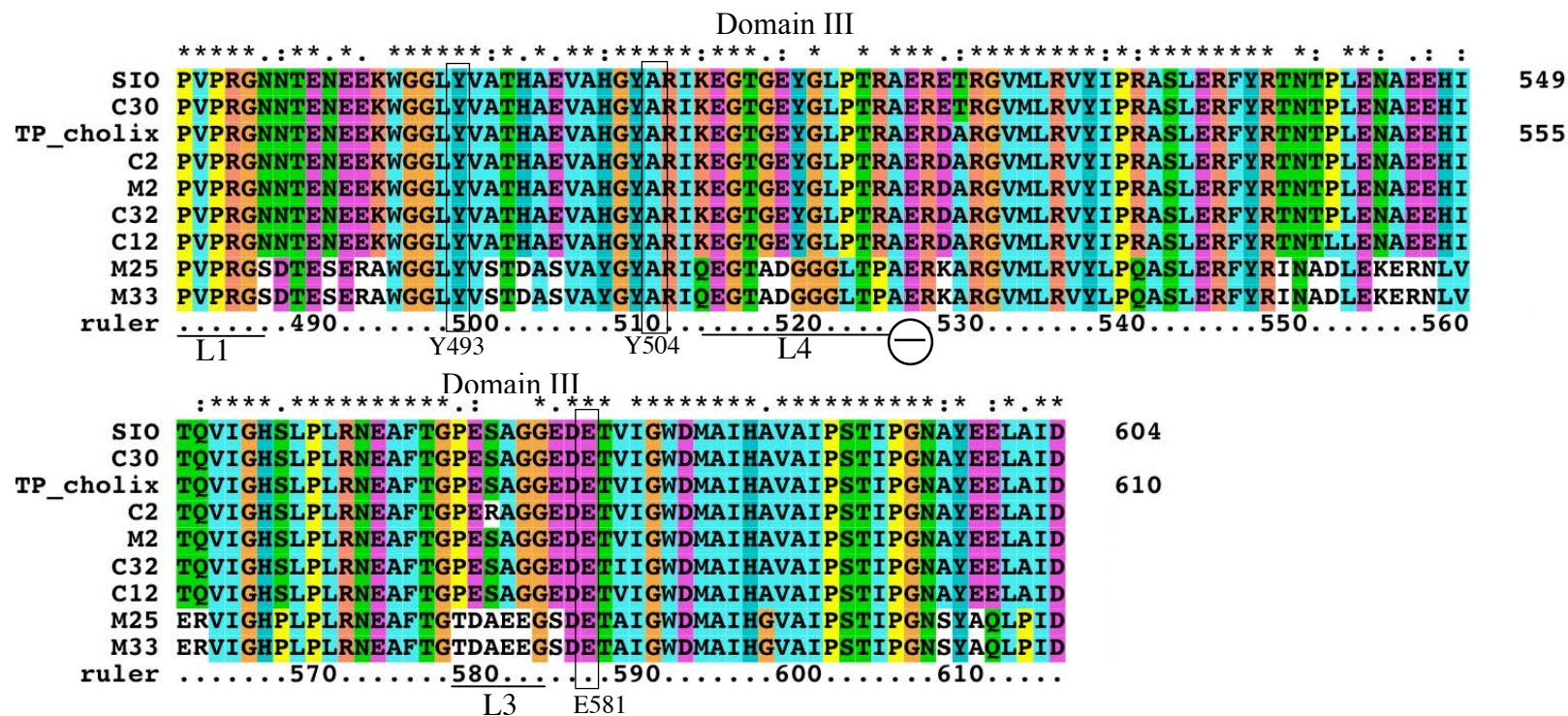


Figure 4.3. Annotated cholix toxin protein sequence alignment, continued. Divisions between protein domains are designated by a vertical double line. Recombination breakpoints (RecBr) assigned by GARD are indicated by a single vertical line. Residues undergoing negative selection are indicated by a circle with a negative sign. Residues of particular importance are outlined. Catalytic loop regions (L1 - L4) are underlined. Clade 1: TP cholix, C2, M2, C32; Clade 2: S10, C30, C12; Clade 3: M25, M33.

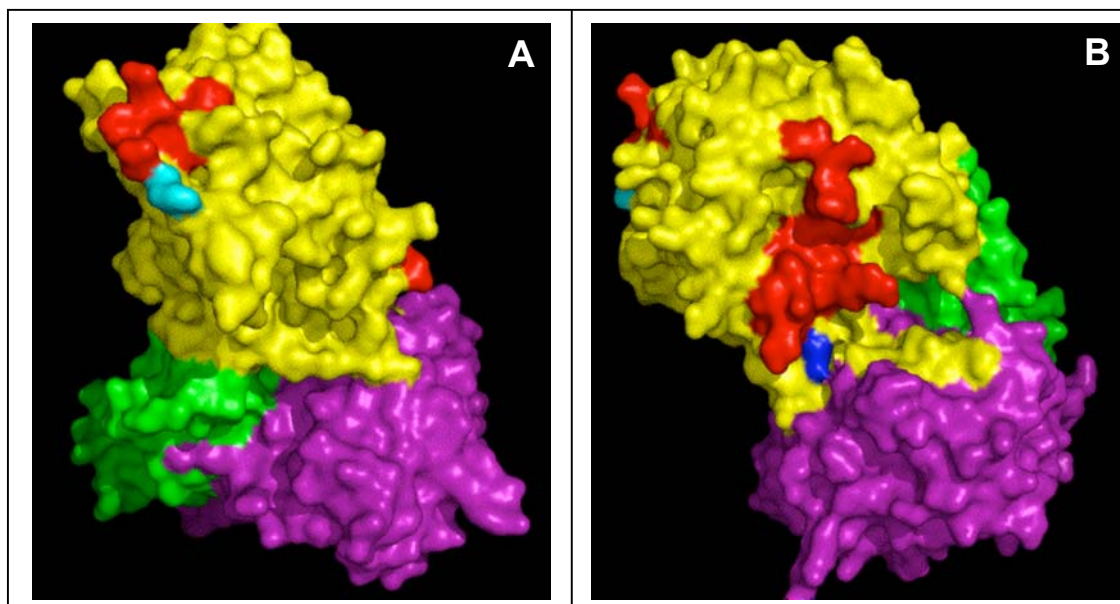


Figure 4.4. Surface view of cholix toxin structure: regions of interest within the receptor binding domain. Domains Ia and Ib are yellow, domain II is green, and domain III is purple. The two insertion-deletion regions subject to variability between sequences in clades 1, 2, and 3 are indicated in red. (A) The lysine residue putatively involved in receptor binding is in cyan. (B) The P386 residue in domain Ib that is subject to negative selection is in blue. The E133 residue subject to negative selection is not visible, as it is not on the surface of the protein.

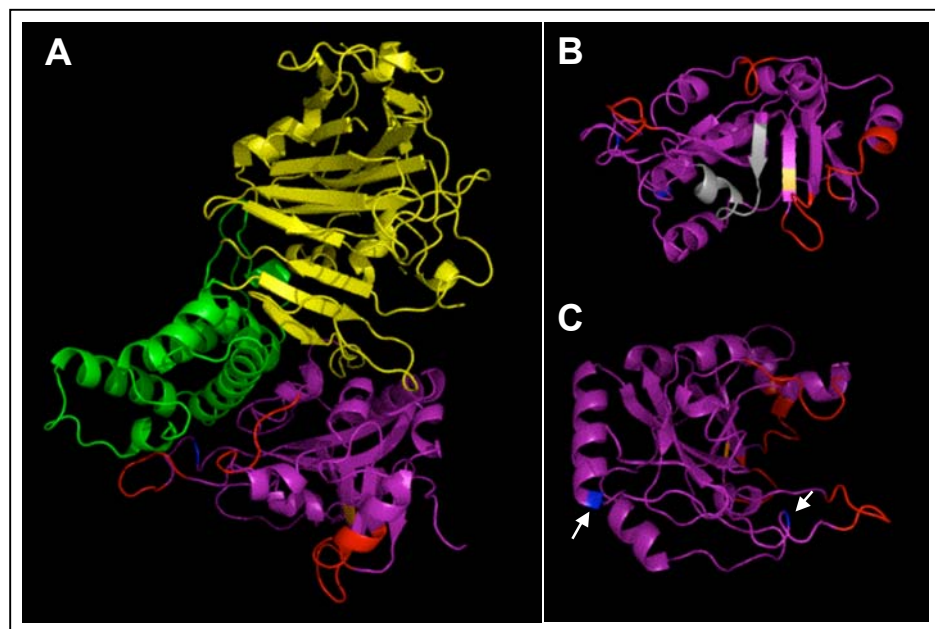


Figure 4.5. Cholix toxin protein structure: regions of interest within the catalytic domain. (A) Domains Ia and Ib are yellow, domain II is green, and domain III is purple. (B) Domain III only. The four loops surrounding the catalytic site are in red. The NAD⁺ binding site between Tyr-493 and Tyr-504 is in grey. The glutamate residue critical to ADPRT function (Glu-581) is in orange. (C) An alternative view of domain III, with arrows pointing out residues subject to negative selection (Tyr-430 and Glu-513) that are in blue.

References

1. **Arita, M., T. Takeda, T. Honda, and T. Miwatani.** 1986. Purification and characterization of *Vibrio cholerae* non-O1 heat-stable enterotoxin. *Infect. Immun.* **52**:45-49.
2. **Boyd, E. F., and M. K. Waldor.** 2002. Evolutionary and functional analyses of variants of the toxin-coregulated pilus protein TcpA from toxigenic *Vibrio cholerae* non-O1/non-O139 serogroup isolates. *Microbiology* **148**:1655-1666.
3. **Byun, R., L. D. H. Elbourne, R. Lan, and P. R. Reeves.** 1999. Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infect. Immun.* **67**:1116-1124.
4. **Chen, Y., J. A. Johnson, G. D. Pusch, J. G. Morris, Jr., and O. C. Stine.** 2007. The genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from those of O1 *Vibrio cholerae*. *Infect. Immun.* **75**:2645-2647.
5. **Colwell, R. R.** 1996. Global climate and infectious disease: the cholera paradigm. *Science* **274**:2025-2031.
6. **Dalsgaard, A., M. J. Albert, D. N. Taylor, T. Shimada, R. Meza, O. Serichantalergs, and P. Echeverria.** 1995. Characterization of *Vibrio cholerae* non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru. *J. Clin. Microbiol.* **33**:2715-2722.
7. **Douglas, C. M., and R. J. Collier.** 1990. *Pseudomonas aeruginosa* exotoxin A: alterations of biological and biochemical properties resulting from mutation of glutamic acid 553 to aspartic acid. *Biochemistry* **29**:5043 - 5049.
8. **Dziejman, M., D. Serruto, V. C. Tam, D. Sturtevant, P. Diraphat, S. M. Faruque, M. H. Rahman, J. F. Heidelberg, J. Decker, L. Li, K. T. Montgomery, G. Grills, R. Kucherlapati, and J. J. Mekalanos.** 2005. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proc. Nat. Acad. Sci. U. S. A.* **102**:3465-3470.

9. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**:1792-1797.
10. **Faruque, S. M., N. Chowdhury, M. Kamruzzaman, M. Dziejman, M. H. Rahman, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2004. Genetic diversity and virulence potential of environmental *Vibrio cholerae* population in a cholera-endemic area. *Proc. Natl. Acad. Sci. U. S. A.* **101**:2123-2128.
11. **Fullner, K. J., J. C. Boucher, M. A. Hanes, G. K. Haines, B. M. Meehan, C. Walchle, P. J. Sansonetti, and J. J. Mekalanos.** 2002. The contribution of accessory toxins of *Vibrio cholerae* O1 El Tor to the proinflammatory response in a murine pulmonary cholera model. *J. Exp. Med.* **195**:1455-1462.
12. **Ghosh, A., D. R. Saha, K. M. Hoque, M. Asakuna, S. Yamasaki, H. Koley, S. S. Das, M. K. Chakrabarti, and A. Pal.** 2006. Enterotoxigenicity of mature 45-kilodalton and processed 35-kilodalton forms of hemagglutinin protease purified from a cholera toxin gene-negative *Vibrio cholerae* non-O1, non-O139 strain. *Infect. Immun.* **74**:2937-2946.
13. **Halpern, M., H. Gancz, and Y. Kashi.** 2003. *Vibrio cholerae* hemagglutinin/protease degrades chironomid egg masses. *Appl. Environ. Microbiol.* **69**:4200-4204.
14. **Jinno, Y., V. K. Chaudhary, T. Kondo, S. Adhya, D. J. FitzGerald, and I. Pastan.** 1988. Mutational analysis of domain I of Pseudomonas exotoxin. Mutations in domain I of Pseudomonas exotoxin which reduce cell binding and animal toxicity. *J. Biol. Chem.* **263**:13203-13207.
15. **Jorgensen, R., A. R. Merrill, S. P. Yates, V. E. Marquez, A. L. Schwan, T. Boesen, and G. R. Andersen.** 2005. Exotoxin A-eEF2 complex structure indicates ADP ribosylation by ribosome mimicry. *Nature* **436**:979-984.
16. **Jorgensen, R., A. E. Purdy, R. J. Fieldhouse, M. S. Kimber, D. H. Bartlett, and A. R. Merrill.** 2007. Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. Submitted.
17. **Karaolis, D. K. R., R. Lan, and P. R. Reeves.** 1995. The sixth and seventh cholera pandemics are due to independent clones separately derived from

- environmental, nontoxigenic, non-O1 *Vibrio cholerae*. J. Bacteriol. **177**:3191-3198.
18. **Kosakovsky Pond, S. L., and S. D. W. Frost.** 2005. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol Biol Evol **22**:1208-1222.
 19. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost.** 2006. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. Mol Biol Evol **23**:1891-1901.
 20. **Kotetishvili, M., O. C. Stine, Y. Chen, A. Kreger, A. Sulakvelidze, S. Sozhamannan, and J. G. Morris, Jr.** 2003. Multilocus sequence typing has better discriminatory ability for typing *Vibrio cholerae* than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness. J. Clin. Microbiol. **41**:2191-2196.
 21. **Kounnas, M. Z., R. E. Morris, M. R. Thompson, D. J. FitzGerald, D. K. Strickland, and C. B. Saelinger.** 1992. The alpha 2-macroglobulin receptor/low density lipoprotein receptor-related protein binds and internalizes Pseudomonas exotoxin A. J. Biol. Chem. **267**:12420 – 12423.
 22. **Markowitz, V. M., F. Korzeniewski, K. Palaniappan, E. Szeto, G. Werner, A. Padki, X. Zhao, I. Dubchak, P. Hugenholtz, I. Anderson, A. Lykidis, K. Mavromatis, N. Ivanova, and N. C. Kyrpides.** 2006. The integrated microbial genomes (IMG) system. Nucl. Acids Res. **34**:D344-348.
 23. **Matz, C., D. McDougald, A. M. Moreno, P. Y. Yung, F. H. Yildiz, and S. Kjelleberg.** 2005. Biofilm formation and phenotypic variation enhance predation-driven persistence of *Vibrio cholerae*. Proc. Natl. Acad. Sci. U. S. A. **102**:16819-16824.
 24. **Mukhopadhyay, A. K., S. Chakraborty, Y. Takeda, G. B. Nair, and D. E. Berg.** 2001. Characterization of VPI pathogenicity island and CTXphi prophage in environmental strains of *Vibrio cholerae*. J. Bacteriol. **183**:4737-4746.

25. **Pond, S. L. K., and S. D. W. Frost.** 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**:2531-2533.
26. **Pukatski, S., A. T. Ma, D. Sturtevant, B. Krastins, D. Sarracino, W. C. Nelson, J. F. Heidelberg, and J. J. Mekalanos.** 2006. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* model system. *Proc. Natl. Acad. Sci. U. S. A.* **103**:1528-1533.
27. **Purdy, A., F. Rohwer, R. Edwards, F. Azam, and D. H. Bartlett.** 2005. A glimpse into the expanded genome content of the *Vibrio cholerae* species through the identification of genes present in environmental strains. *J. Bacteriol.* **187**:2992-3001.
28. **Reguera, G., and R. Kolter.** 2005. Virulence and the environment: a novel role for *Vibrio cholerae* toxin-coregulated pili in biofilm formation on chitin. *J. Bacteriol.* **187**:3551-3555.
29. **Sarkar, B., T. Bhattacharya, y. T. Ramamurth, T. Shimada, Y. Takeda, and G. B. Nair.** 2002. Preferential association of the heat-stable enterotoxin gene (*stn*) with environmental strains of *Vibrio cholerae* belonging to the O14 serogroup. *Epidemiol. and Infect.* **129**:245-251.
30. **Sharma, C., M. Thungapathra, A. Ghosh, A. K. Mukhopadhyay, A. Basu, R. Mitra, I. Basu, S. K. Bhattacharya, T. Shimada, T. Ramamurthy, T. Takeda, S. Yamasaki, Y. Takeda, and G. B. Nair.** 1998. Molecular analysis of non-O1, non-O139 *Vibrio cholerae* associated with an unusual upsurge in the incidence of cholera-like disease in Calcutta, India. *J. Clin. Microbiol.* **36**:756-763.
31. **Singh, D. V., M. H. Matte, G. R. Matte, S. Jiang, F. Sabeena, B. N. Shukla, S. C. Sanyal, A. Huq, and R. R. Colwell.** 2001. Molecular analysis of *Vibrio cholerae* O1, O139, non-O1, and non-O139 strains: Clonal relationships between clinical and environmental isolates. *Appl Environ Microbiol.* **67**:910-921.
32. **Vaitkevicius, K., B. Lindmark, G. Ou, T. Song, C. Toma, M. Iwanaga, J. Zhu, A. Andersson, M.-L. Hammarstrom, S. Tuck, and S. N. Wai.** 2006. A *Vibrio cholerae* protease needed for killing of *Caenorhabditis elegans* has a

role in protection from natural predator grazing. Proc. Natl. Acad. Sci. U. S. A. **103**:9280-9285.

33. **Yates, S. P., R. Jorgensen, G. R. Andersen, and A. R. Merrill.** 2006. Stealth and mimicry by deadly bacterial toxins. Trends Biochem Sci. **31**:123-133.

Chapter 5.

Ecological and pathogenic implications for cholera toxin in *Vibrio cholerae*: future research directions

Cholera disease continues to threaten the world's most vulnerable populations in regions where clean drinking water is unavailable. The disease is caused by toxigenic *Vibrio cholerae*, and it spreads rapidly because the primary symptom, the production of 'rice stool,' returns large concentrations of *V. cholerae* to the environment. The action of cholera toxin, an ADP-ribosylating toxin that affects ion channels, is responsible for the main attributes of the disease, but only pandemic strains of the O1 or O139 serogroups consistently carry the *ctxAB* genes that encode the toxin. Other *V. cholerae* strains, the non-O1, non-O139 strains, can be responsible for sporadic disease outbreaks with symptoms distinct from those of cholera. Though the phylogenetic diversity of these strains had been well documented via several different whole genome PCR methods-based methods (6, 7, 11), Chapter 2 of this thesis was the first work describing the accompanying genomic sequence variation. This study provided a glimpse into the types of genes comprising the 'flexible genome' and placed constraints on the amount of changeable sequence. Unexpectedly, the pan-species genome of *V. cholerae* contains an expanded virulence-like gene repertoire; gene segments with similarity to toxin genes from other bacterial

pathogens were discovered in the genomes of environmental *V. cholerae* strains TP and SIO, isolated nearby to Scripps Institution of Oceanography in La Jolla, CA. From these segments, the full-length genes were then sequenced. The genes included a putative secreted adenylate cyclase and a cytotoxic necrotizing factor (CNF). CNF is a virulence factor of uropathogenic *E. coli* (2, 4), but the sequence recovered from *V. cholerae* suggested that significant portions of the enzyme had been deleted, most likely rendering the toxin non-functional (P. Ghosh, personal communication). The adenylate cyclase is tethered to a putative membrane bound protein, unlike other adenylate cyclase domains, and this gene is being studied in detail by other researchers. However, alignments indicated that a gene with similarity to exotoxin A from *Pseudomonas aeruginosa* had conserved residues required for catalysis, and this toxin was chosen for further study.

The gene encoding this toxin, *chxA*, was first described in *V. cholerae* strains isolated from coastal ocean regions, though it was later discovered in non-O1, non-O139 strains of both clinical and environmental origins. This thesis presents the first steps towards developing a comprehensive understanding of the new toxin, now designated cholix toxin, from biochemical, phylogenetic, ecological, and virulence perspectives. The toxin is enzymatically active, and it is a *bona fide* third member of the eEF2-specific ADP-ribosylating toxins, a small class of toxins consisting only of diphtheria toxin from *Corynebacterium diphtheriae* and exotoxin A from *Pseudomonas aeruginosa*. Cholix toxin inhibits protein synthesis in mammalian tissue culture cell lines, thereby demonstrating that the features required for cell entry

-- the receptor binding domain, furin cleavage site, a particular disulfide bridge, and a KDEL endoplasmic reticulum retention sequence -- were also conserved and functional *in vivo*. Our collaborators at the University of Guelph in Ontario, Canada, have solved the crystal structure for both the catalytic domain and the full length toxin and have shown that despite relatively low sequence conservation, the structure is almost identical to that of exotoxin A from *P. aeruginosa* (8). The importance of cholix toxin to *V. cholerae* is further suggested by the presence of the *chxA* gene in strains from around the world, and in each of these alleles, catalytically-important residues are conserved. Intriguingly, functional diversity is suggested by polymorphisms that are present in regions that could have ecological ramifications; for example, sequence changes present in a surface loop of the receptor-binding domain could alter binding affinity or specificity.

Cholix toxin most likely targets eukaryotic cells. Conserved amino acid sequence features strongly suggest that the primary target of cholix toxin is a eukaryotic organism. Bacteria are not affected because they cannot synthesize diphthamide, the modified histidine residue within eEF2 that is targeted by the ADP-ribosylation reaction (13). Cholix toxin also retains the canonical KDEL endoplasmic reticulum (ER) retention sequence at its C-terminus, suggesting that the infection mechanism involves retrograde transport from the Golgi apparatus to the ER, both of which are organelles specific to the Eukarya. Moreover, cholix toxin most likely targets multicellular eukaryotic organisms that carry a receptor similar to low density lipoprotein receptor-like protein (LRP1), the receptor for exotoxin A (9) due to the

structural similarity between the receptor binding domains of exotoxin A and cholix toxin (8). However, it is entirely possible that another receptor on the surface of a single-celled eukaryotic organism could act as an additional, or even the primary, receptor.

Searching for cholix toxin contributions to (virulence) phenotypes. After discovering a suite of virulence-like genes in our *V. cholerae* strains from southern California, it seemed sensible to explore potential virulence phenotypes conferred by these strains. With help from numerous groups around the country, we investigated the ability of *V. cholerae* strain TP to cause disease in hybrid striped bass, penaeid shrimp, *Drosophila*, Eastern oysters, mice, *C. elegans*, *Artemia*, *Tigriopus* copepods, and abalone larvae. In most cases, TP had no effect on the organisms, but a dramatic virulence phenotype in mice was discovered by our collaborators at the VA Hospital at UCSD. In a pulmonary model of bacterial infection, TP caused extreme fluid accumulation and extensive damage to the lung architecture, most likely due to the effects of secreted toxins, because only ~0.1% of the inoculum could be recovered after 24 hours (10). These results indicate that the strains harbor virulence gene repertoires sufficient to cause disease symptoms *in vivo*. Mortality was also observed in a *Drosophila* model of cholera pathogenesis (1) and in *Artemia* nauplii at high doses. However, most interactions between *V. cholerae* TP and eukaryotes did not result in mortality of the host organism, perhaps due to a lack of expression of virulence genes, an effective host defense response, or the actions of commensal bacteria.

In species in which a virulence phenotype was observed, we concurrently performed experiments using *V. cholerae* TP with an in-frame markerless deletion of the *chxA* gene, but this strain displayed no attenuated virulence properties. The inability to detect a difference in phenotype is most likely due to the presence of other toxins in the genomes of these strains (10) such as the RTX toxin, hemagglutinin/protease (HA/P), hemolysin, the VAS operon encoding a type VI secretion system, and perhaps other toxin-like genes such as the CNF or the adenylate cyclase, and other unknown factors (5). In order to differentiate the actions of cholix toxin, a series of *V. cholerae* TP strains containing in-frame deletions in each of these putative virulence genes should be created. This approach has been taken by researchers studying *P. aeruginosa* secreted toxins, because this species also expresses a multitude of virulence factors (12).

Analyses of the distribution of *chxA* among strains from a global collection of *V. cholerae* suggested that it has been lost in the vast majority of O1 and O139 strains. Is this because, in combination with the O1 and O139 strain backgrounds (which include the TCP pilus and CTX), the toxin confers a disadvantageous phenotype, perhaps invoking a stronger immune response? Or rather could the presence of this toxin increase the pathogenicity of these strains? Because *chxA* is integrated into the *V. cholerae* genome in a single location on chromosome 1, a toxigenic O1 or O139 strain could be constructed in which the *chxA* gene has been recombined into the genome. Using models of *V. cholerae* infection, the virulence and colonization potential of these strains, as well as the host immune response, could be monitored for

changes that might suggest reasons for the near mutual exclusivity of cholix toxin and cholera toxin.

Cholix toxin may be acting in ways more subtle than simple virulence experiments would detect, perhaps in conjunction with other toxins or in very limited or specific circumstances. The toxin may be important in situations aside from virulence, such as colonization of a host organism. These phenotypes might be identified only in the context of competition experiments, in which *V. cholerae* strains with and without the *chxA* gene begin an experiment in a roughly 1:1 ratio, and differences in survival and growth are measured at the finish. The two *V. cholerae* strains should also be differentially labeled with a marker that itself will not affect growth, such as an identical plasmid carrying different drug resistance markers or a small deletion in *lacZ* in one strain. By competing the strains against one another, it is possible to determine circumstances in which the toxin may be conferring an advantage, though the mechanism may still be unclear. Experiments could be performed under highly complex conditions, such as those present within whole seawater microcosms that include zooplankton, phytoplankton, protists, viruses, and other bacteria, in order to gain a sense of an overall growth advantage granted by the toxin under realistic conditions. Alternatively, it is possible that cholix toxin confers an advantage under very specific conditions that might be too subtle to detect in a whole seawater microcosm. Thus designing experiments in which colonization and competition are measured in association with individual zooplankton, phytoplankton or protist species may be more informative.

However, it may be impossible to discover situations when toxin secretion is imperative for bacterial survival or growth using these types of laboratory-based measurements. For example, if the true host is in fact a marine mammal or a larger marine organism, it is impossible or impractical to investigate *V. cholerae* colonization or infection in a controlled experiment.

Expression analyses will suggest new hypotheses. Results not presented here indicate that the *chxA* gene is indeed expressed by *V. cholerae* strain TP in LB at 37°C, and preliminary results obtained by our collaborators suggest that the toxin is expressed *in vivo* in a pulmonary mouse model as well (M. Makri and J. Fierer, personal communication). Expression of *chxA* in response to a variety of conditions such as temperature, pH, nutrient levels, growth phase, salinity, viscosity, iron concentrations, O₂ concentrations, and biofilm formation can be monitored by quantitative reverse transcriptase-PCR (qRT-PCR) or the construction of *lacZ* transcriptional or translational fusions to the *chxA* promoter or gene sequence. Regulatory elements controlling cholix toxin expression could be discovered by screening a Tn5-mutant library in a *V. cholerae* TP *chxA::lacZ* strain for colonies up- or down-regulating beta-galactosidase expression. Furthermore, investigating variations in expression and regulation among strains carrying *chxA* genes that fall into divergent clades, as presented in this thesis, might suggest further hypotheses regarding the significance of these phylogenetic groupings.

Genome-wide studies of gene expression using microarrays will also reveal genes that might be co-expressed with *chxA*. However, microarray slides need to be

printed using oligos that are representative of the *V. cholerae* pan-genome, as currently only microarrays with open reading frames from N16961 are available. Expression profiles of *V. cholerae* strains carrying both the *ctxA* gene and deletions in regulators may be instructive in terms of discerning genes that are under control of similar regulatory systems, and this might prove to be an effective method to hypothesize synergistic interactions between cholix toxin and other toxins.

Toxin expression should also be investigated under conditions found in the aquatic environment, but initial preliminary studies not described here encountered methodological and experimental design challenges, primarily due to the small concentrations of *V. cholerae* needed for these experiments, and the subsequent small concentrations of RNA isolated. Furthermore, investigating gene expression in *V. cholerae* in concert with a eukaryotic host may face some initial methodological hurdles, but the results might be especially informative. Together results from *in vitro* and *in vivo* expression studies might provide clues as to when the toxin may be deployed by *V. cholerae*.

Additional considerations.

Is cytotoxicity the relevant effect of cholix toxin? Cholix toxin halts protein synthesis in eukaryotic cells within minutes to hours, which ultimately leads to cell death. However, the role that this plays in *V. cholerae* infection, colonization, pathogenesis, or survival is not clear because cytotoxicity is not directly observed until 48 hours following toxin addition. This suggests that the most important effect of the toxin may be the immediate cessation of protein synthesis in the targeted cell type.

During infection or colonization, this may affect the adaptive and/or innate immune responses by preventing cell division, cell activation, phagocytosis, or secretion of toxic compounds. Alternatively, the toxin could be changing protein expression or activating a particular cell death pathway that might benefit *V. cholerae*. These are questions best addressed in collaboration with cell biologists, but they may be critical to understanding cholix function.

Receptor binding: novel receptors and structure-function relationships inferred from sequence variation. The discovery of novel features within the receptor binding domains of some *chxA* sequences from southern California may suggest that cholix toxin is targeting different receptors, or has different receptor-binding affinity for the LRP1 receptor relative to the other sequence types.

Interestingly, recombinant proteins consisting of the catalytic domain of exotoxin A fused to the receptor-associated protein (RAP), a known ligand of LRP1, were less toxic than exotoxin A, although binding affinity for the receptor was higher (14). The relatively high binding affinity might prevent dissociation of RAP from LRP1, allowing the toxin to be transported to the lysosomes and degraded (14). Furthermore, new receptors for cholix toxin could be discovered by screening protein extracts from a variety of organisms and tissues to look for ligands binding to the receptor binding domain, using standard protein purification techniques.

Ecological implications for secreted toxins. Cholix toxin is a secreted toxin; it carries an N-terminal signal sequence that is detected by SignalP (3). Unless the toxin is associated with the surface of the bacterium following secretion, expression is

futile in a free-living state in an aquatic environment, due to immediate dilution effects. Therefore it is more likely that the toxin is deployed in a setting where concentrations in a localized microenvironment can be significant, or when effects at a remote site would be advantageous, such as during an infection cycle.

Summary

A role for cholix toxin in *V. cholerae* survival remains undefined. We have shown that it has the potential to be wielded as a cytotoxic factor against eukaryotic organisms. In multicellular eukaryotic organisms, this may result in a virulence phenotype, depending on the expression level and the targeted cell type. It is feasible that cholix toxin may be used to establish parasitic or mutualistic associations that could be subtle or transitory. Alternatively, if the targeted organism is single-celled, the toxin would have to be deployed in a microenvironment in order to prevent immediate dilution.

The presence of virulence and virulence-like genes in bacterial genomes is much more extensive than previously appreciated. Their roles in bacterial survival may extend beyond human infection to include environmental defense mechanisms and colonization of eukaryotic organisms. The diversity and number of virulence-like factors within single genomes (especially *Vibrios* and *P. aeruginosa*) suggests that there may be complex regulatory mechanisms governing their relative expression under different conditions. Furthermore, not all *V. cholerae* strains carry the same complement of virulence factors, and selective pressures affecting these genes are not

well understood. The presence of other toxins may influence the types of genes maintained in the genomes based upon their ability to influence access to specific niches. In the future, high throughput genome sequencing will enable a better understanding of the correlation of genomic backgrounds with specific genes within a species. Deciphering the signals that regulate expression may reveal the selective pressures that lead to the evolution of toxins, and their functions in pathogenesis, environmental survival, or symbiosis.

References

1. **Blow, N. S., R. N. Salomon, K. Garrity, I. Reveillaud, A. Kopin, F. R. Jackson, and P. I. Watnick.** 2005. *Vibrio cholerae* infection of *Drosophila melanogaster* mimics the human disease cholera. PLoS Pathog. **1**:e8.
2. **Boquet, P.** 2001. The cytotoxic necrotizing factor 1 (CNF1) from *Escherichia coli*. Toxicon **39**:1673-1680.
3. **Dyrlov Bendtsen, J., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology **340**:783-795.
4. **Falbo, V., T. Pace, L. Picci, E. Pizzi, and A. Caprioli.** 1993. Isolation and nucleotide sequencing of the gene encoding cytotoxic necrotizing factor type 1. Infect. Immun. **9**:1247-1254.
5. **Faruque, S. M., N. Chowdhury, M. Kamruzzaman, M. Dziejman, M. H. Rahman, D. A. Sack, G. B. Nair, and J. J. Mekalanos.** 2004. Genetic diversity and virulence potential of environmental *Vibrio cholerae* population in a cholera-endemic area. Proc. Natl. Acad. Sci. U. S. A. **101**:2123-2128.
6. **Jiang, S. C., V. Louis, N. Choopun, A. Sharma, A. Huq, and R. R. Colwell.** 2000. Genetic diversity of *Vibrio cholerae* in Chesapeake Bay determined by amplified fragment length polymorphism fingerprinting. Appl. Environ. Microbiol. **66**:140-147.
7. **Jiang, S. C., M. Matte, G. Matte, A. Huq, and R. R. Colwell.** 2000. Genetic diversity of clinical and environmental isolates of *Vibrio cholerae* determined by amplified fragment length polymorphism fingerprinting. Appl. Environ. Microbiol. **66**:148-153.
8. **Jorgensen, R., A. E. Purdy, R. J. Fieldhouse, M. S. Kimber, D. H. Bartlett, and A. R. Merrill.** 2007. Cholix toxin, a novel ADP-ribosylating factor from *Vibrio cholerae*. Submitted.

9. **Kounnas, M. Z., R. E. Morris, M. R. Thompson, D. J. FitzGerald, D. K. Strickland, and C. B. Saelinger.** 1992. The alpha 2-macroglobulin receptor/low density lipoprotein receptor-related protein binds and internalizes *Pseudomonas* exotoxin A. *J. Biol. Chem.* **267**:12420 – 12423.
10. **Makri, S., A. E. Purdy, D. Bartlett, and J. Fierer.** 2007. Pathogenicity of environmental isolates of *V. cholerae* in mice. *Microbes and Infection* **In press**.
11. **Singh, D. V., M. H. Matte, G. R. Matte, S. Jiang, F. Sabeena, B. N. Shukla, S. C. Sanyal, A. Huq, and R. R. Colwell.** 2001. Molecular analysis of *Vibrio cholerae* O1, O139, non-O1, and non-O139 strains: Clonal relationships between clinical and environmental isolates. *Appl Environ Microbiol.* **67**:910-921.
12. **Vallis, A. J., V. Finck-Barbancon, T. L. Yahr, and D. W. Frank.** 1999. Biological effects of *Pseudomonas aeruginosa* type III-secreted proteins on CHO cells. *Infect. Immun.* **67**:2040-2044.
13. **Yates, S. P., R. Jorgensen, G. R. Andersen, and A. R. Merrill.** 2006. Stealth and mimicry by deadly bacterial toxins. *Trends Biochem Sci.* **31**:123-133.
14. **Zdanovsky, A. G., M. V. Zdanovskaia, D. Strickland, and D. J. FitzGerald.** 1996. Ligand-toxin Hybrids Directed to the alpha(2)-Macroglobulin Receptor/Low Density Lipoprotein Receptor-related Protein Exhibit Lower Toxicity than Native *Pseudomonas* Exotoxin. *J. Biol. Chem.* **271**:6122-6128.