Genomics and Population Dynamics of *Vibrio cholerae* Causing Pandemic Cholera

Shah Manzur Rashed
Registration Number: 98 (Old Reg. No. 157)
Session: 2016-2017 (Old Session: 2011-2012)

Department of Microbiology
University of Dhaka, Dhaka-1000
Maryland Pathogen Research Institute
University of Maryland College Park

Dissertation submitted to the Faculty of Biological Sciences at the University of Dhaka in partial fulfillment of the requirements for the degree of Doctor of Philosophy
2019

CERTIFICATE

This is to certify that the thesis entitled "Genomics and population dynamics of *Vibrio cholerae* causing pandemic cholera" was carried out by Shah M. Rashed, registration number: 98, session: 2016-2017 for the fulfillment of the degree of the Doctor of Philosophy (Science), from the Department of Microbiology, University of Dhaka, Bangladesh.

This work was carried out under our supervision and the style and contents of the thesis have been approved and recommended for the award of Ph.D. degree.

M. Mozammel Hoq

Professor M. Mozammel Hoq
Department of Microbiology
University of Dhaka
Dhaka-1000, Bangladesh

Anwar Huq

Professor Anwar Huq
Maryland Pathogen Research Institute
University of Maryland
College Park, Maryland 20742, USA

ABSTRACT

Cholera remains a major public health threat in many developing countries of Asia, Africa and the Americas. The causative agent of cholera, the Gram-negative bacterium *Vibrio cholerae*, is distributed ubiquitously in estuarine and marine water all over the world. Among all enteric pathogens, only cholera bacterium has the potential to cause pandemics. Seasonal cholera outbreaks in Bangladesh and the Ganges delta region of the Bay of Bengal are influenced by several biotic and abiotic factors. Over the past few decades, numerous epidemiological studies have been conducted to elucidate the ecology and epidemiology of *V. cholerae*, with strong emphasis on its survival, transmission, and pathogenesis. However, the global burden of cholera is increasing steadily in many parts of the world and massive cholera outbreaks in non-endemic countries have been reported recently. This study, therefore, employed a holistic approach to understand the emergence and evolution of pathogenic *V. cholerae* and microbial community dynamics in natural reservoir. Conventional microbiological and molecular methods were employed along with genome sequencing of *V. cholerae* isolates and whole genome shotgun sequencing of water samples to achieve the goal. Phenotypic and genetic characterization of 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh, revealed the presence of cholera toxin gene and other virulence factors, and showed six antibiotic resistance patterns with reduced susceptibility to ciprofloxacin and azithromycin. Multilocus variable number tandem repeat analysis (MLVA) of 222 environmental and clinical isolates identified genotypic variability in *V. cholerae* O1 isolated in two distinct areas of Bangladesh, i.e., Mathbaria and Chhatak. Multiple genetic lineages were detected among toxigenic *V. cholerae* O1 strains from Mathbaria; however, the outbreak in Chhatak was attributed to clonal expansion, commonly known as ‘founder flush’. Thus, results of this study support ‘person-to-person’ or accelerated mode of transmission of cholera during an outbreak. Whole genome sequencing and analysis of 165 *V. cholerae* isolates from Bangladesh identified multiple clones of toxigenic *V. cholerae* O1 carrying genotype *ctxB1*, *ctxB3*, and *ctxB7*. Comparative genomics of *V. cholerae* O1 and non-O1/O139 strains revealed a substantial number of GIs and MGEs, suggesting genomic rearrangement plays an important role in the

dynamic of pathogenic *V. cholerae*. Comparative genomics also detected greater genetic diversity in environmental *V. cholerae* O1 strains compared to clinical strains. Extensive genetic analysis of 91 *V. cholerae* O1 isolates from Mexico, isolated between 1991 and 2008, identified a truncated pro-phage in the upstream region of CTX, TCP variant, and predominance of *V. cholerae* with *ctxB3* during 2004 and 2008. Given the fact that very few clones may have been introduced to Mexico from Asia and Africa, core genome phylogeny also identified genetically different clones that might have evolved over time in the aquatic environment of Mexico. Presence of West African-South American genomic island (WASA)-1 in *V. cholerae* O1 isolated in Mexico and lack of SXT/R391 ICE is remarkable compared to atypical El Tor strains isolated in Asia and Africa. Whole genome shotgun sequencing and bioinformatics analysis of 20 pond water samples from Mathbaria, Bangladesh, identified 549 different species of bacteria, 50 viruses, 11 fungi, and 9 protista, commonly found in both fresh and coastal waters. Five *Vibrio* species were detected, of which *V. cholerae* and *V. mimicus* were predominant during both the epidemic and non-epidemic seasons of cholera. Interestingly, CTXΦ and other *Vibrio* phages were consistently detected throughout the study period, 2013 -2014. Like several previous studies, results of this study also showed environmental factors, i.e., temperature and salinity have direct influence in the microbial community dynamics. Another interesting observation was detection and identification of *Vibrio* phages during both epidemic and inter-epidemic periods. Toxigenic and non-toxigenic *V. cholerae*, other *Vibrio* species, symbiotic microbes, and predatory phages were found to exhibit a complex synchronization of the microbial community with environmental factors influencing the abundance of *V. cholerae* related to seasonal epidemics.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge my supervisor at the University of Dhaka, Prof. Md. Mozammel Hoq and at the University of Maryland Prof. Anwar Huq for their meticulous guidance, generous assistance and encouragement for successful completion of the study. I am ever grateful and highly indebted to Dr. Munirul Alam, Senior Scientist, icddr,b, Prof. O. Colin Stine, University of Maryland Baltimore, and all of my beloved teachers, particularly Prof. Sabita Rezwana Rahman, Chairman, Department of Microbiology, University of Dhaka, for their constant support, suggestions and cooperation. My sincere appreciation and thanks also goes to my laboratory partners Seon Yonug Choi, Kyle Brumfield, Arlene Chen, Daniela Ceccarelli, and Phillip Clark for their invaluable help throughout the study.

I would also like to express my highest gratitude and thanks to all my colleagues and collaborators at icddr,b especially Shahnewaz B. Mannan, Abdus Sadique, Marzia Sultana, Kabir U. Ahmed, Mohammed Tarequl Islam, and Mahamud-ur Rashid for all of their help. Next generation sequencing, i.e., whole genome sequencing of the *Vibrio* isolates and shotgun metagenomics of pond water samples would not have been possible without the help of Prof. Najib M. El-Sayed, University of Maryland College Park, his Lab members, and the Core Lab providing Next Generation Sequencing facility at UMD. My sincere gratitude and thanks also goes to Dr. Nur A. Hasan, and all his team members at the CosmosID for their valuable assistance and technical supports with bioinformatics pipeline. I would especially like to thank Victoria Lord for her tireless work to make the laboratory function.

I am delighted to pay tribute to my entire family for their tireless support, encouragement, and love all throughout.

Dhaka University Institutional Repository

© Copyright by

Shah M. Rashed

2019

Table of Contents

Table of Contents i

List of Tables viii

List of Figures x

CHAPTER 1: GENERAL INTRODUCTION 1

1.1. INTRODUCTION 2

1.2. Virulence factors of *V. cholerae* 4

1.3. Biotype classification scheme 4

1.4. Pandemic History of cholera 5

1.5. Ecology of *V. cholerae* 6

1.5.1. Natural Reservoir of *V. cholerae* 6

1.5.2. Survival strategy and persistence in aquatic environment 7

1.5.3. Environmental factors influencing cholera epidemics 8

1.6. Transmission of cholera 9

1.7. Genetic Variants and Genomic Diversity of *V. cholerae* 10

1.7.1. Emergence of O139, El Tor Variants, and atypical El Tor 10

1.7.2. Emergence of multidrug resistant *V. cholerae* 11

1.7. 3. Clonal divergence, evolution, and global dissemination of *V. cholerae* 12

1.8. *Vibrio* species and associated microbial community dynamics 15

1.9. Vision Statement and Hypothesis 16

CHAPTER 2: GENETIC CHARACTERISTICS, VIRULENCE FACTORS, AND ANTIBIOTIC SUSCEPTIBILITY OF *VIBRIO CHOLERAE* O1 ISOLATED IN RURAL COASTAL BANGLADESH 18

2.1. ABSTRACT 19

2.2. INTRODUCTION 19

2.3. MATERIALS AND METHODS 21

2.3.1. Bacterial Strains 21

2.3.2. Serogrouping of *V. cholerae* 22

2.3.3. Antimicrobial susceptibility 23

2.3.4. Genetic screening and PCR assay 23

2.4. RESULTS 25

2.4.1. Phenotypic and geneotypic characteristics 25

2.4.2. Antimicrobial susceptibility 28

2.4.2.1. MIC of ciprofloxacin 30

2.4.2.2. MIC of azithromycin 30

2.5. DISCUSSION 32

2.6. CONCLUSION 35

CHAPTER 3: GENOTYPING OF *VIBRIO CHOLERAE* O1 TO DETERMINE GENETIC CHANGES DURING SEASONAL OUTBREAKS OF CHOLERA 36

3.1. ABSTRACT 37

3.2. INTRODUCTION 37

3.3. MATERIALS AND METHODS 39

3.3.1. Sample Collection 39

3.3.2. Genetic Analysis 40

3.3.3. Statistical Analyses 41

3.4. RESULTS 41

3.4.1. Chhatak 42

3.4.2. Mathbaria 47

3.5. DISCUSSION 51

3.6. CONCLUSION 53

CHAPTER 4:GENOMIC INSIGHT INTO *VIBRIO CHOLERAE* CAUSING ENDEMIC CHOLERA IN BANGLADESH 54

4.1. Toxigenic *Vibrio cholerae* O1 from Bangladesh Possess West African-South American genomic island (WASA)-1 and WASA variant of VSP-II 55

4.1.1. ABSTRACT 55

4.1.2. INTRODUCTION 55

4.1.3. MATERIALS AND METHODS 57

4.1.3.1. Bacterial Strains 57

4.1.3.2. Genome Sequencing, Assembly, and Phylogenetic Analysis 57

4.1.3.3. Multilocus Variable Number Tandem Repeat Analysis 58

4.1.3.4. Antibiotic susceptibility 58

4.1.4. RESULTS AND DISCUSSION 58

4.1.5. CONCLUSION 64

4.2. Phylogenomic Aanalysis of *V. cholerae* isolated from patients and environmental samples in the coastal area of Bangladesh. 66

4.2.1. ABSTRACT 66

4.2.2. INTRODUCTION 66

4.2.3. MATERIAL AND METHODS 69

4.2.3.1. Bacterial Strains 69

4.2.3.2. DNA Extraction, Library preperation, and Illumina Sequencing 69

4.2.3.3. Genome Assembly and Phylogenetic Analysis 69

4.2.3.4. Identification and Annotation of Genomic Islands 71

4.2.4. RESULTS AND DISCUSSION 71

4.2.4.1. Virulence factors and *ctxB* genotypes 71

4.2.4.2. Phylogenetic analysis of *V. cholerae* 72

4.2.4.3. Genomic Islands and Mobile Genetic Elements 77

4.2.5. CONCLUSION 80

CHAPTER 5: GENETIC TRAITS AND GENOMIC DIVERSITY OF *VIBRIO CHOLERAE* O1 CAUSING CHOLERA IN MEXICO 82

5.1. Genetic traits, Molecular typing, and CTX pro-phage array of *V. cholerae* O1 isolated in Mexico, 1998 - 2008 83

5.1.1. ABSTRACT 83

5.1.2. INTRODUCTION 83

5.1.3. MATERIAL AND METHODS 84

5.1.3.1. Bacterial strains 84

5.1.3.2. Serogrouping and Biotype determination 85

5.1.3.3. Extraction of genomic DNA 86

5.1.3.4. Complementation of serogrouping and biotyping results by PCR assays 86

5.1.3.5. Determination of *ctxB* genotype 87

5.1.3.6. DNA sequencing of *ctxB* and *tcpA* 87

5.1.3.7. Genetic analysis of CTXΦ and flanking region 88

5.1.3.8. Pulsed-field gel electrophoresis (PFGE) 88

5.1.4. RESULTS 89

5.1.4.1. Microbiology, serology, and PCR analyses for *V. cholerae* 89

5.1.4.2. Phenotypic and related genotypic characteristics 89

5.1.4.3. Nucleotide sequencing of *ctxB* and *tcpA* gene 92

5.1.4.4. Genome organization of CTXΦ and flanking region 94

5.1.4.5. PFGE and cluster analysis 96

5.1.5. DISCUSSION 98

5.1.6. CONCLUSION 102

5.2. Genomic diversity of *Vibrio cholerae* associated with Endemic Cholera in Mexico, 1991 – 2008. 104

5.2.1. ABSTRACT 104

5.2.2. INTRODUCTION 105

5.2.3. MATERIAL AND METHODS 107

5.2.3.1. Sequencing, Assembly, and Annotation 107

5.2.3.2. Comparative Genomics 107

5.2.3.3. Identification and Annotation of Genomic Islands 108

5.2.3.4. Phylogenetics Based on Genome Sequences 108

5.2.3.5. Nucleotide sequence accession number 108

5.2.4. RESULTS AND DISCUSSION 109

5.2.4.1. Variations in CTXΦ-RS1 109

5.2.4.2. Vibrio Pathogenicity Island 1 and 2 111

5.2.4.3. Vibrio Seventh Pandemic Islands 113

5.2.4.4. Genomic Islands and Integrative Conjugative Elements (ICE) 114

5.2.4.5. Lipopolysaccharide Coding Region 115

5.2.4.6. Phylogenetics of the Mexican isolates 116

5.2.5. CONCLUSION 119

CHAPTER 6: METAGENOMIC ANALYSIS OF ENVIRONMENTAL WATER SAMPLES FROM COASTAL AREA OF BANGLADESH 121

6.1. ABSTRACT 122

6.2. INTRODUCTION 122

6.3. MATERIALS AND METHODS 124

6.3.1. Sample Collection 124

6.3.2. DNA Extraction 125

6.3.3. Library prep and Sequencing 126

6.3.4. Metagenomic Analysis using Bioinformatics 126

6.4. RESULTS AND DISCUSSION 126

6.4.1. Metagenomic profiling of Bacterial community 128

6.4.2. Identification of Viruses 131

6.4.3. Identification of fungi and Protista 132

6.4.4. Resistome profiling 134
6.4.5.. Comparison of conventional culture, PCR, and metagenomic detection, identification, and characterization of pathogenic *Vibrio* species 135
6.5. CONCLUSION 137
CHAPTER 7: CONCLUSION 138
REFERENCES 143
List of Publications 176

List of Tables

Table 2.1. Primer sequences for the genes, *V. cholerae* O1 (*wbe*), *V. cholerae* O139 (*wbf*), and *ctxA*. 23

Table 2.2. Primer sequences for virulence factor associated genes and genetic screening. 24

Table 2.3. Primer sequences for antibiotic resistance marker genes. 25

Table 2.4. Genetic characteristics and drug resistance of *V. cholerae* O1 isolated in Bangladesh.. 26

Table 2.5. Minimum inhibitory concentration of ciprofloxacin and azithromycin for 97 *V. cholerae* O1 isolates 30

Table 3.1. List of targeted loci, primers, and mathematical formula for MLVA. 41

Table 3.2. Genotypes of *V. cholerae* O1 isolated in Chhatak.. 43

Table 3.3. Genotypes of 59 isolates collected from six stool samples from cholera cases in Chhatak (10 isolates/stool) 46

Table 3.4. Genotypes of Mathbaria Clinical and Environmental Isolates. 48

Table 4.1. Nucleotide variation of CTX-prophage mediated genes in *V. cholerae* O1 El Tor strains. 61

Table 4.2. Source, serogroups and number of *V. cholerae* isolates used for genome sequencing and comparative genomic analysis. 69

Table 4.3. CT genotypes of *V. cholerae* O1 isolated in Bangladesh 72

Table 5.1. Phenotypic and genetic characteristics of *V. cholerae* O1 strains (n=91) isolated from diarrheal and environmental samples in Mexico (1998-2008) 90

Table 5.2. Genetic arrangement of CTXΦ and flanking region of representative ctx+ strains (2004 – 2008). 95

Table 5.3. Characteristics of *Vibrio cholerae* strains analyzed in this study. 107

Table 5.4. Sites of nucleotide polymorphisms in CTX prophages................................110

Table 6.1. Antibiotic resistant marker genes of the microbial community. 134

Table 6.2. Identification of different *Vibrio* spp. by conventional culture method, PCR, and NGS based metagenomic analysis..136

List of Figures

Figure 2.1. Distribution of *ctxB* genotypes in 97 *V. cholerae* O1 from Mathbaria, Bangladesh isolated during 2010 and 2014. 27

Figure 2.2. Drug resistance profile of 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh. 29

Figure 2.3. MIC of CIP and AZ for 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh. 31

Figure 3.1. Diagrams of genetic relatedness between genotypes of Chhatak. 44

Figure 3.2. Diagrams of genetic relatedness among MLVA genotypes of Mathbaria. 49

Figure 3.3. MLVA genotypes are plotted with the days of isolation for *V. cholerae* O1...50

Figure 4.1. Neighbor-joining trees showing phylogenetic relationships of 54 *V. cholerae* genomes based on 1,798 orthologs of protein-coding genes. 60

Figure 4.2. SNP-based maximum likelihood tree of 42 *V. cholerae* O1 El Tor isolated in different countries of Asia, Africa, and the Americas. 62

Figure 4.3. West African-South American genomic island WASA1 in *V. cholerae* O1 strains EC-0051 and EM-1727 isolated in Bangladesh in 2011. 63

Figure 4.4. A. SNP based pylogenetic tree showing pre-seventh pandemic clades. 74

Figure 4.4. B. SNP based pylogenetic tree showing seventh pandemic waves. 75

Figure 4.5. A. Distribution of Virulence factors and Genomic Islands in pre-seventh and seventh pandemic *V. cholerae*. 78

Figure 4.5. B. Distribution of Virulence factors and MGEs in pre-seventh and seventh pandemic *V. cholerae*. 79

Figure 5.1. Geographic map of Mexico showing provinces and areas close to the Gulf of Mexico where the national surveillance took place and *V. cholerae* O1 strains collected between 1998 and 2008. 85

Figure 5.2. The *tcpA* Phylogenetic tree constructed using neighbor joining. 93

Figure 5.3. Genetic mapping of CTX prophage lacking RS1 in *V. cholerae* O1 biotype El Tor strains isolated in Mexico during 2004-2008. 96

Figure 5.4. Dendrogram of PFGE patterns of *V. cholerae* O1 ET and transitional CTX− ET possessing TCP of ET or CL biotype. 97

Figure 5.5. Vibrio pathogenicity island -1 (VPI-1) of *V. cholerae* O1 strains isolated in Mexico and reference El Tor strain N16961. 112

Figure 5.6. Vibrio seventh pandemic island -II (VSP-II) of *V. cholerae* O1 strains isolated in Mexico and Brazil. 113

Figure 5.7. 'O' antigen biosynthetic genes of *V. cholerae* O1 strains CP1035, TM11079, and N16961. 116

Figure 5.8. Neighbor-joining trees showing phylogenetic relationships of 124 *V. cholerae* genomes based on 905 orthologs of protein-coding genes. 117

Figure 6.1. Total number of bacteria, virus, and fungi identified by metagenomic analysis of pond water samples collected in Mathbaria, Bangladesh (2013 – 2014) 127

Figure 6.2. Seasonal changes of environmental parameters of pond 10 and cholera outbreak seasons in Mathbaria, Bangladesh during 2013 and 2014. 128

Figure 6.3. Krona visualization showing total bacterial diversity based on relative abundance 129

Figure 6.4. Relative abundance of *Vibrio* species in Pond 10 during the study period 130

Figure 6.5. Krona visualization showing total viral diversity based on relative abundance 132

Figure 6.6. Krona visualization showing diversity of Fungi based on relative abundance 133

CHAPTER 1:

GENERAL INTRODUCTION

1.1. INTRODUCTION

The Gram-negative bacterium, *Vibrio cholerae*, the causative agent of cholera, is autochthonous to the marine and estuarine environment. Cholera is an acute life-threatening diarrheal disease that occurs in many developing countries, particularly in South Asia, Africa, and Latin America (Kaper *et al.*, 1995, Sack *et al.*, 2004). Cholera is transmitted by contaminated water and food, and continues to be a substantial health burden in countries where sanitation and safe drinking water are limited. The disease is endemic in the Ganges Delta region of the Bay of Bengal, historically referred to the 'Home of Asiatic Cholera' (Pollitzer, 1954). The father of modern epidemiology is considered to be John Snow who was the first to map a relationship between cholera and drinking water during a cholera outbreak in London in 1954 (Snow, 1856). Subsequently, microscopic observations of cholera stool by Filippo Pacini in 1954 described the morphology of the cholera bacterium and Robert Koch isolated the bacterium which he named *Vibrio comma* in 1984 in reference to its morphology and etiology of cholera. Both Pacini and Koch are acknowledged in the history of cholera for discovering cholera bacterium (Pacini, 1855, Koch, 1884).

Cholera has been endemic in the Indian subcontinent for centuries; however, the disease occurs elsewhere in Asia, Africa, and the Americas, causing 3-5 million cases globally each year with a substantial number (~100,000) of deaths (Ali *et al.*, 2012, Ali *et al.*, 2015). Cholera is established as an endemic disease in countries of Africa and the Americas, showing an overall increase in the global burden of the disease (Ali *et al.*, 2015). Over the past decade, it has been observed that cholera outbreaks are not limited to endemic countries (Mutreja *et al.*, 2011). The cholera epidemic in Haiti during 2010 and cases thereafter, with the current epidemic in Yemen, which started in 2017, reveals a distribution of the disease beyond endemic areas (Chin *et al.*, 2011, Camacho *et al.*, 2018). In Haiti, the cholera epidemic occurred in late October and November, after the earthquake earlier that year. Warfare disrupted the public health infrastructure in Yemen also before the cholera epidemic. Notably, the drinking water supply and sanitation system were damaged in both countries as a result of these disasters, eventually favoring spread of the disease (Montgomery *et al.*, 2018). Moreover, travel associated cholera cases have also been reported from different parts of the world (Steffen *et al.*, 2003, Kuleshov *et al.*, 2016).

The epidemic strains of *V. cholerae* are reported to travel across countries and continents over time with carriers or their host to give rise to cholera pandemics, according to the interpretation of historical records.

Despite major advancements in the knowledge of infectious disease dynamics and epidemiology of cholera, including environmental determinants triggering seasonal epidemics, many parts of the world have been experiencing recurrent cholera epidemics over the past few decades (Sack *et al.*, 2004, Emch *et al.*, 2008). Thus, cholera is recognized as one of the major emerging and re-emerging global infectious diseases. Two limiting factors, i.e., poverty and ubiquity of the pathogen, are making it almost impossible to eradicate cholera. In 2019, the UNDP reported 600 million people were living in extreme poverty around the world and providing safe drinking water, sanitation and good hygienic practice is a global challenge. Cholera is considered as a disease of developing countries of Asia, Africa, and Latin America. Its traditional home is thought by many to be the Ganges delta region of India and Bangladesh. However, cholera appears to have spread in waves throughout the world over the last two centuries (Mutreja *et al.*, 2011, Domman *et al.*, 2017, Weill *et al.*, 2017). Since its occurrence in Latin America in 1991 and Haiti in 2010, nearly all developing countries can be considered at risk of cholera. Although cholera is a reportable disease, its actual global incidence remains unknown because cases are not reported from those countries of Asia and Africa where it is most prevalent and from Latin American countries where it occurs sporadically. Recurrent cholera strikes in many remote villages of Bangladesh, particularly in the coastal areas of the Bay of Bengal where immediate health care facilities are lacking (Alam *et al.*, 2006, Rashed *et al.*, 2017).

The cholera bacterium inhabits aquatic environment and constitutes part of the normal flora of estuarine and brackish waters (Colwell & Spira, 1992). The species, *V. cholerae*, an important human pathogen, belongs to the genus *Vibrio* and Gram-negative Gammaproteobacteria class. The genus *Vibrio* comprises several bacterial species, of which *V. mimicus, V. parahemolyticus* and *V. vulnificus* have clinical significance as human pathogens. *V. cholerae* has more than 200 serogroups based on the somatic 'O' antigen. However, only two serogroups, O1 and O139 are associated

cholera epidemics (Kaper *et al.*, 1995, Sack *et al.*, 2004). Serogroup O1 comprises two major serotypes, Ogawa and Inaba, while a rarely reported Hikojima serotype has also been reported. Aside from O1 and O139 serogroups, all others are collectively known as *V. cholerae* non-O1/O139 and these occasionally cause gastrointestinal or extra-intestinal illness (Dziejman *et al.*, 2005, Chatterjee *et al.*, 2009, Octavia *et al.*, 2013).

1.2. Virulence factors of *V. cholerae*

Two critical virulence factors, cholera toxin (CT) and toxin co-regulated pilus (TCP), have been found consistently in epidemic strains of *V. cholerae* O1 and O139, and only occasionally in non-O1/O139 strains (Kaper *et al.*, 1995, Octavia *et al.*, 2013). CT, the primary virulence factor responsible for severe diarrhea, is encoded in the filamentous bacteriophage CTXΦ (Waldor & Mekalanos, 1996). The functional role of TCP is to colonize the intestinal mucosa and also serve as a receptor for lysogenic CTXΦ. TCP is encoded on a mobile genetic element, namely Vibrio pathogenicity island (VPI) (Waldor & Mekalanos, 1996). CT is a typical A-B subunit toxin, in which B subunit serves to bind the toxin to the intestinal epithelial cell receptor and the A subunit possesses a specific enzymatic function, causing electrolyte imbalance of the gut. *V. cholerae* also possesses minor virulence factors, i.e., zonula occludens toxin (*zot*), accessory cholera enterotoxin (*ace*), hemolysin (*hlyA*), repeat in toxin (RTX), heat-stable enterotoxin (NAG-ST), and cholix toxin (*chxA*). All these minor virulence factors have been detected in strains of the non-O1/O139 serogroups, including the type three secretion system (T3SS) (Dalsgaard *et al.*, 1995, Kaper *et al.*, 1995, Dziejman *et al.*, 2005).

1.3. Biotype classification scheme

V. cholerae O1 strains are classified into two biotypes: (1) classical and (2) El Tor, based on differences in phenotypic traits and biochemical reactions. These differentiating tests include hemolysis of sheep erythrocytes, agglutination of chicken erythrocytes, Voges–Proskauer reaction, and susceptibility to polymyxin B, and to specific phages (Kaper *et al.*, 1995). In addition, two biotypes possess biotype-specific genetic markers or alleles, e.g., *tcpA* (toxin co-regulated pilin A), *ctxB* (cholera toxin B), *hlyA* (hemolysin), *rstR* (regulatory region for phage lysogeny)

(Faast *et al.*, 1989, Olsvik *et al.*, 1993, Kimsey *et al.*, 1998). Genetic analysis also identified differences in gene clusters among these biotypes, such as VSP-I and II (Vibrio seventh pandemic island I and II), RS1 genetic element and RTX (repeat in toxin) (Waldor *et al.*, 1997, Chow *et al.*, 2001, Dziejman *et al.*, 2002, O'Shea *et al.*, 2004). The VSP-I and II gene clusters are unique to El Tor strains, whereas classical strains carry a truncated RTX (lacking *rtxC*) as opposed to El Tor (RTX containing four genes, *rtxA–rtxD*) (Chow *et al.*, 2001, Dziejman *et al.*, 2002). In addition to these phenotypic and genotypic differences, two biotypes have distinguished pathogenic potential, survivability, and infection patterns in the human host (Kaper *et al.*, 1995, Sack *et al.*, 2004). Notably, El Tor strains are frequently associated with asymptomatic infection and lower fatality and survive relatively better in both the environment and the human host. They are more efficient in host-to-host transmission than classical strains. In contrast, clinical manifestation of the disease caused by the classical biotype is more severe (Kaper *et al.*, 1995, Sack *et al.*, 2004).

1.4. Pandemic History of cholera

Since 1817, seven cholera pandemics have been recorded and notably all originated in Asia. The first six pandemics were caused by the classical biotype, while the El Tor biotype is considered the cause of ongoing seventh pandemic, which began in 1961 in Sulawesi, Indonesia (Barua, 1972, Kaper *et al.*, 1995). Very little is known about the pathogenicity and genetics of classical strains associated with the first five pandemics that occurred during 1817 and 1898 (Kaper *et al.*, 1995, Devault *et al.*, 2014). The sixth cholera pandemic, between 1899 and 1923, originated in India during which more than 800,000 people died of cholera and subsequently spread to the Middle East, North Africa, Eastern Europe, and Russia (Siddique & Cash, 2014).

The seventh pandemic is the largest, considering disease incidence and occurrence in diverse geographical regions over prolonged periods of time. Unlike earlier pandemics, this pandemic is believed to have begun on the island of Sulawesi in Indonesia (Sack *et al.*, 2004). The seventh pandemic occurred throughout Asia during the 1960s and in the Middle East and Africa by the following decade. In 1991, cholera appeared unexpectedly in Peru, on the western coast of South America, where it had not been reported for 100 years (Seas *et al.*, 2000). The epidemic subsequently

reached to Ecuador, Colombia, Brazil, and Chile and northward to Central America and Mexico by the end of 1993 (Guthmann, 1995). Overall, cholera has been reported in 120 countries, of which 69 are considered to be endemic during the seventh pandemic (Ali *et al.*, 2015). Over the past decade (2008-2018), three distinct cholera epidemics occurred in Zimbabwe, Haiti, and Yemen. Like a few of the African countries, cholera is considered to be endemic in Zimbabwe. However, Haiti and Yemen were considered non-endemic prior to the cholera epidemics in those countries. Since its occurrence in October, 2010, in Haiti, the disease has cumulatively sickened more than 810,000 Haitians and killed nearly 10,000 (Guillaume *et al.*, 2018). In Yemen, 1.2 million cholera cases have been recorded since the beginning of the outbreak in April, 2017 (Federspiel & Ali, 2018).

1.5. Ecology of *Vibrio cholerae*

1.5.1. Natural Reservoir of *V. cholerae*

The marine heterotrophic bacterium, *V. cholerae*, is ubiquitously distributed in the aquatic environment around the world, with non-O1/O139 serogroups readily isolated, hence considered to be more prevalent (Colwell & Spira, 1992). While living in a complex aquatic microbial community, several biotic and abiotic factors are involved in determining species abundance and emergence of epidemic clones. Despite the worldwide distribution of *V. cholerae*, the disease is endemic mostly in tropical countries where a low socioeconomic prevailing population is considered a contributing factor in major cholera outbreaks. Countries of South Asia and Sub-Saharan Africa have endured endemic cholera for centuries and outbreaks occur in a seasonal pattern (Sack *et al.*, 2004, Emch *et al.*, 2008). Several environmental parameters, including pH, temperature, salinity, dissolved oxygen concentration, rainfall, and the inorganic and organic content of water and sediment regulate the abundance of *V. cholerae* populations (Huq *et al.*, 1984, Huq *et al.*, 2005, Emch *et al.*, 2008).

In the aquatic environment, *V. cholerae* occurs both in a free-living state and attached to phytoplankton, zooplankton, and benthic invertebrates (Huq *et al.*, 1983, Colwell & Spira, 1992, Huq *et al.*, 1995). During non-epidemic or inter-epidemic periods, attachment to plankton, notably copepods, accounts for persistence of Vibios,

presumably obtaining nutrient from the host. The cholera bacterium has been detected in association with a wide range of aquatic life, including cyanobacteria, diatoms, freshwater filamentous green algae, oysters, water hyacinths, arthropods, and blue crabs (Martin & Bianchi, 1980, Spira *et al.*, 1981, Huq *et al.*, 1983, Islam *et al.*, 1989). *V. cholerae* produces extracellular chitinase enzyme which binds with the chitinacious substance of crustacean shells, degrading this mucopolysaccharide into available nutrients (Colwell & Spira, 1992). Moreover, another extracellular enzyme, mucinase, facilitates attachment of *V. cholerae* to algal surfaces and to diatoms to draw nutrient from the algal host (Islam *et al.*, 1989, Epstein, 1993, Epstein *et al.*, 1993). In an epidemic area like the Ganges Delta, survival of *V. cholerae* is favored by copepod abundance and algal blooms, presumably due to its ability to produce chitinase and mucinase. Furthermore, predator and prey relationships between Vibrios and lytic phages may regulate the population abundance of *Vibrio* species. It was hypothesized that lytic Vibriophages are a component that enforces significant reduction in the total *V. cholerae* population to end a seasonal epidemic, but this hypothesis remains to be proven (Faruque *et al.*, 2005, Faruque & Mekalanos, 2012).

1.5.2. Survival strategy and persistence in the aquatic environment

In response to adverse environmental conditions, i.e., nutrient starvation and cold temperatures, *V. cholerae* enters into a viable but non-culturable state (VBNC) to thrive in the ecosystem (Colwell *et al.*, 1985, Colwell, 2000). While in the VBNC state, the bacterium maintains limited physiological and metabolic activity, however its pathogenicity remains active. It has been demonstrated that VBNC is a survival strategy for *V. cholerae* during inter-epidemic periods and when accessibility to a human host is rare (Alam *et al.*, 2007). Although VBNC cells of *V. cholerae* fail to grow on conventional culture media, they remain pathogenic, as shown by animal model studies, whereas a proportion of the VBNC population regains culturability after access to the animal gut (Alam *et al.*, 2007). Like other aquatic bacteria, *V. cholerae* has the potential to form a biofilm in response to several physicochemical factors (Silva & Benitez, 2016). Investigators have shown that biofilm formation allows toxigenic *V. cholerae* to persist in the estuarine environment of geographic regions where cholera is endemic, contributing to successive seasonal outbreaks of cholera (Sultana *et al.*, 2018). *V. cholerae* can also alter its morphology, producing a

‘rugose variant’, which characteristically produces large amount of an exopolysaccharide that confers resistance to chlorine, acidic pH, and oxidative and osmotic stresses (White, 1938, Morris *et al.*, 1996). Thus, the rugose variant is considered to be a survival phenotype under challenging environmental conditions. Recently, a ‘persister phenotype’ has been proposed for toxigenic *V. cholerae* while growing in a laboratory microcosm containing a very minute quantity of nutrients that allows bacterial growth (Jubair *et al.*, 2012). Both the ‘rugose variant’ and ‘persister phenotype’ appear to be a residual culturable fraction of cells maintained in laboratory culture on media of low nutrient concentration.

1.5.3. Environmental factors influencing cholera epidemics

In endemic countries, occurrence of seasonal cholera epidemics varies widely depending on geographical location and associated environmental factors (Emch *et al.*, 2008). Notably, cholera dynamics in endemic regions demonstrate regular seasonal cycles. An annual bimodal distribution is observed in Bangladesh and Uganda, while annual single peaks occur in India, Pakistan, Mozambique, Congo, and few South American countries (Pascual *et al.*, 2000, Alam *et al.*, 2011, Bwire *et al.*, 2017). Temperature, salinity, rainfall, pH, dissolved organic material, dissolved oxygen concentration, and phytoplankton and zooplankton have been reported to be influential in regulating abundance of *V. cholerae*, thus influencing transmission of the disease in those parts of the world where the majority of the population relies on untreated water as a source of drinking water (Huq *et al.*, 1984, Huq *et al.*, 2005, Emch *et al.*, 2008).

As a consequence of these various external factors, the abundance and diversity of the *Vibrio* population exhibits a complex seasonal distribution driven by fluctuations of physiochemical parameters including temperature and salinity, as well as shifts in *Vibrio* reservoirs (Thompson *et al.*, 2004, Akanda *et al.*, 2013). Extreme climatic events such as prolonged droughts, floods, and cyclones have been implicated in major cholera epidemics (Akanda *et al.*, 2011). Such extreme events are also likely to create unseen changes in the ecosystem that those changes potentially can impact the cholera bacteria (Jutla *et al.*, 2011, Jutla *et al.*, 2013). In an aquatic reservoir, any change that favors the *Vibrio* population can stimulate the bacteria to

multiply rapidly, hence increase the total number in the population (Colwell, 1996). Thus, global warming and climate change is likely to have wide-ranging effects on cholera dynamics by creating a favorable environment for unexpected cholera outbreaks and transmission in different parts of the world where hydrologic, climatic, and ecological extremes and population vulnerability coincide. For example, cholera incidence has increased three-fold in African El Niño sensitive regions, with India and Bangladesh similarly at risk with climate changes and sea level rise (Hashizume *et al.*, 2013, Vezzulli *et al.*, 2016, Moore *et al.*, 2017).

1.6. Transmission of cholera

Transmission of cholera is predominantly through fecal-oral route. The infectious dose of *V. cholerae* in humans is generally 10^8–10^{11} cells or even lower, depending on the virulence potential of the given bacterial strain and immune status of the susceptible host (Sack *et al.*, 2004, Nelson *et al.*, 2009). Infected individuals develop typical signs and symptoms of cholera when the toxigenic *V. cholerae* infectious dose reaches the gut via contaminated food and water. Cholera bacteria multiply rapidly in human the gut and excreted to the environment though cholera-stools. Once *V. cholerae* is excreted outside from the human host, it apparently gains an accelerated virulence and infectivity for at least five hours, commonly known as 'hyper-infectious' state (Nelson *et al.*, 2009). During an outbreak, hyper-infectious *V. cholerae* accelerate their spread in a densely populated vulnerable community, notably household members (Kendall *et al.*, 2010, Chowdhury *et al.*, 2015). This rapid mode of transmission, namely 'person-to-person' transmission, is also fecal-oral. Onset of an outbreak in a new location can result from person-to-person transmission when a symptomatic or asymptomatic carrier enters a susceptible population (Nelson *et al.*, 2009). However, carriers of *V. cholerae* have not been described. Once cholera cases occur in a new geographical area, initial propagation depends on the extent of individual bacterial shedding, host and organism characteristics, and the likelihood of additional persons ingesting an infectious dose. In endemic areas such as Bangladesh, water appears to be the major vehicle of transmission, but in other regions food has also been implicated. In fact, it is very difficult to separate the two sources, since water often contaminates food. *V. cholerae*

has been shown to multiply in food, thereby increasing the number of bacteria and the likelihood of illness if the contaminated food is ingested.

1.7. Genetic Variants and Genomic Diversity of *V. cholerae*

1.7.1. Emergence of O139, El Tor Variants, and atypical El Tor

Since the seventh pandemic began in 1961, the El Tor continues to cause epidemics in Asia and Africa and lately in Latin America during 1991 and thereafter. *V. cholerae* O1 classical biotype, the cause of the first six pandemics gradually was replaced by El Tor in Asia and Africa in 1980s. However, it was present in Mexico until 1997 (Alam *et al.*, 2012). Surprisingly, a second toxigenic non-O1 serogroup, designated *V. cholerae* O139 synonym Bengal, emerged in the neighboring region of Ganges delta in 1992, followed by causing severe outbreaks in coastal regions of India and Bangladesh (Albert *et al.*, 1993). The genetic backbone of *V. cholerae* O1 El Tor and serogroup O139 appears to be highly similar, therefore a strong possibility that El Tor biotype served as progenitor for the emergence of O139 by acquiring the 'O' biosynthetic *wbf* gene-clusters. Despite having a genetic similarity to El Tor biotype and sharing ecological niche, serogroup O139 became transiently predominant over El Tor due to its association with the majority of cholera cases in South Asia and disseminated in Pakistan, Nepal, China, Thailand, Kazakhstan, Afghanistan, and Malaysia (Huq *et al.*, 1995, Faruque *et al.*, 2003, Albert & Nair, 2005). Following its initial appearance and explosive epidemic during late 1992 and early 1993, the occurrence of O139 serogroup declined in many cholera-endemic regions (Faruque *et al.*, 1997). Notably, transient disappearance and sudden re-emergence was the characteristic of the O139 serogroup until 2005, while El Tor biotype steadily persisted both in aquatic environment and in cholera cases. Interestingly, *V. cholerae* O139 was isolated again from surface water in 2011 in Bangladesh and found to have been associated with household cholera outbreaks in Dhaka, Bangladesh during 2013 and 2014 (Rashed *et al.*, 2013, Chowdhury *et al.*, 2015).

Although classical and El Tor biotypes evolved from different lineages (Karaolis *et al.*, 1995), hybrids between the classical and El Tor biotypes, resulting from genetic exchange, also exist in nature (Nair *et al.*, 2002, Safa *et al.*, 2006). New variants of El Tor strains have emerged during the 1990s in Indian subcontinent that

could not be categorized either as El Tor or classical, based on the phenotypic traits and conventional typing schemes (Raychoudhuri *et al.*, 2008). Despite the apparent extinction of classical strains from endemic foci, the appearance of classical biotype specific attributes in El Tor remains a mystery. El Tor variants have acquired classical cholera toxin (*ctxB* genotype 1) and rstR$^{\text{Classical}}$, first identified in *V. cholerae* O1 Matlab Variant stains isolated between 1991 and 1994 in Matlab, Bangladesh (Nair *et al.*, 2002, Safa *et al.*, 2006). Notably, a similar variant of El Tor was identified in Mozambique in 2004, carrying two copies of classical CTX prophage in the small chromosome (Ansaruzzaman *et al.*, 2004, Chun *et al.*, 2009). *V. cholerae* O1 altered El Tor possessing classical cholera toxin (*ctxB* genotype 1) and rstR$^{\text{El Tor}}$ emerged as the predominant cause of cholera in India, Bangladesh, and Thailand during the 1990s (Nair *et al.*, 2002, Raychoudhuri *et al.*, 2009, Na-Ubol *et al.*, 2011). Thus, altered El Tor strains completely displaced the seventh pandemic prototype El Tor in cholera endemic countries during the late-1990s and/or thereafter. Ever since El Tor strains acquired classical cholera toxin, the disease severity has been increased several fold than those of typical seventh pandemic El Tor (Siddique *et al.*, 2010). Furthermore, subtle genetic changes in the *ctxB* gene and identification of multiple *ctxB* alleles in *V. cholerae* have reconstructed an extended *ctxB* genotyping scheme which comprises a total of 12 different *ctxB* genotypes (*ctxB1* to *ctxB12*) (Olsvik *et al.*, 1993, Kim *et al.*, 2015). All these variants of *V. cholerae* O1 El Tor including Haiti variant are collectively known as atypical El Tor (Mutreja *et al.*, 2011, Kim *et al.*, 2014, Kim *et al.*, 2015). However, CTXΦ array, copy number of CTXΦ, and pathogenicity may vary among atypical El Tor stains depending on geographical distribution.

1.7.2. Emergence of multidrug resistant *V. cholerae*

Clinical management of cholera and its prevention are often hindered due to rapid emergence of multiple drug resistant *V. cholerae* in Asia and Africa (Glass *et al.*, 1980, Sack *et al.*, 2004, Saha *et al.*, 2006). Ever since antimicrobial agents were employed as successful therapeutic options in the treatment of cholera, tetracycline was the drug of choice worldwide, except for young children and pregnant women (Greenough *et al.*, 1964, Lindenbaum *et al.*, 1967). In addition, furazolidone, erythromycin, trimethoprim-sulfamethoxazole, and chloramphenicol were also considered to be effective against the cholera bacterium (Greenough *et al.*, 1964).

Emergence of drug resistant *V. cholerae* was first reported from Africa in late-1970s, showing resistance to tetracycline and other antimicrobials (Mhalu *et al.*, 1979). Likewise, in December 1979, multidrug resistant *V. cholerae* O1 showing resistance to tetracycline, ampicillin, kanamycin, streptomycin, and trimethoprim–sulfamethoxazole, were found to have been associated with local cholera outbreak in Matlab, Bangladesh (Glass *et al.*, 1980, Glass *et al.*, 1983). In 1995, nalidixic acid resistant *V. cholerae* O1 was identified in southern India and most recently, the bacterium developed resistance to erythromycin and ciprofloxacin in endemic countries (Jesudason & Saaya, 1997, Mukhopadhyay *et al.*, 1998, Quilici *et al.*, 2010). Recent studies in southwestern India have shown *V. cholerae* to be resistant to several antibiotics, including tetracycline, furazolidone, norfloxacin and ciprofloxacin (Jain *et al.*, 2011).

Ever since multidrug resistant *V. cholerae* emerged in Asia and Africa, the antibiotic susceptibility patterns of epidemic strains have been changed frequently. The genetic basis for such fluctuation in drug resistance was due to lateral acquisition of the self-transmissible transposon-like element (SXT element), which carries multiple antibiotic resistance markers (Waldor *et al.*, 1996). Tetracycline resistance in *V. cholerae* is known to be plasmid-mediated, although plasmids have been shown to be highly unstable in Vibrios (Taneja *et al.*, 2010). Quinolone resistant *V. cholerae* possess the *qnrVC* gene cassette, and mutation in the *gyrA* and *parC* genes (Zhang *et al.*, 2018). Recently, IncA/C plasmid has been identified in *V. cholerae* non-O1/O139 strains and confers resistance to multiple drugs including azithromycin (Folster *et al.*, 2014, Carraro *et al.*, 2016, Wang *et al.*, 2018). Given that antibiotic susceptibility patterns of epidemic *V. cholerae* strains have frequently changed, continuous monitoring of drug resistance is crucial for selecting an effective drug for the treatment of cholera.

1.7. 3. Clonal divergence, evolution, and global dissemination of *V. cholerae*

Evolution of toxigenic *V. cholerae* has been characterized by acquisition, loss, and rearrangement of specific mobile genetic elements in the genome that plays a major role in emergence of epidemics clones (Chun *et al.*, 2009, Mutreja *et al.*, 2011). For example, CT is encoded by filamentous phage, RS1 and TLC are satellite phages, VPI

gene clusters carrying TCP encoding genes and SXT/R391 can excise from the genome, and VSP-I and II are horizontally transferred genetic elements. Notably, all contribute heavily to virulence and ecological fitness of *V. cholerae* (Boyd *et al.*, 2000, Murphy & Boyd, 2008, Faruque & Mekalanos, 2012). Although different molecular typing tools, e.g., ribotyping, pulsed-field gel electrophoresis, restriction fragment length polymorphism, multilocus variable number tandem repeat analysis, multilocus sequence typing, and whole genome sequencing, have been used to characterize environmental and clinical strains (Rahaman *et al.*, 2015), it appears that *V. cholerae* strains are genetically related if isolated from a defined geographical origin during an outbreak (Moore *et al.*, 2014, Robins & Mekalanos, 2014). Several retrospective studies have demonstrated environmental *V. cholerae* O1 strains are genetically more diverse than epidemic strains (Kiiru *et al.*, 2013, Moore *et al.*, 2014, Domman *et al.*, 2018).

Epidemiological investigations and detailed genomic analyses hypothesize that pathogenic *V. cholerae* strains may derive from ancestral environmental strains by acquisition of critical virulence genes and other gene-clusters, although those evolutionary trends are not yet fully understood (Karaolis *et al.*, 1995). Moreover, the possibility of an existing pathogenic strain undergoing subtle genetic change to give rise of a new epidemic clone cannot be ruled out. Recent genomic analysis identified significant genetic ‘inflow’ and ‘outflow’ of genomic islands (GIs), pathogenicity islands (PIs), phage elements, and transposon mediated lateral gene transfer in *V. cholerae* genome (Chun *et al.*, 2009). Despite making significant progress in the molecular epidemiology of cholera, allowing the tracking of newly emerged epidemic clones, the question remains ‘what drives the frequent emergence of new epidemic clones by displacement of existing clones’? From an epidemiological point of view, it certainly appears that newly emerging clones obtain selection advantage over pre-existing clones due to genetic rearrangement (Stine & Morris, 2014). Despite lacking selection advantage, preexisting clones persist in ecological niches for prolonged periods of times unless a resurgence event leads to epidemic clone (Faruque *et al.*, 2003, Chowdhury *et al.*, 2015).

Recent genomic analyses of globally collected *V. cholerae* O1 strains showed that cholera bacterium has been disseminated from Ganges Delta region to other parts

of the word, notably in three independent but overlapping waves during the ongoing seventh pandemic (Mutreja *et al.*, 2011). Mutreja *et al.* categorized these three waves chronologically as wave 1, 2, and 3. The wave 1 (1961-1999) isolates were mostly prototype El Tor containing CT genotype-3 (*ctxB3*) and *rstR*$^{\text{El Tor}}$ on chromosome 1 and may carry the TLC and RS1 genetic element. However, these isolates uniformly lack ICE of the SXT/R391 family. Wave-1 spread from Indonesia to the Bay of Bengal region in the Indian subcontinent and then to East Africa and South-West Americas. Wave-2 (1978-1984) isolates were predominantly Variant El Tor carrying CT genotype-1 (*ctxB1*) and tandem repeat of CTXΦ on chromosome 2. Isolates belonging to this wave may or may not contain RS1 on chromosome 1, *rstR*$^{\text{El Tor}}$ and/or *rstR*$^{\text{Classical}}$, and R391 family ICE-SXT element. These variants along with O139 were first identified in the Indian subcontinent and spread to East Asia and Africa. The wave-3 (1991–2010) isolates harbor *rstR*$^{\text{El Tor}}$, TLC, RS1, and either CT-genotype 1 (*ctxB1*) or CT-genotype 7 (*ctxB7*) on chromosome 1. Almost all the *V. cholerae* isolates belonging to wave 3 were found to be atypical El Tor possessing ICE of the SXT/R391 family, mostly the ICEVchInd5 (Mutreja *et al.*, 2011, Kim *et al.*, 2015). Furthermore, the same retrospective study also explained the pathway of cholera importation from Nepal to Haiti by UN peacekeepers in 2010, which eventually caused the largest cholera epidemics in the history of Haiti (Tappero & Tauxe, 2011). Following the Haiti epidemic, same *V. cholerae* O1 clone or Haiti variant possessing *ctxB7* genotype has been spread into neighboring countries, including Cuba, the Dominican Republic, the United States, and Mexico (Tappero & Tauxe, 2011, Moore *et al.*, 2014). Remarkably, a genetically similar *V. cholerae* O1 clone possessing *ctxB7* is also responsible for the devastating cholera outbreak in Yemen, which started in 2017 (Weill *et al.*, 2019). The Haiti variant has been identified with travel associated cases of cholera in European countries as well (Smirnova *et al.*, 2017). Origin of Haiti variant of *V. cholerae* O1 has been claimed to be Calcutta, India where it was first isolated in 2006 (Ghosh *et al.*, 2014, Ghosh *et al.*, 2014). The Haiti variant and its descendants caused epidemic in several countries of Asia, Africa, and the Americas over the past decades. Therefore, it can be concluded that genetic screening and detailed genomic analysis of *V. cholerae* O1 isolated in Bangladesh and other cholera endemic countries need to be done consistently to track the emergence of toxigenic clones.

1.8. *Vibrio* species and associated microbial community dynamics

The genus *Vibrio* now comprises more than 100 species, of which the vast majority inhabits the marine and estuarine environment and constitutes part of the normal flora of those environments (Colwell & Spira, 1992, Westrich *et al.*, 2018). These marine heterotrophs produce a diverse array of extracellular enzymes that degrade complex nutrients, therefore play an important role in nutrient recycling (Islam *et al.*, 1989, Colwell & Spira, 1992). In the aquatic environment, one well known survival strategy is inter-species or intra-species symbiotic relationship, or antagonism which encourages the famous theory of Charles Darwin: 'survival of the fittest' (Fuhrman *et al.*, 2015). The association of *Vibrio* species with the degradation of chitin, chitinacious substances, and other organic compounds is an integral aspect of *Vibrio* ecology in the marine environment, because the metabolites can attract other microbial species to colonize the same habitat. Thus, Vibrios are involved directly in the mineralization of chitinous exoskeletons of zooplankton, copepods, and other higher plankton (Huq *et al.*, 1983, Tamplin *et al.*, 1990, Colwell & Spira, 1992). Moreover, algal blooms also may have substantial impact on the abundance of *Vibrio* species in the aquatic environment (Epstein, 1993). Other than nutrient, several physicochemical factors and biotic components of the ecosystem regulate the ecological diversity and abundance of different species. However, spatiotemporal changes in microbial communities and species abundance are evident caused by both external variables and internal factors (Fuhrman *et al.*, 2015, Morrison *et al.*, 2017). The term 'dynamics' specifically refers to changes over time, and temporal changes in microbial communities may result from growth and death and import and export of microbes in an existing community. Marine microbial communities are known to be dynamic and it is highly likely that ecologists can predict variations in community composition on a daily, seasonal, and annual basis. Despite seasonal variables enforcing changes in the microbial community, there are internal feedback mechanisms, including competition, viral infection, and predator–prey interactions, that help to maintain a remarkably steady 'average' community year after year (Fuhrman *et al.*, 2015).

A relatively recent development is whole genome shotgun sequencing, a powerful tool for identification of bacterial species and adopted in the new area of

metagenomics (Fiedler *et al.*, 2018, Roy *et al.*, 2018). Considering that only 2% or less of microbes globally are culturable by conventional culture method, direct extraction of DNA from environmental samples, coupled with next generation sequencing, and subsequent bioinformatics analysis potentially can detect, identify, and characterize an entire microbial community composition (Stewart, 2012, Roy *et al.*, 2018). This has been considered as a significant stride in the microbial ecology, because it enables monitoring of microbial community dynamics of diverse ecosystems (Fuhrman *et al.*, 2015, Morrison *et al.*, 2017). Although, amplicon based 16s, 18s, and ITS analyses have been employed to reveal microbial community composition, those methods lack precision and cannot identify species level, reflecting major drawbacks (Buse *et al.*, 2014). Next generation shotgun sequencing coupled with high quality bioinformatics analysis provides better resolution and accuracy in identifying microbes to the species level. Those human pathogens that are natural inhabitants of the aquatic environment, e.g., *V. cholerae*, *V. mimicus*, and *V. parahaemolyticus*, and *V. vulnificus* can easily be detected using metagenomics approaches. From an epidemiological point of view, detailed information about spatiotemporal changes in those pathogenic microbial populations and their associated community can provide an improved understanding of disease dynamics and employ preventive measures prior the epidemic strikes. In cholera endemic country like Bangladesh, where seasonal cholera epidemics are frequent, metagenomic analysis of surface waters from coastal areas could reveal the dynamics of *Vibrio* species, including toxigenic *V. cholerae*. Thus, monitoring spatiotemporal changes during epidemic and inter-epidemic periods is important to identify a possible indicator for forthcoming cholera epidemic.

1.9. Vision Statement and Hypothesis

- ❑ *V. cholerae* O1 continues to acquire resistance to clinically important drugs by horizontal and vertical gene transfer events.
- ❑ Seasonal cholera outbreaks in endemic settings are caused by clonal expansion and the environment harbors multiple clones of toxigenic *V. cholerae*.

- The aquatic reservoir enables the survival of ancient *V. cholerae* clones where few clones maintain genomic integrity and new clones continue to evolve in from environmental progenitors.
- The genomic epidemiology of *V. cholerae* differs from continent to continent due to evolution in local niches and at the same time global dissemination contributes to genetically similar clones.
- Environmental factors drive spatiotemporal changes in microbial community dynamics, and shape *V. cholerae* and related *Vibrio* species during seasonal outbreaks of diseases.

CHAPTER 2:

GENETIC CHARACTERISTICS, VIRULENCE FACTORS, AND ANTIBIOTIC SUSCEPTIBILITY OF *VIBRIO CHOLERAE* O1 ISOLATED IN RURAL COASTAL BANGLADESH

2.1. ABSTRACT

Cholera outbreaks occur each year in remote coastal areas of Bangladesh, and epidemiological surveillance and routine monitoring of cholera in these areas is challenging. In this study, a total of 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh, collected from rectal swabs and surface water samples during 2010 and 2014 were analyzed for phenotypic and genotypic traits, including antimicrobial susceptibility. Of the 97 isolates, 95 possessed CTX-phage mediated genes, *ctxA*, *ace*, and *zot*, and two lacked the cholera toxin gene, *ctxA*. Also both CTX^+ and CTX^- *V. cholerae* O1 isolated in this study carried *rtxC*, $tcpA^{El\ Tor}$, and *hlyA*. The classical cholera toxin gene, *ctxB1,* was detected in 87 isolates, while eight had *ctxB7.* Of 95 CTX^+ *V. cholerae* O1, 90 contained $rstR^{El\ Tor}$ and 5 had $rstR^{Classical}$. All isolates, except two, contained SXT related integrase *intSXT*. Resistance to penicillin, streptomycin, nalidixic acid, sulfamethoxazole-trimethoprim, erythromycin, and tetracycline varied between the years of study. Most importantly, 93% of the *V. cholerae* O1 strains were multidrug resistant. Six different resistance profiles were observed, with resistance to streptomycin, nalidixic acid, tetracycline, and sulfamethoxazole-trimethoprim predominant every year. Ciprofloxacin and azithromycin MIC were 0.003 – 0.75 µg/ml and 0.19 – 2.00 µg/ml, respectively, indicating reduced susceptibility to these antibiotics. Sixteen of the *V. cholerae* O1 isolates showed higher MIC for azithromycin (≥0.5 µg/ml) and were further examined for ten macrolide resistance genes, *erm*(A), *erm*(B), *erm*(C), *ere*(A), *ere*(B), *mph*(A), *mph*(B), *mph*(D), *mef*(A), and *msr*(A), with none testing positive for macrolide resistance genes.

2.2. INTRODUCTION

Vibrio cholerae, the causative agent of cholera, is autochthonous to the estuarine and marine environment worldwide. Of more than 200 O-antigen serogroups identified in *V. cholerae*, only toxigenic O1 and O139 are considered to be primarily associated with epidemics and pandemics (Sack *et al.*, 2004). Cholera, an ancient diarrheal disease, continues to be a serious threat in countries of Asia, Africa, and South America. Even though cholera is underreported in many countries, 3–5 million cases are recorded annually in different parts of the world, with a significant number of deaths (Ali *et al.*,

2012). The case fatality rate of cholera has been reduced over the past few decades, mainly because patients are treated with oral and/or intravenous rehydration therapy together with appropriate dosage of antibiotics. Effective antibiotic treatment shortens the duration of diarrhea and limits the loss of body fluids by ca. 50% (Sack *et al.*, 2004). However, antibiotic resistant enteropathogens, including *V. cholerae,* are emerging rapidly due to the selective pressure of antibiotics existing in the environment and from excessive use (Laxminarayan *et al.*, 2013, Andersson & Hughes, 2014). *V. cholerae,* both O1 and O139, have developed resistance to several antimicrobial drugs, including tetracycline (TE), chloramphenicol (C), furazolidone, ampicillin (AM), and trimethoprim-cotrimoxazole, used successfully to treat cholera over the years (Garg *et al.*, 2001, Kitaoka *et al.*, 2011). As a consequence, multidrug resistant (MDR) *V. cholerae* has been on the rise, causing clinicians to face a serious challenge when deciding a drug of choice and regimen for treating cholera patients.

Two biotypes of *V. cholerae* O1, Classical (CL) and El Tor (ET) are universally recognized, with each possessing distinct phenotypic and genetic traits, including major virulence genes, i.e., toxin coregulated pilus (*tcpA*) and B-subunit of cholera toxin (*ctxB*) (Kaper *et al.*, 1995, Safa *et al.*, 2010). Of the two biotypes, CL is associated with the sixth and presumably the earlier pandemics of cholera that occurred between 1817 and 1923 (Kaper *et al.*, 1995, Devault *et al.*, 2014), while ET is reported to have initiated the ongoing seventh pandemic in the early 1960s, gradually displacing the CL biotype (Kaper *et al.*, 1995, Kim *et al.*, 2015). Over the past two decades, variants of ET with only a few CL attributes (phage-encoded repressor $rstR^{CL}$ and B-subunit of cholera toxin $ctxB^{CL}$) have emerged in Asia and Africa. These variants are collectively known as atypical ET (Safa *et al.*, 2010, Kim *et al.*, 2015). Moreover, based on amino acid substitutions in CtxB, 12 different *ctxB* genotypes have been identified in *V. cholerae* (Kim *et al.*, 2015). In 1992, *V. cholerae* O139 carrying the SXT/R391 family integrative conjugative element (ICE) appeared transiently as the major cause of cholera in Bangladesh and India (Albert *et al.*, 1993, Ramamurthy *et al.*, 1993, Waldor *et al.*, 1996). SXT/R391 ICE was the first MDR marker detected in *V. cholerae*, conferring resistance to streptomycin (S), sulfamethoxazole, and trimethoprim (Waldor *et al.*, 1996,

Hochhut *et al.*, 2001). SXT/R391 ICE also found to provide a selective advantage to *V. cholerae* O1 ET, a strain that has been tracked globally in three overlapping waves during the seventh pandemic (Mutreja *et al.*, 2011).

Interestingly, outbreaks of cholera that occur in coastal areas are seasonal each year in Bangladesh. For example, in Mathbaria, cholera occurs predominantly during the spring, months of March through May, where inhabitants lacking safe drinking water are most susceptible (Emch *et al.*, 2008, Akanda *et al.*, 2013). Several antibiotics are used to treat cholera, including doxycycline, ciprofloxacin (CIP), and azithromycin (AZ), usage of all is greatly influenced by the drug sensitivity pattern of the bacterium reported in the contemporary literature (Harris *et al.*, 2012). In Bangladesh, a single dose of AZ or CIP currently is used prophylactically. Not surprising, *V. cholerae* O1 is now reported to have reduced susceptibility to CIP in Bangladesh, India, Vietnam, Haiti, Zimbabwe, and Western Africa (Islam *et al.*, 2009, Quilici *et al.*, 2010, Sjolund-Karlsson *et al.*, 2011, Tran *et al.*, 2012, Kumar *et al.*, 2014, Khan *et al.*, 2015). MDR *V. cholerae* O1 resistant to TE, AM, S, sulfonamides, norfloxacin, gentamicin, furazolidone, kanamycin (K), sulfamethoxazole-trimethoprim (SXT), and erythromycin (E), is currently circulating in cholera endemic countries of Asia and Africa (Finch *et al.*, 1988, Faruque *et al.*, 2007, Jain *et al.*, 2011, Rashed *et al.*, 2012, Dixit *et al.*, 2014). Furthermore, genes conferring resistance to CIP and AZ have been shown to be transferred to *V. cholerae* via plasmids, gene cassettes, and mobile genetic elements with horizontal gene transfer mechanisms in environmental reservoir implicated in transforming sensitive bacteria to resistant (Kitaoka *et al.*, 2011). Considering these phenotypic and genetic modifications reported previously, a study of 97 *V. cholerae* O1 isolates was undertaken to determine the antibiotic resistance/susceptibility status of *V. cholerae* O1 isolated from recent cholera cases in cholera endemic Mathbaria, Bangladesh.

2.3. MATERIALS AND METHODS

2.3.1. Bacterial Strains

In this study, a total of 97 *V. cholerae* O1 stains isolated from rectal swabs and surface water samples collected in the coastal villages of Mathbaria, Bangladesh, between June,

2010 and December, 2014, as a part of epidemiological surveillance conducted by the International Centre for Diarrheal Disease Research, Bangladesh (icddr,b) were analyzed for antibiotic susceptibility and genotypic traits. Mathbaria is geographically adjacent to the Bay of Bengal, located approximately 165 km south-west of Dhaka city. Clinical isolates (n = 52) were obtained from rectal swabs of suspected cholera patients seeking treatment at the local health center during cholera peak and off-peak season. Environmental isolates (n = 45) were obtained from water and plankton samples collected periodically at six different ponds and a river in the same area where the clinical samples were also collected. The clinical and environmental samples were collected, transported, and subjected to bacteriological analysis for *V. cholerae,* following standard procedures (Alam *et al.*, 2006, Huq *et al.*, 2012). Isolation and identification were performed according to standard methods (Alam *et al.*, 2006, Alam *et al.*, 2006, Huq *et al.*, 2012). All samples were collected according to protocols approved by institutional review boards at the Johns Hopkins University, University of Maryland (College Park, MD, USA), and icddr,b. Informed consent was obtained from the patients, and parents or legal guardians of children who participated in this study. Genomic DNA was prepared from the presumptively identified *V. cholerae* isolates using the boiling lysis method of Park et al. (2013) and *V. cholerae* species-specific *ompW* PCR was done to confirm identity of the isolates (Nandi *et al.*, 2000).

2.3.2. Serogrouping of *V. cholerae*

Serogroups of the *V. cholerae* isolates were confirmed by slide agglutination using specific polyvalent antisera for *V. cholerae* O1 and O139. Bacterial isolates, showing positive agglutination with O1 antisera, were tested further using a serotype-specific monoclonal antibody, i.e., Inaba and Ogawa (Alam *et al.*, 2006). Serogroups of these isolates were reconfirmed by multiplex PCR (**Table 2.1**), targeting O1- (*wbe*) and O139- (*wbf*) specific O biosynthetic genes, together with the cholera toxin gene (*ctxA*) (Hoshino *et al.*, 1998).

Table 2.1. Primer sequences for the genes, *V. cholerae* O1 (*wbe*), *V. cholerae* O139 (*wbf*), and *ctxA*.

Target	Primer	Sequence (5`-3`)	Amplicon (bp)	Reference
ctxA	VCT1	ACAGAGTGAGTACTTTGACC	192	Hoshino *et al.*, 1998
	VCT2	ATACCATCCATATATTTGGGAG		
O1-*wbe*	O1F2-1	GTTTCACTGAACAGATGGG	308	Hoshino *et al.*, 1998
	O1R2-2	GGTCATCTGTAAGTACAAC		
O139- *wbf*	O139F2	AGCCTCTTTATTACGGGTGG	449	Hoshino *et al.*, 1998
	O139R2	GTCAAACCCGATCGTAAAGG		

2.3.3. Antimicrobial susceptibility

Susceptibility to antimicrobials was determined by standard disc diffusion on Muller-Hinton agar (BD, USA) according to Clinical and Laboratory Standards Institute guidelines for *V. cholerae* (CLSI, 2010) and *Enterobacteriaceae* (CLSI, 2010). *Escherichia coli* ATCC 25922 was used as a control for antimicrobial susceptibility. All strains of *V. cholerae* O1 were tested for resistance to AM (10 µg), CIP (5 µg), C (30 µg), E (15 µg), K (30 µg), S (10 µg), TE (30 µg), nalidixic acid (NA, 30 µg), penicillin (P, 10 µg), and SXT (23.75 and 1.25 µg, respectively) using commercially available discs (BD BBL Sensi-Disc). Minimum inhibitory concentrations (MIC) of CIP and AZ were determined using E-test strips (bioMérieux-USA), according to the manufacturer's instructions. In brief, overnight-grown bacterial cultures were resuspended in normal saline and the cell concentration was adjusted to 0.5 McFarland standards. Using a sterile cotton swab, bacterial cells were streaked on Mueller-Hinton agar plate to facilitate consistent bacterial growth. A sterile forceps was used to place the antibiotic disc and E-test strips on the surface of the streaked agar plates. The plates were incubated at 37° C for 14-20 hours, after which antibiotic susceptibility was determined. Cut-off for assessing resistance were determined following the CLSI document M45 guidelines (CLSI, 2010).

2.3.4. Genetic screening and PCR assay

PCR assays were carried out to detect genes encoding accessory cholera enterotoxin (*ace*) (Shi *et al.*, 1998), zonula occludens toxin (*zot*), hemolysin (*hlyA*) (Rivera *et al.*, 2001), SXT-related integrase (*int*SXT) (Hochhut *et al.*, 2001) and biotype-specific (ET and CL)

toxin coregulated pilus (*tcpA*) (Rivera *et al.*, 2001), phage-encoded repressor (*rstR*) (Kimsey *et al.*, 1998), and repeat in toxin (*rtxC*) (Chow *et al.*, 2001) using primers and conditions described previously (**Table 2.2**). Double mismatch amplification mutation assay (DMAMA)-PCR was performed to identify three genotypes of the cholera toxin gene, i.e. *ctxB*1, *ctxB*3, and *ctxB*7, based on nucleotide substitutions at position 58, 115, and 203 (Naha *et al.*, 2012). *V. cholerae* O1 strains O395 (CL), N16961 (ET) and 2010EL-1786 were used as controls for PCR analysis. *V. cholerae* O1 isolates showing MIC for AZ ≥ 0.5 μg/mL were analyzed further for the macrolide resistance genes: *erm*(A), *erm*(B), and *erm*(C), that encode methylase; *ere*(A) or *ere*(B) encoding esterases; *mph*(A), *mph*(B), and *mph*(D) encoding phosphotransferases; and *mef*(A) and *msr*(A) encoding efflux pumps (**Table 2.3**) (Phuc Nguyen *et al.*, 2009).

Table 2.2. Primer sequences for virulence factor associated genes and genetic screening.

Target	Primer	Sequence (5`-3`)	Amplicon (bp)	Reference
ctxB	Rv-cla	CCTGGTACTTCTACTTGAAACG		Morita *et al*., 2008
	ctxB-F3	GTTTTACTATCTTCAGCATATGCGA	191	Naha *et al*., 2012
	ctxB-F4	GTTTTACTATCTTCAGCATATGCGC	191	Naha *et al*., 2012
tcpA	72F	CACGATAAGAAAACCGGTCAAGAG		Rivera *et al*., 2001
	477R (ET)	CGA AAGCACCTTCTTTCACGTTG	451	
	647R (CL)	TTACCAAATGCAACGCCGAATG	620	
ace	ace-F	TAAGGATGTGCTTATGATG GACACCC	316	Shi *et al*., 1998
	ace-B	CGTGATGAATAAAGATACT CATAGG		
zot	225F	TCGCTTAACGATGGCGCGTTTT	947	Rivera *et al*., 2001
	1129R	AACCCCGTTTCACTTCTACCCA		
hlyA	489F	GGCAAACAGCGAAACAAATACC	481	Rivera *et al*., 2002
	744F	GAGCCGGCATTCATCTGAAT	738	
	1184R	CTCAGCGGGCTAATACGGTTTA		
rtxC	rtxC-F	CGACGAAGATCATTGACGAC	263	Chow *et al*., 2001
	rtxC-R	CATCGTCGTTATGTGGTTGC		
intSXT	INT1	GCTGGATAGGTTAAGGGCGG	592	Hochhut *et al*., 2001
	INT2	CTCTATGGGCACTGTCCACATTG		
rstR	rstR1 F	CTTCTCATCAGCAAAGCCTCCATC	500	Kimsey *et al*., 1998
	rstR2 F	GCACCATGATTTAAGATGCTC	500	
	rstR3A R	TCGAGTTGTAATTCATCAAGAGTG		

ET, El Tor; CL, Classical

Table 2.3. Primer sequences for antibiotic resistance marker genes.

Target	Primer	Sequence (5`-3`)	Amplicon (bp)	Reference
mph(A)	mphAF	GTGAGGAGGAGCTTCGCGAG	403	Phuc Nguyen *et al.*, 2009
	mphAR	TGCCGCAGGACTCGGAGGTC		
mph(B)	mphBF	GATATTAAACAAGTAATCAGAATAG	494	Phuc Nguyen *et al.*, 2009
	mphBR	GCTCTTACTGCATCCATACG		
erm(A)	ermAF	TCTAAAAAGCATGTAAAAGAAA	503	Phuc Nguyen *et al.*, 2009
	ermAR	CGATACTTTTTGTAGTCCTTC		
erm(B)	ermBF	GAAAAAGTACTCAACCAAATA	639	Phuc Nguyen *et al.*, 2009
	ermBR	AATTTAAGTACCGTTACT		
erm(C)	ermCF	TCAAAACATAATATAGATAAA	642	Phuc Nguyen *et al.*, 2009
	ermCR	GCTAATATTGTTTAAATCGTCAAT		
ere(A)	ereAF	GCCGGTGCTCATGAACTTGAG	420	Phuc Nguyen *et al.*, 2009
	ereAR	CGACTCTATTCGATCAGAGGC		
ere(B)	ereBF	TTGGAGATACCCAGATTGTAG	537	Phuc Nguyen *et al.*, 2009
	ereBR	GAGCCATAGCTTCAACGC		
mef(A)	mefAF	AGTATCATTAATCACTAGTGC	345	Phuc Nguyen *et al.*, 2009
	mefAR	TTCTTCTGGTACTAAAAGTGG		
msr(A)	msrAF	GCACTTATTGGGGGTAATGG	384	Phuc Nguyen *et al.*, 2009
	msrAR	GTCTATAAGTGCTCTATCGTG		

2.4. RESULTS

2.4.1. Phenotypic and geneotypic characteristics

All 97 isolates produced colonies typical of *V. cholerae* on both taurocholate tellurite gelatin agar (TTGA) and thiosulfate citrate bile-salts sucrose (TCBS) agar. These isolates gave biochemical reactions characteristic of *V. cholerae* and reacted to polyvalent antibody specific for *V. cholerae* serogroup O1. Of 97 isolates, 89 gave positive agglutination with monovalent Ogawa antisera, while the remaining eight reacted positively with monovalent Inaba antisera. All isolates were serologically identified as *Vibrio cholerae* O1. Notably, the serotype of 89 strains was determined to be Ogawa and eight to Inaba (**Table 2.4**).

Table 2.4. Genetic characteristics and drug resistance of *V. cholerae* O1 isolated in Bangladesh.

Year of isolation	Number of strains	Source	Serotype	*wbeO1*	*ctxA*	*ace*	*zot*	*tcpA*	*rtxC*	*ctxB* type	*rstR*	*hlyA*	int_{SXT}	Drug Resistance Profile
2010	7	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	3	Env	OGET	+	+	+	+	ET	+	*B7*	ET	+	+	S, NA, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
2011	3	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	1	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, SXT
	11	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	4	Clinical	OGET	+	+	+	+	ET	+	*B7*	ET	+	+	S, NA, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, E, SXT
2012	5	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	3	Env	INET	+	+	+	+	ET	+	*B1*	CL	+	+	S, SXT
	1	Env	INET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, SXT
	4	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	2	Clinical	INET	+	+	+	+	ET	+	*B1*	CL	+	+	S, SXT
	1	Clinical	INET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	1	Clinical	INET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B7*	ET	+	+	S, NA, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	P, S, NA, TE, SXT
2013	10	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	2	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	P, S, NA, TE, SXT
	10	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	P, S, NA, TE, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	-	NA
2014	9	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	1	Env	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	P, S, NA, TE, SXT
	10	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	+	S, NA, TE, SXT
	2	Clinical	OGET	+	-	-	-	ET	+	-	-	+	+	S, NA, TE, SXT
	1	Clinical	OGET	+	+	+	+	ET	+	*B1*	ET	+	-	NA

Env, environmental; OGET, Ogawa El Tor; INET, Inaba El Tor; ET, El Tor; CL, classical

Genomic DNA of all isolates (n = 97) amplified *V. cholerae* species-specific genes, namely *ompW* and O-antigen biosynthetic-*wbe* (O1) confirming identification as *V. cholerae* O1. None amplified the O-antigen biosynthetic-*wbf* (O139). As shown in **Table 2.4**, except for two isolates, all amplified CTX-phage mediated genes *ctxA*, *ace*, and *zot*, suggesting 95 of the isolates were toxigenic *V. cholerae* O1. The PCR assay results also showed *hlyA* gene present in all isolates (**Table 2.4**). Of 97 *V. cholerae* O1, *intSXT* was identified in 95 of the isolates, while two lacked *intSXT*. All *V. cholerae* O1 isolates contained ET biotype specific *tcpA* and *rtxC,* reflecting ET attributes. Among the 95 toxigenic *V. cholerae* O1 isolates, 90 possessed *rstR* of the ET biotype (*rstR*^{ET}), while the remaining five revealed CL biotype specific *rstR*^{CL}. Unlike hybrid ET strains, none of the toxigenic isolates contained both *rstR*^{ET} and *rstR*^{CL}. DMAMA-PCR detected the cholera toxin gene of CL biotype (*ctxB*1) in 87 of the *V. cholerae* O1 isolates, while eight had Orissa variant or Haiti variant cholera toxin *ctxB*7 (**Table 2.4**).

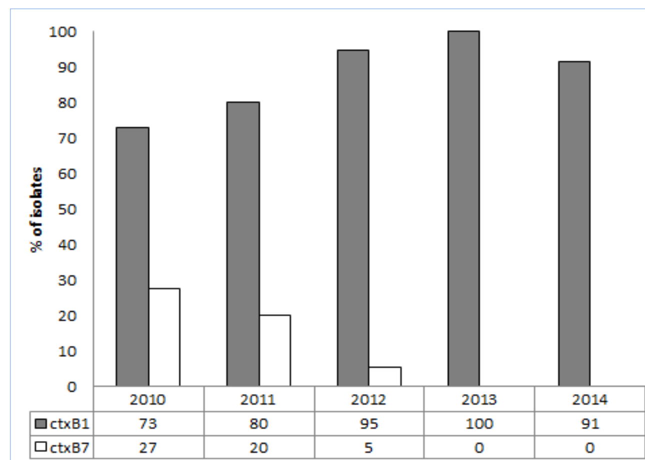

Figure 2.1. Distribution of *ctxB* genotypes in 97 *V. cholerae* O1 from Mathbaria, Bangladesh isolated during 2010 and 2014. In both clinical and environmental isolates, *ctxB1* (grey bars) was predominant each year. Although *ctxB7* in the *V. cholerae* O1 isolates (white bars) was detected at relatively low percentage in 2010, 2011, and 2012, it was not detected in 2013 and 2014.

Overall, the PCR results confirmed 90 of the *V. cholerae* O1 isolates were atypical ET, possessing *rstR*^{ET} and either the *ctxB1* or *ctxB7 ctxB* gene. Five toxigenic *V. cholerae* O1 possessing *rstR*^{CL} and *ctxB1* are designated as variant ET and their genetic

attributes were similar to the Matlab variant (MJ1236) isolated in 1994 in Matlab, Bangladesh. As shown in **Table 2.4**, *V. cholerae* O1 variant ET was isolated from both clinical and environmental sources in Mathbaria, Bangladesh only in 2012. *V. cholerae* O1 atypical ET was associated with cholera cases in June, 2010, and December, 2014, in Mathbaria, Bangladesh and these strains were also isolated frequently from environmental sources (**Table 2.4**).

As shown in **Figure 2.1**, the CL type cholera toxin genotype, *ctxB1,* was predominant, having been detected in 73%, 80%, 95%, 100%, and 91% *V. cholerae* O1 isolates in 2010, 2011, 2012, 2013, and 2014, respectively. In contrast, Orissa or Haiti variant cholera toxin genotype *ctxB7* was found in 27%, 20%, and 5% of the isolates in 2010, 2011, and 2012, respectively. Remarkably, *ctxB7* was not detected in *V. cholerae* O1 isolated in 2013 and thereafter (**Table 2.4**). Although, 9% of the *V. cholerae* O1 were non-toxigenic in 2014, *ctxB1* was the only genotype prevailed among toxigenic isolates in 2013 and 2014.

2.4.2. Antimicrobial susceptibility

Antimicrobial susceptibility tests, using ten different antibiotics revealed that 93% of the total set of *V. cholerae* O1 isolates were MDR, i.e., resistant to at least three different antibiotics (**Table 2.4**). As shown in **Figure 2.2(A)**, six different resistance profiles were observed, with a range of resistance to one to five antibiotics during 2010 and 2014. *V. cholerae* O1 showing resistance to S, NA, TE, and SXT was the dominant pattern (53 - 91%) each year between 2010 and 2014 (**Figure 2.2(A)**). Interestingly, resistance of *V. cholerae* O1 to P, S, NA, SXT, E, and TE varied during the years of the study period.

As shown in **Figure 2.2(B)**, 100% of the *V. cholerae* O1 showed resistance to S and SXT in 2010, 2011, and 2012. However, S and SXT resistance fell to 96% the following two years, 2013 and 2014. The SXT-related integrase (*intSXT*) was detected in all isolates resistant to S and SXT, suggesting the SXT/R391 ICE mediated the resistance to S and SXT. Except for five of the *V. cholerae* O1 variant ET isolated in 2012, all were resistant to NA (**Figure 2.2(B)**), an indicator of reduced susceptibility to CIP.

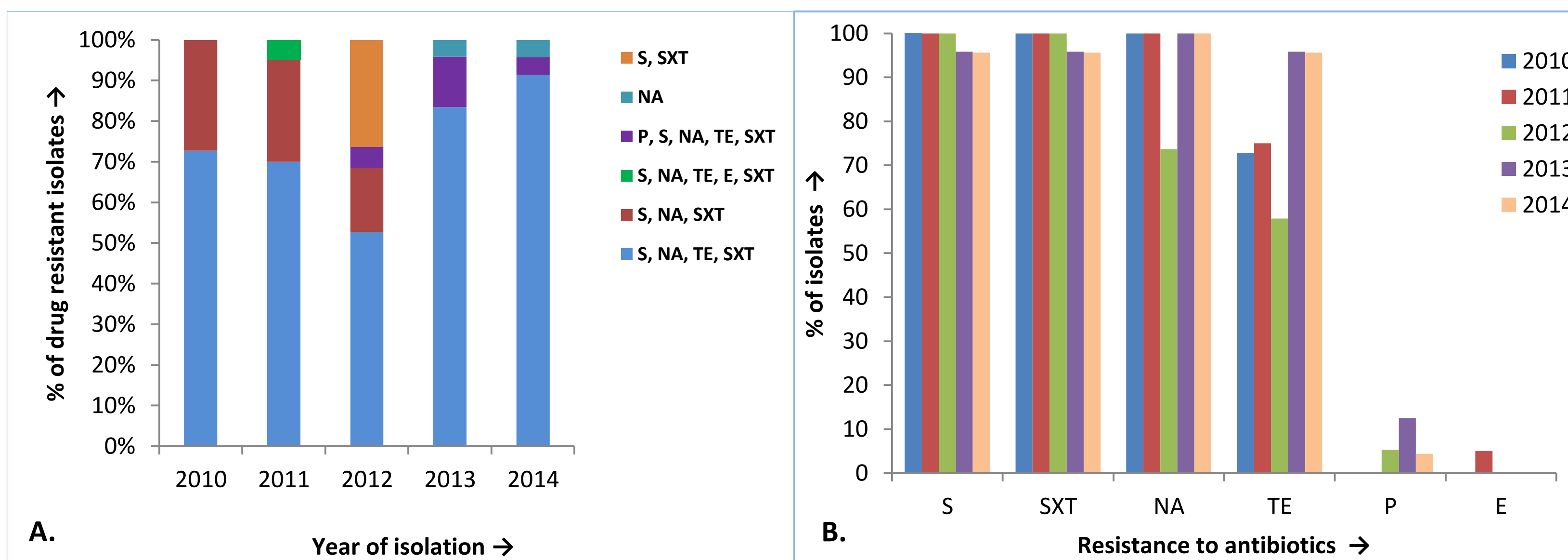

Figure 2.2. (A). Drug resistance profile of 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh. Six different resistance profiles were observed for *V. cholerae* O1 isolated during 2010 and 2014, of which four profiles were multidrug resistance: resistance to S, NA, TE, and SXT was most abundant and predominant in *V. cholerae* O1. (B) *V. cholerae* O1 isolated during 2010 and 2014 showing resistance to six different drugs. The majority of the isolates were resistant to S, SXT, TE, and NA, while only a few were resistant to P and E.

TE resistant *V. cholerae* O1 comprised 73%, 75%, 58%, 96%, and 96% in 2010, 2011, 2012, 2013, and 2014, respectively (**Figure 2.2(B)**). Of 97 *V. cholerae* O1, five showed resistance to P during 2012 and 2014, while only one isolate showed resistance to E in 2011. Notably, all 97 *V. cholerae* O1 isolates were uniformly sensitive to AM, CIP, C, and K.

2.4.2.1. MIC of ciprofloxacin

The MIC of CIP of all 97 *V. cholerae* O1 isolates was determined to be 0.003 – 0.75 μg/ml during 2010 and 2014. As shown in **Table 2.5**, MIC_{50} and MIC_{90} of CIP was 0.5 μg/ml in 2010. However, the MIC_{50} and MIC_{90} were 0.38 μg/ml in 2011, maintained consistently over the following years until 2014. Five of the *V. cholerae* O1 variant ET isolated in 2012 had an MIC for CIP of 0.003 μg/ml. Only 1% of the total set of strains had the highest MIC, 0.75 μg/ml, while 77% had an MIC of 0.38 μg/ml (**Figure 2.3(A)**).

Table 2.5. Minimum inhibitory concentration of ciprofloxacin and azithromycin for 97 *V. cholerae* O1 isolates

Year of isolation	No. of strains	Ciprofloxacin			Azithromycin		
		MIC Range (μg/ml)	MIC_{50}	MIC_{90}	MIC Range (μg/ml)	MIC_{50}	MIC_{90}
2010	11	0.38 - 0.75	0.5	0.5	0.19 - 0.75	0.25	0.38
2011	20	0.25 - 0.5	0.38	0.38	0.25 - 1.5	0.25	0.5
2012	19	0.003 - 0.38	0.38	0.38	0.25 - 0.75	0.38	0.75
2013	24	0.25 - 0.38	0.38	0.38	0.25 - 1	0.25	0.38
2014	23	0.25 - 0.38	0.38	0.38	0.19 - 2	0.25	0.5

2.4.2.2. MIC of azithromycin

The MIC of AZ for 97 of the *V. cholerae* O1 isolates was 0.19 – 2.00 μg/ml. As shown in **Table 2.5**, except for the 2012 isolates, MIC_{50} was consistently 0.25 μg/ml. Year-wise data revealed that the lowest MIC_{90} was 0.38 μg/ml in 2010 and 2013 and increased to 0.5 μg/ml and 0.75 μg/ml in 2011 and 2012, respectively. As shown in **Figure 2.3(B)**, 52% and 27% of the total isolates had MIC 0.25 and 0.38 μg/ml, respectively.

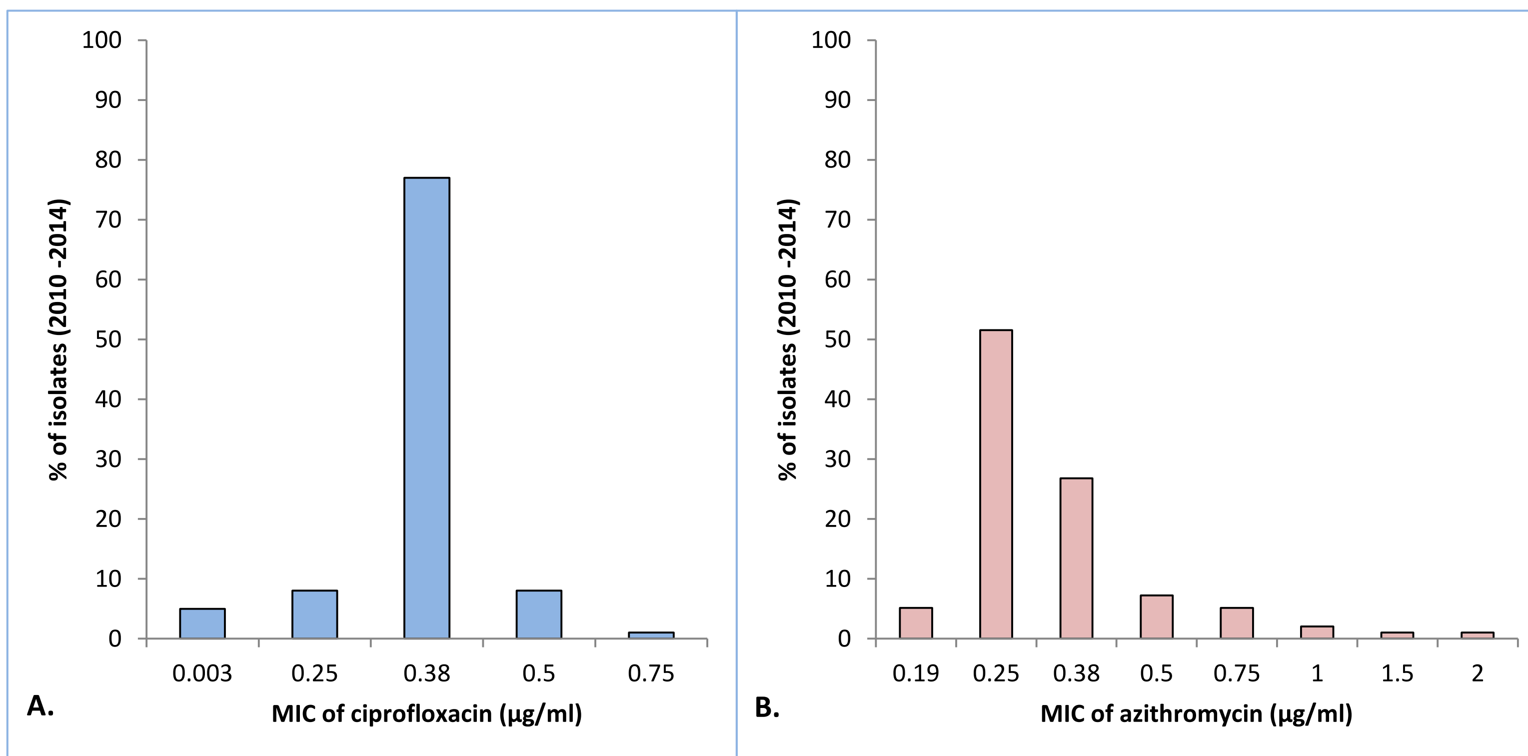

Figure 2.3 (A). MIC of CIP for 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh. The MIC range for CIP was 0.003 µg/ml to 0.75 µg/ml. The majority of the isolates showed MIC 0.38 µg/ml, while a low percentage had MIC of 0.5 µg/ml and 0.75 µg/ml. (B) MIC of AZ for 97 *V. cholerae* O1 isolates from Mathbaria, Bangladesh. The range of MIC for AZ was 0.19 µg/ml to 2 µg/ml.

The highest MIC of 2.00 μg/ml occurred in 1% of the *V. cholerae* O1 isolates (**Figure 2.3(B)**). Sixteen (16%) *V. cholerae* O1 with an MIC of ≥ 0.5 μg/ml were analyzed further for 10 macrolide resistance genes. PCR assay results revealed that none of the isolates contained the following macrolide resistance genes: *erm*(A); *erm*(B); *erm*(C); *ere*(A); *ere*(B); *mph*(A); *mph*(B); *mph*(D); *mef*(A); and *msr*(A).

2.5. DISCUSSION

Endemic cholera occurs in many geographic locations of Bangladesh each year, with two distinct seasonal peaks, one in the spring (March - May) and the other in the fall (September-November) (Emch *et al.*, 2008, Akanda *et al.*, 2013). The Ganges delta region of the Bay of Bengal is a well-known reservoir of *V. cholerae* where it has established residence for centuries (Sack *et al.*, 2004). Historically, this part of Asia has been affected severely by both CL and ET cholera during the seventh pandemic up to 1991, prior to the disappearance of CL strains from Bangladesh (Siddique & Cash, 2014). Epidemiological data suggest that most of the recorded epidemics struck the coastal populations first (Jutla *et al.*, 2010), a pattern typical of recent epidemics in Bangladesh as well before reaching inland (Akanda *et al.*, 2013). Since the beginning of the ongoing seventh pandemic, *V. cholerae* O1 strains have undergone multiple genetic changes, with the evolution of new clones and also atypical ET strains (Chun *et al.*, 2009, Safa *et al.*, 2010, Kim *et al.*, 2014). Results of this study show association of *V. cholerae* O1 atypical ET with cholera occurring in the coastal areas of Mathbaria, Bangladesh during 2010 and 2014. Since 2001, atypical ET emerged as the major cause of cholera in Bangladesh superseding prototype ET, although isolated a decade earlier in the 1990s in Matlab, Bangladesh (Nair *et al.*, 2006, Safa *et al.*, 2006). This transition was considered remarkable for cholera epidemiology, mainly because atypical ET strains possessing CL cholera toxin (*ctxB1*) cause a more severe cholera than prototype ET (Siddique *et al.*, 2010). In recent years, several non-synonymous mutations have been detected in the *ctxB* gene, although, correlation of these mutations with clinical outcomes of the disease remains to be clarified (Kim *et al.*, 2015). The *ctxB7* genotypes have an amino acid substitution at position 20 [histidine (H)→asparagine (N)] first reported in a cholera outbreak in Orissa, India (Kumar *et al.*, 2009). Later, *V. cholerae* O1 carrying *ctxB7* was

determined to be associated with cholera in Haiti, Zimbabwe, India, Bangladesh, Nepal, Nigeria, and Cameroon (Quilici *et al.*, 2010, Chin *et al.*, 2011, Hasan *et al.*, 2012, Rashed *et al.*, 2012, Marin *et al.*, 2013, Dixit *et al.*, 2014). In Mathbaria, Bangladesh, *ctxB1* was consistently dominant to *ctxB7* during 2010 and 2012, whereas *ctxB7* was not detected thereafter. The alternating dominance of *ctxB1* and *ctxB7*, i.e., one genotype disappearing transiently for two or three years and reappearing in the following years with remarkable dominance, was previously observed in *V. cholerae* O1 causing cholera in Kolkata, India, and Dhaka, Bangladesh (Rashed *et al.*, 2012, Mukhopadhyay *et al.*, 2014, Rashid *et al.*, 2015).

Bacterial resistance to antimicrobial drugs is a serious public health concern worldwide and antibiotic therapy constitutes a major component of the clinical management of cholera. An antimicrobial drug considered to be a successful therapeutic agent may not be successful in the future and notably so if *V. cholerae* acquires resistance to drugs of choice. Resistance can arise from single or multiple mutations in target genes or by acquisition of resistance genes carried by mobile genetic elements, such as plasmids, transposons, integrons, and ICEs (Kitaoka *et al.*, 2011). Prior to the use of macrolides and fluoroquinolone drugs for treatment of cholera, TE was the drug of choice, except for young children and pregnant women (Greenough *et al.*, 1964, Sack *et al.*, 2004). However, tetracycline was limited as a drug of choice because of the emergence of *V. cholerae* O1 resistant to AMP, KN, S, and SXT, as well as TE, in Asia and Africa (Mhalu *et al.*, 1979, Glass *et al.*, 1980). In this study, 93% of *V. cholerae* O1 strains proved to be multidrug resistant, mostly resistant to S, NA, TE, and SXT. Despite having spatio-temporal variation in the resistance profiles, multidrug resistant *V. cholerae* O1 was consistently identified as the etiological agent of cholera epidemics in Asia and Africa, and most recently in Haiti (Mhalu *et al.*, 1979, Glass *et al.*, 1980, Dalsgaard *et al.*, 2001, Jain *et al.*, 2011, Sjolund-Karlsson *et al.*, 2011, Rashed *et al.*, 2012, Tran *et al.*, 2012). A recent study from China reported that the prevalence of multidrug resistant *V. cholerae* O1 strains has been increased rapidly since 1993, showing resistance to AMP, NA, TE, and SXT (Wang *et al.*, 2012). The same study also revealed relatively low numbers of *V. cholerae* O1 with reduced susceptibility to azithromycin in China isolated

only in 1965, 1998, and 2006 (Wang *et al.*, 2012). Drug resistant markers, such as SXT ICE, class 1 integrons, and low molecular weight plasmids are commonly found in multidrug resistant *V. cholerae* O1 (Kitaoka *et al.*, 2011). Interestingly, a recent study reported the presence of a transmissible multidrug resistant plasmid (IncA/C) in Haitian *V. cholerae* isolates possessing several multidrug resistance determinants, i.e., *aac(3)-IIa*, bla_{CMY-2}, $bla_{CTX-M-2}$, $bla\text{-}_{TEM-1}$, *dfrA15*, *mphA*, *sul1*, *tetA*, *floR*, *strAB*, and *sul2* (Folster *et al.*, 2014).

Among the fluoroquinolone antibiotics, only CIP has been recommended by the Pan American Health Organization and International Centre for Diarrhoeal Disease Research, Bangladesh for treatment of cholera. Although, *V. cholerae* O1 has not shown complete resistance to CIP, current epidemiological data confirm a gradual increase in the MIC of CIP has been occurring (Khan *et al.*, 2015). *V. cholerae* O1 with reduced susceptibility to CIP has been reported in different parts of the world and appears to be disseminating globally (Islam *et al.*, 2009, Quilici *et al.*, 2010, Sjolund-Karlsson *et al.*, 2011, Tran *et al.*, 2012, Khan *et al.*, 2015). A recent study showed that the MIC of CIP for *V. cholerae* O1 has increased 45 fold in a 19 year time-span in Bangladesh. That is, the MIC was 0.010 µg/ml in 1994 and has increased dramatically to 0.475 µg/ml in 2012 (Khan *et al.*, 2015). In our study, 95% of the *V. cholerae* O1 isolated in Mathbaria, Bangladesh, showed reduced susceptibility to CIP during 2010 and 2014. Notably, the CIP MIC_{50} and MIC_{90} did not show rapid change over the five year of our study period and MIC remained below the susceptibility breakpoint (≤ 1 µg/ml) according to CLSI guidelines (CLSI, 2010). It is important to note that all *V. cholerae* O1 isolates were resistant to NA, another drug in the quinolone group. *V. cholerae* O1 resistance to NA is an indicator of reduced susceptibility to CIP (Khan *et al.*, 2015). The genetic basis of quinolone drug resistance in *V. cholerae* is the accumulation of mutations in *gyrA* ($83_{Ser \rightarrow Ile}$) and *parC* ($85_{Ser \rightarrow Leu}$) (Kitaoka *et al.*, 2011). These point mutations have been detected in currently circulating *V. cholerae* O1 associated with cholera epidemics in Bangladesh, India, Nepal, Nigeria, Cameroon, and Haiti (Quilici *et al.*, 2010, Sjolund-Karlsson *et al.*, 2011, Hasan *et al.*, 2012, Dixit *et al.*, 2014).

Frequent use of a specific group of antibiotics for treatment of cholera over a prolonged period will increase the likelihood of resistance. Global dissemination of *V. cholerae* O1 with reduced CIP sensitivity raises a serious concern for clinical management of cholera in countries where the disease is endemic. Results of a recent study showed that single-dose CIP used to treat cholera was not as effective as it was in the past because of the emergence of *V. cholerae* O1 less susceptible to CIP and NA (Khan *et al.*, 2015). Single dose AZ has been introduced as an alternative treatment for cholera in India and Bangladesh. However, the sensitivity breakpoint guidelines for the AZ disc diffusion assay has not yet been published by the CLSI for *V. cholerae* (CLSI, 2010). In this study, all *V. cholerae* O1 isolates showed reduced susceptibility to AZ and the MIC for 1% of the isolates was at the sensitivity breakpoint borderline (≤ 2 µg/ml). Interestingly, none of the *V. cholerae* O1 (MIC ≥ 0.5 µg/ml) possessed macrolide resistance genes that have been reported for the *Enterobacteriaceae* (Phuc Nguyen *et al.*, 2009). Although at a relatively low incidence, E and AZ resistant *V. cholerae* O1 have been reported in Bangladesh and India (Faruque *et al.*, 2007, Bhattacharya *et al.*, 2012).

2.6. CONCLUSION

Reduced susceptibility to CIP and AZ is alarming for cholera-endemic countries of Asia and Africa. Environmental factors trigger seasonal cholera in endemic countries including Bangladesh, but cholera cases have occurred in other countries immediately after a devastating natural calamity, e.g., floods, earthquakes, typhoons, and cyclones. The morbidity and mortality rates of cholera, which were under control for several decades, can be expected to increase if *V. cholerae* O1 acquires total resistance to currently used drugs. Considering the global burden of cholera, it is important that the appropriate antibiotic and appropriate concentration be used to treat cholera. Indiscriminate use of antibiotics, for example in agriculture and animal husbandry for disease management should be controlled to assure continued success of antibiotic for the treatment of disease in humans, including cholera. Therefore, global monitoring of antimicrobial sensitivity of *V. cholerae* O1 is essential to assess clinical efficacy worldwide.

CHAPTER 3:

GENOTYPING OF *VIBRIO CHOLERAE* O1 TO DETERMINE GENETIC CHANGES DURING SEASONAL OUTBREAKS OF CHOLERA

3.1. ABSTRACT

Cholera remains a major public health problem in countries of south Asia due to lack of sanitation and safe drinking water. In cholera endemic areas, several biotic and abiotic factors of the ecosystem contribute to the disease dynamics of cholera. The aquatic reservoir of toxigenic *V. cholerae* O1 serves as the primary source of pathogen at the beginning of a cholera outbreak. Understanding the contribution of strains from the environment compared to those strains isolated from cases during outbreaks has proven difficult. In this study, Multilocus variable tandem repeat analyses (MLVA) were performed on 222 environmental and clinical isolates of *V. cholerae* O1 systematically collected during outbreaks during 2010 and 2011 in two geographically distinct areas of Bangladesh, e.g., Chhatak, and Mathbaria. MLVA revealed a total of 51 five-locus genotypes in 222 *V. cholerae* O1 isolates. However, only three genotypes were exclusively common in Mathbaria and Chhatak regardless of the sample sources. Genetic analyses of isolates from Chhatak identified 30 MLVA genotypes, whereas, 24 MLVA genotypes were detected in *V. cholerae* isolates from Mathbaria. In Chhatak, 75 of 79 isolates belong to the same clonal complex, in which extensive differentiation was found in a temporally consistent pattern of successive mutations at single loci. Fifty-nine isolates were collected from six individuals and most isolates from an individual differed by sequential single locus mutations. In Mathbaria, 60 of 84 isolates represented two separate clonal complexes, while remaining isolates were singletons. The small number of genetic lineages in cases, compared with strains isolated from the environment is consistent with accelerated transmission of some strains among humans during an outbreak.

3.2. INTRODUCTION

Cholera remains a major public health problem in many areas of the world and affects millions of people each year, causing a significant number of deaths (Sack *et al.*, 2006, Ali *et al.*, 2012). Cholera transmission is seasonal in Bangladesh, with two annual peaks initiated by emergence of *V. cholerae* from environmental reservoirs (Huq *et al.*, 1983, Huq *et al.*, 1984). The infectious dose of *V. cholerae* is estimated to be 10^5 to 10^8 CFU

with the lower estimates associated with a buffered stomach (Lombardo *et al.*, 2007). The change in niche, from environment to human intestine, induces a physiological change that alters the expression of majority of genes in *V. cholerae* (Merrell & Camilli, 2000). One outcome of this alteration is that, after exiting the colon, the infectious dose is reduced, referred to as the "hyperinfectious" state that persist only for a brief period of time (Merrell, 2002). After returning to the aquatic environment, *V. cholerae* reverts to a standard infectious state (Merrell, 2002), with ability to cause illness potentially reduced even further after 24 hours (Nelson *et al.*, 2008). The relative contribution of strains from the environment and recently shed hyperinfectious strains to propagation of an outbreak of cholera remains controversial. The terms "slow" and "fast" have been used to distinguish between the two modes of transmission where slow refers to the human-to-aquatic environment-to-human pathway (which does not have time constraints) and fast to presumed person-to-person or person-to-household environment-to-person in which transfer is more likely immediately after fecal shedding, i.e., when strains are in a hyperinfectious state (Morris, 2011).

Microbiological, genetic, and modeling approaches have been taken to understand the relative contribution of the hypothesized slow and fast routes of transmission in outbreaks. A major problem with the microbiological and genetic approaches has been lack of genetic diversity to track strains. Many methods, including pulse field gel electrophoresis which is often used to analyze outbreaks, reveal too few genetic differences between isolates to be useful in tracking the microdynamics of strains. This problem was resolved, in part, by identification of loci containing a variable number of tandem repeats, which provided enough genetic variability to permit tracking of specific strains (Danin-Poleg *et al.*, 2007, Ghosh *et al.*, 2008, Stine *et al.*, 2008). However, initial studies using multilocus variable tandem repeat analysis (MLVA) did not sample intensively enough to distinguish optimally between slow and fast transmission. One study carried out in rural Bangladesh, in which isolates were collected every fortnight, showed isolates from different geographic locations collected in different seasons and from clinical and environmental sources had only a few genotypes in common (Stine *et*

al., 2008). When isolates from the same geographic location were compared, genotypes collected within the same season or from the same source tended to be similar to one another compared to those from different seasons or sources. Although this could be interpreted as evidence in support of fast transmission, most environmental isolates were not collected in the same month as were most of the clinical isolates, making interpretation of data difficult. In a second study, isolates from one hundred index cases and their household relatives were analyzed (Kendall *et al.*, 2010). Remarkably, individuals within a single household often had genetically unrelated isolates, implying either different sources of infection or a single source with multiple genetic lineages. The observed unexpected variability was reinforced by the observation that unrelated genetic isolates were often isolated from a single stool. However, the study design of sampling three households per month did not provide sufficient resolution to address transmission pathways. Mathematical modeling of incidence data from outbreaks has been used to estimate the contribution of fast and slow transmission (Mukandavire *et al.*, 2011, Mukandavire *et al.*, 2013). While promising, these estimates have been based on clinical surveillance data without more detailed underlying epidemiologic information. Furthermore, it may not be possible to estimate the contribution of two transmission mechanisms from incidence data alone, when the time scale of slow is similar to that of fast transmission (Eisenberg *et al.*, 2013).

Here we characterized 222 clinical and environmental *V. cholerae* O1 isolates from outbreaks in two geographically distinct locations using MLVA. The objective was to determine genetic relatedness among the isolates, notably clonal relationships between environmental and clinical isolates and to further explore the relative contribution of different transmission pathways to disease occurrence.

3.3. MATERIALS AND METHODS

3.3.1. Sample Collection

Clinical and environmental *V. cholerae* were collected in Chhatak and Mathbaria, Bangladesh, from October 2010 to May 2011. Chhatak is a village located 180 km NE of

Dhaka in the central region and Mathbaria is a rural area 150 km SSE of Dhaka in the coastal region. Rectal swabs were collected from all cases presenting to the health center with symptoms of suspected cholera during three consecutive days every week, representing ca. 40% of suspected cases per week during the peak season of cholera. To characterize within host variability of *V. cholerae*, nine to ten isolates were selected from a culture of each of six cases presenting at the Chhatak clinic. Environmental isolates were obtained from water and plankton samples collected at six sites (pond or river sites) in each community, as previously described (Alam *et al.*, 2006). The same environmental sites were sampled and the same sampling methods employed across all collection times. Water and plankton samples were collected, concentrated, and analyzed for the isolation of *V. cholerae* following standard procedures (Alam *et al.*, 2006, Huq *et al.*, 2006). Isolation was done using standard culture methods and *V. cholerae* identified using standard methods (Tkachenko, 1982, Alam *et al.*, 2006). DNA was prepared from broth cultures of *V. cholerae* using PrepMan Ultra (ABI, Foster City, CA, USA), according to manufacturer's instructions. All samples were collected according to protocols approved by Institutional Review Boards at Johns Hopkins University, University of Maryland, and icddr,b.

3.3.2. Genetic Analysis

All PCR was conducted by a single technician. Fluorescently labeled PCR primers were used to amplify five loci with variable length tandem repeats (**Table 3.1**), and a new reverse primer was used for VCA0283 (AGCCTCCTCAGAAGTTGAG) (Kendall *et al.*, 2010). The reaction yielded products which were separated, detected, and sized using a model 3730xl Automatic Sequencer, internal lane standards (Liz600) and GeneScan program (ABI, Foster City). Genotypes were determined using published formulas to calculate the number of repeats from the length of each allele except for VC0437, now (x-252)/7 (Kendall *et al.*, 2010). Five loci were ordered, VC0147, VC0437, VC1650, VCA0171, and VCA0283. A genotype e.g. 9-4-6-19-11 was interpreted as the isolate having alleles of 9, 4, 6, 19, and 11 repeats at the five loci, respectively. Relatedness of isolates was assessed using eBURSTv3 (http://eburst.mlst.net). Genetically related

genotypes are defined as possessing identical alleles at four of the five loci and groups of related genotypes termed clonal complexes. Additional sensitivity analyses were done using only the three more stable loci from the large chromosome (VC0147, VC0437, VC1650), with relatedness determined by having identical alleles at two of the three loci, as was done previously (Kendall *et al.*, 2010, Lam *et al.*, 2012).

Table 3.1. List of targeted loci, primers, and mathematical formula for MLVA

Locus	Dye-primer	Expected rangea (bp)	Formula
VC0147	Tet -ACGTGCAGGTTCAACCGTG TTGTCATGGCTTGGATTTGG	186–224	(x -150)/6
VC0437	Tet-GTTGCCGCCATCACCAGCTTG CGTTAGCATCGAAACTGCTG	265–301	(x - 252)/7
VC1650	Tet-CCGCTAACTGAGTGACCGC CTACCAAGCGGCGGTTAAGCTG	370–440	(x - 307)/9
VCA0171	Fam -AGGCGCCTGATGACGAATCC GCTGAAGCCTTTCGCGATCC	316–442	(x - 270)/6
VCA0283	Fam-GGAGGTAGCTACGAATTCTAC GTACATTCACAATTTGCTCACC	118–244	(x - 95)/6

3.3.3. Statistical Analyses

A Chi-squared test was used to determine whether the null hypothesis that the genotype isolated from each clinical case has an equal probability of being from each of the genetic lineages (for five-locus analysis, from one of the observed clonal complexes or singletons; or for three-locus analysis, the same genotype) from the environment ($p_1 = p_2 = \ldots = p_n$), which if true, would be evidence in support of a strong role of the slow transmission route in these outbreaks. The Simpson Index was calculated as a measure of genotype diversity (low values indicating more diversity) within each location as $D = \sum_{i=1}^{G} p_i^2$, where G is the total number of genotypes at a particular location and p_i is the proportion of genotypes of type i.

3.4. RESULTS

A total of 222 *V. cholerae* O1 Ogawa isolates were collected from both cases and from environmental samples in Chhatak and Mathbaria from October 2010 to May 2011. In

Chhatak, the environmental sample collection from six ponds yielded five isolates and stools of 74 cases were sampled with a single isolate from each and nine or ten additional isolates from six of these clinical specimens. In Mathbaria, sampling yielded 28 isolates from six ponds and single isolates were obtained from 56 cases. A total of 51 five-locus genotypes were observed, of which only three genotypes were from both Mathbaria and Chhatak, a result consistent with prior observation of geographic differences (Stine *et al.*, 2008). Thus, isolates from the two locations were analyzed separately.

3.4.1. Chhatak

Extensive genetic variation was detected among the 138 isolates (**Table 3.2**, Simpson Index 0.10). The number of distinct alleles at loci VC0147, VC0437, VC1650, VCA0171 and VCA0283 was 6, 3, 1, 12, and 8, respectively, with a minimum of four and up to 27 repeat units. Thirty different genotypes (each a unique combination of alleles at the five loci) were detected. Only one genotype was found in both clinical and environmental isolates. An additional 26 genotypes were found among the clinical and three additional among the environmental isolates.

Analysis of the genotypes by eBURST showed that 26 genotypes clustered into one clonal complex (**Figure 3.1(A)**) and four genotypes were singletons. The 'founder' of a clonal complex is defined as the genotype with the largest number of single locus variants (SLV). In this clonal complex, the founder genotype 9-4-14-21-17 was present in seven clinical and two environmental isolates. The ten SLV genotypes diverging from the founder genotype were all clinical isolates. Three of the ten SLV genotypes (9-4-14-23-17; 9-4-14-22-17; 9-4-14-21-18) differentiated further into nine double locus variants (DLV). Three DLV genotypes were differentiated even further into six additional variants, often a variant of a locus already noted as a variant.

Table 3.2. Genotypes of *V. cholerae* O1 isolated in Chhatak.

Genotypes	Source	No. of Isolates[a] (Single/stool or water)	No. of Isolates[b] (Ten or nine/stool)	Clonal complex
8-4-14-21-18	Clinic	7		1
8-4-14-22-18	Clinic	1		1
9-4-14-14-16	Clinic	8		1
9-4-14-14-17	Clinic	1		1
9-4-14-9-17	Clinic	1		1
9-4-14-20-17	Clinic	2		1
9-4-14-21-12	Clinic	11		1
9-4-14-21-13	Clinic	1		1
9-4-14-21-19	Clinic	1		1
9-4-14-21-18	Clinic	5	4	1
9-4-14-22-12	Clinic	2		1
9-4-14-22-16	Clinic	2		1
9-4-14-23-16	Clinic	1	1	1
9-4-14-23-17	Clinic	1	3	1
9-4-14-23-18	Clinic	5		1
9-4-14-23-19	Clinic	1		1
9-4-14-25-16	Clinic	1		1
9-4-14-25-17	Clinic	15	14	1
9-4-14-21-17*	Clin & Envi	5 & 2	18	1
9-4-14-27-18	Environment	1		1
8-4-14-21-11	Environment	1		1
7-4-14-23-17	Clinic		5	1
9-4-14-13-16	Clinic		1	1
9-4-14-18-16	Clinic		9	1
9-4-14-22-18	Clinic		2	1
9-4-14-22-17	Clinic		1	1
7-9-14-14-6	Clinic		1	Singleton, Oct 13
11-9-14-14-17	Clinic	2		Singleton, Oct 21
12-8-14-15-17	Clinic	1		Singleton, Oct 11
10-8-14-17-18	Environment	1		Singleton, Nov 1

[a] – Single isolates chosen from the 79 samples
[b] – Multiple isolates from 6 individuals
* Designates the founder genotype defined by eBURST as the genotype that differs from the largest number of genotypes at a single locus.

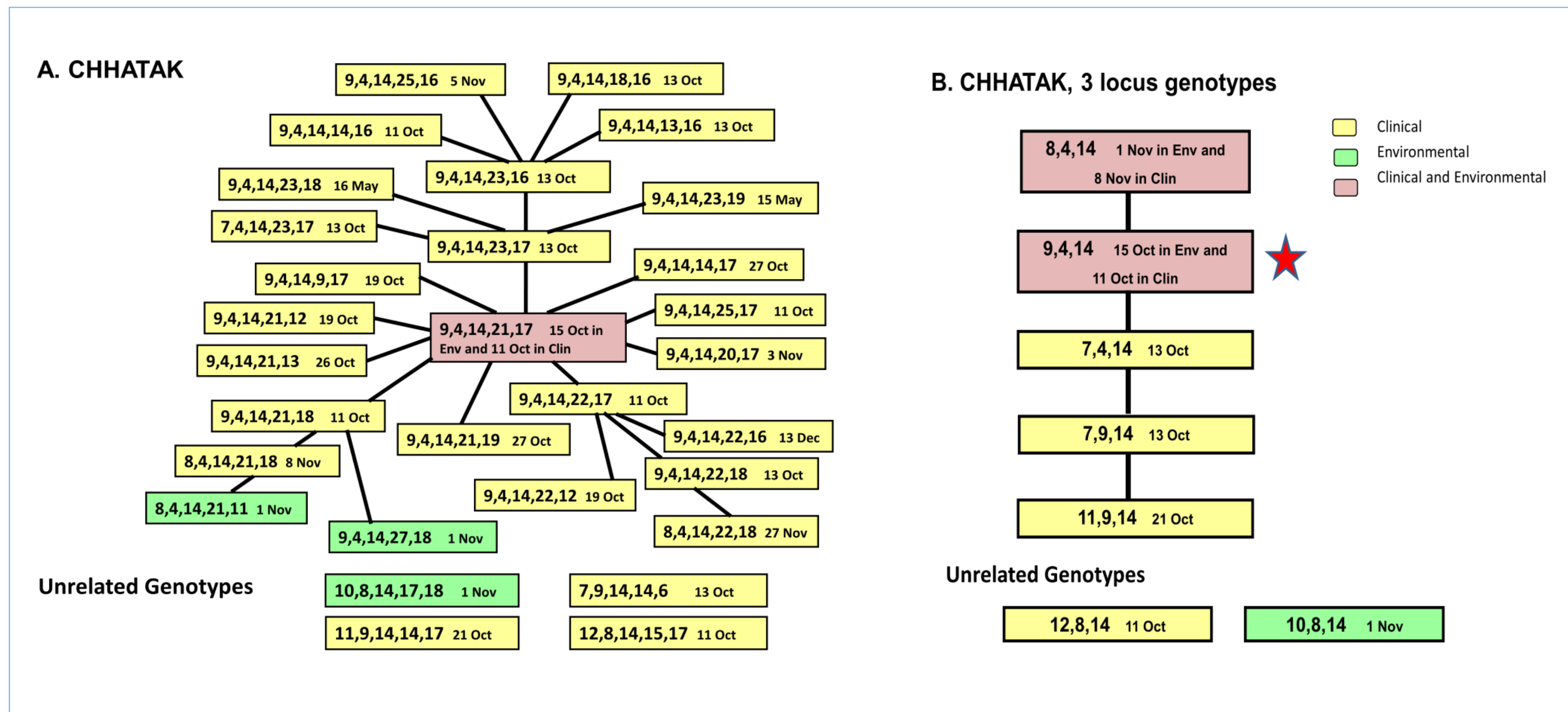

Figure 3.1. Diagrams of genetic relatedness between genotypes of Chhatak. Each genotype is identified by the number of repeats in the allele at the five loci VC0147, VC0437, VC1650, VCA0171 and VCA0283. The earliest date of detection is recorded in the box after the 5th allele. The background of the box indicates whether the genotype was clinical only (yellow), environmental isolates (green) or both (pink). A. Clonal complex of Chhatak genotypes and five unrelated genotypes based on a five-locus genotype. B. Analysis of three-locus genotypes of Chhatak.

In any outbreak, time is a critical variable (the earliest date of detection is incorporated into **Figure 3.1, 3.2, and 3.3**). The founder was isolated from multiple cases on the first day, 11th October, as were four related genotypes. Six additional related genotypes were observed thereafter on the 13th October. The founder was recovered from the environment on 15th October. The ten SLVs were collected between 11th October and 3rd November. Twenty-three of the 25 variants were isolated on the same day or after the genotype next closest to the founder - the expected result if the genotypes are derived successively from the founder. If instead of evolving during the outbreak, each genotype existed initially (in the environment) then we would expect the probability of finding related genotypes before vs. after the founder (or next closest to the founder) to be 0.50. However, we observed 23 of 25 after the founder (or next closest) and, if derived from a random (binomial) process, we would expect to see 23 or more with probability of 10^{-5}. Among the genotypes distant from the founder, only two were isolated from the environment, one on November 1st and one on November 8th, postdating the first observation of the SLV genotype closer to the founder, which was observed in a clinical isolate.

Most of the variation occurs in two loci (the fourth and fifth of the genotype) on the small chromosome. If we analyze only the more stable loci, each isolate is described by a three-locus genotype and only seven genotypes are observed. Five genotypes can be related to another genotype by a mutation at a single locus (**Figure 3.1(B)**), while two genotypes are unrelated. The 9-4-14 genotype occurs in 86% (64/74) of the clinical cases sampled.

Variation among isolates from a single stool was also observed (**Table 3.3**). Only one of the six stools from which multiple isolates were tested had a single genotype for all isolates, two stools yielded two genotypes, two yielded three genotypes and one had four genotypes. In five of six stools, all genotypes of the isolates were related variants with variations seen only in loci of the small chromosome. In one stool, one isolate was

unrelated to any of the other three genotypes observed in the other nine isolates, evidence of multiple isolates involved in producing disease. Additional evidence for multiple isolates being involved in disease is the observation that five of the six genetic lineages in Chhatak were observed in clinical samples.

Table 3.3. Genotypes of 59 isolates collected from six stool samples from cholera cases in Chhatak (10 isolates/stool)

Patient ID	Date	No. of isolates	Genotypes	Number of genotypes	Relation to Founder
1	11th Oct, 2010	10	9-4-14-***25***-17	10	SLV
2	11th Oct, 2010	9	**9-4-14-21-17**	5	Founder
			9-4-14-***25***-17	4	SLV
3	11th Oct, 2010	10	**9-4-14-21-17**	7	Founder
			9-4-14-21-***18***	2	SLV
			9-4-14-***22***-17	1	SLV
4	13th Oct, 2010	10	***7***-4-14-***23***-17	5	DLV
			9-4-14-**23**-17	3	SLV
			9-4-14-***23-16***	1	DLV
			7-9-14-14-6	1	unrelated
5	13th Oct, 2010	10	9-4-14-***18-16***	9	DLV
			9-4-14-***13-16***	1	DLV
6	13th Oct, 2010	10	**9-4-14-21-17**	6	Founder
			9-4-14-21-***18***	2	SLV
			9-4-14-***22-18***	2	DLV

Variant alleles from founder genotype are in bold and italicized

The genetic diversity observed in these environmental and clinical isolates allows only crude inference about whether there is more than one mode of transmission. If all genotypes found in the clinical isolates were randomly sampled from the five genetic lineages of the five-locus genotypes, or from one of the seven three-locus genotypes, this would provide evidence for a strong role for slow transmission. However, we observed one lineage predominated (p-value < 10^{-5}, χ^2-test, five-locus set and p-value <10^{-5}, χ^2-

test, three-locus set) consistent with the presence of a non-random mode of transmission among human cases in Chhatak.

3.4.2. Mathbaria

Significant genetic variation was also observed among the 84 isolates collected in Mathbaria (**Table 3.4**, Simpson Index = 0.26). The number of distinct alleles at loci VC0147, VC0437, VC1650, VCA0171 and VCA0283 was 6, 5, 1, 8, 5, respectively. A total of 24 genotypes, of which 14 were present in 56 clinical isolates and twelve in 28 environmental isolates, with two present in both clinical and environmental isolates. As shown in **Figure 3.2(A)**, five clonal complexes and five singletons were determined, almost twice the number of genetic lineages detected in Chhatak. Two clonal complexes comprised only two genotypes. Among the unrelated (singleton) genotypes, four were found only in environmental isolates and one in a clinical isolate.

In Mathbaria, *V. cholerae* O1 isolates were obtained during the last three months of 2010 and during an outbreak in March through May, 2011. In 2010, the clonal complex comprised three genotypes (**Figure 3.2(A)**). All three were in environmental isolates prior to identification in an isolate from a clinical sample in December. The founder genotype was one of the two genotypes isolated from the environment on October 13th. An isolate with an unrelated genotype was also collected from the environment at that time.

The annual (seasonal) outbreak of cholera in Mathbaria occurred in March through May, 2011 with three distinct clonal complexes and four unrelated genotypes detected. The smallest clonal complex was detected in two clinical isolates in March. In the largest clonal complex (April-May complex I in **Figure 3.2(A)**), the founder genotype was observed in isolates from the environment on the 4th of April. On the 6th, the founder genotype and two others were identified in isolates from cases. The founder genotype was detected in 42 isolates (35 from clinical cases and seven from the environment) over the next 48 days.

Table 3.4. Genotypes of Mathbaria Clinical and Environmental Isolates

Genotypes	Source	No. of Isolates	Clonal complex
9-4-14-23-18 *	Clinic	3	1
10-4-14-23-18	Clinic	1	1
10-4-14-9-18	Environment	1	1
9-4-14-17-18	Clinic	1	1
9-4-14-22-18	Clinic	2	1
9-4-14-23-17	Clinic	1	1
10-4-14-9-17	Environment	1	1
11-9-14-15-18 *	Clin and Envi	35 & 7	2
11-9-14-15-17	Clinic	2	2
11-9-14-15-16	Clinic	1	2
11-9-14-15-19	Clinic	4	2
12-9-14-15-18	Clinic	1	2
11-7-14-14-15	Clin and Envi	1 & 5	3
10-7-14-14-15 *	Environment	1	3
10-7-14-14-16	Environment	1	3
9-4-14-14-16	Environment	3	4
9-4-14-11-16	Environment	1	4
11-8-14-14-19	Clinic	1	5
11-8-14-13-19	Clinic	1	5
9-5-14-14-17	Clinic	2	Singleton, Apr 19
6-5-14-17-18	Environment	1	Singleton, Apr 18
8-4-14-14-17	Environment	1	Singleton, Nov 3
9-9-14-19-16	Environment	5	Singleton, Apr 4
10-8-14-17-16	Environment	1	Singleton, May 9

* - identifies the founder genotype defined as the genotype that has the most single locus variants.

Two additional genotypes from this clonal complex were observed in May, 2011. The intermediate size clonal complex (April-May complex II in **Figure 3.2(A)**) was first observed in an isolate from a patient on the 26th of April and subsequently in two isolates from the environment on May 2nd. The founder genotype was observed in three isolate

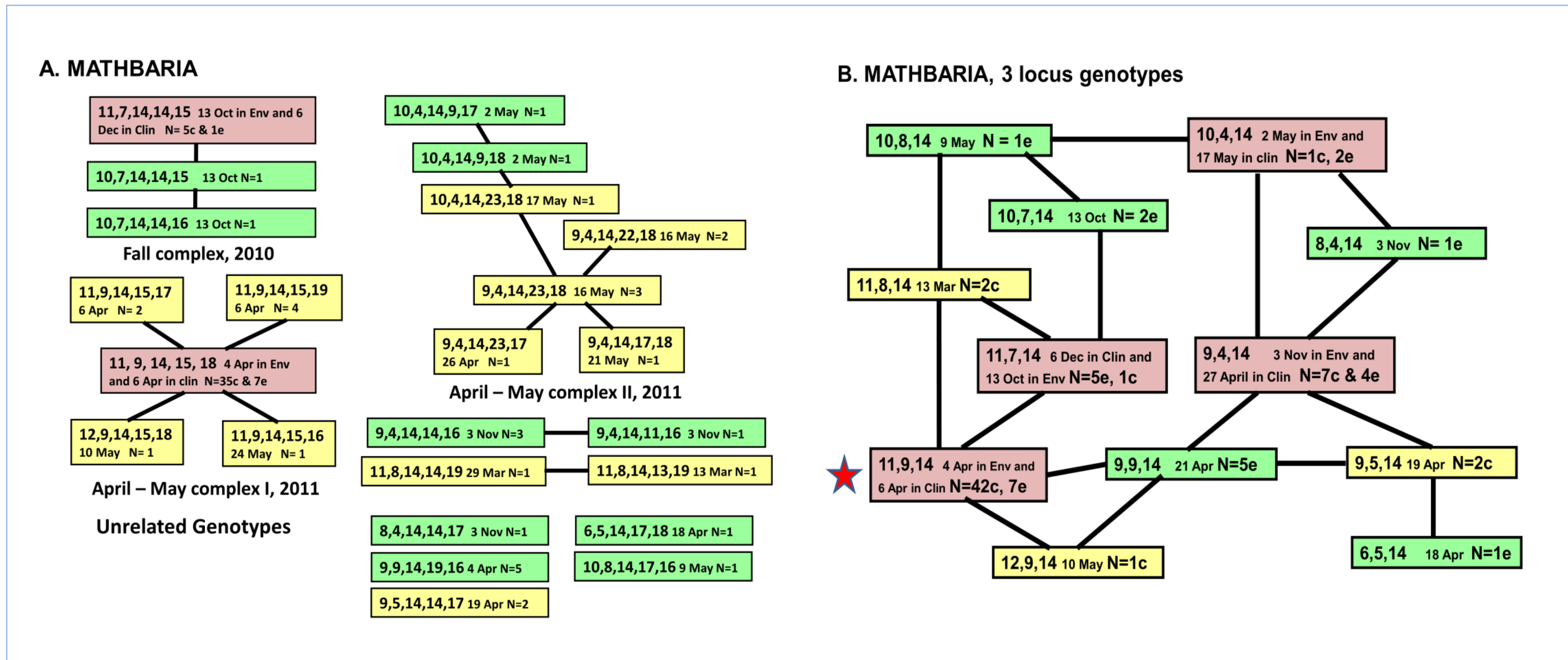

Figure 3.2. Diagrams of genetic relatedness among MLVA genotypes of Mathbaria. Each genotype is identified by the number of repeats in the allele at the five loci VC0147, VC0437, VC1650, VCA0171 and VCA0283. The earliest date of detection is recorded in the box after the 5th allele. The background of the box indicates whether the genotype was clinical only (yellow), environmental isolates (green) or both (pink). A. Clonal complex of Mathbaria genotypes and five unrelated genotypes based on a five-locus genotype. B. Analysis of three-locus genotypes of Mathbaria.

from cases on May 16th. Subsequently, three other SLV genotypes were observed in isolates from cases. A clear pattern of the founder being detected initially and “derived” genotypes later did not hold for this clonal complex.

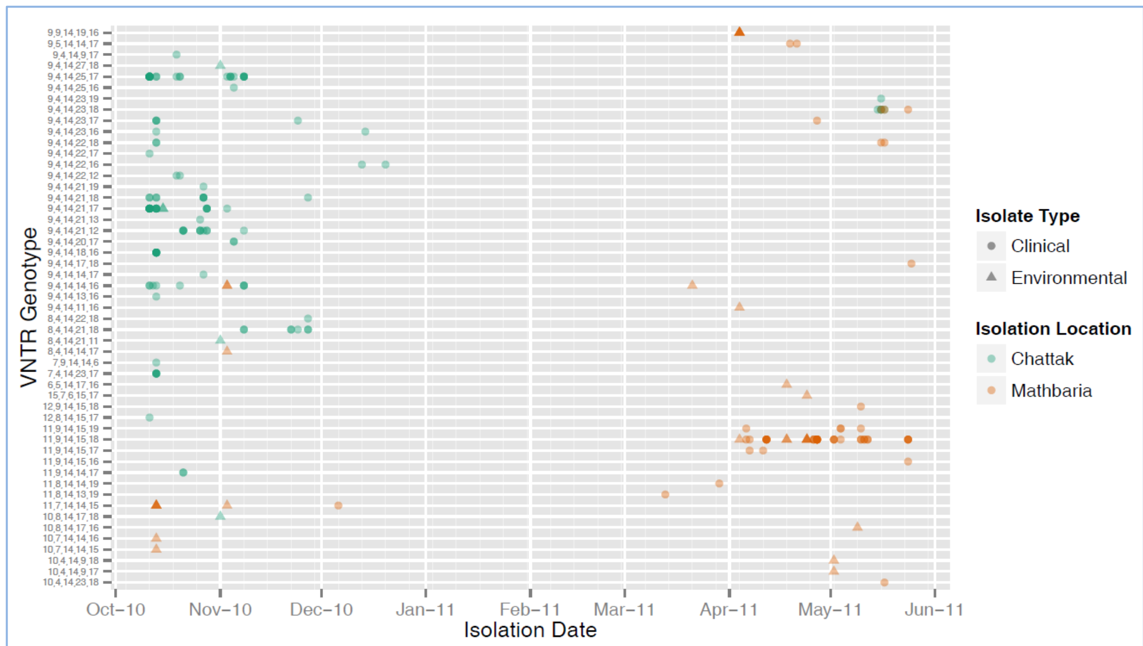

Figure 3.3. MLVA genotypes are plotted with the days of isolation for *V. cholerae* O1. Clinical isolates are represented by circles, while isolates from the ponds are represented by triangles. Isolates from Chhatak are blue, and those from Mathbaria are orange. Multiple genetically identical isolates on the same day increase the darkness of the shape.

When the genotypes were defined only by the three large chromosome loci, twelve genotypes were observed, each related to one or more of the other genotypes. As shown in **Figure 3.2(B)**, a network best describes the relatedness, since there is no a priori method to assign which mutation is more likely to have happened. Two genotypes vary by one mutation from four other genotypes and six genotypes vary by one from three other genotypes. In the seasonal outbreak, the 11-9-14 genotype was observed in 76% (42/55) of the isolates.

The genetic lineages observed in the clinical isolates was a non-random sample of those found in Mathbaria. Although five of the ten lineages were observed from clinical isolates, one lineage accounted for 77% (43/56) of the isolates. Assuming that any genetic lineage could cause a clinical case with equal probability, then the observed results are non-random (p-value $< 10^{-5}$, χ^2-test, five-locus set). For the three-locus analysis, twelve genotypes were observed and one occurred in 42 of 56 isolates derived from cases, a significantly non-random distribution (p-value $< 10^{-5}$, χ^2-test, three-locus set) and, as at Chhatak, an accelerated mode of transmission (i.e., human-to-human) is likely for certain genotypes.

3.5. DISCUSSION

The results of this study show multiple genetic lineages of *V. cholerae* are naturally occurring in the environment with both geographic and seasonal genetic variation observed. The genotype patterns of environmental and clinical isolates in the two rural Bangladeshi communities reveal: i) identical genotypes were detected in isolates from the environment and humans, and ii) the genotypes in human isolates were not a random sample of those in the environment. The result of a simple chi-squared test provides evidence that an accelerated, human-to-human mode is likely to have contributed to a process that generates genetic uniformity among clinical isolates. Whether the accelerated mode incorporates the hyperinfective state or involves massive numerical increases of a genetic lineage from the earliest cases cannot be distinguished from the data.

Rapid expansion of *V. cholerae* in Chhatak fits the pattern of a founder flush event (Wallace, 1981). The founder flush principle asserts that a rapid expansion in the size of a population may be accompanied by a relaxation of selection pressure so that genotypes otherwise not detected may be observed. This principle was applied to the presence of novel multilocus sequence genotypes in *V. cholerae* O139 (Garg *et al.*, 2003). In Chhatak, after no clinical cases were recovered in September and early October, on October 11th and 13th seven cases presented at the clinic and eleven genotypes were

observed among the isolates obtained from the rectal swabs. The expansion continued, with 50 cases yielding eleven more genotypes identified over the next 26 days. The spring outbreak in Mathbaria showed two clonal complexes exhibiting additional genetic differentiation, but not as dramatic as in Chhatak. However, the size of the outbreak was smaller in Mathbaria than in Chhatak.

Appearance of novel MLVA genotypes different from the founder over time, as observed in the Chhatak outbreak, is a microcosm of the evolution seen previously. In Kolkata among *V. cholerae* O139 isolates, the founder genotype appeared shortly after initial mutation from O1 to O139 and then mutated into multiple novel genotypes, mutations continued to occur over the course of several years (Ghosh *et al.*, 2008). Similarly, in Dhaka, O1 Ogawa, O1 Inaba and O139 isolates mutated and a clear progression of genotypes was documented from one year to the next (Kendall *et al.*, 2010). In the previous studies, differentiation occurred over several years of time, while in Chhatak, the differentiation occurred within three months.

Our analyses of three loci from the first chromosome demonstrate that the outbreak has a dominant genotype drawn from many observed in the area. These large (first) chromosome loci are considered to be more stable than the small (second) chromosome loci (Kendall *et al.*, 2010, Lam *et al.*, 2012). Although the second chromosome loci may be less useful in the context of evolution across decades (Lam *et al.*, 2012), in this context the increased resolution of local genetic linages in Chhatak reveals that during rapid expansion many genetic variants may occur, although data from Mathbaria indicate increased variation is not an obligate part of the expansion.

The presence of multiple genotypes of *V. cholerae* in a single stool was observed, as previously reported (Lam *et al.*, 2012). However, unlike the previous finding, where a minority (3 of 9) of stools contained isolates differing only in successive single allelic changes, here there was a majority (5 of 6). Thus, it is impossible to determine whether the different genotypes are part of the same inoculum or differentiated during infection.

In the earlier study, six of nine stools yielded isolates with unrelated genotypes, i.e. different clonal complexes. It should be noted that this study was carried out in a rural area. Also clinical samples were analyzed during a seasonal outbreak, as opposed to ongoing infections throughout the year. Finally the majority of isolates from cases from the surrounding community were of the same clonal complex, which was not the case in the previous study.

3.6. CONCLUSION

Molecular typing of *V. cholerae* using multilocus variable number tandem repeat analysis has proven useful for outbreak investigations. Apart from whole genome sequencing and analysis, every typing tool has limitation. Therefore, results of molecular typing need to be interpreted carefully when determining clonal relationships of bacterial isolates. Overall, our results provide evidence that could be used in support an accelerated mode of transmission and multiple strains comprising an infective dose causing cholera. As shown from analysis of clinical isolates, a single isolate does not sufficiently describe a single clinical sample, an important observation for future clinical studies.

CHAPTER 4:

GENOMIC INSIGHT INTO *VIBRIO CHOLERAE* CAUSING ENDEMIC CHOLERA IN BANGLADESH

4.1. Toxigenic *Vibrio cholerae* O1 from Bangladesh possess West African-South American genomic island (WASA)-1 and WASA variant of VSP-II

4.1.1. ABSTRACT

Phylogenomic analysis of two *Vibrio cholerae* O1 El Tor strains carrying *ctxB*$^{\text{El Tor}}$ (genotype *ctxB3*) isolated from Bangladesh in 2011 demonstrated close genetic relatedness to 1990s Latin American epidemic strains. The two isolates were found to be genetically distant from *V. cholerae* O1 atypical El Tor strains responsible for cholera in Asia, Africa, and Haiti. Both strains were susceptible to penicillin (P), ampicillin (AMP), streptomycin (S), chloramphenicol (C), sulfamethoxazole-trimethoprim (SXT), tetracycline (TE), kanamycin (K), erythromycin (E), nalidixic acid (NA), and ciprofloxacin (CIP) and devoid of quinolone resistance-associated mutations in *gyrA* and *parC* genes. *V. cholerae* O1, EC-0051 and EM-1727, possesseed virulence genes *ctxAB*, *tcpA*, *hylA*, *ace*, and *zot* and showed identical MLVA genotype 10-6-6-17-11. Both El Tor strains lacked SXT-ICE and contained West African-South American genomic island (WASA)-1 and WASA variant of VSP-II which had been detected previously in 1990s Latin American epidemic strains and *V. cholerae* O1 strains isolated in Angola, Africa. Overall, identification of the WASA1 and WASA variants of VSP-II in *V. cholerae* O1 isolated in Bangladesh and their genetic relatedness to Latin American strains isolated in the 1990s are novel findings that are of epidemiological importance in understanding the genetic diversity of *V. cholerae* O1 in cholera endemic countries.

4.1.2. INTRODUCTION

V. cholerae O1 El Tor strains possessing the El Tor biotype specific cholera toxin (genotype *ctxB3*) were shown to be associated with a substantial number of cholera cases during pre-2001 in different parts of the world, including the massive Latin American epidemic that occurred in the 1990s (Wachsmuth *et al.*, 1993, Mutreja *et al.*, 2011). Prior to 1991, except for a few sporadic cases, cholera outbreaks or epidemics had not been reported in countries of Latin America for more than hundred years (Sepulveda *et al.*, 2006). The *V. cholerae* O1 El Tor strains responsible for the 1990s Latin American

epidemic was found to be genetically similar to African outbreak strains (Lam *et al.*, 2010) and thus believed to have been transmitted from Africa to Latin America in the 1980s or earlier (Lam *et al.*, 2010, Mutreja *et al.*, 2011). However, results of recent genomic analyses nullify the hypothesis of a single source transmission of cholera in Mexico by demonstration of the association of indigenous *V. cholerae* O1 with cholera outbreaks during the period pre and post-1990 (Choi *et al.*, 2016). Previously, single nucleotide polymorphism (SNPs) analyses of 154 *V. cholerae* O1 genomes identified a West African–South American (WASA) clade which comprised 1990s Latin American epidemic strains from Peru, Mexico, Columbia, Argentina, and Bolivia. Notably, they differed from those isolated in the rest of the world by the presence of a unique genomic island, WASA1, and WASA variant of VSP-II in the genome (Mutreja *et al.*, 2011). WASA1 and WASA variant of VSP-II were also found in *V. cholerae* O1 strains isolated in Angola, Africa (Mutreja *et al.*, 2011, Valia *et al.*, 2013). Therefore, it was hypothesized that WASA1 and WASA variant of VSP-II could serve as regional marker for *V. cholerae* O1 strains isolated in West Africa and South America (Morais *et al.*, 2013, Dutilh *et al.*, 2014).

During the course of the ongoing seventh pandemic, emergence of new variants and selective sweeps due to natural section have been observed in *V. cholerae* (Stine & Morris, 2014). Subsequent dominance of a different *V. cholerae* sub-population has been reported over the past few decades. For example, *V. cholerae* O1 El Tor displaced Classical in 1961, *V. cholerae* O139 transiently displaced El Tor in 1992, and, lastly, the atypical El Tor displaced prototype El Tor (Mukhopadhyay *et al.*, 2014). Although *V. cholerae* O1 causing cholera outbreaks is a nonrandom sample of the total population existing in a natural reservoir, multiple clones can thrive in the ecosystem of endemic settings (Rashed *et al.*, 2014). Isolation of *V. cholerae* O139 from surface water samples collected in Bangladesh during 2011 and 2012, followed by *V. cholerae* O139 associated cholera cases during 2013 and 2014 in Dhaka, Bangladesh, provide strong evidence that pre-existing clones of *V. cholerae* can thrive in the natural aquatic reservoir for a long time without entry to human hosts, attesting to the aquatic ecosystem as its natural habitat (Rashed *et al.*, 2013, Chowdhury *et al.*, 2015). Although, *V. cholerae* O1 prototype El

Tor has rarely been isolated over the past ten years or more, a recent study reported isolation of four El Tor strains possessing *ctxB3,* in Bangladesh, during 2011 and 2012 (Rashed *et al.*, 2013). In this study, two *V. cholerae* O1 El Tor (*ctxB3*) strains isolated in 2011 from surface water samples collected in Bangladesh were subjected to whole genome sequencing and their genome sequences were compared with a set of globally collected *V. cholerae* genomes to determine phylogenetic relatedness with currently circulating and archived *V. cholerae* strains.

4.1.3. MATERIALS AND METHODS

4.1.3.1. Bacterial Strains

V. cholerae O1 strains were isolated from rectal swabs and surface water samples collected in Bangladesh between June, 2010, and December, 2014, during an epidemiological surveillance conducted by the international centre for diarrheal disease research, Bangladesh (icddr,b) (Rashed *et al.*, 2013). All samples were collected according to protocols approved by institutional review boards at the Johns Hopkins University, University of Maryland (College Park, MD, USA), and icddr,b. Informed consent was obtained from the patients, and parents or legal guardians of the children who participated in this study. Isolation and identification of *V. cholerae* were performed using standard culture methods followed by serological and molecular tests (Alam *et al.*, 2006, Alam *et al.*, 2006, Huq *et al.*, 2012). Four *V. cholerae* O1 El Tor carrying *ctxB3* were identified among isolates from Bangladesh (Rashed *et al.*, 2013), of which the genomes of two environmental strains, EC-0051 and EM-1727, were sequenced.

4.1.3.2. Genome Sequencing, Assembly, and Phylogenetic Analysis

Genomic DNAs were prepared using Qiagen Kit according to the manufacturer's recommendation. Concentration of genomic DNA was measured and 50 ng of DNA (2.5 ng/µl) was used for genomic DNA library construction, employing the Nextera DNA library kit. Genomic DNA libraries were subjected to next-generation whole-genome Illumina and hybrid Illumina/454 sequencing. Hybrid and Illumina sequences were assembled using Celera and Velvet assemblers, as described previously (Chun *et al.*, 2009, Zerbino, 2010). Genome sequences of both *V. cholerae* O1 strains have been

deposited in NCBI GenBank under accession numbers APFS00000000 and APFZ00000000. All other genome sequences of *V. cholerae* included in this study were retrieved from the NCBI GenBank for comparative analysis. Genome-to-genome comparisons, identification of mobile genetic elements (MGEs) and core genome phylogeny were accomplished, as described previously (Chun *et al.*, 2009, Choi *et al.*, 2016). Single nucleotide polymorphism (SNP) based phylogeny was performed using default settings in CSI Phylogeny1.4 (http://cge.cbs.dtu.dk/services/CSIPhylogeny/), as described previously (Kaas *et al.*, 2014, Rashid *et al.*, 2016). A maximum likelihood tree was constructed using high quality SNPs and was visualized using MEGA 7 (Tamura *et al.*, 2011).

4.1.3.3. Multilocus Variable Number Tandem Repeat Analysis

Multilocus variable number tandem repeat analysis (MLVA) was performed targeting five loci, VC0147 [cell division protein (*ftsY*)], VC0436-7 [non-coding (intergenic)], VC1650 (collagenase), VCA0171 (hypothetical protein), and VCA0283 (hypothetical protein) using primers and conditions, as described elsewhere (Choi *et al.*, 2010). In brief, each locus was amplified by PCR and Sanger sequencing was also performed to determine the number of repeat sequences. The repeat numbers were assigned against each locus sequentially to designate the MLVA genotype (Choi *et al.*, 2010, Rashed *et al.*, 2014).

4.1.3.4. Antibiotic susceptibility

Antibiotic susceptibility of both *V. cholerae* O1 strains was tested using commercially available discs (BD BBL Sensi-Disc) for penicillin (P, 10 μg), chloramphenicol (C, 30 μg), kanamycin (K, 30 μg), streptomycin (S, 10 μg), and nalidixic acid (NA, 30 μg), following standard CLSI guidelines (CLSI, 2010, CLSI, 2010).

4.1.4. RESULTS AND DISCUSSION

The phylogeny of two *V. cholerae* O1 El Tor, namely EC-0051, and EM-1727, was determined by constructing a genome-relatedness neighbor-joining tree, using homologous alignment of 1,657 orthologs of protein-coding genes (~1,664,756 bp bp) of

55 *V. cholerae* genomes (**Figure 4.1**). Phylogenetic analysis placed both strains into a monophyletic clade together with 1990s Latin American epidemic strains from Peru, Brazil, and Mexico. Within this monophyletic clade, variable branch lengths suggest none of these strains were identical, except EC-0051 and EM-1727. Notably, they differed from contemporary *V. cholerae* O1 atypical El Tor strains from Asia, Africa, and Haiti (**Figure 4.1**). In the phylogenetic tree, EC-0051 and EM-1727 appeared to be basal to reference prototype *V. cholerae* O1 El Tor N16961, which was isolated in 1975 in Bangladesh. As shown in **Figure 4.1**, EC-0051 and EM-1727 are genetically distant from concurrent *V. cholerae* O1 atypical El Tor strains isolated in 2010 in Bangladesh. A maximum likelihood tree of *V. cholerae* O1 El Tor strains was constructed, using 705 high quality SNPs, which complemented these interpretations, revealing close genetic relatedness among EC-0051, EM-1727, and 1990s Latin American epidemic strains (**Figure 4.2**). EC-0051 and EM-1727 possess virulence genes, *ctxAB*, *tcpA*, *hylA, ace*, and *zot*, as described previously (Rashed et al., 2013). MLVA genotyping of both strains revealed identical MLVA genotype 10-6-6-17-11. Phylogenetic analyses and MLVA clearly indicate that EC-0051 and EM-1727 are clonally related *V. cholerae* O1 strains.

As shown in **Table 4.1**, the *rstA* gene of EC-0051 and EM-1727 is identical to that of N16961, whereas polymorphisms were detected in the *rstB* gene. A trinucleotide $(GTA)_{77\text{-}79}$ deletion was detected in the *rstB* gene of both stains and is absent in *V. cholerae* O1 El Tor N16961. Deletion of this trinucleotide $(GTA)_{77\text{-}79}$ was also detected in *V. cholerae* O1 strains isolated in Haiti, Mexico, Peru, and Brazil (**Table 4.1**). In addition, EC-0051 and EM-1727 showed three polymorphisms ($A_{90}\rightarrow T$, $T_{96}\rightarrow C$, and $G_{108}\rightarrow A$) in the *rstB* gene, commonly found in *V. cholerae* O1 ET strains isolated in Latin America (**Table 4.1**) (Choi *et al.*, 2016). Both EC-0051 and EM-1727 contained four copies of ToxR binding heptamer repeats (TTTTGAT) in the promoter region of the cholera toxin encoding gene (*ctxAB*), located between *zot* and *ctxA*. Likewise, the 1990s Latin American epidemic strains contain four copies of ToxR binding repeats (TTTTGAT) (**Table 4.1**). ToxR regulates transcription of several of the virulence genes of *V. cholerae* (Ottemann & Mekalanos, 1995). The toxR binding repeat sequences is

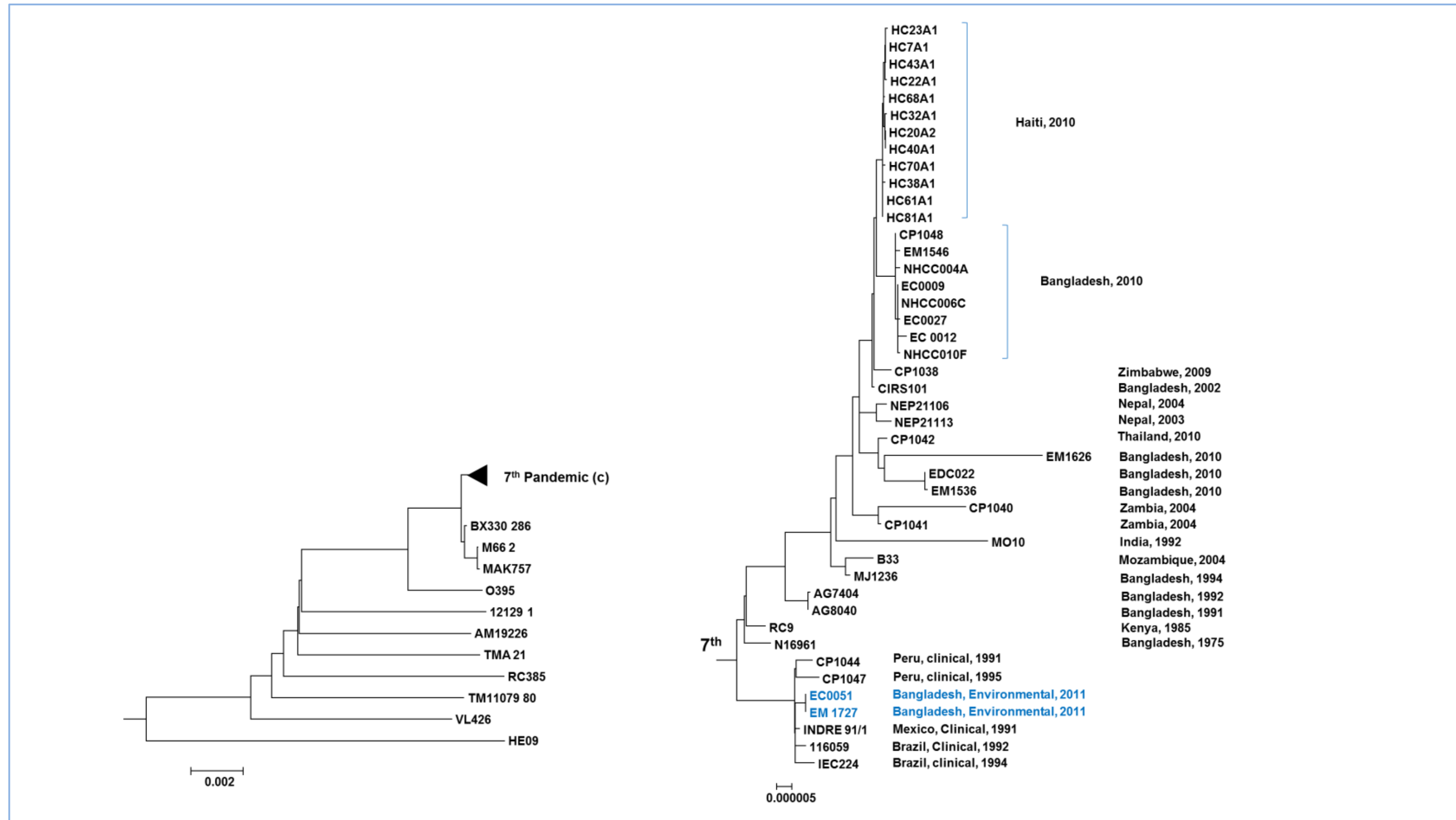

Figure 4.1. Neighbor-joining trees showing phylogenetic relationships of 54 *V. cholerae* genomes based on 1,798 orthologs of protein-coding genes (~1,794811 bp). The Haitian *V. cholerae* non-O1/O139 strain HE09, isolated from surface water samples during 2010, was used as an outgroup of the tree. *V. cholerae* O1 strains EC-0051 and EM-1727 are shown in blue, indicating placement in a monophyletic clade with Latin American epidemic strains isolated in Mexico, Peru, and Brazil during the 1990s.

Table 4.1. Nucleotide variation of CTX-prophage mediated genes in *V. cholerae* O1 El Tor strains

Strain	Country of Origin	Year of Isolation	CTXϕ	*rstR* type	*rstA* polymophic positions				*rstB* polymophic positions						*zot-ctxA*$^{\Psi}$	*ctxB* genotypes
					315	927	933	942	77-79	90	96	108	192	363	(TTTTGAT)	
N16961	Bangladesh	1975	CTX^{ET}	ET	T	T	C	G	GTA	A	T	G	A	A	4	*B3*
CIRS101	Bangladesh	2002	CTX^{HYB}	ET	*	*	*	*	*	*	*	*	*	*	3	*B1*
HCO1	Haiti	2010	CTX^{HYB}	ET	*	*	*	*	–	*	*	*	*	*	5	*B1*
CP 1032	Mexico	1991	CTX^{HYB}	ET & CL	C	C	T	T	–	T	C	A	*	*	4	*B1*
CP 1033	Mexico	2000	CTX^{HYB}	ET & CL	*	C	T	T	–	T	C	A	G	G	4	*B1*
CP1044	Peru	1991	CTX^{ET}	ET	*	*	*	*	–	T	C	A	G	*	4	*B3*
CP1047	Peru	1995	CTX^{ET}	ET	*	*	*	*	–	T	C	A	G	*	4	*B3*
116059	Brazil	1992	CTX^{ET}	ET	*	*	*	*	–	T	C	A	*	*	4	*B3*
ICE224	Brazil	1994	CTX^{ET}	ET	*	C	*	*	–	T	C	A	*	*	4	*B3*
EM1727	Bangladesh	2011	CTX^{ET}	ET	*	*	*	*	–	T	C	A	G	*	4	*B3*
EC0051	Bangladesh	2011	CTX^{ET}	ET	*	*	*	*	–	T	C	A	*	*	4	*B3*

ET, El Tor; HYB, hybrid; CL, classical; *, indicates sequence identical to that of *V. cholerae* N16961; –, GTA deletion; *zot-ctxA*$^{\Psi}$, numbers of copies of the heptamer (TTTTGAT) repeat sequence

essential for ToxR binding and activation of *ctxAB* promoter (Pfau & Taylor, 1996). The number of toxR binding repeat sequences is variable in *V. cholerae* O1 Classical strains

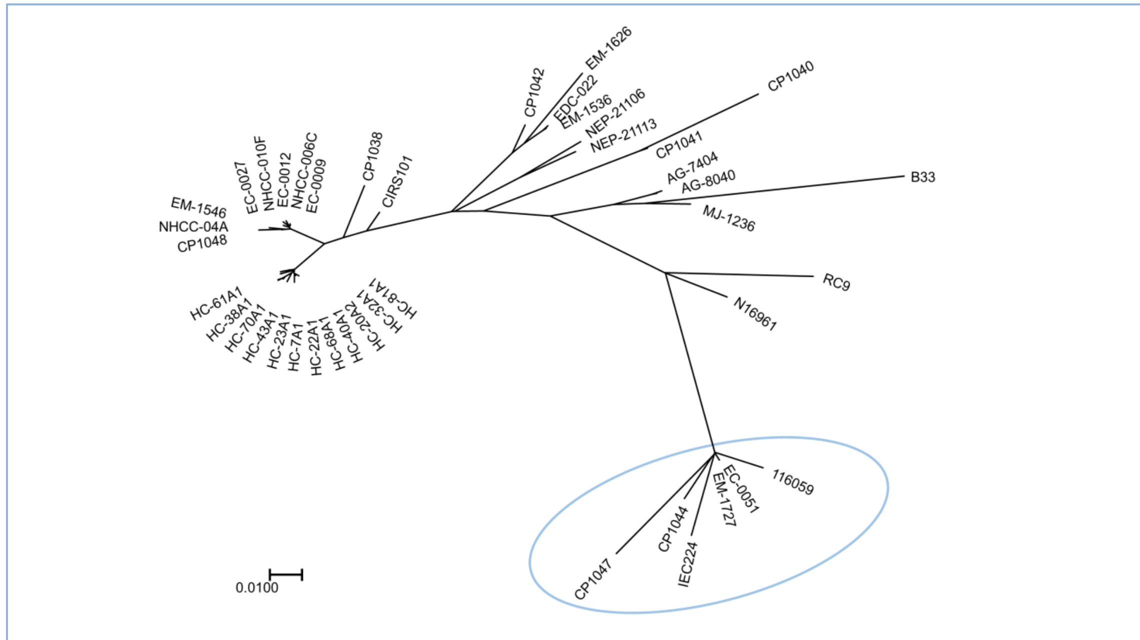

Figure 4.2. SNP-based maximum likelihood tree of 42 *V. cholerae* O1 El Tor isolated in different countries of Asia, Africa, and the Americas. The blue circle includes two *V. cholerae* El Tor strains (EC-0051 and EM-1727) carrying *ctxB3* isolated in Bangladesh and the 1990s Latin American epidemic strains, suggesting genetically relatedness.

(6 – 7 copies) and also in Haitian *V. cholerae* O1 El Tor strains (5 copies) isolated during the devastating cholera epidemic in 2010 (Hasan *et al.*, 2012, Choi *et al.*, 2016).

EC-0051 and EM-1727 possess *Vibrio* Pathogenicity Island (VPI)-1 and 2, VSP-I, and WASA variant of VSP-II which contains a putative recombination between VC0510 and VC0516 (Mutreja *et al.*, 2011). Furthermore, both strains lack SXT/R391 ICE and carry genomic islands: GI-1 to 10, GI-85, and WASA1 (**Figure 4.3**). As shown in **Figure 4.3**, WASA1 contains 44 coding sequences (CDSs) (Mutreja *et al.*, 2011), encoding hypothetical proteins, phage related proteins, and several enzymes, i.e., DNA polymerase, recombinase, exonuclease, helicase, and deoxyguanosine kinase. Notably, SXT/R391 ICE was also absent in the genomes of Latin American strains that clustered together with EC-0051 and EM-1727 (**Figure 4.1**). Lack of SXT/R391 ICE in toxigenic

V. cholerae O1 strains isolated in Latin America has also been reported previously (Mutreja *et al.*, 2011, Choi *et al.*, 2016).

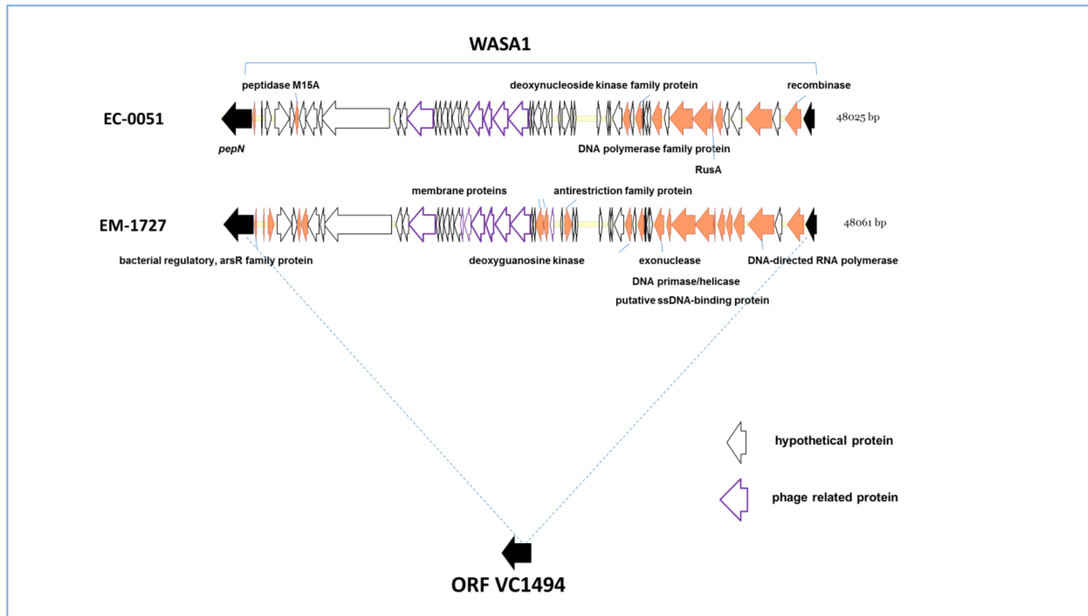

Figure 4.3. West African-South American genomic island WASA1 in *V. cholerae* O1 strains EC-0051 and EM-1727 isolated in Bangladesh in 2011. WASA1 is inserted in ORF VC1494 and absent in *V. cholerae* O1 El Tor reference strain N16961.

Antibiotic susceptibility tests revealed both strains to be sensitive to penicillin (P), streptomycin (S), chloramphenicol (C), kanamycin (K), and nalidixic acid (NA). As described previously (Rashed *et al.*, 2013), both strains were also susceptible to ampicillin (AMP), trimethoprim-sulfamethoxazole (SXT), tetracycline (TE), erythromycin (E), and ciprofloxacin (CIP). Conversely, drug resistance due to SXT/R391 ICE has been observed frequently in currently circulating *V. cholerae* O1 atypical El Tor strains from Bangladesh, India, Thailand, Cameroon, Nigeria, Zimbabwe, and Haiti (Mutreja *et al.*, 2011, Reimer *et al.*, 2011, Hasan *et al.*, 2012, Rashed *et al.*, 2012). The gyrase (*gyrA*) and topoisomerase (*parC*) genes of EC-0051 and EM-1727 were devoid of quinolone resistance–associated mutations, while concurrent atypical El Tor strains possess point mutations in the *gyrA* (Ser83Ile) and *parC* (Ser85Leu) genes (Hasan *et al.*, 2012). Accumulations of mutations in *gyrA* and *parC* genes confer resistance to

fluoroquinolone drugs, including ciprofloxacin, the commonly used drug for treatment of cholera in endemic countries. *V. cholerae* O1 stains possessing mutations in *gyrA* (Ser83Ile) and *parC* (Ser85Leu) genes show resistance to NA and reduced susceptibility to ciprofloxacin (Quilici *et al.*, 2010, Sjolund-Karlsson *et al.*, 2011).

The presence of WASA1 genomic island and WASA variant of VSP-II in two environmental *V. cholerae* O1 El Tor stains isolated in Bangladesh is an interesting observation because these two genetic elements were proposed as potential markers for *V. cholerae* O1 El Tor isolated in West Africa and South America (Mutreja *et al.*, 2011, Morais *et al.*, 2013, Dutilh *et al.*, 2014). Our findings clearly indicate that WASA1 and WASA variant of VSP-II are not confined to West Africa and South America isolates of *V. cholerae* O1 El Tor. Therefore, the concept of WASA as a regional marker must be carefully re-interpreted (Morais *et al.*, 2013, Dutilh *et al.*, 2014). Here we have shown distribution of WASA1 and WASA variant of VSP-II in *V. cholerae* O1 El Tor isolated in Bangladesh.

4.1.5. CONCLUSION

Our findings indicate a close phylogenetic relatedness of two environmental *V. cholerae* O1 El Tor carrying *ctxB3*, both of which had been isolated from Bangladesh, and 1990s Latin American epidemic strains. Notably, *V. cholerae* O1 strains isolated in the 1970s and 1980s in Bangladesh were found to be genetically more similar to seventh pandemic prototype El Tor (N16961) than to 1990s Latin American epidemic strains (Dutilh *et al.*, 2014). A previous study by Mutreja *et al*. proposed global dissemination of seventh pandemic cholera in three overlapping waves, including trans-continental transmission from South America to Southeast Asia during 1986-87 (Mutreja *et al.*, 2011). In our study, despite a lack of sufficient metadata to support transcontinental transmission, the likelihood of introduction of *V. cholerae* O1 El Tor with WASA1 and WASA variant of VSP-II from Latin America by human activity is plausible. Alternatively, subsequent genetic shift or drift in the genome of ancient *V. cholerae* O1 El Tor would allow this *V. cholerae* O1 to evolve in local niches of the Ganges delta in the Bay of Bengal, mimicking evolution of *V. cholerae* O139 in this region in late 1992 (Chun *et al.*, 2009).

The epidemiological significance of the presence of this *V. cholerae* O1 El Tor in cholera endemic areas of Bangladesh has yet to be clarified because *V. cholerae* O1 atypical El Tor has been the predominant agent of cholera in Bangladesh in recent years. Considering the emergence of new variants and diverse genotypes, this genetically divergent *V. cholerae* O1 appearing in the era of *V. cholerae* O1 atypical El Tor very likely contributes to the changing epidemiology of global cholera.

4.2. Phylogenomic Aanalysis of *V. cholerae* isolated from patients and environmental samples in the coastal area of Bangladesh.

4.2.1. ABSTRACT

In this study, a total of 165 strains of *V. cholerae* isolated in Mathbaria, Bangladesh, during 2004 and 2014, were subjected to whole genome sequencing and their genome sequences were compared with 59 globally collected *V. cholerae* strains. Of 165 isolates, 158 were *V. cholerae* O1, and the remaining 7 were non-O1/O139. Among the O1 isolates, 146 contained *ctxB1*, six contained *ctxB7*, one contained *ctxB3* and five lacked cholera toxin. All non-O1/O139 isolates lacked the cholera toxin gene (*ctxAB*) and toxin coregulated pilus (*tcpA*). Except for the one toxigenic isolate (Vc57CM-0088), all possessed RS1, TLC, *ace*, and *zot*, along with *ctxAB*. Furthermore, multiple genomic islands (GIs), mobile genetic elements (MGEs) and a truncated SXT were identified in all *V. cholerae* O1 isolates. SNP based phylogenomic analysis revealed that non-O1/O139 clustered with basal strains, while *V. cholerae* O1 isolates clustered with isolates of the seventh pandemic clade. Of 158 *V. cholerae* O1 strains, isolated between 2004 and 2014, 156 belonged to the Wave 3 of seventh pandemic clade; while Wave 1 and Wave 2 contained one isolate each from Mathbaria, Bangladesh. All of the Wave 3 isolates were atypical El Tor including Haitian variants. Within Wave 3, Bangladesh isolates clustered into three sub-clades. A total of 129 atypical El Tor isolates from Mathbaria, Bangladesh belonged to the major sub-clade which was designated as Bangladesh sub-clade. Clinical and environmental isolates within the Bangladesh clade showed close clonal relatedness. The Bangladesh sub-clade appeared to be a clonal expansion that occurred between 2004 and 2014. Moreover, one isolate in Wave 2 was clonally linked to Matlab variant, and another isolate in Wave 1 was genetically similar to reference El Tor N16961. Phylogenomic analyses identified multiple clones causing endemic cholera in the coastal region of Bangladesh.

4.2.2. INTRODUCTION

Cholera is an ancient disease that has been endemic in the Ganges delta region of the Bay of Bengal for more than centuries. The Bay of Bengal, which serves as a natural reservoir

of *V. cholerae*, is also referred to be as the home of Asiatic cholera (Sack *et al.*, 2004). In Bangladesh, the disease annually affects 100,000 people, causing approximately 4,500 deaths each year (Ali *et al.*, 2015). The incidence of cholera shows a bimodal seasonal distribution in Bangladesh: the first peak (March–May) occurs before the monsoon and the second (September–November) occurs at the end of the monsoon (Akanda *et al.*, 2009, Alam *et al.*, 2011). In coastal areas of Bangladesh, a larger outbreak of cholera occurs during the spring season (March – May). Epidemiological studies have shown strong correlation between incidence of cholera and several environmental factors, i.e., rainfall, river level, water temperature, conductivity, and algal and plankton blooms (Epstein, 1993, Huq *et al.*, 2005, Jutla *et al.*, 2011). Therefore, global warming and climate change pose a huge threat to the people living in coastal areas because it could cause frequent cholera flare-ups in neighboring regions of the Bay of Bengal. Although, modest health care facilities are available in rural coastal areas, lack of sanitation, inadequate supply of safe drinking water, and poor personal-hygienic practices, the coastal communities are considered higher risk groups for cholera. Once a cholera outbreak occurs in a coastal community, large amount of toxigenic *V. cholerae* O1 is released in the environment since domestic waste remains untreated (Nelson *et al.*, 2009). Toxigenic *V. cholerae* O1 can, therefore, transmit via contaminated food and/or water to cause subsequent cholera cases.

Transmission of the cholera bacterium from the environment to humans and vice-versa plays a substantial role in clonal selection and population dynamics of *V. cholerae* (Stine & Morris, 2014). The genome of *V. cholerae* is highly prone to genetic change due to genomic shift and genomic drift (Chun *et al.*, 2009). These two phenomena comprise all kind of horizontal and vertical gene transfer events occurring in the genome. In the aquatic reservoir, *V. cholerae* acquires mobile genetic elements, plasmids, and genomic islands, adapting to diverse ecological niches. Over the past few decades, several new epidemic lineages of *V. cholerae* O1 El Tor have emerged and reemerged in the Ganges delta region (Safa *et al.*, 2010, Mukhopadhyay *et al.*, 2014). In the process of evolution and natural selection, the classical biotype was replaced by 7th pandemic strains of El Tor that possess *ctxB3* and this occurred in the 1960s (Safa *et al.*, 2010). Despite the

disappearance of classical strains from South Asia and Africa during the 1990s, classical cholera toxin (*ctxB1*) appeared in *V. cholerae* O1 El Tor between 1988 and 1993 (Nair *et al.*, 2002, Raychoudhuri *et al.*, 2009, Na-Ubol *et al.*, 2011). The atypical El Tor strains had a significant impact on the disease dynamics of cholera since the severity of the disease increased several-fold compared to typical El Tor stains (Siddique *et al.*, 2010).

While El Tor biotype continued as the major cause of cholera worldwide, in 1992-1993 a non-O1 serogroup, designated O139 synonym Bengal, emerged as the cause of epidemic cholera in the coastal areas of the Bay of Bengal (Albert *et al.*, 1993, Albert & Nair, 2005). Notably, classical cholera toxin (*ctxB1*) was also detected in the O139 serogroup strains, replacing *ctxB3*. The serogroup O139 strains co-existed with *V. cholerae* O1 El Tor for a decade and gradually faded out during 2005 and thereafter. Retrospective genetic analysis of the O139 serogroups isolates revealed the presence of SXT genetic elements encoding resistance to several antibiotics, i.e., sulfamethoxazole, trimethoprim, chloramphenicol, and streptomycin (Hochhut *et al.*, 2001). *V. cholerae* O1 atypical El Tor have become predominant in cholera endemic countries of Asia, Africa and Latin America over the past twenty years (Safa *et al.*, 2008, Na-Ubol *et al.*, 2011, Alam *et al.*, 2014). Recently, cholera outbreaks are not confined in endemic foci, as massive cholera epidemics have been observed in Haiti and Yemen (Mutreja *et al.*, 2011, Hasan *et al.*, 2012, Camacho *et al.*, 2018, Weill *et al.*, 2019). Genetic characterization and phylogenomic analysis of outbreak strains have proven critical for determining clonal relatedness of strains isolated in different geographical regions and over time. Genome sequencing and comparative analyses have been useful for evaluating phylogenetic relations by identifying single nucleotide polymorphisms (SNPs), which potentially elucidate the phylodynamics of *V. cholerae* associated with ongoing pandemic (Mutreja *et al.*, 2011, Reimer *et al.*, 2011). In addition, genome sequencing data can be used to characterize pathogenic strains and track emergence and spread of *V. cholerae* in both endemic and non-endemic areas. In this study, a total of 165 *Vibrio cholerae* strains (**Table 4.2**), isolated from both clinical and environmental samples collected in Mathbaria, Bangladesh, were sequenced and compared with genomes of a set of globally collected strains to understand the global epidemiology of cholera.

4.2.3. MATERIALS AND METHODS

4.2.3.1. Bacterial Strains

V. cholerae O1 strains were isolated from rectal swabs and surface water samples collected in Mathbaria, Bangladesh between 2004 and 2014, during two phases of an epidemiological surveillance conducted by the international centre for diarrheal disease research, Bangladesh (icddr,b) (Alam *et al.*, 2006, Rashed *et al.*, 2013, Rashed *et al.*, 2017). All samples were collected according to protocols approved by institutional review boards at the Johns Hopkins University, University of Maryland (College Park, MD, USA), and icddr,b. Bacterial isolates were cultured on selective media followed by isolation and identification of *V. cholerae* using standard culture methods, and both serological and molecular tests (Alam *et al.*, 2006, Alam *et al.*, 2006, Huq *et al.*, 2012).

Table 4.2. Source, serogroups and number of *V. cholerae* isolates used for genome sequencing and comparative genomic analysis.

Source	Serotype	Number of Isolates
Environmental	*V. cholerae* non-O1/O139	4
Environmental	*V. cholerae* O1	60
Clinical	*V. cholerae* non-O1/O139	3
Clinical	*V. cholerae* O1	98
Total		165

4.2.3.2. DNA Extraction, Library preparation, and Illumina Sequencing

Genomic DNA was extracted from overnight pure cultures using Qiagen Kit according to the manufacturer’s guideline. Quantification of genomic DNA was performed using AccuBlue dye and SpectraMax® Gemini™ EM Spectrofluorometer. Once quantified, 50 ng of Genomic DNA was used to construct DNA libraries using Nextera Library Prep Kit (Illumina). All DNA libraries were quantified and the Agilent 2100 Bioanalyzer was used to check the quality prior sequencing. Finally, DNA libraries were sequenced using a 100 bp paired-end run on an Illumina HiSeq1000 Instrument (Illumina, San Diego, CA).

4.2.3.3. Genome Assembly and Phylogenetic Analysis

Trimmomatic was used to remove Illumina adapter sequences from paired reads and to trim bases off the start or end of the read when the quality score fell below a threshold of

15 (Bolger *et al.*, 2014). All reads less than 30 bp in length were discarded. Sequence quality metrics were assessed using FastQC. To limit the presence of human contamination in the downstream analysis, all reads were mapped to the human reference genome hs38 (contaminant database) and 50 published *V. cholerae* genomes (keep database) using DeconSeq (Schmieder & Edwards, 2011). Reads were paired using Pairfq. For each sample, sequence reads were mapped against reference genome, *V. cholerae* O1 El Tor N16961 (Genbank assession numbers NC_002505.1 and NC_002506.1), using the default filtering thresholds of Snippy. Read libraries free of human contamination were assembled with SPAdes software (v.3.12.0) (Bankevich *et al.*, 2012), using the options '—careful', to reduce the number of misassembles, and '—cov-cutoff auto', to remove potentially mis-assembled low coverage contigs. Contigs were ordered with Progressive Mauve using the genome sequence of N16961 as reference and annotated using Prokka (Darling *et al.*, 2010, Seemann, 2014). Virulence factors (VFs) and Mobile Genetic Elements (MGEs) were determined using TBLASTX thresholds of identity (>30%), alignment (>80%) and e-value (<10-36).

Genome-to-genome comparison was performed by using different approaches because completeness and quality of the nucleotide sequences varied from strain to strain. First, ORFs of a given pair of genomes were identified and reciprocally compared with each other using the BLASTN, BLASTP, and TBLASTX programs (ORF-dependent comparison). Second, a bioinformatic pipeline was constructed to identify homologous regions of a given query ORF. Initially, a segment on the target contig, homologous to a query ORF, was identified using the BLASTN program. This potentially homologous region was expanded in both directions by 2,000 bp. Nucleotide sequences of the query ORF and selected target homologous regions were aligned using a pairwise global alignment algorithm and the resulting matched region in the subject contig was extracted and saved as a homologue (ORF independent comparison). Orthologues and paralogues were differentiated by reciprocal comparison. In most cases, ORF dependent and ORF-independent comparisons yielded the same orthologues, although the ORF-independent method performed better for draft sequences of low quality, in which sequencing errors, albeit rare, hampered identification of correct ORFs. High quality SNPs were identified

on the core genome using ParSNP to construct the phylogenetic tree (Treangen *et al.*, 2014).

4.2.3.4. Identification and Annotation of Genomic Islands

The genomic islands (GIs) are defined as a continuous array of five or more coding sequences (CDSs) discontinuously distributed in genomes of test strains. Correct transfer or insertion of GIs was readily differentiated from a deletion event by comparing the genome based phylogenetic tree and full matrices showing pairwise detection of orthologous genes between test strains. Identified GIs were designated and annotated using the BLASTP search of its member CDSs against the GenBank NR database, as described elsewhere (Chun *et al.*, 2009, Hasan *et al.*, 2012).

4.2.4. RESULTS AND DISCUSSION

4.2.4.1. Virulence factors and *ctxB* genotypes

Toxigenic *V. cholerae* possess several known virulence factors, of which the cholera toxin (CT) and toxin coregulated pilus (TCP) are considered the most significant due to their direct correlation with the disease syndrome. In addition, other virulence factors, i.e., accessory cholera enterotoxin (*ace*), zonula occludens toxin (*zot*), repeat in toxin (RTX), and hemolysin (*hlyA*) are frequently found in the cholera bacterium. Of 158 *V. cholerae* O1 isolates, genomic analysis revealed the presence of CT and TCP in 153, while the remaining 5 isolates lacked these two major virulence factors. All non-O1/O139 isolates uniformly lacked CT and TCP. Except one toxigenic isolate (Vc57CM-0088), all contained RS1, TLC, *ace*, and *zot*, along with CT. All isolates possessed RTX and *hylA* in their genome. As shown in **Table 4.3**, 146 isolates contained *ctxB1*, 6 isolates contained *ctxB7*, and *ctxB1* was detected in only one isolate. Among the CT genotypes, *ctxB1* was predominant (92%) in both clinical and environmental isolates from Mathbaria, Bangladesh, during the study period. Our results provide ample evidence of the prevalence of atypical El Tor in the coastal area of Bangladesh. Although *ctxB7* or Haitian type CT were associated with the cholera epidemics in Haiti, Yemen, Nepal, India, and Bangladesh during 2006 and 2017 (Mutreja *et al.*, 2011, Hasan *et al.*, 2012,

Naha *et al.*, 2012, Rashed *et al.*, 2012, Shakya *et al.*, 2012), the abundance of this CT genotype appears to be lower in recent years.

Table 4.3. CT genotypes of *V. cholerae* O1 isolated in Bangladesh.

***ctxB* Type**	**2004**	**2005**	**2006**	**2007**	**2010**	**2011**	**2012**	**2013**	**2014**	**Total**
ctxB1(Classical)	4	11	28	14	2	15	11	54	7	146
ctxB3 (El Tor)	0	0	0	0	0	0	1	0	0	1
ctxB7 (Haitian)	0	0	0	0	1	4	1	0	0	6
ctxB (-)	0	0	1	0	0	1	0	1	2	5
Isolates/year	4	11	29	14	3	20	13	55	9	158

Genetic analysis of seven *V. cholerae* non-O1/O139 revealed the presence of *hlyA* and RTX. In addition, the cholix toxin gene (*chxA*) was detected in two strains, Vc33 and Vc109, and the type three secretion system (T3SS) was present in one strain, VC111. Despite lacking cholera toxin, presence of minor virulence factors, e.g., NAG-ST, hemolysin, cholix toxin, and T3SS, in *V. cholerae* non-O1/O139 isolates has been reported from different countries of Asia, Europe, Australia and Americas (Chatterjee *et al.*, 2009, Octavia *et al.*, 2013, Ceccarelli *et al.*, 2015, Schwartz *et al.*, 2019). Notably, these virulence factors occasionally can cause gastrointestinal disturbance and diarrhea depending on the immune status of the human host (Dalsgaard *et al.*, 1999, Dutta *et al.*, 2013).

4.2.4.2. Phylogenetic analysis of *V. cholerae*

A total of 224 *V. cholerae* genomes were used to construct a SNP based phylogenetic tree (**Figure 4.4**). The phylo-core genome of these globally collected strains appeared to be ~2.8 Mbp which was used to identify high quality SNPs. Genome sequences of Haitian *V. cholerae* non-O1/O139 isolates, HE09 and HE16, were chosen as an outgroup of the phylogenetic tree. As shown in **Figure 4.4(A)**, seven non-O1/O139 isolates from Mathbaria, Bangladesh, clustered with the other basal isolates. This basal group comprised of non-O1/O139 isolates from Bangladesh, India, Australia, Mexico, Peru, and Haiti. Although, GIs and their orthologs occasionally identified in basal *V. cholerae* non-

O1/O139 isolates, genomes of the basal group appeared to be highly diverse compared to those of the sixth and seventh pandemic isolates **Figure 4.4(A)**.

All *V. cholerae* O1 El Tor isolates from Mathbaria, Bangladesh, were found to belong to the seventh pandemic clade, with a set of globally collected strains from Haiti, Zimbabwe, Thailand, and Mozambique **Figure 4.4(B)**. Seventh pandemic clade was further differentiated into multiple small sub-clades, suggesting a minor divergence between *V. cholerae* El Tor and its variants. The El Tor "era" has been continuing globally for more than 50 years and these strains have undergone a series of genetic events during this long time-frame. Several variants of El Tor have emerged that possess altered phenotypic and genetic attributes (Chun *et al.*, 2009, Safa *et al.*, 2010). As shown in **Figure 4.4(B)**, small sub-clades represent the typical seventh pandemic El Tor, Matlab variant (MJ-1236), Mozambique variant (B-33), altered ET (CIRS-101), atypical El Tor or Haitian variant. Despite having all of the recognized virulence factors, only few toxigenic clones disseminated successfully all around the world, e.g., Haitian Variant (Mutreja *et al.*, 2011). Based on extensive genomic analyses, a recent study showed the transcontinental transmission model of seventh pandemic cholera from the Bay of Bengal that involves at least three overlapping waves, with a common ancestor in the 1950s (Mutreja *et al.*, 2011). Here, we adopted this transmission model and the genomic analysis scheme to evaluate differentiation of environmental and clinical isolates from Mathbaria, Bangladesh, within the three waves of the seventh Pandemic clade.

As shown in **Figure 4.4(B)**, a majority (97%) of the *V. cholerae* O1 isolates from Mathbaria, Bangladesh, that had been isolated between 2004 and 2014, belonged to Wave 3 of the seventh pandemic clade. All of the Wave 3 isolates are atypical El Tor, including Haiti variant. Within Wave 3, atypical El Tor isolates of Bangladesh formed three sub-clusters and showed a few polytomies. The major sub-cluster is designated to be Bangladesh sub-clade comprising 129 atypical El Tor isolates from patients and environmental samples collected during 2004-2014.

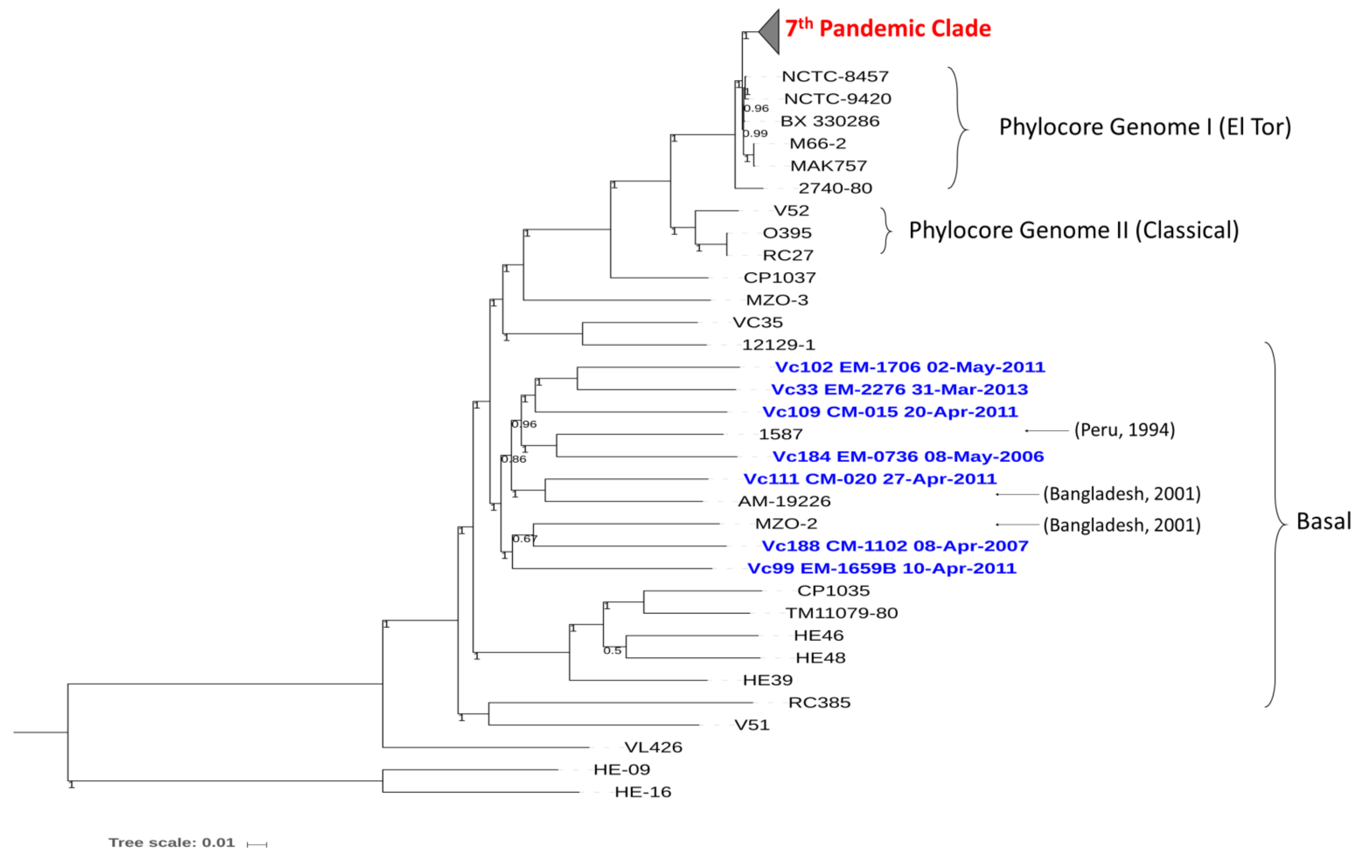

Figure 4.4. A. SNP based pylogenetic tree showing pre-seventh pandemic clades. All of the *V. cholerae* non-O1/O139 isolates from Bangladesh clustered with basal isolates of other countries.

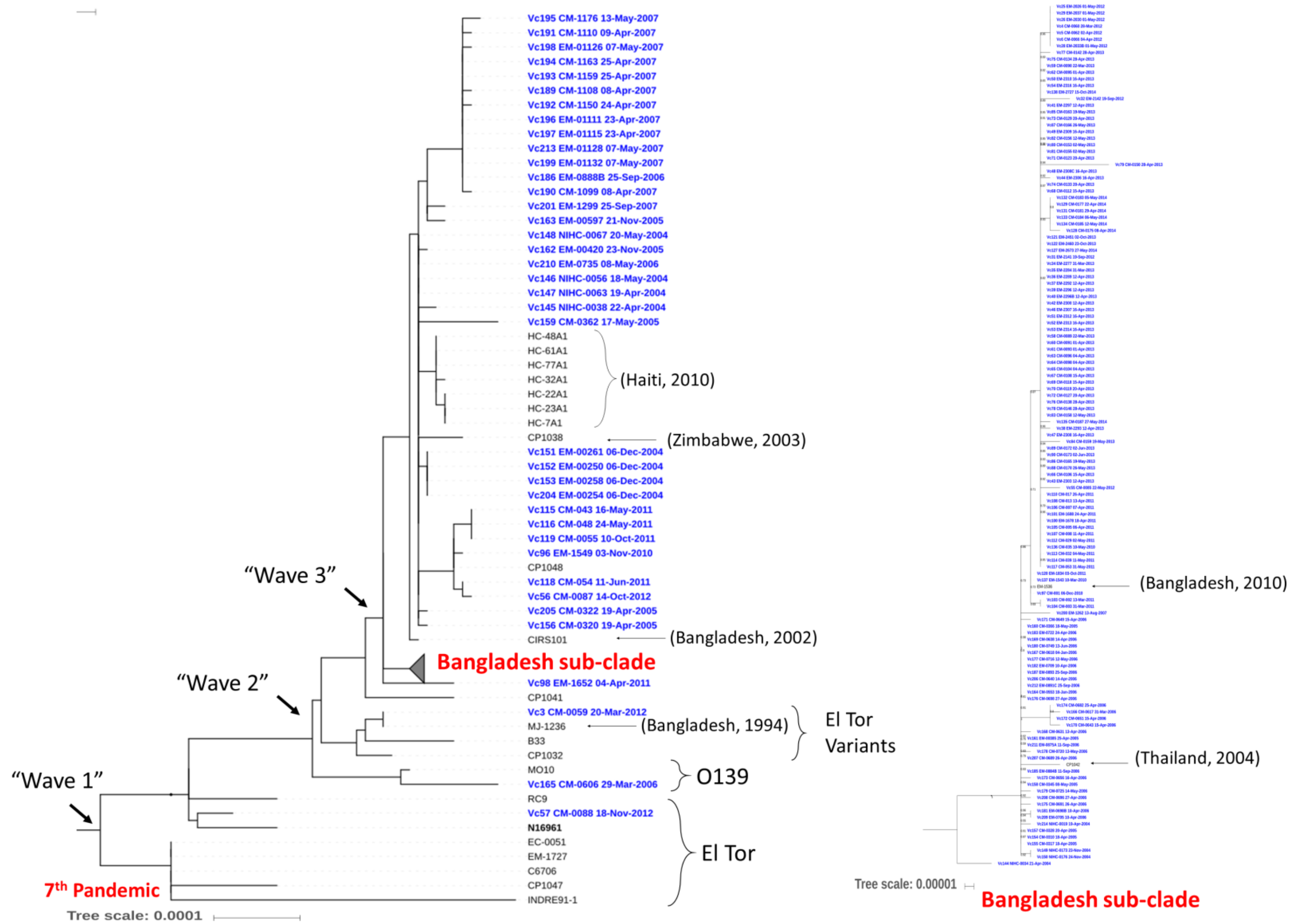

Figure 4.4. B. SNP based pylogenetic tree showing seventh pandemic waves. All *V. cholerae* O1 isolates from Bangladesh (2004 – 2014) clustered in the seventh pandemic waves with a set of globally collected strains from Asia, Africa, and Americas.

Topology of the Bangladesh sub-clade suggests an expansion of a single epidemic clone over time (2004 – 2014). Interestingly, subtle differentiation of the Bangladesh clade was observed based on isolation years for atypical El Tor isolates. Considering this time-dependent differentiation, we classified the isolation years into two groups: 2004-2007 and 2010-2014. Of 129 isolates, 92 atypical El Tor belonged to group 2010-2014, and group 2004 – 2007 contained the remaining 37 atypical El Tor. All are genetically closely related. In the Bangladesh sub-clade, the atypical El Tor of group 2004-2007 clustered at the base and group 2010-2014 expanded from the base. The results clearly demonstrate that Bangladesh sub-clade is a clonal expansion of a single clone that occurred in patients and the environment during 2004 and 2014. Within the Bangladesh sub-clade, a high degree of genetic similarity was observed between clinical and environmental isolates suggesting circulation of strains between patients and the environment. Notably, this observation was not limited to a single year or a set of successive years.

Another sub-cluster in Wave 3 comprised of 13 atypical El Tor isolated between 2004 and 2007. This sub-cluster is a near neighbor of Haitian variants responsible for the devastating cholera epidemic in Haiti in 2010, followed by an outbreak along the coast of Mexico (Mutreja *et al.*, 2011, Diaz-Quinonez, 2017). Several retrospective studies reported that the Haitian Variant or similar atypical El Tor originally emerged in Kolkata, India in 2006 and caused epidemics in India and Nepal, prior to its transmission to Haiti (Mutreja *et al.*, 2011, Shakya *et al.*, 2012, Ghosh *et al.*, 2014). Furthermore, a small sub-cluster of 12 atypical El Tor and the CIRS101 from Bangladesh was identified in Wave 3. Notably, one isolate from Bangladesh, VcCM-0059, clustered with Wave 2 with Matlab variant (MJ-1236) and another isolate, Vc57CM-0088, which cluster within Wave 1 of the seventh cholera pandemic clade, along with reference *V. cholerae* ET, N16961.

In this study, we identified 306 high quality SNPs among the clinical isolates, while 292 SNPs were detected among the environmental isolates when compared with the reference N16961 genome. Our results show the presence of at least five clones of *V. cholerae* O1 in the cholera endemic coastal area of Bangladesh during 2004 and 2014. These clones are associated with seasonal outbreaks of cholera in Mathbaria, Bangladesh.

Despite the coexistence of multiple *V. cholerae* O1 clones in the environment, only a few get the selection advantage to cause cholera in successive years. Remarkably, transient disappearance and reemergence of *V. cholerae* O1 clones in Bangladesh was observed in this study. Clonal integrity of the *V. cholerae* O1 Matlab variant was maintained for approximately two decades in Bangladesh, even though this clone is outcompeted by concurrent atypical ET. Overall, our findings are consistent with those of previous studies showing genomic diversity in different endemic countries of Asia and Africa (Mutreja *et al.*, 2011, Reimer *et al.*, 2011, Weill *et al.*, 2017).

4.2.4.3. Genomic Islands and Mobile Genetic Elements

Integrative mobile genetic elements (MGEs) such as transposons and genomic islands (GIs), contribute significantly to genetic rearrangement in the *V. cholerae* genome. GIs carrying multiple genes are inserted in the genome during a recombination event that may confer a fitness advantage to the bacterium and allow it to acclimate in different niches. Although, GIs encode mostly hypothetical proteins, GIs can encode virulence determinants or colonization factors promoting pathogenicity are commonly known as pathogenicity islands (Dziejman *et al.*, 2002, Chun *et al.*, 2009). In our study, all *V. cholerae* O1 isolates carried Vibrio Pathogenicity Island VPI-1 and VPI-2, whereas, all non-O1/O139 isolates lacked VPI-1 and VPI-2 (**Figure 4.5(A)**). Vibrio Pathogenicity Islands (VPI-1 and VPI-2) are considered a marker for epidemic and pandemic serogroups of *V. cholerae*, i.e., O1 and O139 (Karaolis *et al.*, 1998). Unlike epidemic strains and O141 serogroup, the majority of non-O1/O139 strains do not carry VPI-1 and VPI-2. However, GIs homologous to VPI-1 and VPI-2 have been identified in non-O1/139 serogroups in different parts of the world (Chatterjee *et al.*, 2009, Octavia *et al.*, 2013).

Seventh pandemic El Tor strains are characterized by the presence of two gene clusters, namely Vibrio Seventh Pandemic (VSP) I and II. These two gene clusters have an unusually lower G+C content (40%) compared to the entire *V. cholerae* genome (47%) and are absent in Classical and pre-pandemic *V. cholerae* El Tor strains (Dziejman *et al.*, 2002). As shown in **Figure 4.5(A),** all *V. cholerae* O1 strains contained (VSP)-I and

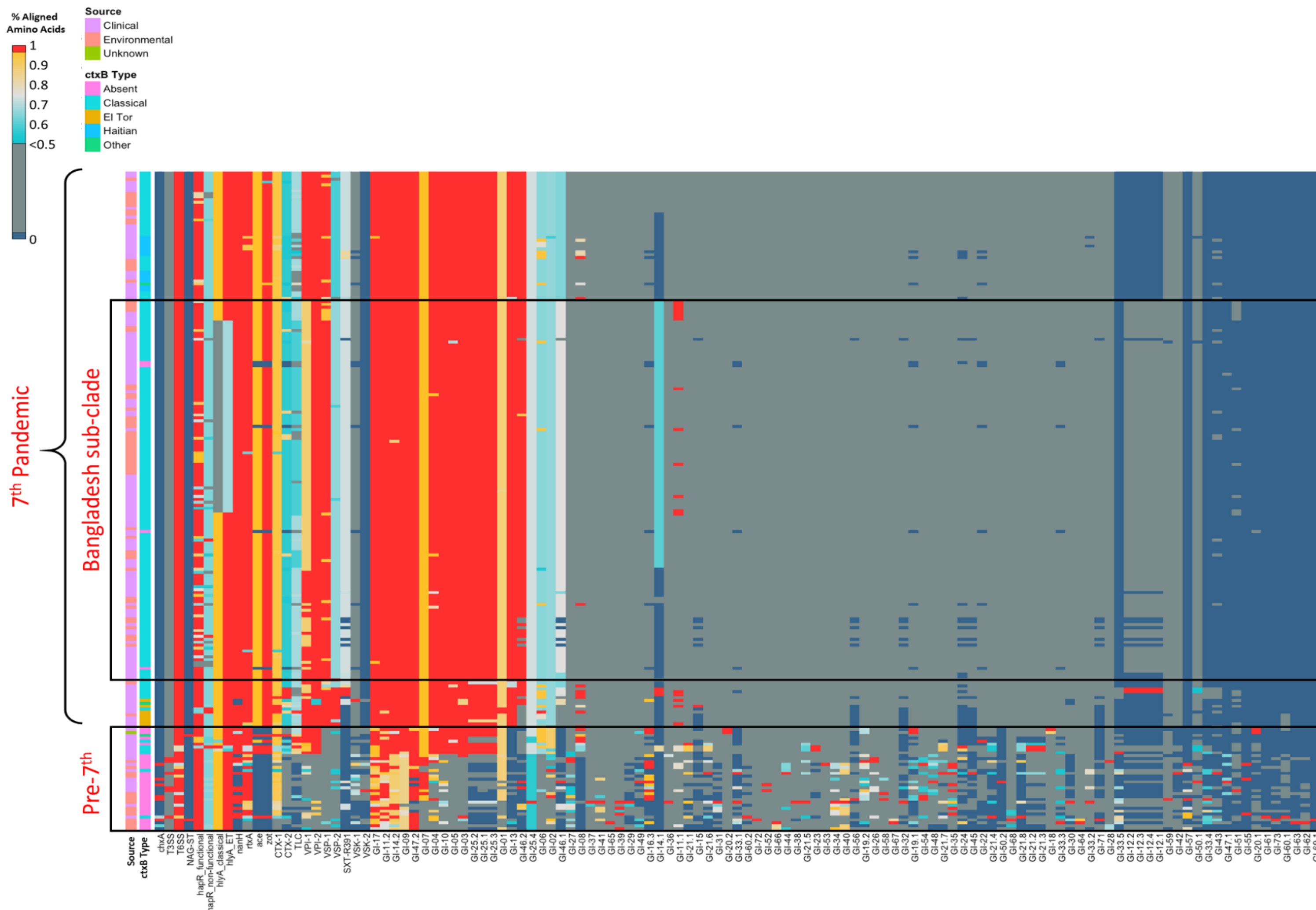

Figure 4.5. A. Distribution of virulence factors and genomic islands in pre-seventh and seventh pandemic *V. cholerae*. Color codes are shown at the top left where red indicates 100% match, i.e., identical.

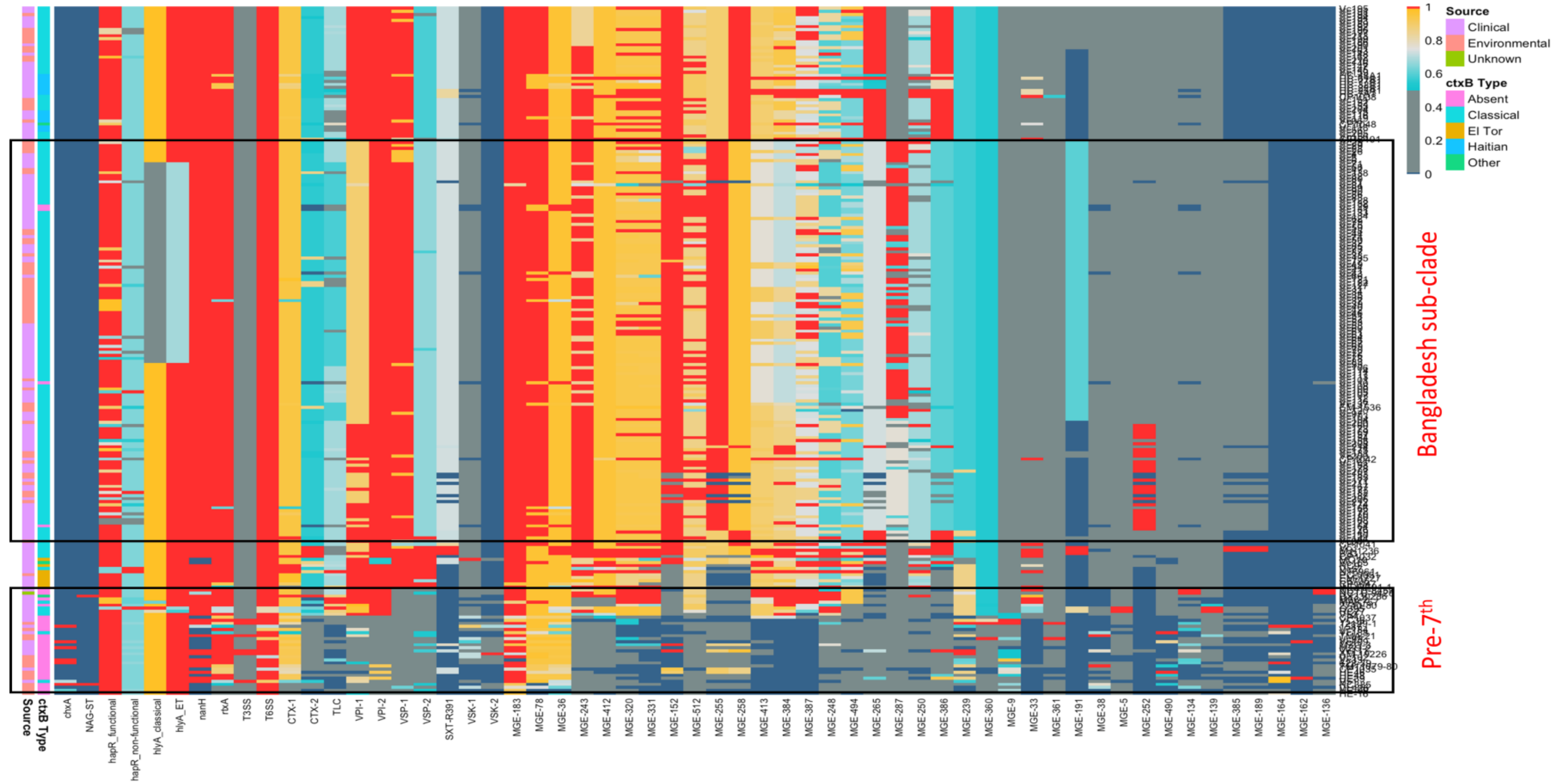

Figure 4.5. B. Distribution of Virulence factors and MGEs in pre-seventh and seventh pandemic *V. cholerae*. Color codes are shown at the top right where red means 100% match or identical.

II. However, a deletion was detected in VSP-II from ORF495 to ORF502 compared to prototype El Tor. The same deletion has been previously reported in *V. cholerae* CIRS101 and also observed in atypical El Tor and/or altered El Tor circulating in cholera endemic areas of Asia, Africa, and the Americas (Taviani *et al.*, 2010, Mutreja *et al.*, 2011). Furthermore, polymorphism in the ORF content in VSP-II was evident in El Tor isolates that caused the 1990s Latin American epidemic (Nusrin *et al.*, 2009). These genetic re-assortment, insertion or deletions of large gene clusters may provide ecological fitness and selection advantage to atypical El Tor for promoting itself as one of the dominant epidemic clones all over the world. However, it remains an interesting research topic that should be explored in future.

As shown in **Figure 4.5(A)**, the comparative genomics analyses identified several GIs in *V. cholerae* El Tor stains, e.g., GI-1, GI-2, GI-3, GI-4, GI-5, GI-6, GI-7, GI-9, GI-10, GI-11, GI-13, GI-14, GI-17, GI-25, GI-46, and GI-47. All of these GIs have been characterized previously in seventh pandemic clade associated El Tor strains isolated in Haiti, South Asia, and Africa (Chun *et al.*, 2009, Hasan *et al.*, 2012). Interestingly, these GIs or orthologs are uniformly distributed in prototype El Tor and atypical El Tor regardless of origin and source. Along with these GIs, all *V. cholerae* O1 isolates from Mathbaria, Bangladesh, possess a truncated SXT/R91. Moreover, a total of 37 MGEs have been identified among *V. cholerae* isolates from Bangladesh **Figure 4.5(B)**. Of these 37 MGEs, 14 are well-characterized (Chun *et al.*, 2009), while detailed characterization and the functional role of remaining 23 MGEs and/or their orthologs have yet to be determined. Our findings are consistent with reports of the genetic attributes of concurrent *V. cholerae* atypical El Tor causing pandemic cholera in both cholera endemic and non-endemic countries around the globe.

4.2.5. CONCLUSION

Toxigenic *V. cholerae* O1 has been evolving in cholera endemic regions of the world. The devastating cholera epidemic in Haiti alarmed the whole would to rethink about the ecology, epidemiology and spread of cholera. Any part of the world, where poor sanitation is coupled with an inadequate supply of safe drinking water will be at serious

risk of a cholera epidemic. Pathogenic clones from endemic regions can be introduced to an ecosystem at any time where it could cause massive outbreaks and establish itself as endemic pathogen. Furthermore, a natural calamity, such as an earthquake, cyclone, or flood, will have the potential to create an imbalance in a water treatment and a sanitation facility resulting in a spread of cholera. Therefore, tracking new epidemic strains and their virulence potential is crucial to better understanding of the epidemiology of global cholera. New epidemic strains with enhanced virulence are likely to evolve in endemic foci and will certainly disseminate beyond political boundaries. Genomic rearrangement and genetic changes, e.g., mutations and acquisitions of GIs and MGEs, provide a platform for natural selection to shape the cholera bacterium. Thus, it is important to create a robust genomic database for *V. cholerae* which will promote epidemiological tracking of the pathogen. Overall, the findings of this study have contributed to a better understanding of the emergence, evolution, and genomic diversity of *V. cholerae*.

CHAPTER 5:

GENETIC TRAITS AND GENOMIC DIVERSITY OF *VIBRIO CHOLERAE* O1 CAUSING CHOLERA IN MEXICO

5.1. Genetic traits, Molecular typing, and the CTX pro-phage array of *V. cholerae* O1 isolated in Mexico, 1998 - 2008

5.1.1. ABSTRACT

Vibrio cholerae classical biotype (CL) faded from Asia and Africa during the 1980s; however, it continued to co-exist with prototype El Tor (ET) and atypical ET in Mexico until 1997. Based on results of microbiological, molecular, and phylogenetic analyses of 91 *V. cholerae* O1 isolated in Mexico between 1998 and 2008, significant genetic events were confirmed, namely ET carrying *ctxB3* was predominant over CL and atypical ET in Mexico during 2004 and thereafter. Notably, *V. cholerae* atypical ET was the dominant type in Mexico between 1991 and 2000, and the majority of *V. cholerae* O1 isolated between 2001 and 2004 lacked the CTX prophage (Φ) mediated genes *ctxA*, *ctxB*, $rstR^{ET}$, and $rstR^{CL}$. Despite a large proportion of $CTXΦ^-$ *V. cholerae* isolates lacked *tcpA*, some carried the *tcpA* allele, either ET or CL type in the ET host. All $CTXΦ^-$ ET possessed $hylA^{ET}$ and RTX. DNA sequencing confirmed the $tcpA^{ET}$ allele to be identical to that of ET reference strain *V. cholerae* N16961 and $tcpA^{CL}$ allele is a variant showing sequence different from CL reference strain O395. The TCP variants were not detected in 2005 after $CTXΦ^+$ET became dominant. All *V. cholerae* O1 strains isolated from both clinical and environmental samples collected during 2005-2008 in Mexico were $CTXΦ^+$ ET carrying truncated CTX prophage instead of RS1 element. *V. cholerae* $CTXΦ^-$ ET strains exhibited heterogeneity whereas $CTXΦ^+$ ET strains isolated in Mexico during 2004 – 2008 displayed homogeneity in PFGE (*Not*I) patterns and clonal relationship with *V. cholerae* ET reference strain N16961 and *V. cholerae* ET isolated in Peru.

5.1.2. INTRODUCTION

Although cholera has been endemic in Asia for centuries and sporadic cases had been recorded in the Americas, the presence of *V. cholerae* CL biotype and its identification as the causative agent of cholera in the Americas have not been clarified. A massive epidemic of cholera occurred in South America during 1991 – 1992, first reported in Peru in January, 1991, after which cholera appeared in other countries of Latin America,

notably Mexico, by June, 1991. Characteristic features of the Latin American epidemic strains of *V. cholerae* O1 biotype ET distinguished them from 7th pandemic *V. cholerae* ET isolated in Asia (Olsvik, 1992). They appeared to be clonal, leading some investigators to conclude the 1991 Latin American cholera epidemic was simply an extension of the 7th pandemic, with *V. cholerae* ET transported by maritime traffic to and from the Western hemisphere (Wachsmuth *et al.*, 1993).

A significant recent development in cholera epidemiology is the emergence of toxigenic variants of *V. cholerae* ET carrying traits of the *V. cholerae* CL biotype isolated in Asia during 2001 and thereafter (Nair *et al.*, 2006, Morita *et al.*, 2010). Similarly, in Africa during the 1970s the atypical variant of *V. cholerae* ET became dominant (Morita *et al.*, 2010). Genetic changes in Latin American strains of *V. cholerae* were described and the Peruvian *V. cholerae* O1 isolated between 1991 and 2003 was shown to be ET 7th pandemic prototype, carrying a distinct signature in the VSP-II region that distinguished it from the Asian ET prototype (Nusrin *et al.*, 2009). A recent study of *V. cholerae* isolated between 1991 and 1997 from diarrhea patients and surface water sources in Mexico showed both *V. cholerae* CL and ET biotype strains were present with the atypical *V. cholerae* ET involved in epidemic cholera globally (Alam *et al.*, 2010). In this study, *V. cholerae* isolated between 1998 and 2008 from diarrheal patients and from samples collected from the aquatic environment in Mexico were characterized using microbiological, molecular, and phylogenetic techniques.

5.1.3. MATERIALS AND METHODS

5.1.3.1. Bacterial strains

Vibrio cholerae O1 strains included in the present study (n=91), listed in **Table 5.1**, together with source, location, and year of isolation, were provided by the Department of Public Health, Faculty of Medicine, National Autonomous University of Mexico (UNAM) and Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California, Mexico. *Vibrio cholerae* O1 was isolated from cholera patients (n=36) and surface water samples (n=55) as part of a nationwide cholera surveillance program

conducted between 1998 and 2008 (Beltran *et al.*, 1999, Lizarraga-Partida & Quilici, 2009), and transported in soft agar, transferred to bacteriological media employing standard culturing methods, and identified using a combination of biochemical and molecular procedures, as described previously (Alam *et al.*, 2006, Alam *et al.*, 2007, Huq *et al.*, 2012). The geographical distribution of *V. cholerae* in Mexico between 1998 and 2008 was extensive. *V. cholerae* was isolated from hospitalized diarrheal patients and from surface water sources (**Table 5.1**) collected from nearly every province of Mexico, including those bordering the US in the north and Guatemala in the south (**Figure 5.1**).

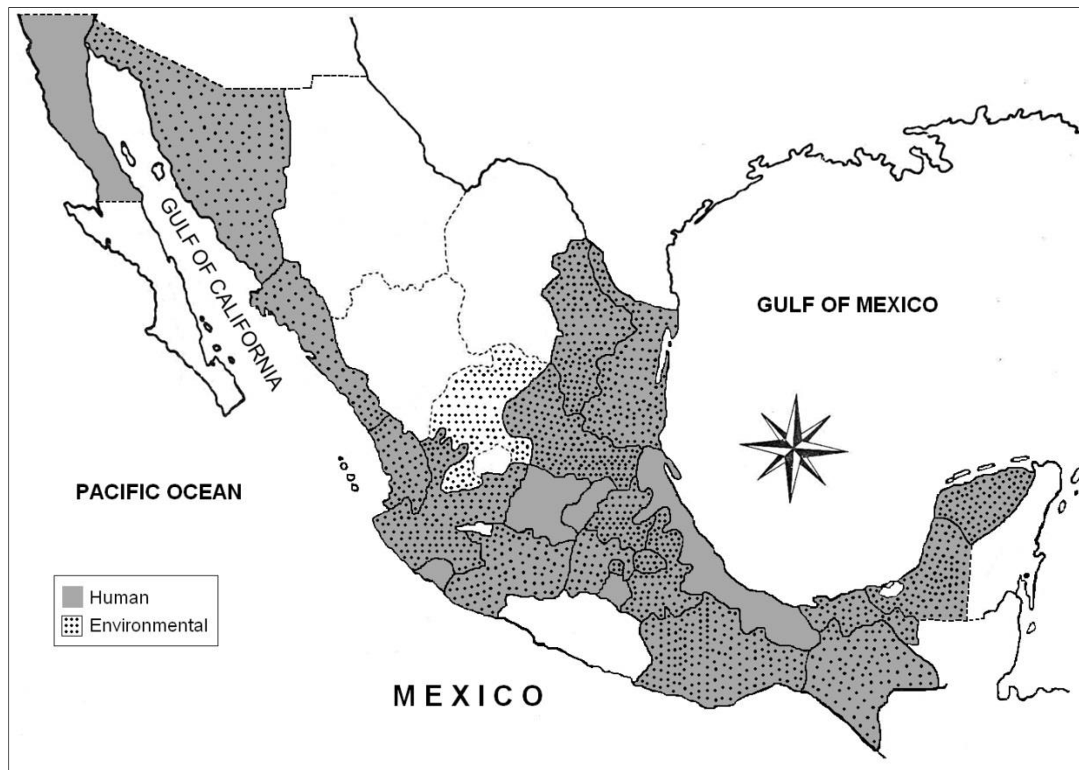

Figure 5.1. Geographic map of Mexico showing provinces and areas close to the Gulf of Mexico where the national surveillance took place and *V. cholerae* O1 strains collected between 1998 and 2008. Shading and dots indicate where *V. cholerae* O1 isolation from patients and environmental samples was done, respectively.

5.1.3.2. Serogrouping and Biotype determination

V. cholerae strains identified by standard biochemical and molecular methods were screened serologically by slide agglutination, using polyvalent antisera specific for *V.*

cholerae O1 or O139, followed by monoclonal antibody specific for each serogroup (Kaper *et al.*, 1995). Serogroups were reconfirmed using polyvalent O1 and monovalent Inaba and Ogawa antisera (Alam *et al.*, 2006, Alam *et al.*, 2010). For biotype determination, *V. cholerae* O1 strains were screened for chicken erythrocyte agglutination, sensitivity to polymyxin B (50 U), and Mukerjee classical (CL)-specific phage IV and ET (ET)-specific phage V tests (Kaper *et al.*, 1995). All *V. cholerae* strains identified using biochemical, serological, and molecular methods, as described above, were sub-cultured on gelatin agar (GA) and a single representative colony from each GA plate was inoculated into T1N1 broth (1% Trypticase and 1% NaCl), incubated at 37ºC for 4-6 hrs, and stored at -80°C with 15% glycerol for future use.

5.1.3.3. Extraction of genomic DNA

In brief, bacterial cells were harvested from 3 ml overnight culture in LB broth. Cells were then subjected to alkaline lysis followed by proteinase K and RNase treatment. Therefore, genomic DNAs were extracted using phenol/chloroform extraction method, as described previously (Nusrin *et al.*, 2009). Excess salt was removed by a 70% alcohol wash and the nucleic acid was air-dried and resuspended in sterile TE buffer. The concentration and purity of the DNA were assayed using a spectrophotometer (GeneQuant) and the DNA was stored at −20 °C for subsequent experiments. (Alam *et al.*, 2010).

5.1.3.4. Complementation of serogrouping and biotyping results by PCR assay

V. cholerae O1 strains identified primarily by phenotypic characteristics employing methods cited above were reconfirmed by PCR targeting the species-specific *ompW* gene (Nandi *et al.*, 2000) and multiplex PCR of serogroup-specific O biosynthetic *wbe*O1 (serogroup O1), *wbf*O139 (serogroup O139), and cholera toxin *ctxA* (Hoshino *et al.*, 1998). Furthermore, PCR based genetic screening was performed using primers for biotype-specific marker genes, *tcpA* (CL or ET), *hlyA* (CL or ET), and *rstR* (CL or ET) (Kimsey *et al.*, 1998, Rivera *et al.*, 2001).

5.1.3.5. Determination of *ctxB* genotype

Biotype-specific *ctxB* (*ctxB1* and *ctxB3*) was determined using the recently developed mismatch amplification mutation assay (MAMA)-PCR which utilizes sequence polymorphism of *ctxB* by focusing on nucleotide position 203 (Morita *et al.*, 2008). MAMA-PCR assay was done to test for presence of *ctxB* specific for CL (*ctxB1*) and ET (*ctxB3*) biotypes. A conserved forward primer (Fw-con, 5'-ACTATCTTCAGCATATGCACATGG -3') and two allele-specific polymorphism detection primers, Rv-cla (5'-CCTGGTACTTCTACTTGAAACG-3') and Rv-elt (5'-CCTGGTACTTCTACTTGAAACA -3') were used. PCR conditions were as follows: after initial denaturation at 96°C for 2 min, 25 cycles of denaturation at 96°C for 10 sec, annealing at 50°C for 10 sec, extension at 72°C for 30 sec; and final extension at 72°C for 2 min. *V. cholerae* O1 O395 CL and N16961 ET were included as reference strains.

5.1.3.6. DNA sequencing of *ctxB* and *tcpA*

Nucleotide sequencing of *ctxB* and *tcpA* of randomly selected representative strains of *V. cholerae* O1 ET isolated from endemic cholera cases in Mexico during 1998-2008 was carried out using an ABI PRISM BigDye Terminator Cycle Sequencing Reaction kit (Perkin-Elmer Applied Biosystems, Foster City, CA, USA) on an ABI PRISM 310 automated sequencer, as described previously (Olsvik *et al.*, 1993, Nair *et al.*, 2006). Nucleotide sequences of representative strains were compared with the corresponding sequences of *V. cholerae* ET reference N16961 (GenBank accession no. NC-002505), and *V. cholerae* CL reference 569B (GenBank accession no. U25679), retrieved from GenBank by Basic Local Alignment Search Tool (BLAST). To determine genetic relatedness of *tcpA* in the Mexican *V. cholerae* O1 strains, *tcpA* of representative ET, classical, and TCP variants was sequenced, using primers and PCR conditions as described previously (Kumar *et al.*, 2011). Multiple sequence alignment of trimmed *tcpA* (600bp, encoding mature protein) and phylogenetic analysis of *tcpA* alleles were done using Molecular Evolutionary Genetic Analysis program version 5.05 (MEGA5) (Tamura *et al.*, 2011). For comparison and construction of *tcpA*-phylogenetic tree, a total of 35 *tcpA* sequences were analyzed, of which 31 *tcpA* sequences of serogroup O1, non-

O1/O139 and others from a previous study were included (Kumar *et al.*, 2011), as well as *tcpA* of two nontoxigenic *V. cholerae* O1 strains from China (ZJ59 and LN93097, Accession no. EU622531 and AF512425 in GenBank), and two strains (Mex-2058 and Mex-3065) from the present study. A neighbor-joining tree was generated with 1,000 bootstrap to determine phylogenetic relationships on the basis of *tcpA* sequences.

5.1.3.7. Genetic analysis of CTXΦ and flanking region

Mexican ET strains possessing *ctxB3* and *rstR*^{ET} allele were further analyzed using primers *rstA*F/*rst*CR and *rstR*^{ET}F/*rstA3*R targeting the RS1 element. *V. cholerae* N16961, possessing RS1 downstream of CTX prophage and *V. cholerae* O395, lacking RS1, were used as controls, considering N16961 has CTX prophage only in its large chromosome, whereas O395 possesses CTX prophage in both the large and small chromosomes. The genetic structure of CTX prophage and its flanking region in representative ET strains harboring *ctxB3* and *rstR*^{ET} was analyzed using similar primers and primer walking and sequencing as described previously (Nguyen *et al.*, 2009, Mohapatra *et al.*, 2011). Primers cepF/rstR^{ET}R, cepF/rstAR and cepF/rstB were used to detect truncated CTX prophage in strains lacking RS1 element and tandem repeat of the CTX prophage. The ~1200bp amplicon for the cepF/rstA3R primers was sequenced for confirmation of truncated CTX prophage. Nucleotide sequences of ~1200bp amplicon for the cepF/rstA3R primers were deposited in GenBank under accession no. KC952008.

5.1.3.8. Pulsed-field gel electrophoresis (PFGE)

PFGE was carried out using a contour-clamped homogeneous electrical field (CHEF-DRII) apparatus (Bio-Rad), following procedures as described elsewhere (Cameron *et al.*, 1994, Nusrin *et al.*, 2009). Conditions for separation were 2 to 10s for 13h, followed by 20 to 25s for 6h. An electrical field of 6V/cm was applied at an included field angle of 120°. Genomic DNA of the *V. cholerae* strains was digested by *Not*I restriction enzyme (GIBCO-BRL, Gaithersburg, Md.) and *Salmonella braenderup* DNA by *Xba*I, with fragments used as molecular size markers. The restriction fragments were separated in 1% pulsed-field-certified agarose in 0.5X TBE (Tris-borate-EDTA) buffer. In the post

electrophoresis gel treatment step, each gel was stained and de-stained. DNA was visualized using a UV trans-illuminator and images were digitized employing the 1D Gel documentation system (Bio-Rad). Gel fingerprint patterns were analyzed using the Bionumeric Software Package (Applied Maths, Belgium). After background subtraction and gel normalization, fingerprint patterns were subjected to typing based on banding similarity and dissimilarity, employing the Dice similarity coefficient and UPGMA clustering, as recommended by the manufacturer. Results were graphically represented as dendrogram.

5.1.4. RESULTS

5.1.4.1. Microbiology, serology, and PCR analyses of *V. cholerae*

All isolates (n=91) showed characteristic colonies on taurocholate-tellurite-gelatin agar (TTGA), reacted to polyvalent O1-specific antisera, and were agglutinated with monoclonal antibodies specific for O1, but not for O139, reacting either in monovalent Inaba or Ogawa antisera, with the conclusion all belonged to serogroup O1. Among the 91 *V. cholerae* O1 strains, 48 were Inaba and 43 Ogawa (**Table 5.1**). Year-wise data revealed a predominance of Ogawa serotype during 1998 - 2004 and Inaba during 2005 - 2008. All strains amplified primers for *V. cholerae* species-specific gene *ompW* and O-biosynthetic gene *wbe*O1, confirming *V. cholerae* serogroup O1. The *ctxA* gene encoding A subunit of cholera toxin (CT) was present in 58 strains, confirming toxigenic (CTX+) *V. cholerae* O1 (**Table 5.1**) and 33 were lacking cholera toxin (CTX -).

5.1.4.2. Phenotypic and related genotypic characteristics

Phenotypic and related genetic characteristics of *V. cholerae* O1 strains are presented in **Table 5.1**. All *V. cholerae* O1 strains were resistant to CL biotype-specific phage IV, all but eleven were susceptible to ET-specific phage V, and all showed ET biotype specific phenotypic traits, including ability to agglutinate chicken blood cells (CCA) and resistance to polymyxin B (50 U) (**Table 5.1**). All isolates amplified primers for *rtxC*, an ET specific marker gene, and *hlyA*^{ET}, confirming ET biotype. As shown in **Table 5.1**, all but thirteen amplified primers for *tcpA*, 74 isolates amplified primers for *tcpA*^{ET}, and four

Table 5.1. Phenotypic and genetic characteristics of *V. cholerae* O1 strains (n=91) isolated from diarrheal and environmental samples in Mexico (1998-2008).

Country	Year of isolation	No. of isolates	Serotype	Source	Phenotypic properties				Genetic screening by PCR									Deduced Biotype
					Chicken Cell Aggl	Poly B (50U)	Phage IV Classical	Phage V El Tor	*ompW*	*wbeO1*	*ctxA*	**CtxB* type	*tcpA*	*rtxC*	*hylA*	*rstC*	*rstR* type	
Mexico	1998	3	Ogawa	Human	+	R	R	S	+	+	+	C	E	E	E	+	E,C	Alt-ET
	1999	4	Ogawa	Human	+	R	R	S	+	+	+	C	E	E	E	+	E,C	Alt-ET
		5	Ogawa	Human	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
		1	Ogawa	Human	+	R	R	R	+	+	-	-	E	E	E	-	-	ET
		1	Inaba	Human	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
	2000	2	Ogawa	Human	+	R	R	S	+	+	+	E	E	E	E	+	E	ET
		8	Ogawa	Human	+	R	R	S	+	+	+	C	E	E	E	+	E,C	Alt-ET
		1	Ogawa	Human	+	R	R	R	+	+	+	C	E	E	E	+	E,C	Alt-ET
		1	Ogawa	Human	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
		4	Inaba	Env.	+	R	R	S	+	+	-	-	-	E	E	-	-	ET
	2001	1	Ogawa	Human	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
		2	Ogawa	Env.	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
		2	Ogawa	Env.	+	R	R	S	+	+	-	-	C	E	E	-	-	TCP-Var
		3	Ogawa	Env.	+	R	R	R	+	+	-	-	E,C	E	E	-	-	TCP-Var
		1	Inaba	Env.	+	R	R	R	+	+	-	-	E,C	E	E	-	-	TCP-Var
	2002	1	Ogawa	Human	+	R	R	S	+	+	-	-	E,C	E	E	-	-	TCP-Var
		1	Ogawa	Human	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
		1	Inaba	Env.	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
	2003	1	Ogawa	Human	+	R	R	S	+	+	-	-	E,C	E	E	-	-	TCP-Var
		1	Ogawa	Env.	+	R	R	S	+	+	-	-	E,C	E	E	-	-	TCP-Var
		1	Inaba	Env.	+	R	R	S	+	+	-	-	E,C	E	E	-	-	TCP-Var
		1	Ogawa	Env.	+	R	R	S	+	+	-	-	C	E	E	-	-	TCP-Var
	2004	1	Ogawa	Human	+	R	R	S	+	+	-	-	C	E	E	-	-	TCP-Var
		1	Ogawa	Human	+	R	R	R	+	+	-	-	E,C	E	E	-	-	TCP-Var
		3	Ogawa	Env.	+	R	R	S	+	+	-	-	E	E	E	-	-	ET
		4	Inaba	Env.	+	R	R	S	+	+	+	E	E	E	E	-	E	ET
	2005	10	Inaba	Env.	+	R	R	S	+	+	+	E	E	E	E	-	E	ET
	2006	11	Inaba	Env.	+	R	R	S	+	+	+	E	E	E	E	-	E	ET
	2007	1	Inaba	Human	+	R	R	S	+	+	+	E	E	E	E	-	E	ET
		3	Inaba	Human	+	R	R	R	+	+	+	E	E	E	E	-	E	ET
		2	Inaba	Env.	+	R	R	S	+	+	+	E	E	E	E	-	E	ET
	2008	8	Inaba	Env.	+	R	R	S	+	+	+	E	E	E	E	-	E	ET
		1	Inaba	Env.	+	R	R	R	+	+	+	E	E	E	E	-	E	ET
N16961	1975	ET	Inaba	Human	+	R	R	S	+	+	+	E	E	E	E	+	E	El Tor
O395	1965	CL	Ogawa	Human	-	S	S	R	+	+	+	C	C	-	C	-	C	Classical

* Determined by mismatch amplification mutation assay (MAMA) PCR (Morita et al., 2008) ;**Serologically non-O1/non-O139, Chicken cell aggl, chicken cell agglutination; Poly B, polymixin B; R, resistant; S, sensitive; ET, El Tor; E, El Tor; Alt, altered; C, classical

for *tcpA*^{CL}. Of 58 CTX+ *V. cholerae* O1 ET strains, 16 amplified primers for *tcpA*^{ET}, *ctxB*^{CL}, *rstR*^{ET}, and *rstR*^{CL}, indicating *V. cholerae* atypical ET, as reported previously (Alam *et al.*, 2010). All atypical ET strains amplified primers specific for *rstC* of the RS1 satellite phage flanking ET CTXΦ. The remaining CTX+ *V. cholerae* O1 ET strains amplifed primers for *rstR*^{ET} and *ctxB*^{ET}, confirming *V. cholerae* ET 7th pandemic prototype, although these strains failed to amplify *rstC* of the RS1 element. Thirty-three *V. cholerae* ET strains lacked *ctxA*, *ctxB*, and *rstR*, confirming those to be CTXΦ^{-}. Of 33 CTXΦ^{-} *V. cholerae* strains, 13 lacked *tcpA*, while the remaining strains carried either ET-specific *tcpA* (*tcpA*^{ET}) or CL-specific *tcpA* (*tcpA*^{CL}) in their *V. cholerae* ET biotype background.

Year-wise analysis showed most of the clinical *V. cholerae* strains isolated between 1998 and 2004 were Ogawa (**Table 5.1**), with Inaba dominant among the environmental isolates. Ogawa was the only serotype isolated in Mexico between 2005 and 2008. The three *V. cholerae* strains isolated in 1998 were CTXΦ^{+} atypical ET and all seven isolated in 1999 were CTXΦ^{-} *V. cholerae* ET. *V. cholerae* CTXΦ^{+} atypical ET was isolated in 2000 in relatively larger proportions, together with *V. cholerae* ET strains with and without CTX-prophage, but was not consistently isolated from clinical cases in Mexico thereafter, until 2008 (**Table 5.1**). *V. cholerae* O1 biotype ET isolated between 2001 and 2003 lacked CTX prophage. However, most remarkable was the isolation in 2004 of CTX^{+} *V. cholerae* ET and thereafter, between 2005 and 2008, all Mexican isolates were CTXΦ^{+}*V. cholerae* ET of the seventh pandemic (**Table 5.1**).

Chronologically, the majority of CTXΦ^{-} *V. cholerae* ET isolated between 1999 and 2000 possessed ET biotype *tcpA* (**Table 5.1**), whereas all *V. cholerae* ET strains isolated during 2001 and thereafter were CTXΦ^{-}. The majority of CTXΦ^{-} *V. cholerae* isolated in Mexico between 2001 and 2003 lacked *tcpA*, while some CTXΦ^{-} *V. cholerae* carried the *tcpA* allele of either ET or CL type in an ET host; the latter were designated TCP variant. Only *V. cholerae* CTXΦ^{-} ET was isolated from environmental sources between 2001 and 2003. *V. cholerae* ET CTXΦ^{-} was no longer isolated in Mexico

between 2004 and 2005, indicating dramatic clonal switching. During 2005 - 2008 CTXΦ$^+$ *V. cholerae* O1 ET (Inaba serotype) was dominant, replacing CTXΦ$^-$*V. cholerae* ET (mostly Ogawa, Inaba, including transient *V. cholerae* TCP variants) (**Table 5.1**). Notably, the Mexican CTXΦ$^+$*V. cholerae* O1 ET isolated during 2005-2008 exhibited phenotypic and molecular traits of the 7th pandemic ET.

5.1.4.3. Nucleotide sequencing of *ctxB* and *tcpA* gene

Data on the *ctxB* (460 bp) sequence of CTXΦ$^+$ strains from each year class (1998 to 2008) revealed that *V. cholerae* atypical ET strains isolated before 2004 presented amino acid sequences identical to *V. cholerae* CL biotype CT (*ctxB1*), with histidine and threonine at positions 39 and 68, respectively. However, *ctxB* sequences of ET strains isolated in 2004 and thereafter matched ET biotype CT (*ctxB3*). Sequencing of the *tcpA*^{ET} allele of ten Mexican ET strains possessing only ET type *tcpA* showed the sequence was identical to that of seventh pandemic *V. cholerae* ET N16961(Accession No. AF536868). The nucleotide sequences of *tcpA*^{CL} of four *V. cholerae* ET strains amplifying only the CL biotype-specific *tcpA* primers (TCP variant ET, Mex-2174, Mex-3358, Mex-3065, and Mex-2058) were identical, despite the sequence similarity being 81% with reference CL and 74% with ET biotype *tcpA* alleles. The nucleotide sequence of *tcpA*^{CL} allele of the Mexican *V. cholerae* ET strains matched exactly the *tcpA* sequence in GenBank for non-toxigenic *V. cholerae* O1 ZJ59 (Accession no. EU622531) and LN93094 (Accession no. AF512422) reported from China. The deduced amino acid sequences revealed substitutions at positions 74 [valine to isoleucine (V→ I)], 99 [proline to serine (P→ S)], 145 [alanine to glycine (A→ G)], and 189 [threonine to alanine (T→ A)], suggesting genetic differences from CL, ET, and other types of *tcpA* alleles described previously (Mukhopadhyay *et al.*, 2001).

A phylogenetic tree was constructed using the *tcpA* sequences of two Mexican strains (Mex-2058 and Mex-3065) together with *tcpA* gene sequences from GenBank (**Figure 5.2**). The Mexican and Chinese strains with analogous *tcpA* sequences formed a separate cluster and ten distinctly different clusters were also observed, based on major

tcpA alleles reported to date (Kumar *et al.*, 2011). Overall, results of the analysis clearly indicate *tcpA* of Mex-2058 and Mex-3065 have a lineage distinct from the seventh pandemic ET.

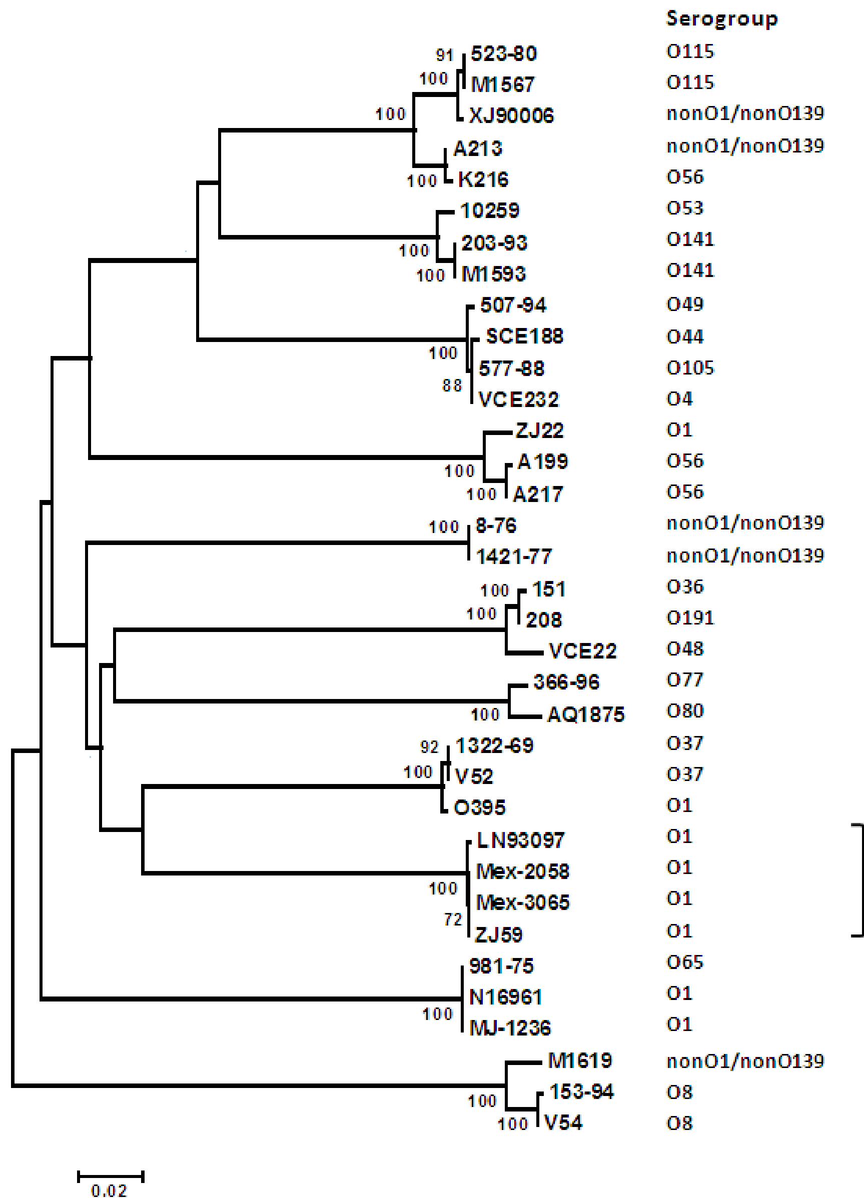

Figure 5.2. The *tcpA* Phylogenetic tree constructed using neighbor joining. Mex- 2058 and Mex-3065 strains represent the TcpA variant of *V. cholerae* O1 biotype ET isolated in Mexico during 2000-2004 and share the cluster with TcpA of *V. cholerae* O1 strains isolated in China obtained from GenBank. All other TcpA sequences were from

GenBank. Bootstrap values are percentages of 1000 replications and if 50% or greater are indicated at the nodes.

5.1.4.4. Genome organization of CTXΦ and flanking region

The Mexican CTXΦ^{+} *V. cholerae* ET strains isolated between 2004 and 2008 failed to amplify *rstC* of the RS1 element. To follow up on this finding, representative ET biotype specific *ctxB*^{ET} and *rstR*^{ET} were tested for genetic arrangements in CTXΦ and its flanking region by primer walking. Results showed that, unlike ET reference strain N16961, none of the *rstC* negative strains amplified the ca. 1.6kb and 2.1kb targeted RS1 sequences primed by *rstA*-F/*rstC*-R and *rstR*^{ET}-F/*rstC*-R, respectively (**Table 5.2**), confirming absence of RS1 in the genome of the Mexican *V. cholerae* ET strains. Since the Mexican *V. cholerae* ET lacked RS1, a phenomenon previously observed in the CL biotype, variant from Matlab, Bangladesh, and Mozambique variant strains (Chun *et al.*, 2009), PCR assay was performed, targeting *dif2* and using primers CIIF and CIIR to determine the presence of CTXΦ in the small chromosome (Chr II). As shown in **Table 5.2**, all strains yielded a ca. 747 bp amplicon, confirming absence of CTXΦ in Chr II. As the ET strains were CTXΦ+, primer set ctxAF/RTX5R and TLC3F/ rstAR were used to locate the exact position of CTX prophage. As shown in **Table 5.2**, amplification of ca. 1400 bp and 1800 bp DNA fragments, respectively, confirmed CTX prophage to be present in the large chromosome (Chr I). Presence or absence of CTX prophage tandem repeat was determined using the primer pair ctxBF/cepR. However, the amplicon was not found using template DNA prepared from the Mexican strains, indicating a tandem repeat of CTX prophage was not present in either chromosome.

Primer sets cepF/rstR^{ET}R, cepF/rstAR and cepF/rstBR were used to detect truncated CTX prophage at the upstream region of CTX prophage for strains lacking RS1. All yielded amplicons, ca. 1200 bp, 2200 bp and 2700 bp for cepF/rstR^{ET}R, cepF/rstAR and cepF/rstBR primers, respectively, as did CL reference strain O395, confirming presence of a truncated CTX prophage at the upstream region of CTX. DNA sequence analysis of the ca. 1200 bp amplicon, using cepF/rstR^{ET}R primers, confirmed

Table 5.2. Genetic arrangement of CTXΦ and flanking region of representative ctx+ strains (2004 – 2008)

Year of isolation	No. of strains (representative)	RS1		CIIF-CIIR (0.75 kb)	Primer sets targeting TLC-CTXΦ-RTX region						
		***rstC* (238 bp)**	***rstA*F-*rstC*R (~1.6 kb)**		**TLC3F-*rstA*R (~1.8 kb)**	***ctxA*F-RTX5R (~1.4 kb)**	***ctxB*F-*cep*R (~3.0 kb)**	***rstA*F-*rstR*^{ET}R (~2.8 kb)**	***cep*F-*rstR*^{ET}R (~1.2 kb)**	***cep*F-*rstA*R (~2.2 kb)**	***cep*F-*rstB*R (~2.7kb)**
2004	2	-	-	+	+	+	-	+	+	+	+
2005	2	-	-	+	+	+	-	+	+	+	+
2006	2	-	-	+	+	+	-	+	+	+	+
2007	2	-	-	+	+	+	-	+	+	+	+
2008	2	-	-	+	+	+	-	+	+	+	+
N16961 (ET Reference)		+	+	+	+	+	-	-	-	-	-
O395 (CL Reference)		-	-	-	+	+	-	-	-	+	+
B-33 (Mozambique Variant)		-	-	-	-	-	+	-	-	-	-

the presence of a truncated orfU subunit (438 bp) within the truncated CTX prophage. Thus, the results showed *V. cholerae* O1 ET strains isolated in Mexico between 2004 and 2008 had a truncated CTX prophage in addition to CTX prophage in the large chromosome, with neither an RS1 element nor a tandem array of CTX prophage.

A schematic genetic map displaying the deduced chromosomal localization of CTX prophage and its flanking genetic region is provided in **Figure 5.3**.

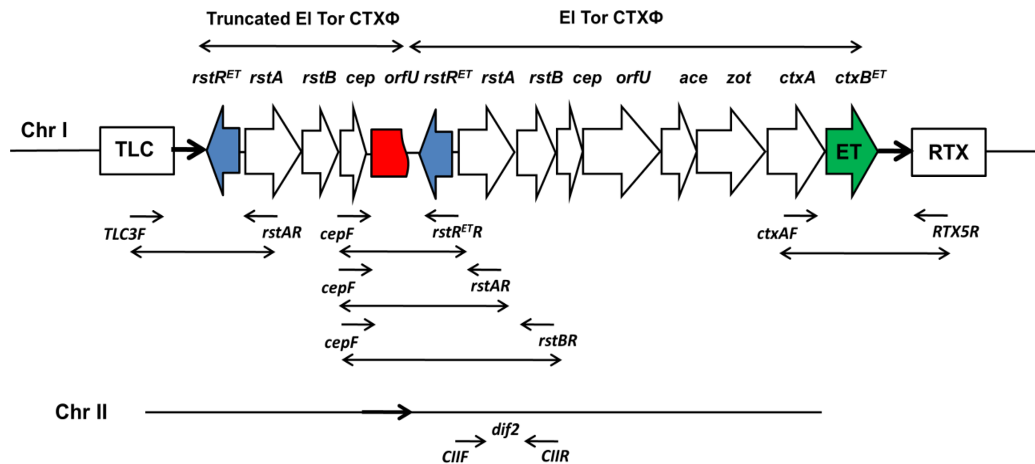

Figure 5.3. Genetic mapping of CTX prophage lacking RS1 in *V. cholerae* O1 biotype El Tor strains isolated in Mexico during 2004-2008. The deduced genetic organization shows integration of an additional truncated CTX prophage in the upstream region of the CTX prophage in the large chromosome (Chr I) of *V. cholerae*, whereas *dif2* site in the small chromosome (Chr II) of *V. cholerae* O1 biotype El Tor strains isolated in Mexico during 2004-2008 did not harbor CTX prophage. Arrows indicate transcription directions of each of the genes in the CTX prophage and flanking regions.

5.1.4.5. PFGE and cluster analysis

Results of genomic fingerprinting determined by pulsed-field gel electrophoresis (PFGE) showed Mexican *V. cholerae* O1 ET strains carrying *tcpA*^{ET} and ET biotype-specific CTX-prophage genes yielded an overall banding pattern characteristic of 7th pandemic *V. cholerae* ET N16961 (**Figure 5.4**). *V. cholerae* CTXΦ^{-} ET and TCP variants revealed

significantly different PFGE patterns from signature banding patterns of the *V. cholerae* ET isolated in Mexico between 1998 and 2008. As shown in **Figure 5.4**, cluster analysis separated genetically heterogeneous CTXΦ$^-$ ET and TCP variants from pre- and post-2004 (1991 and 2008) *V. cholerae* O1 ET, the latter comprising a cluster with Peruvian *V. cholerae* ET and *V. cholerae* N16961.

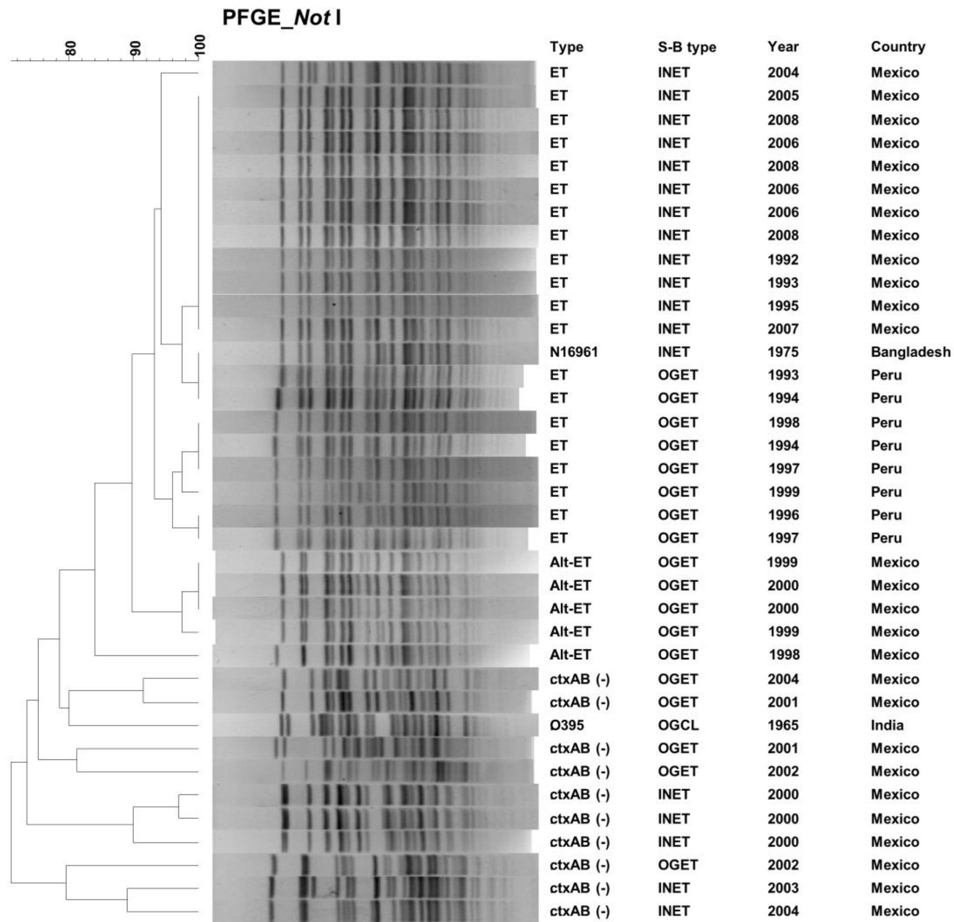

Figure 5.4. Dendrogram of PFGE patterns of *V. cholerae* O1 ET and transitional CTX− ET possessing TCP of ET or CL biotype. The tight clustering of Mexican *V. cholerae* O1 prototype ET strains indicates clonality whereas CTX− ET and transitional TCP variants of ET demonstrate heterogeneous patterns, indicating divergence. Peruvian prototype ET strains clustered differently from Mexican *V. cholerae* ET. S-B type, sero-biotype, Alt-ET, altered ET; INET, Inaba ET; OGET, Ogawa ET

The Mexican *V. cholerae* ET strains (1991 - 2008) formed a tight cluster containing *V. cholerae* ET N16961 and closely linked to *V. cholerae* ET strains from Peru. However, tight-clustering was not observed for *V. cholerae* O1 CTX- ET and transitional TCP variants of ET since their patterns were heterogeneous, indicating divergence.

5.1.5. DISCUSSION

Reports of the emergence and pandemic spread of an atypical *V. cholerae* El Tor (ET) suggest that the pattern of global cholera is changing, a phenomenon that appears to be occurring in Mexico, where the *V. cholerae* CL biotype of pre-1991 cholera has been isolated, as well as atypical *V. cholerae* ET, the latter dominant among *V. cholerae* CL and ET biotype progenitors associated with cholera outbreaks between 1991 and 1997 (Alam *et al.*, 2010). To date, the CL biotype may no longer be associated with human disease in Mexico; and atypical ET, previously considered the major cause of endemic cholera (Alam *et al.*, 2010), was no longer isolated after emergence in 2000 and predominance of the $CTX\Phi^{-}$ population, including a unique TCP variant of *V. cholerae* ET. Characterization of the *V. cholerae* population in subsequent years (2004–2008) suggests another major clonal shift, with the emergence and predominance of a CTX^{+} *V. cholerae* O1 ET lacking the RS1 element, with a truncated CTXΦ in the upstream region of the CTXΦ in the large chromosome.

Although the *V. cholerae* ET strains varied in PCR primer amplification for *ctxB*, *tcpA*, and *rstR* biotype-specific alleles (Keasler & Hall, 1993, Kimsey *et al.*, 1998), overall serobiotyping results confirmed that the strains were *V. cholerae* O1 ET seventh pandemic. The genetic transitions in Asia from *V. cholerae* biotypes CL to ET (Siddique *et al.*, 1991) and ET prototype to atypical ET are considered chronological events occurring in the 1980s and in 2001, respectively (Nair *et al.*, 2002, Nair *et al.*, 2006, Raychoudhuri *et al.*, 2009). In Mexico, both *V. cholerae* CL and ET biotype progenitors were isolated during cholera outbreaks, together with atypical ET, dominant between 1991 and 1997 (Alam *et al.*, 2010). The consistent association of *V. cholerae* CL, ET, and atypical ET with local cholera outbreaks in the Americas between

1991 and 1997 and their occurrence in the natural aquatic environment, together with non-O1/O139 phenotypically and genetically O1 CL, suggesting the emergence of atypical *V. cholerae* ET has occurred locally in Mexico (Alam *et al.*, 2010, Alam *et al.*, 2012) and is not clonal expansion from endemic regions of Asia, as has been proposed for Africa (Mutreja *et al.*, 2011). According to our data, *V. cholerae* atypical ET was not isolated in Mexico after 2000 nor was the CL biotype reported in Mexico until 1997 (Alam *et al.*, 2010). However, *V. cholerae* atypical ET continues to be routinely isolated from clinical cholera cases in Asia and Africa (Nair *et al.*, 2002, Ansaruzzaman *et al.*, 2004, Nair *et al.*, 2006, Safa *et al.*, 2008), and both are reported to be capable of causing a more severe disease and of spreading globally (Siddique *et al.*, 2010, Chin *et al.*, 2011, Mutreja *et al.*, 2011). The clonal shift observed in this study may well have implications for global cholera, considering that atypical ET strains, such as those isolated from cholera cases in Asia and Africa, are dominant pandemic pathogens (Ansaruzzaman *et al.*, 2004, Safa *et al.*, 2008, Chin *et al.*, 2011, Islam *et al.*, 2011).

The toxin coregulated pili (TCP), a receptor for lysogenic CTXΦ (Waldor & Mekalanos, 1996), is encoded by a 40-kb gene cluster, *Vibrio* pathogenicity island (VPI) (Faruque *et al.*, 1999). Like CTXΦ, TCP is biotype-specific, with two distinct alleles of *tcpA* encoding the major protein “pilin” (Mukhopadhyay *et al.*, 2001, Kumar *et al.*, 2011). In Mexico, the TCP variant of *V. cholerae* O1 strains carries a *tcpA* sequence that is different from *tcpA* types reported to date, including CL and ET (Mukhopadhyay *et al.*, 2001, Kumar *et al.*, 2011), although an analog of the Mexican type *tcpA* deposited in GenBank is from a nontoxigenic O1, reportedly isolated in 2006 from China. *V. cholerae* O1 strains from Mexico carrying the variant *tcpA* are nontoxigenic. Thus, nontoxigenic O1 strains with this *tcpA* can occur in geographically different ecosystems, namely Mexico and China.

As in the case of lysogenic CTXΦ that propagate by infecting susceptible nontoxigenic strains of both *V. cholerae* and *Vibrio mimicus* (Waldor & Mekalanos, 1996, Faruque *et al.*, 1999), converting the host to $CTX\Phi^{+}$, excision can also occur,

transforming the host to CTXΦ^{-} (Alam *et al.*, 2007). In the present study, *V. cholerae* O1 strains isolated in Mexico during 2000–2004 were predominantly CTXΦ^{-}, TCP^{-}. Although VPI is known to be horizontally transferred in *V. cholerae* (Karaolis *et al.*, 1998), it is not generally considered a virion. In any case, CTXΦ^{-}, TCP^{-} *V. cholerae* in Mexico can be considered to have originated from a toxigenic progenitor via excision of both CTXΦ and VPI.

Bacterial clonal switching in *V. cholerae* can have a profound epidemiological influence (Siddique *et al.*, 1991, Siddique *et al.*, 2010). Factors playing a role in clonal selection can be biotic, abiotic, or both, in natural aquatic ecosystems where *V. cholerae* is an autochthonous presence (Nelson *et al.*, 2009, Mukhopadhyay *et al.*, 2014, Stine & Morris, 2014). In the study reported here, a unique clonal shift was observed whereby toxigenic *V. cholerae* O1 in Mexico was replaced, starting in 2001, with CTXΦ^{-}, TCP^{-} *V. cholerae* and superseded by toxigenic *V. cholerae* in 2005. It can be hypothesized that nontoxigenic *V. cholerae* in the aquatic environment will outnumber pathogenic strains when fecal–oral transmission is brought under control by application of stringent public health measures, as was done following the 1991–1997 epidemics in Mexico (Lizarraga-Partida *et al.*, 2009, Lizarraga-Partida & Quilici, 2009, Alam *et al.*, 2010). Also important to note is that many bacterial pathogens including *V. cholerae* become nonculturable after release from human host into the aquatic environment, becoming noncompetitive with environmental *V. cholerae*, a phenomenon that may have occurred in Mexico between 2001 and 2004 (Alam *et al.*, 2007, Nelson *et al.*, 2009). An alternative hypothesis is that the predominance of a toxigenic or nontoxigenic clone may occur if phages provide selective advantage to a *V. cholerae* subpopulation (Nelson *et al.*, 2008, Nelson *et al.*, 2009).

Nontoxigenic *V. cholerae* O1 is associated with a mild but rather broad spectrum of human disease. A clinical link for nontoxigenic *V. cholerae* O1 ET was first reported by the US Centers for Disease Control and Prevention (CDC) between 1977 and 1991 (Rodrigue *et al.*, 1994). During the past three decades, there have been reports of

isolation of nontoxigenic *V. cholerae* O1 from clinical, sewage, oyster, and surface-water samples collected in several countries, including Bangladesh, Guam, Brazil, Peru, Japan, England, the United States, and Mexico (Rodrigue *et al.*, 1994, Alam *et al.*, 2010). CT^{-} *V. cholerae* O1 strains occur sporadically as the causative agent of cholera outbreaks, but not at as high a frequency as in Mexico between 2000 and 2004. It is important to note that nontoxigenic *V. cholerae* O1 was associated with a large cluster of cases of cholera in India (Saha *et al.*, 1996). In addition to *V. cholerae* O1 CT^{-} strains, currently available data suggest that a significant proportion of non-O1 strains possessing CT or non–O1-specific heat-stable enterotoxin (NAG-ST) are associated with cases of diarrhea (Bagchi *et al.*, 1993).

In toxigenic *V. cholerae* O1 ET and O139 Bengal strains, the CTXΦ genome often is flanked by satellite phage RS1 carrying genes *rstA, rstB, rstC*, and *rstR* that determine integration of plasmids transporting portions of CTXΦ into the bacterial genome. The RS2 genetic element also is a satellite phage but differs from RS1 in lacking *rst*C (Waldor & Mekalanos, 1996, Faruque *et al.*, 1999). *V. cholerae* O1 CL biotype CTX prophages lack RS elements and exist either as a solitary prophage or as arrays of two truncated, fused prophages (Chun *et al.*, 2009). *V. cholerae* O1 ET and O139 Bengal strains carrying RS1 generally yield infectious CTXΦ, but CL biotype strains do not. *V. cholerae* atypical ET carrying the *ctxB* allele of the CL biotype recently was reported to have been isolated in Bangladesh (Matlab variant) and Mozambique (Mozambique variant), lacking RS1 (Ansaruzzaman *et al.*, 2007, Chun *et al.*, 2009). Pathogenic *V. cholerae* ET strains isolated in Mexico between 2004 and 2008 lack RS1, but, unlike the Matlab and Mozambique variants (Ansaruzzaman *et al.*, 2007, Chun *et al.*, 2009), the *ctxB* allele of these strains is an ET biotype. In addition, ET biotype strains isolated in Mexico during 2004–2008 exhibited a novel genetic array, with a truncated CTX prophage instead of RS1 element in the upstream region of the ET-specific CTXΦ located in the large chromosome (Chr I) of *V. cholerae*. It has been proposed that gene capture via plasmids or phages contributes to rapid adaptation and evolution in *V. cholerae*, a probable situation in Mexico, with clonal CTX^{+} ET displacing all other

preexisting subtypes (Manning *et al.*, 1999, Chun *et al.*, 2009). Although the epidemiological significance of the newly emerged *V. cholerae* CTX^+ ET in Mexico is yet to be understood, the truncated CTXΦ upstream could surrogate RS elements in facilitating replication of the lysogenic CTXΦ genome, producing infectious phage particles.

An aquatic reservoir of *V. cholerae* in the Americas has previously been documented, including association of *V. cholerae* with plankton in coastal waters of Peru and Mexico (Gil *et al.*, 2004, Lizarraga-Partida *et al.*, 2009, Lizarraga-Partida & Quilici, 2009). Although the *V. cholerae* ET isolates causing epidemic cholera in Latin America in 1991 were initially considered to be homogeneous (Wachsmuth *et al.*, 1993), divergence was demonstrated soon thereafter (Dalsgaard *et al.*, 1995, Evins *et al.*, 1995, Beltran *et al.*, 1999). Divergence was clearly evident among *V. cholerae* ET in Peru and Mexico (Nusrin *et al.*, 2009, Alam *et al.*, 2010), and *V. cholerae* associated with the 1991 epidemic and subsequent endemic cholera in Mexico comprises a diverse population of serogroup O1 strains, including biotypes CL and ET, together with atypical variant ET (Alam *et al.*, 2010). *V. cholerae* O1 endemic cholera along the Gulf of Mexico coast after 1997 recently has been confirmed, with changes in ribosomal patterns separating strains into two distinct chronological groups, those isolated before and after 1997. Results presented here are in agreement with those of Lizarraga-Partida et al. (Lizarraga-Partida *et al.*, 2009, Lizarraga-Partida & Quilici, 2009), who described a new CTX prophage^- *V. cholerae* lineage emerging shortly after 1997 in Mexico and continuing to displace preexisting *V. cholerae* O1 CL and ET strains, including atypical ET strains. The CTX^+ *V. cholerae* ET with a truncated CTX prophage isolated during endemic cholera in Mexico may prove to be a historical point in the global epidemiology of cholera.

5.1.6. CONCLUSION

Our findings regarding the emergence of *V. cholere* O1 in Mexico are in agreement with the report of Lizarraga-Partida and colleagues (Lizarraga-Partida & Quilici, 2009), who described a new $CTX\Phi^-$ *V. cholerae* lineage emerging shortly after 1997 in Mexico

continuing to displace preexisting *V. cholerae* O1 CL and ET strains, including atypical ET strains (Alam *et al.*, 2010). Their finding, corroborated by PFGE analysis of CTXΦ$^-$ *V. cholerae* strains isolated in Mexico between 1999 and 2004, showed the latter grouped separately from Mexican *V. cholerae* ET 7th pandemic prototype strains isolated in 2004 and continuing as the cholera pathogen lineage until 2008. The epidemiological significance of *V. cholerae* atypical ET not having been isolated in a recently reported cholera epidemic in Mexico remains unclear. The CTXΦ$^+$ *V. cholerae* ET with a truncated CTX prophage having been isolated during endemic cholera in Mexico, in contrast, may prove to be an historical point in the epidemiology of cholera globally.

5.2. Genomic diversity of *Vibrio cholerae* associated with endemic cholera in Mexico, 1991 – 2008

5.2.1. ABSTRACT

Pandemic cholera occurred in Mexico in 1991 where cholera had not been reported for more than a century. Thereafter, cholera has become endemic in Mexico, with coexistence of *V. cholerae* O1 prototype ET, classical, and atypical ET (1991-1997), followed by predominance of non-toxigenic (CTX^-) *V. cholerae* El Tor over toxigenic (CTX^+) strains (2001-2003). Interestingly, *V. cholerae* CTX^+ variant-El Tor was isolated during 2004-2008, outcompeting CTX^- *V. cholerae*. In this study, the genomes of six Mexican *V. cholerae* O1 strains, isolated during 1991-2008 were sequenced and compared with contemporary isolates and archived strains of *V. cholerae*. Three of the six were CTX^+ El Tor, two CTX^- El Tor, and the remaining a CTX^+ classical isolate. Whole genome sequence analysis showed the six isolates belonged to five distinct phylogenetic clades. One of the two CTX^- isolates appeared to be ancestral to all 6th and 7th pandemic CTX^+ *V. cholerae*. The other CTX^- isolate placed into a basal clade of CTX^- non-O1/O139 isolates from Haiti and seroconverted O1 isolates from Brazil and Amazonia. Interestingly, one of the CTX^+ isolate was phylogenetically placed into sixth pandemic classical isolates clade along with the *V. cholerae* O395 classical reference strain. Two of the CTX^+ El Tor strains possessing an intact VSP-II are closely related to hybrid El Tor strains from Mozambique and Bangladesh. The third CTX^+ El Tor had West African-South American (WASA) recombination in the VSP-II and showed close clonal relatedness with isolates from Peru and Brazil. All of the Mexican isolates were found to lack SXT/R391 Integrative Conjugative Elements (ICE) and were sensitive to penicillin (P), ampicillin (AMP), chloramphenicol (C), trimethoprim-sulfamethoxazole (SXT), tetracycline (TE), kanamycin (K), erythromycin (E), nalidixic acid (NA), and ciprofloxacin (CIP). None of the six isolates showed relatedness to contemporary isolates from Asia, Africa, or Haiti, but instead indicated strong phylogenetic diversity.

5.2.2. INTRODUCTION

Cholera has been endemic in the Ganges delta region of South Asia for centuries. However, several countries of Sub-Saharan Africa and Latin America were severely affected during the 7th pandemic and subsequently are now considered endemic (Sack *et al.*, 2004). That is, cholera appeared in Mexico in June, 1991, after the Latin American epidemic had begun along the Peruvian coast in January, 1991 (Sepulveda *et al.*, 2006). The disease soon broke out in neighboring countries by 1992, with the exception of Uruguay and French Guyana (Olsvik, 1992). In Mexico, a total of 43,536 cholera cases were reported between 1991 and 1996, with a substantial number of deaths (Sepulveda *et al.*, 2006). Epidemiological investigations confirmed the association of *Vibrio cholerae* O1 biotype El Tor with the majority of those cholera cases, although the classical biotype was isolated from some cases in Mexico during subsequent years until 1997 (Olsvik, 1992, Wachsmuth *et al.*, 1993, Alam *et al.*, 2010, Alam *et al.*, 2014).

It has long been established that *V. cholerae* O1 has caused seven pandemics since 1817, of which the 7th pandemic is the largest, considering its longevity and geographical distribution. *V. cholerae* El Tor replaced the classical biotype of the 6th pandemic and presumably earlier pandemics (Sack *et al.*, 2004, Mutreja *et al.*, 2011). Variants of El Tor (hybrid El Tor and/or atypical El Tor) possessing classical biotype specific traits have been reported in Asia, Africa, and Latin America (Nair *et al.*, 2002, Safa *et al.*, 2008, Alam *et al.*, 2010). Genetic changes, i.e. gain or loss of mobile genetic elements and genomic islands occur in *V. cholerae* due to its genomic plasticity (Chun *et al.*, 2009). An example is the emergence of *V. cholerae* O139 in late 1992 in India, a non-O1 serogroup that caused a massive outbreak in South Asia and beyond (Albert *et al.*, 1993, Albert & Nair, 2005). Since 2001, variants of El Tor have been associated with cholera epidemics globally, including the recent epidemic in Haiti and previously Zimbabwe (Islam *et al.*, 2011, Hasan *et al.*, 2012, Eppinger *et al.*, 2014). Although, significant advances have been made in the understanding of genetics, epidemiology, and ecology of *V. cholerae* over the past two decades, the lack of an extensive genomic database severely limits source attribution for some of the recent outbreaks.

The cholera epidemic in Latin America was hypothesized to have been imported from endemic areas since Latin America had not reported cholera for more than 100 years prior to 1991 (Guthmann, 1995). Three hypotheses have been offered: (1) international trade ships from Asia discharged the pathogen into Peruvian ports in ballast water (McCarthy & Khambaty, 1994), (2) immigrants from Africa into Latin America in the 1970s brought the pathogen with them Lam et al 2010, (Wachsmuth *et al.*, 1993); and (3) environmental factors, e.g., El nino, played a significant role (Mourino-Perez, 1998, Seas *et al.*, 2000). Preliminary analysis using molecular typing indicated *V. cholerae* strains isolated in Latin America during the 1990s epidemic were clonal and represented intrusion of the seventh pandemic El Tor strain into the Western hemisphere (unrelated to the US Gulf Coast clone) (Wachsmuth *et al.*, 1993). However, subsequent genomic analysis of 30 single nucleotide polymorphisms (SNPs) indicated close relatedness of the 1990s Latin American isolates to African strains isolated in the 1970s and 1990s (Lam *et al.*, 2010). This finding was supported by a recent phylogenetic analysis showing isolates from the Latin American epidemic in the 1990s were related to a *V. cholerae* strain from Angola, the study that analyzed only seventh pandemic El Tor strains from the Latin American epidemic that carried the *ctxB*3 genotype (*B3* allele) (Mutreja *et al.*, 2011). However, *V. cholerae* atypical El Tor has been found to coexist with classical and prototype El Tor in Mexico since the Latin American epidemic began (Alam *et al.*, 2010). A serious limitation of that retrospective epidemiological study was that the analysis included only a limited number of strains collected spatio-temporally, thereby masking the full genetic diversity of the Mexican *V. cholerae* population. Phenotypic and genotypic characteristics of 182 *V. cholerae* O1 strains from Mexico that had been isolated between 1983 and 2008, previously had been reported to have several unique features (Alam *et al.*, 2010, Alam *et al.*, 2012, Alam *et al.*, 2014). In this study, six *V. cholerae* O1 isolates from Mexico were selected (**Table 5.3**) based on previously published data (Alam *et al.*, 2010, Alam *et al.*, 2012, Alam *et al.*, 2014), for whole genome sequencing to compare genomes of these six with genomes of 124 *V. cholerae* archival and recent isolates to elucidate the evolutionary dynamics of *V. cholerae* in Mexico.

5.2.3. MATERIALS AND METHODS

5.2.3.1. Sequencing, Assembly, and Annotation

Genomic DNA of six *V. cholerae* strains was subjected to next generation, whole genome Illumina and hybrid Illumina/454 sequencing and closure strategies, as previously described (Chun *et al.*, 2009, Hasan *et al.*, 2012). Libraries were constructed with target insert sizes of 3 kb and paired-end 100 bp. Hybrid and Illumina sequences were assembled using Celera and Velvet assemblers, respectively (Zerbino, 2010) and all chromosomes were manually annotated using the Manatee system (http://manatee.sourceforge.net/).

Table 5.3. Characteristics of *Vibrio cholerae* strains analyzed in this study

Strain	Serogroup	Serotype	Biotype	Country of origin	Source	Year of isolation	CTXϕ	Accession Number
CP1032	O1	Ogawa	El Tor	Mexico	Human	1991	+	ALDA00000000
95412	O1	Inaba	Classical	Mexico	Human	1997	+	APFM00000000
CP1033	O1	Ogawa	El Tor	Mexico	Human	2000	+	AJRL00000000
CP1037	O1	Ogawa	El Tor	Mexico	Environment	2003	-	ALDB00000000
CP1035	O1	Ogawa	El Tor	Mexico	Human	2004	-	AJRM00000000
CP1030	O1	Inaba	El Tor	Mexico	Environment	2008	+	ALCZ00000000

5.2.3.2. Comparative Genomics

Genome-to-genome comparison was performed using different approaches since completeness and quality of the nucleotide sequences varied from strain to strain. First, ORFs of a given pair of genomes were identified and reciprocally compared with each other using the BLASTN, BLASTP, and TBLASTX programs (ORF-dependent comparison). Second, a bioinformatic pipeline was constructed to identify homologous regions of a given query ORF. Initially, a segment on the target contig, which is homologous to a query ORF, was identified using the BLASTN program. This potentially homologous region was expanded in both directions by 2,000 bp. Nucleotide sequences of the query ORF and selected target homologous regions were aligned using a pairwise global alignment algorithm and the resultant matched region in the subject contig was extracted and saved as a homologue (ORF independent comparison). Orthologues and paralogues were differentiated by reciprocal comparison. In most cases, ORF dependent

and ORF-independent comparisons yielded the same orthologues, although the ORF-independent method performed better for draft sequences of low quality, in which sequencing errors, albeit rare, hampered identification of correct ORFs.

5.2.3.3. Identification and Annotation of Genomic Islands

We defined genomic islands (GIs) as a continuous array of five or more coding sequences (CDSs) that were discontinuously distributed among genomes of test strains. Correct transfer or insertion of GIs was readily differentiated from a deletion event by comparing the genome based phylogenetic tree and full matrices showing pairwise detection of orthologous genes between test strains. Identified GIs were designated and annotated using the BLASTP search of its member CDSs against the GenBank NR database, as described elsewhere (Chun *et al.*, 2009).

5.2.3.4. Phylogenetics Based on Genome Sequences

Orthologous regions of *V. cholerae* N16961 were identified by comparisons based on similarity and were used to generate phylogenetic trees (Hasan *et al.*, 2012). The set of orthologous regions for each CDS of a reference genome was identified according to nucleotide similarity and aligned using CLUSTALW2. The resultant multiple alignments were concatenated to form genome scale alignments, which were then used to generate the neighbor-joining phylogenetic trees (Saitou & Nei, 1987).

5.2.3.5. Nucleotide sequence accession number

*Whole-genome sequences of CP1030, CP1032, CP1033, CP1035, CP1037, and 95412 have been deposited in DDBJ/EMBL/GenBank under accession Nos. ALCZ00000000, ALDA00000000, AJRL00000000, AJRM00000000, ALDB00000000, and APFM00000000, respectively.

5.2.4. RESULTS AND DISCUSSION

5.2.4.1. Variations in CTXΦ-RS1

Four of the six isolates of *V. cholerae* O1: 95412, CP1030, CP1032, and CP1033 were lysogenic CTXΦ positive, while the remaining two isolates (CP1035 and CP1037) lacked CTXΦ (**Table 5.3**). Lysogenic CTXΦ contains two gene clusters, core region and RS2 element (Waldor & Mekalanos, 1996, Safa *et al.*, 2010). The core region comprises *ctxAB*, encoding cholera toxin (CT), and five other genes, namely *psh*, *cep*, *orfU*, *ace*, and *zot*, that are required for phage morphogenesis. The RS2 element encodes proteins associated with CTXΦ replication (RstA), integration (RstB), and regulation (RstR) (Waldor & Mekalanos, 1996, Safa *et al.*, 2010). Satellite phage RS1 carries an additional *rstC* gene (encoding anti-repressor protein), along with the entire RS2 element that is usually present in the flanking region of CTXΦ in *V. cholerae* El Tor (Safa *et al.*, 2010). The chromosomal location of CTXΦ and its orientation, and copies of CTXΦ may differ among toxigenic *V. cholerae* strains (Lee *et al.*, 2009, Nguyen *et al.*, 2009, Choi *et al.*, 2010). The CTXΦ-RS1 array of CP1030 has been shown to be unique, lacking RS1 and carrying a truncated CTXΦ instead of RS1 in the upstream region of CTXΦ (*B3* allele) in the large chromosome (Chr I) (Alam *et al.*, 2014). The *V. cholerae* O1 El Tor strains isolated in Mexico between 2004 and 2008, show the same CTXΦ array (TLC-truncated CTX- CTXΦ^{B3}) (Alam *et al.*, 2014). Moreover, predicted CTXΦ mapping of El Tor isolates associated with the 1990s Latin American epidemic in Peru, Mexico, Bolivia, Columbia, and Argentina showed two copies of CTXΦ (*B3* allele) together with TLC and RS1 in Chr I (TLC-CTXΦ^{B3}-CTXΦ^{B3}-RS1) (Mutreja *et al.*, 2011). CTXΦ arrays, either TLC-truncated CTX- CTXΦ^{B3}, or TLC-CTXΦ^{B3}-CTXΦ^{B3}-RS1 detected in Latin American isolates was not found in El Tor, atypical El Tor, or El Tor variants from Asia, Africa, and Haiti that have been studied to date (Lee *et al.*, 2009, Nguyen *et al.*, 2009, Choi *et al.*, 2010, Mutreja *et al.*, 2011). However, an isolate from Sweden was found to contain the latter. Recently, genomic analysis of *V. cholerae* O1 showed close relatedness between isolates from Latin America and Angola, but the CTXΦ array was different (Mutreja *et al.*, 2011).

Table 5.4. Sites of nucleotide polymorphisms in CTX prophages

Strain	Country of Origin	Year of Isolation	Gene	*rstR*	*rstA*														*rstB*								*zot-ctxA*	*ctxB*
			Position	→	27	162	183	258	315	345	516	540	579	609	774	927	933	942	77 - 79	90	96	108	192	288	291	363	(TTTTGAT)	
N16961	Bangladesh	1975	CTX^{ET}	ET	C	C	C	G	T	G	G	A	T	T	C	T	C	G	GTA	A	T	G	A	A	C	A	4	*B3*
O395	India	1965	CTX^{CL}	CL	T	T	A	C	.	T	A	G	C	C	T	*	*	*	–	T	C	*	*	G	T	*	7	*B1*
CIRS101	Bangladesh	2002	CTX^{HYB}	ET	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	3	*B1*
HCO1	Haiti	2010	CTX^{HYB}	ET	*	*	*	*	*	*	*	*	*	*	*	*	*	*	–	*	*	*	*	*	*	*	5	*B1*
95412	Mexico	1997	CTX^{CL}	CL	T	T	A	C	*	T	A	G	C	C	T	*	*	*	–	T	C	*	*	G	T	*	6	*B1*
CP1030	Mexico	2008	CTX^{ET}	ET	*	*	*	*	*	*	*	*	*	*	*	C	T	T	–	T	C	A	*	*	*	*	4	*B3*
CP 1032	Mexico	1991	CTX^{HYB}	ET & CL	*	*	*	*	C	*	*	*	*	*	*	C	T	T	–	T	C	A	*	*	*	*	4	*B1*
CP 1033	Mexico	2000	CTX^{HYB}	ET & CL	*	*	*	*	*	*	*	*	*	*	*	C	T	T	–	T	C	A	G	*	*	G	4	*B1*

*, identical; -, deletion, ET, El Tor, CL, Classical, HYB, hybrid

As shown in **Table 5.4**, the *rstA* and *rstB* gene sequence of *V. cholerae* 95412 classical is identical to that of the reference *V. cholerae* O395 classical, whereas variation was observed in *V. cholerae* CTX^{+} El Tor isolates. *V. cholerae* CP1030, CP1032, and CP1033 contained three unique base substitutions in the *rstA* gene at 927 (T→C), 933 (C→T), and 942 (G→T), compared to *V. cholerae* N16961, CIRS101, and a recent Haitian isolate, HCO1. In addition, CP1032 had a base substitution at 315 (T→C) in the *rstA* gene. Interestingly, all point mutations are synonymous for RstA. DNA sequence analysis of CP1030, CP1032, and CP1033 at the *rstB* gene showed a GTA deletion at position 77 to 79 and polymorphism at 90 (A→T), 96 (T→C), 108 (G→A), and 192 (A→G), unlike *V. cholerae* El Tor strains except GTA deletion, which had been reported in Haitian isolates (Hasan *et al.*, 2012).

Virulence gene expression in *V. cholerae* is regulated by ToxR, a transcriptional regulator that binds with the promoter region (in between *zot* and *ctxA*) located upstream of *ctxAB*. The heptamer repeat sequences (TTTTGAT)) directly influence the affinity of ToxR binding and promote binding of ToxR, which is followed by activation of the *ctxAB* promoter (Pfau & Taylor, 1996). As shown in **Table 5.4,** *V. cholerae* CP1030, CP1032, and CP1033. contain four copies of the heptamer repeat, like El Tor, atypical El Tor, and hybrid variants from Asia and Africa. However, they differ from the Haitian isolates in having five repeats (Hasan *et al.*, 2012, Ghosh *et al.*, 2014). *V. cholerae* 95412 classical isolate contains six copies of the heptamer repeat, unlike the classical *V. cholerae* reference strain O395 which possesses seven copies of the repeat (**Table 5.4**).

5.2.4.2. *Vibrio* Pathogenicity Island 1 and 2

Vibrio pathogenicity island-1 (VPI-1) encodes the toxin coregulated pilus (TCP) that promotes colonization of intestinal mucosal epithelium, is involved in biofilm formation, and serves as the receptor for the lysogenic CTXΦ (Boyd *et al.*, 2000). Five of the six *V. cholerae* O1 isolates from Mexico contained VPI-1, but CP1035 lacked this gene cluster. As shown in **Figure 5.5**, *V. cholerae* CP1030, CP1032, and CP1033 possess VPI-1 of the seventh pandemic *V. cholerae* El Tor, whereas the genetic organization of VPI-1of

CP1037 is homologous to that of *V. cholerae* 95412 (classical) despite having a genomic island, GI-47, in the upstream region**.** Interestingly, the *tcpA* gene encoding the major pilin subunit (TcpA) of CP1037 is different from the classical and El Tor *tcpA*. The TCP region showed highest level of sequence polymorphism in VPI-1, with *tcpA* having the most divergence (Tay *et al.*, 2008).

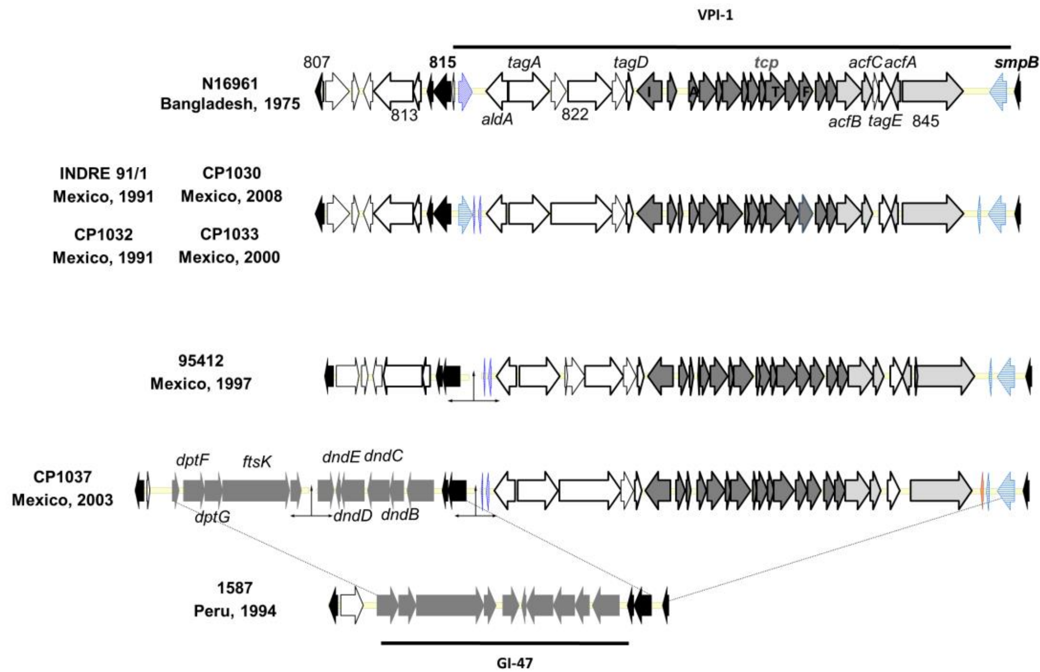

Figure 5.5. *Vibrio* pathogenicity island -1 (VPI-1) of *V. cholerae* O1 strains isolated in Mexico and reference El Tor strain N16961. Mexican CTX$^-$ *V. cholerae* O1 strain CP1037 contains GI-47 in the upstream region of VPI-1.

Previous studies reported TcpA had significant differences in the epitope or antigenic structure when classical and El Tor biotype strains were compared (Jonson *et al.*, 1992). Four of the *V. cholerae* O1 isolates, CP1030, CP1032, CP1033, and 95412, contain the complete VPI-2, whereas the other isolates lack VPI-2. VPI-2 comprises several genes including those encoding sialidase, the type I restriction modification system, and mu-like prophage protein genes.

5.2.4.3. *Vibrio* Seventh Pandemic Islands

The seventh pandemic islands I and II in *V. cholerae* are characteristically found in El Tor strains and they serve as a distinguishing marker from classical strains (Dziejman *et al.*, 2002). However, a variant of the VSP-II gene cluster has also been detected in *V. cholerae* non-O1/O139 strains and in *Vibrio mimicus* (Dziejman *et al.*, 2005, Taviani *et al.*, 2010). *V. cholerae* El Tor strains, CP1032 and CP1033, from Mexico contained all of the ORFs in VSP-I and II, whereas the CTX$^-$ isolates, CP1035 and CP1037, lack VSP-I and II, as does the classical strain 95412.

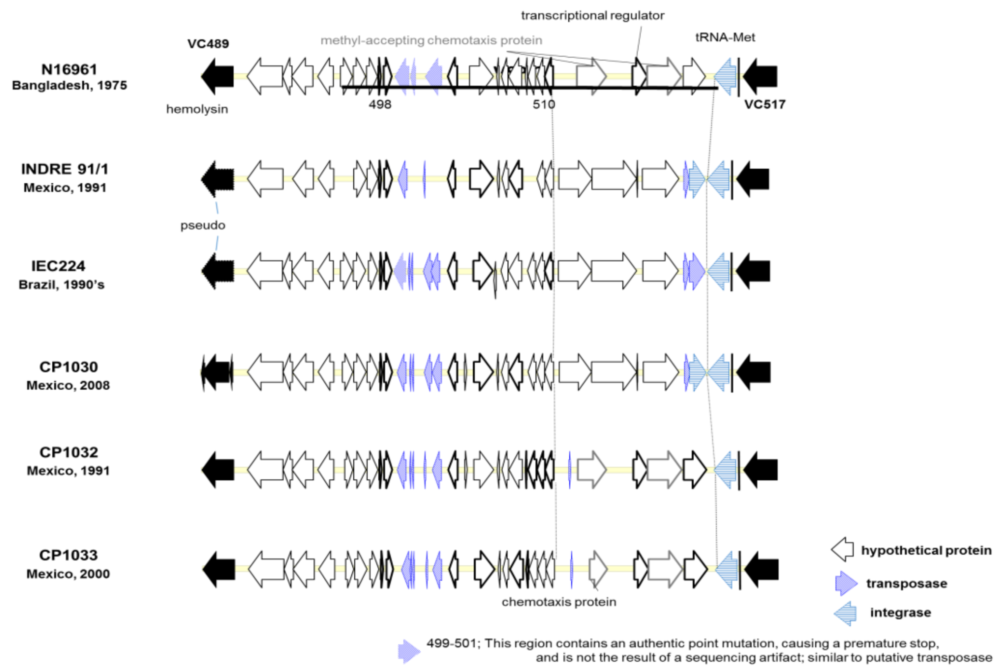

Figure 5.6. *Vibrio* seventh pandemic island -II (VSP-II) of *V. cholerae* O1 strains isolated in Mexico and Brazil.

CP1030 possesses a variant VSP-II with an insertion between VC0510 and VC0516 (**Figure 5.6**), commonly referred to as the West African and South American (WASA) insertion (Mutreja *et al.*, 2011). An identical VSP-II gene cluster has been reported in *V. cholerae* isolated in Peru and Angola (Nusrin *et al.*, 2009, Valia *et al.*, 2013). Conversely, the VSP-II gene cluster in contemporary *V. cholerae* isolates from Asia and Haiti has a 14.4 kb deletion that spans the ORF VC0495 - VC0512 (CIRS101 type VSP-II) (Taviani

et al., 2010, Chin *et al.*, 2011, Hasan *et al.*, 2012). Distribution of the variant VSP-II among the *V. cholerae* isolates suggests this island contains hot spots highly prone to genetic rearrangement by recombination (Taviani *et al.*, 2010).

5.2.4.4. Genomic Islands and Integrative Conjugative Elements (ICE)

V. cholerae O1 isolates from Mexico contain diverse genomic islands (GIs) that differ among the El Tor, classical, and CTX$^-$ strains. *V. cholerae* El Tor isolates, CP1030, CP1032, and CP1033, uniformly contained GI-1 to GI-10 and GI-85. *V. cholerae* CP1033 (Hasan *et al.*, 2012), on the other hand, contains GI-15 in the large chromosome, which encodes the putative integrase found in the Mozambique variant of *V. cholerae* (B-33) and also in hybrid isolates CP1067 from Bangladesh, that had been isolated in 1991. Moreover, *V. cholerae* CP1030 contains the WASA1 genomic island, which has been reported previously in West African and South American strains (Mutreja *et al.*, 2011). *V. cholerae* classical 95412 has GIs typical of the reference classical strain O395, along with GI-11 and GI-21 in the small chromosome. GI-11 encodes the kappa prophage, whereas the function of GI-21 (~34 kb) has not yet been identified. *V. cholerae* CP1035 contains genomic islands that are similar to those of *V. cholerae* non-O1/O139 and differ from classical and El Tor strains. CP1035 contains several previously described genomic islands, including GI-125 and GI-126 encoding a type I restriction modification system and integrase. Interestingly, CP1037 carry GI-36 that has been detected previously in *V. cholerae* non-O1/O139 TM11079-80 and Amazonia, isolated in Brazil. CP1037 also possesses GI-47 in the upstream region of VPI-1 as previously observed in Peruvian *V. cholerae* isolated in 1994 (**Figure 5.5**) and a unique genomic island, GI-112 encoding *umu*CD and nucleotidyltransferase (Chun *et al.*, 2009, Hasan *et al.*, 2012).

The integrating and conjugative elements (ICEs) are self-transmissible mobile genetic elements in bacteria that confer resistance to various antibiotics. SXT is a ~100 Kb ICE originally discovered in *V. cholerae* O139 (Waldor *et al.*, 1996). Since emergence of *V. cholerae* O139 on the Indian subcontinent in 1992, the SXT/R391 ICE has been reported to be present in most clinical *V. cholerae* O1 or O139 strains isolated in

Asia and Africa (Mutreja *et al.*, 2011). *V. cholerae* isolates carrying the SXT/R391 ICE are resistant to streptomycin, chloramphenicol, sulfamethoxazole, and trimethoprim (Waldor *et al.*, 1996). Results of recent phylogenetic analysis suggest *V. cholerae* O1 acquired SXT/R391 ICE sometimes between 1978 and 1984, before its discovery in *V. cholerae* O139, and it is hypothesized that it provides selective advantage to *V. cholerae* O1, allowing it to be globally disseminated (Mutreja *et al.*, 2011). In the present study, all of the Mexican isolates lacked the SXT/R391 ICE. Moreover, the genome sequences of Latin American isolates [INDRE 91/1 (Mexico); CP1044, CP1046, and CP1047 (Peru); and IEC224, RC144, and 116059 (Brazil)] are devoid of SXT/R391 ICE. This observation was confirmed by PCR, i.e., none of the Mexican isolates amplified DNA fragments for primers targeting the SXT integrase gene (*intSXT*) (Thungapathra *et al.*, 2002). Lack of SXT/R391 ICE in epidemic strains isolated in Latin America in the 1990s has been reported (Mutreja *et al.*, 2011). Absence of SXT/R391 ICE among *V. cholerae* isolates has also been reported in a recent cholera outbreak in the Philippines (Klinzing *et al.*, 2015). Antibiotic susceptibility analyses of the five Mexican isolates, CP1030, CP1032, CP1035, CP1037, and 95412, revealed all were sensitive to penicillin (P), ampicillin (AMP), streptomycin (S), chloramphenicol (C), trimethoprim-sulfamethoxazole (SXT), tetracycline (TE), kanamycin (K), erythromycin (E), nalidixic acid (NA), and ciprofloxacin (CIP). *V. cholerae* CP1033 shows resistance only to streptomycin (S). However, *V. cholerae* El Tor strains isolated in 1977 in Africa were resistant to multiple drugs, including tetracycline (Mhalu *et al.*, 1979) and classical strains from Bangladesh isolated during 1982-1989 were resistant to ampicillin (AMP), furazolidone (FR), and trimethoprim-sulfamethoxazole (SXT) (Alam *et al.*, 2012).

5.2.4.5. Lipopolysaccharide Coding Region

The lipopolysaccharide (LPS) of *V. cholerae* is comprised of three main regions: lipid A; the core oligosaccharide (OS); and the O antigen. *V. cholerae* synthesizes the core OS and O antigen using the *wav* and *wb** gene clusters, respectively (Nesper *et al.*, 2002). The *wav* gene cluster (VC0223 -240) of the Mexican isolates is similar to that of *V. cholerae* N16961, except for CP1035, which is different in seven of the ORFs (**Figure**

5.7). *V. cholerae* CP1035 has a *wav* gene cluster homologous to *V. cholerae* TM11079-80, an environmental strain isolated in Brazil in 1980 (**Figure 5.7**). Interestingly, both strains are phenotypically typical El Tor, but lack two major virulence-associated genomic islands, namely the lysogenic CTXϕ encoding CtxAB, and *Vibrio* pathogenicity island VPI-1 that contains the genes for biosynthesis of toxin coregulated pilin (TCP).

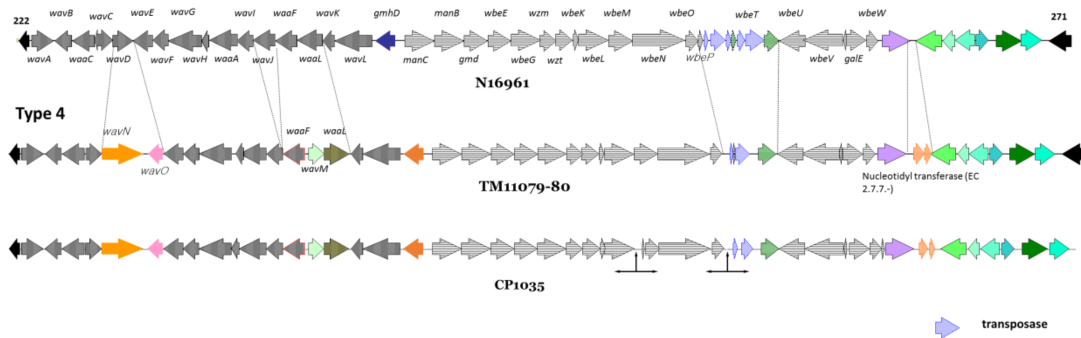

Figure 5.7. 'O' antigen biosynthetic genes of *V. cholerae* O1 strains CP1035, TM11079, and N16961. The *wav* and *wb** gene clusters of CP1035 are homologous to those of TM11079 and different from reference El Tor N16961.

5.2.4.6. Phylogenetics of the Mexican isolates

Phylogeny of the *V. cholerae* strains isolated in Mexico was determined by constructing a genome relatedness neighbor-joining tree using homologous alignment of 905 orthologous protein-coding genes (~897,461 bp) of 124 *V. cholerae* genomes (**Figure 5.8**), which placed El Tor, classical, and non-toxigenic *V. cholerae* isolated from Mexico into distinct phylogenetic clades. CP1035, a CTX$^-$ isolate, was placed into a basal clade with other non-toxigenic non-O1/O139 isolates from Haiti and O1 isolates from Brazil and Amazonia. The other CTX$^-$ isolate, *V. cholerae* CP1037, were phylogenetically placed in an independent node ancestral to all of the sixth and seventh pandemic isolates. The presence of ancestral isolate in the Latin American region is indicative of greater phylogenetic diversity and succession of indigenous *V. cholerae* populations in that

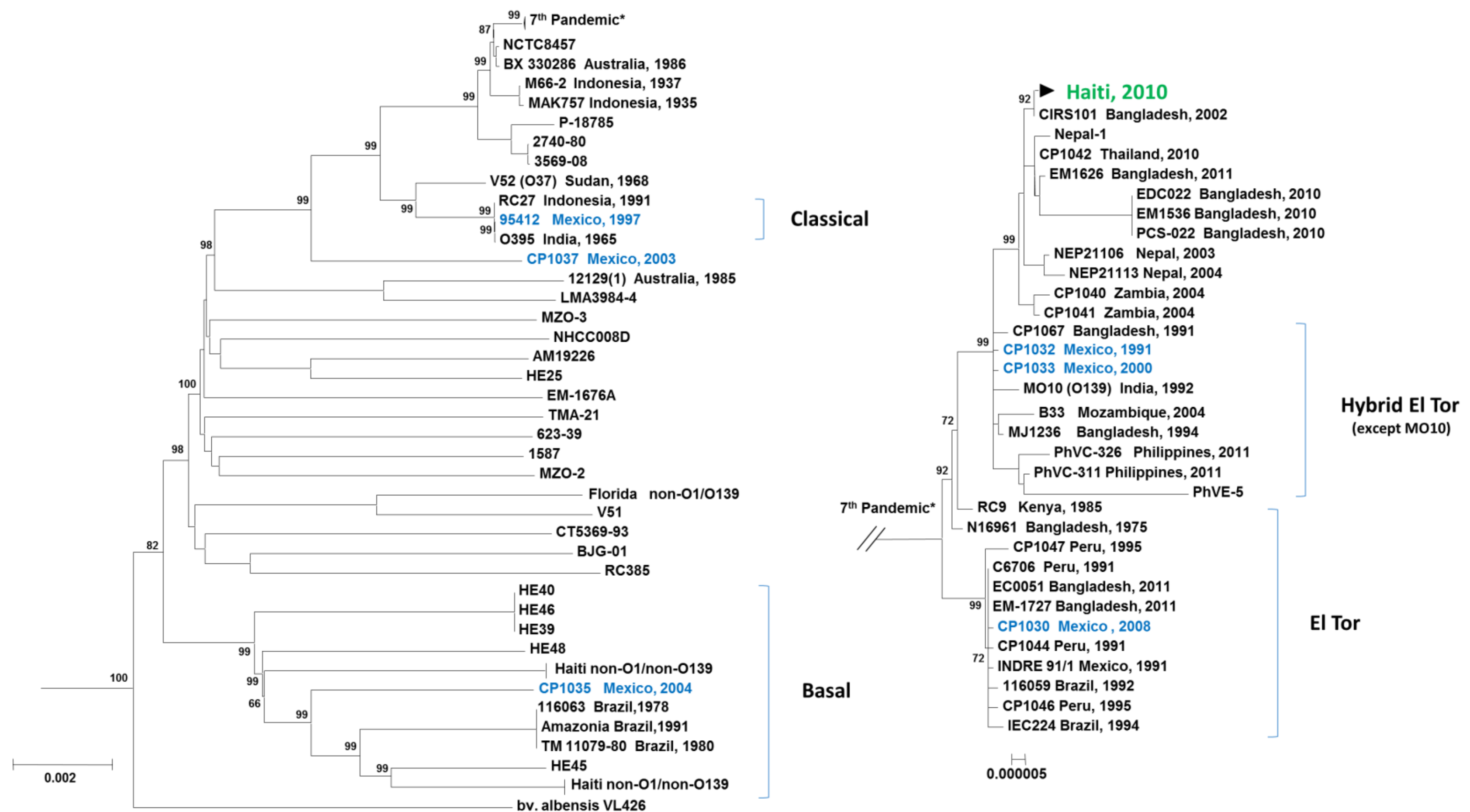

Figure 5.8. Neighbor-joining trees showing phylogenetic relationships of 124 *V. cholerae* genomes based on 905 orthologs of protein-coding genes (~897,461 bp). The two *V. cholerae* non-O1/O139 strains (HE09 and HE16) isolated from surface water during the 2010 cholera epidemic in Haiti were used as an outgroup of the tree and bootstrap values are percentages of 1,000 replications. Haitian node consists of *V. cholerae* O1 isolates from Haiti isolated in 2010.

ecosystem. The classical biotype isolate of *V. cholerae* 95412, isolated from Mexico in 1997, was placed into a monophyletic clade with other sixth pandemic reference *V. cholerae* strain O395, and RC27. Classical biotype strains are considered to have been outcompeted by seventh pandemic *V. cholerae* El Tor strains in the 1980s and have not been isolated in Asia and Africa after 1990 (Alam *et al.*, 2012). In contrast, *V. cholerae* classical strains have been isolated in Mexico until 1997, even though *V. cholerae* El Tor strains were dominant at the beginning of the Latin American epidemic and during the years following, indicates Mexican ecosystem as a reservoir of the classical biotype of *V. cholerae* (Alam *et al.*, 2010).

V. cholerae CP1032 and CP1033 isolated in Mexico were placed into paraphyletic hybrid El Tor clade along with Mozambique and Matlab variants of *V. cholerae* El Tor, namely B-33 and MJ-1236, together with *V. cholerae* O139 isolate MO10 (**Figure 5.8**). These isolates also showed close relatedness with 1991 hybrid *V. cholerae* El Tor strains CP1067 isolated from patients in Bangladesh. Phylogenetic analysis of *V. cholerae* hybrid strains isolated in Mexico clearly shows a separation from contemporary *V. cholerae* El Tor and atypical El Tor strains from Asia, Africa, and Haiti. The Matlab variant strains, isolated in 1994 in Bangladesh, were the first to have been reported in the literature as “hybrid”, showing classical biotype specific traits in an El Tor genetic background (Nair *et al.*, 2002). A decade later, genetically similar hybrid variants were isolated in Mozambique during the 2004 cholera outbreak (Ansaruzzaman *et al.*, 2004, Chun *et al.*, 2009). Isolation of *V. cholerae* CP1032 in 1991 in Mexico suggests hybrid El Tor *V. cholerae* was present at the same time in two different continents, i.e., Asia and America. *V. cholerae* CP1030 also belongs to the seventh pandemic clade. However, it clustered tightly into the monophyletic El Tor clade with *V. cholerae* strains isolated in Mexico, Peru, and Brazil during the Latin American epidemics of the 1990s, but distant from recent isolates from Bangladesh, India, Nepal, and Thailand. Furthermore, Zambia, Zimbabwe, and Haiti isolates are also separated from CP1030 suggesting a conserved *V. cholerae* O1 clone that carry a truncated CTXΦ instead of RS1 in the upstream region of CTXΦ, circulating in the Mexican ecosystem during 2004 - 2008. Since 2000, variants of

V. cholerae O1 El Tor have prevailed in cholera endemic regions of Asia and Africa, with *V. cholerae* prototype El Tor strains rarely isolated (Rashed *et al.*, 2013).

Overall, results of our study and previous studies (Alam *et al.*, 2010, Alam *et al.*, 2014) show the existence of genetically diverse *V. cholerae* O1 in Mexico during 1991-2008. Considering the global epidemiology of cholera, and somewhat puzzling succession of *V. cholerae* O1 in Mexico, the observations reported here clearly contradict the hypothesis of global transmission of cholera from Africa to Latin America, as proposed by other investigators (Mutreja *et al.*, 2011). During the 1990s Latin American epidemic, Peru was the first to have been affected by cholera and a clonal CTX^+ *V. cholerae* O1 El Tor was concluded to be the etiological agent, which was present in Peruvian coastal waters for at least several months prior to the onset of the cholera epidemic (Seas *et al.*, 2000). Furthermore, CTX^- *V. cholerae* O1 El Tor had been isolated from two patients with diarrhea in Lima, Peru in 1988 (Batchelor & Wignall, 1988, Seas *et al.*, 2000) and from sewage in Brazil in1982 (Levine *et al.*, 1982, Seas *et al.*, 2000). An environmental stimulus for *V. cholerae*, i.e., increased air and water temperatures and in phytoplankton abundance, associated with the El Nino phenomenon, or changes in salinity and/or nutrient concentrations, may have triggered the existing CTX^+ *V. cholerae* O1 El Tor to upsurge rapidly during the 1990s in Peru (Seas *et al.*, 2000). Molecular typing and phylogenetic analysis of 1990s Latin American *V. cholerae* O1 isolates has been done by several investigators, none of whom found significant correlation with isolates of Asia and Latin America (Lam *et al.*, 2010, Mutreja *et al.*, 2011). Therefore, the hypothesis of cholera transmission from Asia to Latin America via ballast water lacks evidence, hence is no longer considered valid. Mexican hybrid isolates did show a close relatedness to Bangladesh isolates at a parallel time, but metadata are not sufficient for concluding a direction of transmission from either form Asia to Latin America or vice versa.

5.2.5. CONCLUSION

The results of this study provide important information on the molecular epidemiology of cholera in Mexico. Phylogenetic analysis of the isolates shows cholera in Mexico during

1991 to 2008 was caused by genetically diverse *V. cholerae* O1 belonging to distinct phylogenetic clades. In addition, the antibiotic susceptibility patterns and CTX arrangements of the Mexican isolates strongly contradict the notion of a single source transmission of *V. cholerae* O1 into Mexico from African countries. The lack of SXT/R391 ICE in the Latin American *V. cholerae* isolates is an interesting observation which requires further study, since concurrent Asian and African isolates generally possess SXT/R391 ICE. Genetic events occurring in *V. cholerae* O1 strains associated with endemic cholera in Mexico are different from those of Asian and African countries (Alam *et al.*, 2010, Alam *et al.*, 2014). Results provided in this study are concordant with those of previous investigations (Alam *et al.*, 2010, Alam *et al.*, 2012, Alam *et al.*, 2014), in fact they suggest an association of indigenous populations of *V. cholerae* playing a significant role in the dynamics of cholera in Mexico.

CHAPTER 6:

METAGENOMIC ANALYSIS OF ENVIRONMENTAL WATER SAMPLES COLLECTED FROM THE COASTAL AREA OF BANGLADESH

6.1. ABSTRACT

Metagenomic analysis of environmental samples using next generation sequencing (NGS) is a modern approach to determine microbial community composition and the population dynamics of infectious agents. In this study, deep shotgun sequencing of 20 pond water samples, collected during 2013 and 2014 from Mathbaria, Bangladesh generated an average of 37,803,873 ± 14149013 reads per samples for metgenomic characterization of microbes. Our analysis revealed the presence of 549 different species of bacteria, 50 viruses, 11 fungi, and 9 protista. Higher diversity was observed for the bacterial community (Simpson diversity index 0.7) compared to the virus, fungi, and Protista communities. All of the bacterial species were concluded to belong to 228 genera, 119 families, 30 classes, and 14 phyla. The predominant phylum detected comprised the gram-negative Proteobacteria, representing 74% of the total bacteria, with Firmicutes, Actinobacteria, and Bacteroidetes accounting for 18%, 4%, and 3%, respectively. Among the thirty identified Classes, Gammaproteobacteria comprised 70%, Bacilli 17%, Actinobacteria 4%, and Alphaproteobacteria 3% of the total bacteria. Five *Vibrio* species were detected, of which *V. cholerae* and *V. mimicus* were predominant during both the epidemic and non-epidemic seasons of cholera. Interestingly, CTXΦ and other *Vibrio* phages were consistently detected throughout the study period, 2013 -2014. The results showed temporal changes in total abundance of microorganisms in natural pond water. Resistome analysis identified resistance-marker genes for nine groups of antimicrobials: aminoglycoside; beta-lactumase; macrolide; chloramphenicol; tetracycline; trimethoprim; quinolones; and sulphonamides. Since metagenomic analysis is a culture independent method, results of this study demonstrate strength in providing better resolution and precision in identifying infectious pathogens, e.g., *Vibrio* spp., compared to conventional culture and PCR methods.

6.2. INTRODUCTION

Microorganisms constitute a major part of life on this planet and directly involve maintaining the Earth ecosystem. The microbial community composition of the ecological habitat varies, depending on biotic and abiotic variables. All components of the ecosystem synchronize at best when microbes interact to contribute to ecological

niche. In general, microbes are responsive to environmental change and adapt rapidly, either by altering physiological properties or metabolic activities, whenever the situation demands (Colwell *et al.*, 1985, Oliver, 2005). Therefore, the fittest microbial species will survive in an ecosystem, while unfit species are lost from the ecological niche (Koskella & Vos, 2015, Lenski, 2017). Completion among microbes for available niches is recognized as a driving force, bringing spatiotemporal change to a microbial community. However, the dynamics of a microbial community will depend heavily on environmental factors and seasonal changes (Stevenson *et al.*, 2014, Westrich *et al.*, 2016). Despite the fact that pathogenic microbes constitute a very small portion (< 1%) of the total microbial population, community composition of pathogens must be elucidated to understand infectious disease dynamics and determine preventive measures that should be implemented (Nature Publishing Group, 2011). Metagenomic analysis of environmental samples offers a huge advantage in revealing complex relationships between pathogenic and non-pathogenic microbes in ecosystems, especially in relation to emergence, survival, and transmission of pathogens.

The genus *Vibrio* comprises more than 100 species, all of which are autochthonous to the aquatic environment and ubiquitously present in estuarine and marine waters all over the world (Colwell & Spira, 1992, Westrich *et al.*, 2018). Although, only a few species of *Vibrio* cause human disease, most species of the genus are actively involved in nutrient and chemical cycling in the marine environment (Thompson *et al.*, 2004, Westrich *et al.*, 2018). Pathogenic Vibrios including toxigenic *V. cholerae*, can exist in the aquatic environment in a viable but non-culturable (VBNC) state for long term during the nutrient limiting conditions; however, *V. cholerae* remains an active pathogen and capable of causing an epidemics (Colwell *et al.*, 1985, Oliver, 2005). Global warming, climate change, and El Nino have a remarkable influence on the abundance of *Vibrio* populations, particularly those species causing human diseases throughout the world (Pascual *et al.*, 2000, Lipp *et al.*, 2002, Hashizume *et al.*, 2013, Moore *et al.*, 2017, Cavicchioli *et al.*, 2019). Therefore, any spatiotemporal shift in microbial community and infectious disease dynamics caused by climate change should

be monitored to address potential adverse effects of global warming on human health (Colwell, 1996, Cavicchioli *et al.*, 2019).

Metagenomic analysis involves extraction of DNA directly from environmental samples, e.g., soil and water, followed by shotgun sequencing and bioinformatics analysis to detect, identify, and characterize all components of the microbial community (Stevenson *et al.*, 2014, Toyama *et al.*, 2016, Westrich *et al.*, 2018). Over the past 10 years, the scope of metagenomic analysis has advanced significantly, especially in next generation sequencing (NGS) technology, DNA extraction methods, and development of robust genomic databases. Reduced cost of NGS, rapid turnaround time, and higher precision of shotgun sequencing for microbial community profiling of different niches and geographical locations is highly encouraging for microbial ecology (Quince *et al.*, 2017). Metagenomic analyses make it possible to reveal the entire microbial community composition of water and soil samples collected in remote areas like the amazon forest. It can also be used to detect antibiotic resistance genes, i.e., the resistome, of a microbial community (Bai *et al.*, 2019, Chen *et al.*, 2019).

A seasonal pattern of cholera has been observed in the coastal areas of Bangladesh. Despite the many long term epidemiological studies that have been done in the rural coastal areas of Bangladesh to understand the ecology, occurrence and transmission of cholera (Longini *et al.*, 2002, Huq *et al.*, 2005, Emch *et al.*, 2008), very little is known about population dynamics of *V. cholerae* in the estuarine waters of those coastal areas. Therefore, this study was conducted to monitor the microbial community during pre- and post-outbreak seasons of cholera in Mathbaria, Bangladesh. Here, we report results of shotgun sequencing and metagenomic analysis of twenty surface water samples collected from a pond in Mathbaria, Bangladesh during 2013 and 2014.

6.3. MATERIALS AND METHODS

6.3.1. Sample Collection

Water and plankton samples were collected between October, 2010, and December, 2014, from six ponds and one canal in Mathbaria, Bangladesh. Sampling was done each

week during the epidemic period (February - May), while a monthly sampling strategy was carried out during the rest of the year (June - January). All samples were pre-processed in the research laboratory at icddr,b, Dhaka, Bangladesh. At each sampling site, 100 liters of water were filtered successively through 64-µm and 20-µm mesh nylon nets (Millipore Corp., Bedford, MA) (placing the 64-µm net sequentially in front of the 20-µm nylon net, with each having a collection bucket at the base), and 50 ml of the concentrate was collected initially as a crude measure of zooplankton and phytoplankton, respectively. Also, during the process of filtration, 500 ml of filtrate from the 20-µm mesh net was collected as representative water to be analyzed for free-living bacteria (Alam *et al.*, 2006, Huq *et al.*, 2012). In the laboratory, 500 ml filtrates were filtered through 0.22 µm polycarbonate filters to trap all free living bacteria. Both the 64-µm and 20-µm plankton samples were further concentrated in the laboratory using respective plankton nets (specially devised, netted plastic beakers) to a final volume of 5.0 ml each. The plankton samples were crushed using a glass homogenizer (Elberbach Corp., Ann. Arbor, Mich.) to release attached bacteria. Finally, 3.0 ml of crushed plankton and the 0.22 µm polycarbonate filters were transferred into cryovials for storage at −80°C for metagenomic DNA extraction. All samples were shipped with dry ice to the University of Maryland, College Park for metagenomic analysis.

6.3.2. DNA Extraction

A total of 120 samples from Mathbaria, Bangladesh, collected from two ponds and one canal during 2013 and 2014, were primarily selected for metagenomic analysis, based on isolation frequency of *V. cholerae*. Homogenized plankton samples (1.5 ml) were centrifuged at 13,000 X g for 10 minutes to from pellet prior extraction. Polycarbonate filters (0.22 µm) were pre-processed by chopping them into small pieces using a sterile surgical razor. DNA was extracted using MOBIO Power Soil kit, following the manufacturer's instruction. A Metagenomic DNA pool was constructed for each sample by combining DNA from plankton and the respective filters, representing both free living and plankton-attached bacteria. NanoDrop was used to measure the concentration of DNA. A large proportion of metagenomic DNA failed QC because of low DNA yield (<50 ng of total DNA). Therefore, the metagenomic study plan was revised. Finally, 20

metagenomic DNA from one pond (Pond 10) collected between January, 2013 and May, 2014, were sequenced. The purified metagenomic DNA samples were stored at −80°C until library preparation for NGS.

6.3.3. Library preparation and Sequencing

Ca. 50 ng of metagenomic DNA was used to construct DNA libraries using Nextera Library Prep Kit (Illumina, San Diego, CA). All DNA libraries were quantified and the Agilent 2100 Bioanalyzer was used to confirm quality prior to sequencing. DNA libraries were sequenced using a 100 bp paired-end run on the Illumina HiSeq1000 Instrument (Illumina, San Diego, CA). Since the microbial community of environmental samples is known to be complex, very-deep shotgun sequencing was employed for 20 surface water samples.

6.3.4. Metagenomic Analysis using Bioinformatics

Unassembled whole genome shotgun sequencing reads were directly analyzed using the CosmosID, Inc. bioinformatics software package (CosmosID Inc., Rockville, MD, USA), as described previously (Hasan *et al.*, 2014, Hourigan *et al.*, 2018, Roy *et al.*, 2018), to achieve microbial identification to species, subspecies, and/or strain level. The CosmosID cloud app. also screened the entire resistome and virulence genes of the microbial community using raw sequences. Briefly, the system utilizes a high performance data-mining k-mer algorithm, and highly curated dynamic comparator databases (GeneBook®, CosmosID, Inc., Rockville, MD, USA) that rapidly disambiguates millions of short reads into the discrete genomes or those genes engendering the particular sequences. The GeneBook® databases are composed of over 150,000 microbial genomes and gene sequences, representing over 10000 bacterial, 5000 viral, 250 protista, and 1500 fungal species, as well as over 5500 antibiotic resistant and virulence associated genes. The web portal is hosted at AWS cloud and can be accessed at https://app.cosmosid.com/login.

6.4. RESULTS AND DISCUSSION

Next Generation Sequencing of twenty surface water samples generated an average of 37,803,873 ± 14149013 reads per sample for downstream bioinformatics analysis.

CosmosID metagenomics software (CosmosID cloud app.) identified bacteria, virus, fungi, and protist species in the surface water community. Notably, the bacterial species accounted for 80-94 % of the total microbial species in the pond water samples from Mathbaia, Bangladesh. The abundance of bacterial the population was substantially higher than the virus, fungi, and protist populations in the surface water samples (**Figure 6.1**).

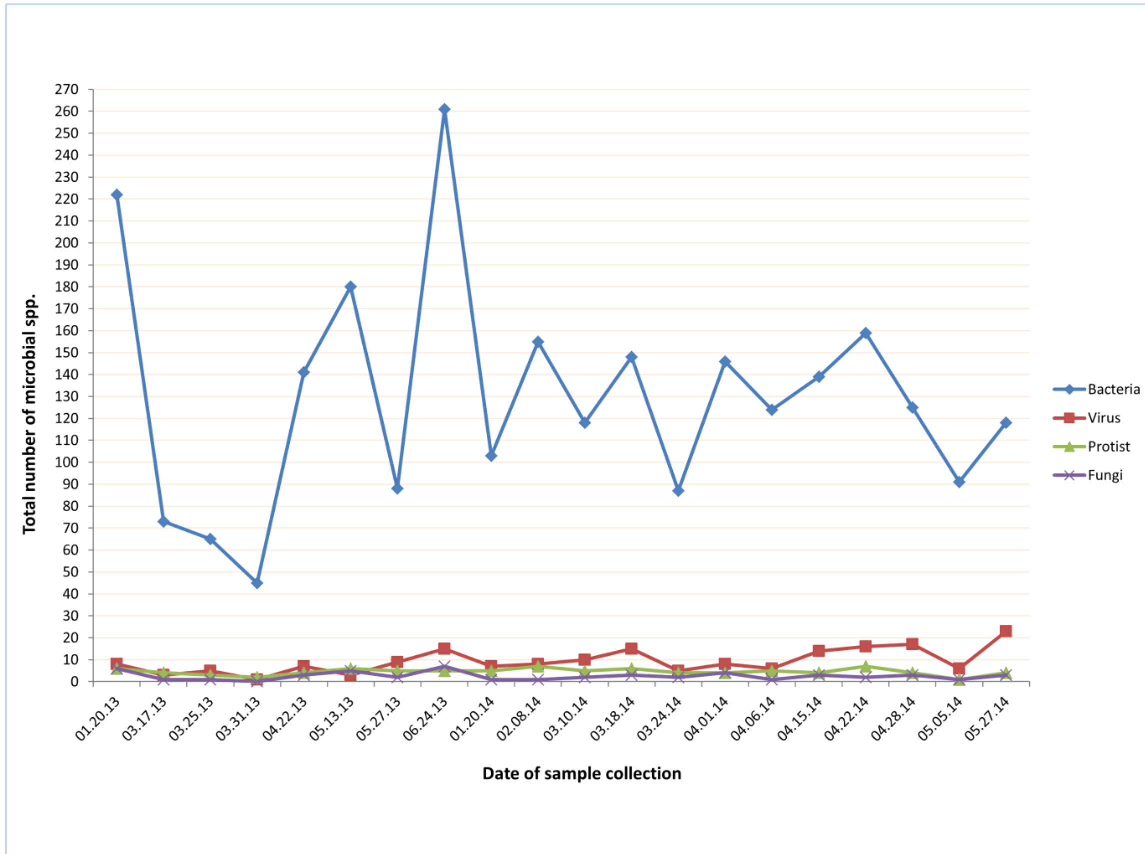

Figure 6.1. Total bacteria, virus, fungi, and protista identified using metagenomic analysis of pond water samples collected in Mathbaria, Bangladesh (2013 – 2014).

Although, the bacterial population constitutes major proportion of the total microbial community, the number of bacterial species fluctuated over time (**Figure 6.1**). The observed difference in species abundance most likely is due to seasonal variables, e.g., pH, water temperature, salinity, and total dissolved oxygen (**Figure 6.2**). As shown in **Figure 6.2**, temporal changes in environmental parameters, particularly water

temperature and dissolved oxygen concentration were observed during 2013 and 2014 for pond 10.

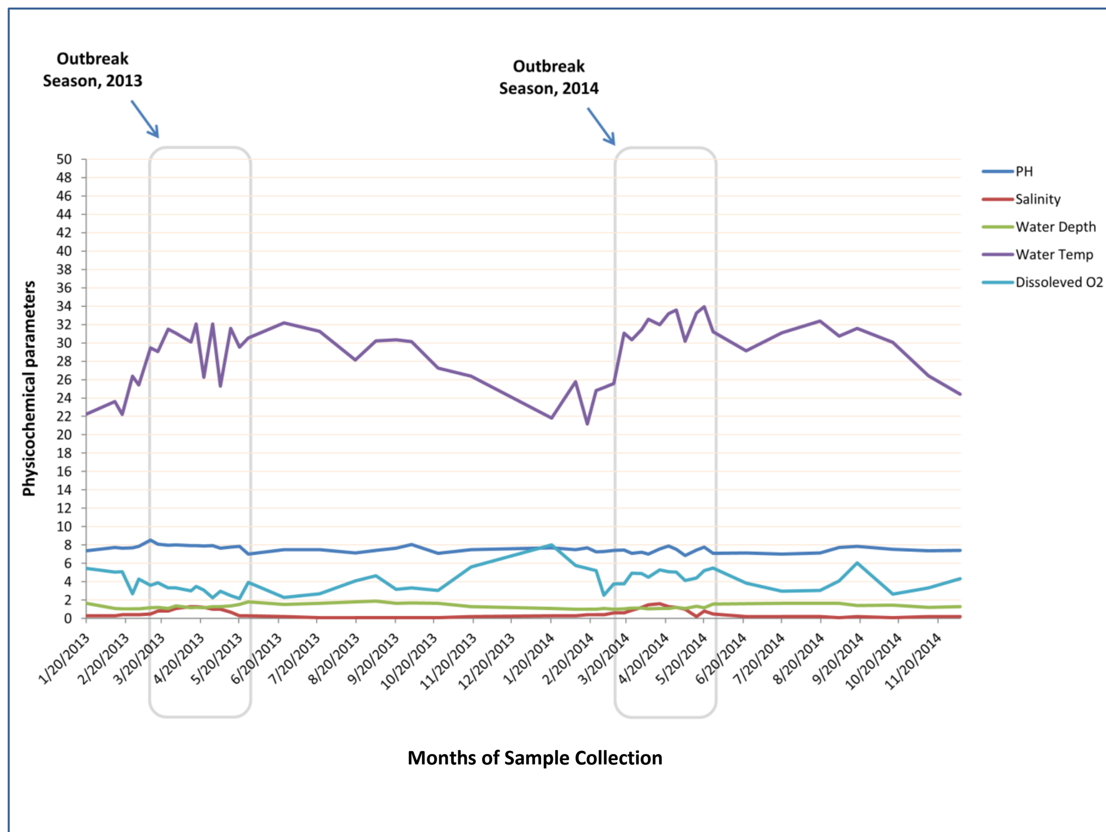

Figure 6.2. Seasonal changes in environmental parameters for pond 10 during cholera outbreak seasons in Mathbaria, Bangladesh (2013 and 2014).

Several previous studies reported environmental variables had a marked influence on microbial community composition and species abundance of the aquatic ecosystem (Mlejnkova & Sovova, 2010, Kaevska *et al.*, 2016). As shown in **Figure 6.1**, the largest number (n=288) of microbial species, i.e., bacteria, virus, fungi, and protista together, was observed in June, 2013 and the lowest (n=48) in March, 2013.

6.4.1. Metagenomic profiling of the Bacterial community

NGS analysis identified a total of 549 bacterial species (Simpson diversity index 0.7), suggesting extensive diversity in the bacterial community of surface water. All of the bacterial species belong to 228 Genera, 119 Families, 30 Classes, and 14 Phyla (**Figure**

6.3). The predominant phylum was gram-negative Proteobacteria representing 74% of the total bacterial diversity; while Firmicutes, Actinobacteria, and Bacteroidetes accounted 18%, 4%, and 3%, respectively. Among 30 identified Classes of bacteria, Gammaproteobacteria comprised 70%, Bacilli 17%, Actinobacteria 4%, and Alphaproteobacteria 3% of the total bacteria. Based on the relative abundance, Pseudomonadales and Bacillales were the dominant order of Gammaproteobacteria and Bacilli, respectively. The class Gamaproteobacteria includes pathogenic bacteria common to the aquatic ecosystem, e.g., *Vibrio* spp.

Figure 6.3. Krona visualization showing total bacterial diversity based on relative abundance.

Bacterial species analysis revealed that *Acinetobacter* spp. and *Bacillus* spp. were most abundant; notably, the relative abundance was higher for *Acinetobacter lwoffii*, *A. junii, A. johnsonii*, and *Bacillus vietnamensis* (**Figure 6.3**). Five *Vibro* spp. were detected

in the pond 10, i.e., *V. cholerae*, *V. mimicus*, *V. fluvialis*, *V. vulnificus*, and *V. albensis* (**Figure 6.4**). Remarkably, *V. cholerae* and *V. mimicus* were consistently detected in all surface water samples, while the other species were infrequently identified, without showing a seasonal pattern. Presence of *V. cholerae* and *V. mimicus* in the surface water of coastal areas is expected, because tidal flows from the Bay of Bengal carry these bacteria into the neighboring villages.

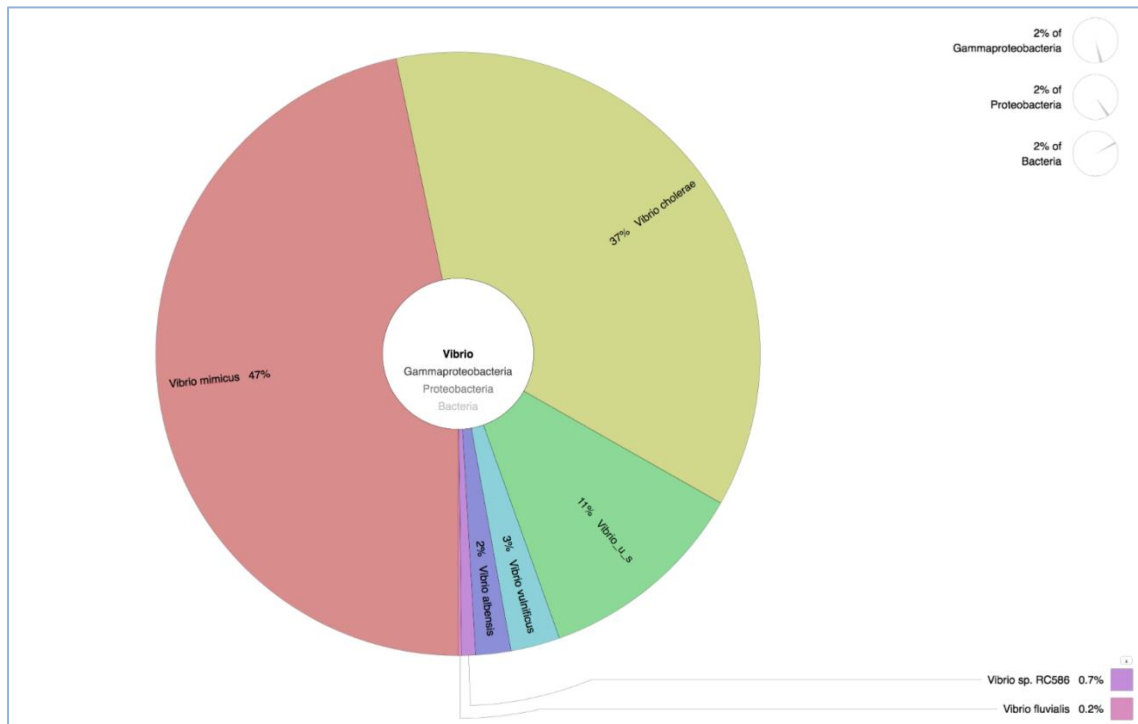

Figure 6.4. Relative abundance of *Vibrio* species in Pond 10 during the study period. *V. mimicus* and *V. cholerae* accounted for 47% and 37% of the total *Vibrio* population, respectively.

Thus, environmental variables are directly connected to the dynamics of cholera in the coastal region of Bangladesh (Akanda *et al.*, 2011, Rahman *et al.*, 2018). In addition to *Vibrio* species, other human pathogens were also detected, including *Acinetobacter baumannii*, *Aeromonas hydrophila*, *A. veronii*, *Bacillus cereus*, *Enterococcus faecalis*, *Shigella dysenteriae*, *Salmonella enterica*, and opportunistic *Escherichia coli*. However, the distribution frequencies of these pathogens were low compared to the entire microbial community (**Figure 6.3**).

The bacterial diversity and species abundance of pond water represent a complex composition of the bacterial community resident in the aquatic environment. The higher abundance of *Acinetobacter* spp. in surface water has been documented previously in the Ganges delta region, Haiti, and other parts of the world (Newton *et al.*, 2011, Roy *et al.*, 2018, Zhang *et al.*, 2019). However, the major difference here is the richness of Gammaproteobacteria, unique to the Ganges delta region of the Bay of Bengal.

6.4.2. Identification of Viruses

Metagenomic analysis identified a total of 50 viral species belonging to four families, namely the Podoviridae (57%), Inoviridae (19%), Myoviridae (13%), and Siphoviridae (8%) and two unidentified families. The viruses detected were mostly bacteriophages, e.g., phages specific for species of *Enterobacteria*, *Vibrio*, *Salmonella*, *Acinetobacter, Clostridium, Escherichia,* and *Klebsiella.* As shown in **Figure 6.5,** Vibrio virus VP2 and VP5, and CTX phage were highly dominant among the viral species. Lytic phages are a predator of host bacteria, therefore, directly regulates the abundance of bacterial populations in the environment (Chibani-Chennoufi *et al.*, 2004). The phage-bacteria dynamics and their interactions in an ecosystem are of great interest because of potential correlation with infectious disease dynamics, including cholera (Chibani-Chennoufi *et al.*, 2004, Merikanto *et al.*, 2018). It has been hypothesized that the environmental abundance of *Vibrio* phages has inverse correlation with seasonal epidemics of cholera (Faruque *et al.*, 2005). This hypothesis was favored by the findings of several microcosm studies and some other small scale studies done in remote settings (Nelson *et al.*, 2008, Burks *et al.*, 2017). However, the question remains whether phages are the only factor or one of several factors, along with multiple environmental variables, that regulates seasonal cholera. Nonetheless, all predictive models for cholera consider lytic *Vibrio* phages as one of the major components that can have some influence on seasonal cholera in endemic countries (Nelson *et al.*, 2008, Nelson *et al.*, 2009, Yen & Camilli, 2017). What is more useful, however, is the confirmation that presence of specific bacterial phages selective for the bacterial species detected and identified in the metagenomic analysis results.

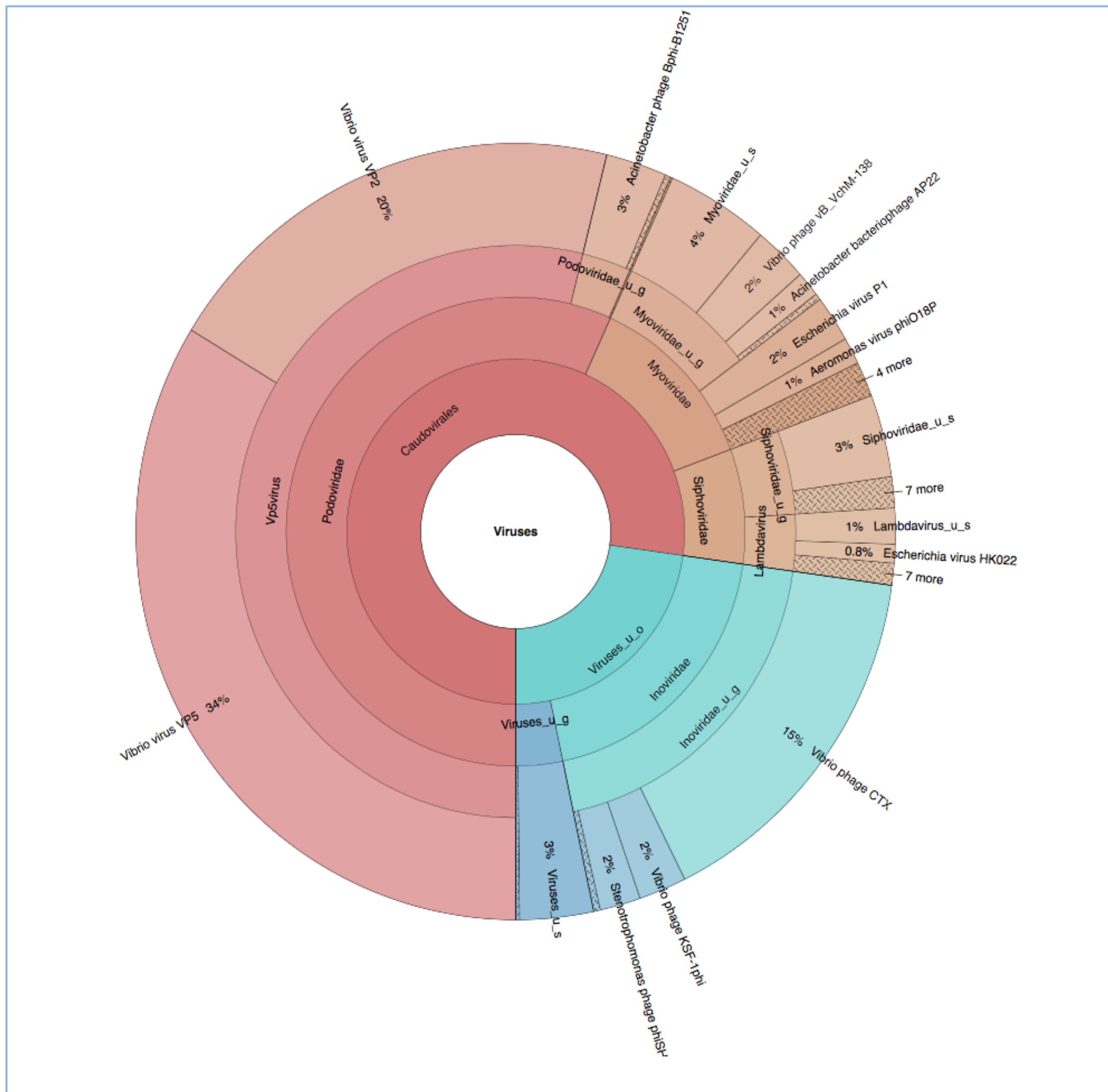

Figure 6.5. Krona visualization showing total viral diversity based on relative abundance.

6.4.3. Identification of fungi and Protista

Fungi and Protista accounted for a small portion of the pond water microbial community compared to bacteria and viruses (**Figure 6.6**). A total of 11 fungi and 9 Protista were identified by the CosmosID cloud based metagenomic analysis. Among the fungi, relative abundance was higher for *Onygenales* sp., *Saccharomyces cerevisiae*, *Amauroascus niger*, and *Enterocytozoon bieneusi*. Moreover, Thalassiosira, Neobalantidium, and Nannochloropsis were the dominant Protista detected and identified in the pond water.

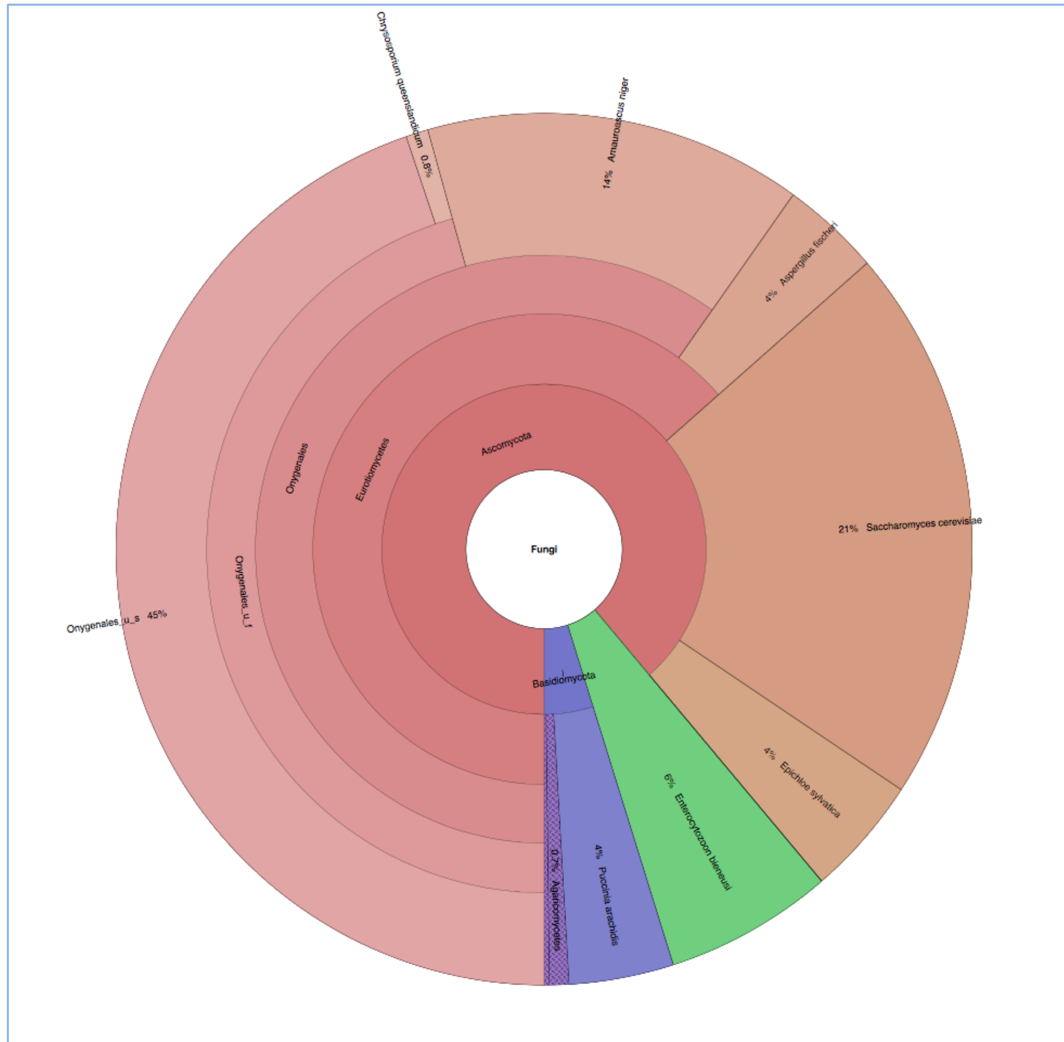

Figure 6.6. Krona visualization showing diversity of Fungi based on relative abundance.

Although, *Entamoeba histolytica* and *Giardia lamblia* are frequently associated with diarrhea in nutrition deprived children and adults in South Asia (Haque *et al.*, 2003, Haque *et al.*, 2006, Mukherjee *et al.*, 2009), pathogenic Fungi and Protista were not identified in the pond water microbial community of Mathbaria, Bangladesh, requiring further study.

6.4.4. Resistome profiling

Antibiotic resistance of infectious pathogens is a major public health concern globally. Emergence of multidrug resistant "superbugs" and their dissemination in Asia, Africa, and the Americas have been reported over the past decade. Thus, screening for antibiotic resistance profiles and associated marker genes is essential for tracking horizontal and vertical transmission of antibiotic resistant determinants. Moreover, monitoring antibiotic resistance of infectious pathogens is important for clinicians in selecting effective drugs for treatment of disease. Metagenomic analysis allowed profiling the antibiotic resistance gene pool of the entire microbial community of the pond water samples.

Table 6.1. Antibiotic resistant marker genes of the microbial community

Antibiotics	Resistance Genes
Aminoglycoside	*aac6, aadA5, aadA10, aadA16, aph(3') Ia, aph(3') Via, aph(6) Id*
Beta-lactumase	*amp, cphA, blaACT-28, blaB-2, blaCARB, blaCME-1, blaDHA, blaGOB-1, blaL-1, bla-MIR, bla-MOX, blaNDM, blaOCH-5, blaOKP, blaOXA-181, blaOXA-211, blaOXA-212, blaOXA-252, blaOXA-266, blaOXA-277, blaOXA-278, blaOXA-280, blaOXA-281, blaOXA-283, blaOXA-285, blaOXA-333, blaOXA-334, blaOXA-335, blaOXA-360, blaOXA-361, blaOXA-362, blaOXA-363, blaOXA-373, blaOXA-421, blaPER-7*
Trimethoprim	*dfrA1, dfrA12, dfrA26, dfrG*
Phenicol	*catA2, catB5, catB9, fexA*
Macrolide	*mefA, mphE*
Quinolone	*qnrVC4, qnrS*
Tetracycline	*tetR, terX*
Chloramphenicol	*cmx*
Sulphonamide	*sul2*
MDR-Efflux-complex	*msrE, robA, smeD, smeE, smeF, abeM, abeS, acrA, acrB, adeF, adeG, adeH, adeI, adeJ, adeK, marA*

In this study, antibiotic resistant markers conferring resistance to nine groups of antibiotics were identified in the microbial community of pond water. The antibiotics

included aminoglycoside, beta-lactumase, macrolide, chloramphenicol, tetracycline, trimethoprim, quinolone, and sulphonamides. As shown in **Table 6.1**, presence of these resistance genes in the environment is alarming, because co-existence of multiple pathogenic and non-pathogenic bacterial species in the pond may acquire these resistance genes and become resistant to clinically import drugs (Laxminarayan *et al.*, 2013, Bai *et al.*, 2019). In Bangladesh, several antibiotic resistance markers against macrolide, quinolone, beta-lactumase and other drugs have been detected in pathogens (Ahmed *et al.*, 2019). Abuse of antibiotics in the public health sector, poultry farming, and veterinary applications are a leading cause of the spread of antibiotic resistance. Two factors: (1) selection pressure due to antibiotic pollution; (2) antibiotic resistant markers in the environment create conditions suitable for bacteria acquiring resistance against antimicrobials. Emergence of antibiotic resistant "superbugs" becomes highly probable in this environment (Davies & Davies, 2010, Laxminarayan *et al.*, 2013, Ahmed *et al.*, 2019, Danner *et al.*, 2019).

6.4.5.. Comparison of conventional culture, PCR, and metagenomic detection, identification, and characterization of pathogenic *Vibrio* species

Conventional culture techniques together with polymerase chain reaction (PCR) targeting species-specific genes, are established methods for isolation and identification of *Vibrio* species (Huq *et al.*, 2012). In the present study, we compared the efficiency of NGS based metagenomics analysis, conventional culture, and PCR for detecting *V. cholerae* and other *Vibrio* species (**Table 6.2**). Traditional culture methods and serological tests consistently detected *V. cholerae* non-O1/O139 in 20 water samples, while *V. cholerae* O1 was identified only in four samples collected during April and May, 2014. Similarly, PCR analysis detected toxigenic *V. cholerae* O1 in five samples collected between March and May, 2014. Notably, one culture-positive sample appeared to be negative for *V. cholerae* O1 by PCR. Culture methods failed to identify *V. cholerae* O1 in two samples that were PCR- positive. As shown in **Table 6.2**, *V. cholerae* O1 was detected by PCR and culture only during the cholera epidemic season, when cholera patients discharge huge amount of infectious bacteria in the environment.

Table 6.2. Identification of *Vibrio* spp. by conventional culture, PCR, and NGS metagenomic analysis.

Location	Sampling site	Year	Month	Date of sampling	Culture Results *V. cholerae*		PCR Results VC O1	NGS Based Metagenomic Identification						
					VC non-O1/O139	VCO1		CTXΦ	VCO1	VC non-O1/O139	*V. mimicus*	*V. vulnificus*	*V. fluvialis*	*V. albensis*
Mathbaria	Pond 10	2013	January	01.20.13	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2013	March	03.17.13	Pos (+)	-	-	-	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2013	March	03.25.13	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2013	March	03.31.13	Pos (+)	-	-	Pos (+)	-	Pos (+)	Pos (+)	-	-	-
		2013	April	04.22.13	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2013	May	05.13.13	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	Pos (+)
		2013	May	05.27.13	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-
		2013	June	06.24.13	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)
		2014	January	01.20.14	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-
		2014	February	02.08.14	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-
		2014	March	03.10.14	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2014	March	03.18.14	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-
		2014	March	03.24.14	Pos (+)	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2014	April	04.01.14	Pos (+)	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	Pos (+)
		2014	April	04.06.14	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2014	April	04.15.14	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	Pos (+)
		2014	April	04.22.14	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-
		2014	April	04.28.14	Pos (+)	-	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2014	May	05.05.14	Pos (+)	Pos (+)	-	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-	-
		2014	May	05.27.14	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	Pos (+)	-	-

VC, *V. cholerae*, +, positive, -, negative

During the non-epidemic period, isolation of *V. cholerae* O1 from environmental samples has proven to be difficult because of their persistence in the VBNC state (Colwell & Spira, 1992, Alam *et al.*, 2006, Huq *et al.*, 2012). However, metagenomic analysis is culture independent and highly reliable for detecting *V. cholerae* O1 and other *Vibrio* species in environmental samples. Except two pond water samples, NGS analysis identified *V. cholerae* O1, non-O1/O139, *V. mimicus*, and CTXΦ regardless of whether the samples were collected in epidemic and non-epidemic periods (**Table 6.2**). In addition, *V. vulnificus*, *V. albensis*, and *V. fluvialis* were also detected at relatively low percentages in the pond water samples. Results of our study clearly demonstrate that NGS, together with metagenomic analysis, is more accurate in detecting and identifying *V. cholerae* and other *Vibrio* species for environmental surveillance than either conventional culture or PCR.

6.5. CONCLUSION

High throughput NGS and bioinformatics provide accurate microbiome characterization directly from environmental samples. Results of this study have shown that the microbial community of surface water is highly diverse, comprising a large number of bacteria, and low but significant number of viruses, fungi, and protista. The presence of multiple *Vibrio* species, including *V. cholerae*, *V. vulnificus*, and *V. mimicus* has epidemiological significance considering their pathogenic attributes. Thus, the population dynamics of *Vibrio* species, their seasonal abundance, and resistome profiling is possible for all cholera endemic countries and globally. However, two factors limit environmental metagenomics and those are the low DNA yield of water samples and lack of a robust curated database for eukaryotes, i.e., fungi and Protista. Recent advancement in NGS and low cost of high throughput sequencing will eventually resolve this limitation. It is crucial, however, that metagenomic analysis be used in environmental monitoring and surveillance of pathogens and to elucidate the community dynamics of microbes in the natural environment.

CHAPTER 7:

CONCLUSIONS

Endemic cholera remains a major public health threat in many countries of Asia, Africa, and Latin America. While the pathogenic agent continues to evolve in endemic foci, *V. cholerae* O1 is disseminated within non-endemic parts of the world, e.g., Haiti, and Yemen and subsequently has caused massive epidemics in recently years. The genome of *V. cholerae* is highly flexible in accepting novel genetic elements by lateral gene transfer events. Several factors, including chitin induced competence and lysogenic bacteriophages, facilitate transfer of genomic islands, mobile genetic elements, virulence factors, and pathogenicity islands into the genome. Incorporation of these novel elements is an ongoing process and numerous novel recombinations have been detected recently with the advent of genomic sequencing. Therefore, these genetic changes give rise to atypical El Tor and new variants, which eventually increase the total diversity of toxigenic *V. cholerae* population. However, these new variants continue to exist as epidemic clones if they are favored by natural selection. Otherwise, the ultimate fate is disappearance from its ecological niche. Whole genome shotgun sequencing has been employed in this study to understand the microevolution and intercontinental genomic epidemiology of *V. cholerae*, accomplished by analyzing a number of strains isolated in two different cholera endemic countries, Bangladesh and Mexico. Next generation sequencing based metagenomic analyses was performed to reveal the dynamics of *V. cholerae* in cholera endemic coastal area of Bangladesh.

In this study, we demonstrated that the antibiotic susceptibility pattern of *V. cholerae* has been changing rapidly due selection pressure of antibiotics, both in the environment and human host, and horizontal transfer of antibiotic resistant marker genes and gene cassettes. The growing resistance to ciprofloxacin and azithromycin is alarming because fluoroquinolone and macrolide drugs are still considered to be effective against several enteropathogens. Although, the *V. cholerae* O1 isolates did not possess macrolide resistance genes, the possibility of acquiring these genes horizontally in the aquatic environment cannot be ruled out. Historically, prophylaxis of cholera has faced a consistent challenge because of development of resistance against effective drugs. Therefore, periodic change or alteration in the choice of drug in the treatment of cholera is suggested. Results of our study highlight the importance of continuous monitoring of antibiotic susceptibility and resistant markers in the epidemiology of cholera.

Molecular typing, i.e., multilocus variable number tandem repeat analysis (MLVA) showed genotypic variability among cholera outbreak strains in two distinct areas of Bangladesh. Interestingly, multiple genetic lineages were identified among toxigenic *V. cholerae* O1 strains causing seasonal cholera in the coastal area. However, the outbreak in Chhatak was attributed to clonal expansion, commonly known as 'founder flush'. Thus, our data supports 'person-to-person' or accelerated mode of transmission of cholera in a densely populated area during an outbreak. We also identified overlapping genotypes between clinical and environmental isolates, suggesting transmission of *V. cholerae* O1 from environment to human in endemic areas. During the cholera outbreak, 'person-to-person' or accelerated mode of transmission is highly likely within members of a household. Therefore, proper hygienic practice needs to be adopted in the household to reduce the risk of cholera transmission when an index case occurs in a household.

Genomic analyses of *V. cholerae* isolated in Bangladesh provided insight into the genomic epidemiology of this microorganism, notably in cholera endemic areas of the Ganges delta where the cholera bacterium continues to evolve in its ecological niche. We identified multiple toxigenic *V. cholerae* O1 clones including the Matlab variant, an ancient clone thriving in aquatic reservoir of Bangladesh for more than two decades without causing major outbreaks. Similarly, *V. cholerae* O1 carrying *ctxB3*, WASA genomic island, and WASA variant of VSP-II also has an epidemiological significance, considering the fact that a similar type of strain was associated with the Latin American epidemic in 1991. Since emergence and re-emergence is a phenomenon characteristic of *V. cholerae*, these observations demonstrate that environmental persistence over a prolonged period is possible in endemic areas even though other clones (atypical El Tor) are favored by natural selection. Therefore, transient disappearance of a clonal population does not represent a complete selection sweep. Moreover, it has been a challenge for environmental surveillance to represent the total genomic diversity of environmental *V. cholerae* populations because of the limitation of conventional culture to isolate VBNC cells. Genomic analyses of *V. cholerae* O1 and non-O1/O139 strains revealed a substantial number of GIs and MGEs, suggesting genomic rearrangement plays an important role in the dynamic of pathogenic *V. cholerae*. We also observed selective amplification of a single clone and its descendants during seasonal epidemics in subsequent years

although the environmental population showed a greater genomic diversity. However, the mystery remains whether the environment or the human gut selects the epidemic clone. Both clinical and environmental isolates possess virulence genes, genomic islands, and mobile genetic elements important for ecological fitness of the bacterium in its native aquatic environment.

Results of our study remarkably demonstrated microevolution of *V. cholerae* O1 in the environment of Mexico. The genomic epidemiology of Mexico is substantially different from that of cholera endemic countries of Asia and Africa. Identification of truncated pro-phage in the upstream region of CTX, existence of TCP variant, and predominance of *V. cholerae* with *ctxB3* make the case very strong regarding evolution within the local niche. However, major cholera epidemic did not occur in Latin American countries for more than a century prior to 1991 when a cholera outbreak struck Peru, followed by dissemination of cholera in neighboring countries, including Mexico. Once cholera became endemic in Mexico, atypical El Tor, El Tor, and classical *V. cholerae* O1 coexisted there during the 1990s, a unique scenario in the global epidemiology of cholera. Given the fact that very few clones may have been introduced to Mexico from Asia and Africa, core genome phylogeny also identified genetically different clones that might have evolved over time in the aquatic environment of Mexico. Antibiotic susceptibility patterns are another interesting feature of *V. cholerae* O1 isolated in Mexico because all of those strains uniformly lacked SXT/R391 ICE. However, contemporary atypical El Tor strains of Asia and Africa possess SXT/R391 ICE and mostly show multidrug resistance. In summary, genetically diverse *V. cholerae* O1 strains were associated with the cholera dynamics of Mexico during 1991 and 2008.

In the present study, NGS based metagenomic analyses successfully characterized the diverse and complex microbial community of surface water samples collected in the coastal area of Bangladesh. The microbial community of Mathbaria pond water includes bacteria, viruses, fungi, and Protista that are commonly found as primary members of both fresh and coastal waters. This was expected because tidal waves force frequent intrusion of coastal water into the fresh water system of Mathbaria, Bangladesh. Numerous bacterial species were detected and identified along with multiple *Vibrio* species pathogenic for human and marine animals.

Notably, we observed spatiotemporal changes in the microbial community known to be regulated by environmental factors, i.e., temperature and salinity. Another interesting observation was detection and identification of *Vibrio* phages during both epidemic and inter-epidemic periods. Toxigenic and non-toxigenic *V. cholerae*, other *Vibrio* species, symbiotic microbes, and predatory phages were found to exhibit a complex synchronization of the microbial community with environmental factors influencing the abundance of *V. cholerae* related to seasonal epidemics.

Future studies are recommended to monitor antibiotic resistance and search for novel antibiotic resistant markers, not only in *V. cholerae* O1, but also in the non-O1/O139 serogroup which can serve as a reservoir for antibiotic resistance genes. In addition, a major question remains to be fully resolved: does clonal selection of toxigenic *V. cholerae* occur in the environment or human gut or both'. Natural selection is a complicated process, microcosms and animal models may be able to meet this challenge and provide a fundamental understanding of cholera dynamics.

REFERENCES

Ahmed, I., Rabbi, M.B., and Sultana, S. (2019). Antibiotic resistance in Bangladesh: A systematic review. *Int J Infect Dis* 80**,** 54-61.

Akanda, A.S., Jutla, A., and Islam, S. (2009). Dual peak cholera transmission in Bengal Delta: A hydroclimatological explanation. *GEOPHYSICAL RESEARCH LETTERS* 36.

Akanda, A.S., Jutla, A.S., Alam, M., Magny, G.C., Siddique, A.K., Sack, R.B., Huq, A., Colwell, R., and Islam, S. (2011). Hydroclimatic influences on seasonal and spatial cholera transmission cycles: Implications for public health intervention in the Bengal Delta. *WATER RESOURCES RESEARCH* 47**,** doi:10.1029/2010WR009914.

Akanda, A.S., Jutla, A.S., Gute, D.M., Sack, R.B., Alam, M., Huq, A., Colwell, R.R., and Islam, S. (2013). Population vulnerability to biannual cholera outbreaks and associated macro-scale drivers in the Bengal Delta. *Am J Trop Med Hyg* 89**,** 950-959.

Alam, M., Hasan, N.A., Sadique, A., Bhuiyan, N.A., Ahmed, K.U., Nusrin, S., Nair, G.B., Siddique, A.K., Sack, R.B., Sack, D.A., Huq, A., and Colwell, R.R. (2006a). Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl Environ Microbiol* 72**,** 4096-4104.

Alam, M., Islam, A., Bhuiyan, N.A., Rahim, N., Hossain, M.A., Khan, G.Y., Ahmed, D., Watanabe, H., Izumiya, H., Faruque, A.S., Akanda, A.S., Islam, S., Sack, R.B., Huq, A., Colwell, R.R., and Cravioto, A. (2011). Clonal transmission, dual peak, and off-season cholera in Bangladesh. *Infection Ecology and Epidemiology* 1**,** 7273 - DOI: 7210.3402/iee.v7271i7270.7273.

Alam, M., Islam, M.T., Rashed, S.M., Johura, F.T., Bhuiyan, N.A., Delgado, G., Morales, R., Mendez, J.L., Navarro, A., Watanabe, H., Hasan, N.A., Colwell, R.R., and Cravioto, A. (2012). *Vibrio cholerae* classical biotype strains reveal distinct signatures in Mexico. *J Clin Microbiol* 50**,** 2212-2216.

Alam, M., Nusrin, S., Islam, A., Bhuiyan, N.A., Rahim, N., Delgado, G., Morales, R., Mendez, J.L., Navarro, A., Gil, A.I., Watanabe, H., Morita, M., Nair, G.B., and Cravioto, A. (2010). Cholera between 1991 and 1997 in Mexico was associated with infection by classical, El Tor, and El Tor variants of *Vibrio cholerae*. *J Clin Microbiol* 48**,** 3666-3674.

Alam, M., Rashed, S.M., Mannan, S.B., Islam, T., Lizarraga-Partida, M.L., Delgado, G., Morales-Espinosa, R., Mendez, J.L., Navarro, A., Watanabe, H., Ohnishi, M., Hasan, N.A., Huq, A., Sack, R.B., Colwell, R.R., and Cravioto, A. (2014). Occurrence in Mexico, 1998-2008, of *Vibrio cholerae* CTX+ El Tor carrying an additional truncated CTX prophage. *Proc Natl Acad Sci U S A* 111**,** 9917-9922.

Alam, M., Sultana, M., Nair, G.B., Sack, R.B., Sack, D.A., Siddique, A.K., Ali, A., Huq, A., and Colwell, R.R. (2006b). Toxigenic *Vibrio cholerae* in the aquatic environment of Mathbaria, Bangladesh. *Appl Environ Microbiol* 72**,** 2849-2855.

Alam, M., Sultana, M., Nair, G.B., Siddique, A.K., Hasan, N.A., Sack, R.B., Sack, D.A., Ahmed, K.U., Sadique, A., Watanabe, H., Grim, C.J., Huq, A., and Colwell, R.R. (2007). Viable but nonculturable *Vibrio cholerae* O1 in biofilms in the aquatic environment and their role in cholera transmission. *Proc Natl Acad Sci U S A* 104**,** 17801-17806.

Albert, M.J., and Nair, G.B. (2005). *Vibrio cholerae* O139 Bengal-10 years on. *Reviews in Medical Microbiology* 16**,** 135-143.

Albert, M.J., Siddique, A.K., Islam, M.S., Faruque, A.S., Ansaruzzaman, M., Faruque, S.M., and Sack, R.B. (1993). Large outbreak of clinical cholera due to *Vibrio cholerae* non-O1 in Bangladesh. *Lancet* 341**,** 704.

Ali, M., Lopez, A.L., You, Y.A., Kim, Y.E., Sah, B., Maskery, B., and Clemens, J. (2012). The global burden of cholera. *Bull World Health Organ* 90**,** 209-218A.

Ali, M., Nelson, A.R., Lopez, A.L., and Sack, D.A. (2015). Updated global burden of cholera in endemic countries. *PLoS Negl Trop Dis* 9**,** e0003832.

Andersson, D.I., and Hughes, D. (2014). Microbiological effects of sublethal levels of antibiotics. *Nat Rev Microbiol* 12**,** 465-478.

Ansaruzzaman, M., Bhuiyan, N.A., Nair, B.G., Sack, D.A., Lucas, M., Deen, J.L., Ampuero, J., and Chaignat, C.L. (2004). Cholera in Mozambique, variant of *Vibrio cholerae*. *Emerg Infect Dis* 10**,** 2057-2059.

Ansaruzzaman, M., Bhuiyan, N.A., Safa, A., Sultana, M., Mcuamule, A., Mondlane, C., Wang, X.Y., Deen, J.L., Von Seidlein, L., Clemens, J.D., Lucas, M., Sack, D.A., and Balakrish Nair, G. (2007). Genetic diversity of El Tor strains of

Vibrio cholerae O1 with hybrid traits isolated from Bangladesh and Mozambique. *Int J Med Microbiol* 297**,** 443-449.

Bagchi, K., Echeverria, P., Arthur, J.D., Sethabutr, O., Serichantalergs, O., and Hoge, C.W. (1993). Epidemic of diarrhea caused by *Vibrio cholerae* non-O1 that produced heat-stable toxin among Khmers in a camp in Thailand. *J Clin Microbiol* 31**,** 1315-1317.

Bai, Y., Ruan, X., Xie, X., and Yan, Z. (2019). Antibiotic resistome profile based on metagenomics in raw surface drinking water source and the influence of environmental factor: A case study in Huaihe River Basin, China. *Environ Pollut* 248**,** 438-447.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., and Pevzner, P.A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19**,** 455-477.

Barua, D. (1972). The global epidemiology of cholera in recent years. *Proc R Soc Med* 65**,** 423-428.

Batchelor, R.A., and Wignall, F.S. (1988). Nontoxigenic O1 *Vibrio cholerae* in Peru: a report of two cases associated with diarrhea. *Diagn Microbiol Infect Dis* 10**,** 135-138.

Beltran, P., Delgado, G., Navarro, A., Trujillo, F., Selander, R.K., and Cravioto, A. (1999). Genetic diversity and population structure of *Vibrio cholerae*. *J Clin Microbiol* 37**,** 581-590.

Bhattacharya, D., Sayi, D.S., Thamizhmani, R., Bhattacharjee, H., Bharadwaj, A.P., Roy, A., and Sugunan, A.P. (2012). Emergence of multidrug-resistant *Vibrio cholerae* O1 biotype El Tor in Port Blair, India. *Am J Trop Med Hyg* 86**,** 1015-1017.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30**,** 2114-2120.

Boyd, E.F., Moyer, K.E., Shi, L., and Waldor, M.K. (2000). Infectious CTXPhi and the vibrio pathogenicity island prophage in *Vibrio mimicus*: evidence for recent horizontal transfer between *V. mimicus* and *V. cholerae*. *Infect Immun* 68**,** 1507-1513.

Burks, D.J., Norris, S., Kauffman, K.M., Joy, A., Arevalo, P., Azad, R.K., and Wildschutte, H. (2017). Environmental vibrios represent a source of antagonistic compounds that inhibit pathogenic *Vibrio cholerae* and *Vibrio parahaemolyticus* strains. *MicrobiologyOpen* 6**,** e00504.

Buse, H.Y., Lu, J., Lu, X., Mou, X., and Ashbolt, N.J. (2014). Microbial diversities (16S and 18S rRNA gene pyrosequencing) and environmental pathogens within drinking water biofilms grown on the common premise plumbing materials unplasticized polyvinylchloride and copper. *FEMS Microbiology Ecology* 88**,** 280-295.

Bwire, G., Munier, A., Ouedraogo, I., Heyerdahl, L., Komakech, H., Kagirita, A., Wood, R., Mhlanga, R., Njanpop-Lafourcade, B., Malimbo, M., Makumbi, I., Wandawa, J., Gessner, B.D., Orach, C.G., and Mengel, M.A. (2017). Epidemiology of cholera outbreaks and socio-economic characteristics of the communities in the fishing villages of Uganda: 2011-2015. *PLoS Negl Trop Dis* 11**,** e0005407.

Camacho, A., Bouhenia, M., Alyusfi, R., Alkohlani, A., Naji, M.a.M., De Radigues, X., Abubakar, A.M., Almoalmi, A., Seguin, C., Sagrado, M.J., Poncin, M., Mcrae, M., Musoke, M., Rakesh, A., Porten, K., Haskew, C., Atkins, K.E., Eggo, R.M., Azman, A.S., Broekhuijsen, M., Saatcioglu, M.A., Pezzoli, L., Quilici, M.L., Al-Mesbahy, A.R., Zagaria, N., and Luquero, F.J. (2018). Cholera epidemic in Yemen, 2016-18: an analysis of surveillance data. *Lancet Glob Health* 6**,** e680-e690.

Cameron, D.N., Khambaty, F.M., Wachsmuth, I.K., Tauxe, R.V., and Barrett, T.J. (1994). Molecular characterization of *Vibrio cholerae* O1 strains by pulsed-field gel electrophoresis. *J Clin Microbiol* 32**,** 1685-1690.

Carraro, N., Rivard, N., Ceccarelli, D., Colwell, R.R., and Burrus, V. (2016). IncA/C Conjugative Plasmids Mobilize a New Family of Multidrug Resistance Islands in Clinical *Vibrio cholerae* Non-O1/Non-O139 Isolates from Haiti. *mBio* 7**,** e00509-00516.

Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., Behrenfeld, M.J., Boetius, A., Boyd, P.W., Classen, A.T., Crowther, T.W., Danovaro, R., Foreman, C.M., Huisman, J., Hutchins, D.A., Jansson, J.K., Karl, D.M., Koskella, B., Mark Welch, D.B., Martiny, J.B.H., Moran, M.A.,

Orphan, V.J., Reay, D.S., Remais, J.V., Rich, V.I., Singh, B.K., Stein, L.Y., Stewart, F.J., Sullivan, M.B., Van Oppen, M.J.H., Weaver, S.C., Webb, E.A., and Webster, N.S. (2019). Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*.

Ceccarelli, D., Chen, A., Hasan, N.A., Rashed, S.M., Huq, A., and Colwell, R.R. (2015). Non-O1/non-O139 *Vibrio cholerae* carrying multiple virulence factors and *V. cholerae* O1 in the Chesapeake Bay, Maryland. *Appl Environ Microbiol* 81**,** 1909-1918.

Chatterjee, S., Ghosh, K., Raychoudhuri, A., Chowdhury, G., Bhattacharya, M.K., Mukhopadhyay, A.K., Ramamurthy, T., Bhattacharya, S.K., Klose, K.E., and Nandy, R.K. (2009). Incidence, virulence factors, and clonality among clinical strains of non-O1, non-O139 *Vibrio cholerae* isolates from hospitalized diarrheal patients in Kolkata, India. *J Clin Microbiol* 47**,** 1087-1095.

Chen, H., Bai, X., Jing, L., Chen, R., and Teng, Y. (2019). Characterization of antibiotic resistance genes in the sediments of an urban river revealed by comparative metagenomics analysis. *Sci Total Environ* 653**,** 1513-1521.

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.L., and Brussow, H. (2004). Phage-host interaction: an ecological perspective. *J Bacteriol* 186**,** 3677-3686.

Chin, C.S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C., Jean-Charles, R.R., Bullard, J., Webster, D.R., Kasarskis, A., Peluso, P., Paxinos, E.E., Yamaichi, Y., Calderwood, S.B., Mekalanos, J.J., Schadt, E.E., and Waldor, M.K. (2011). The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364**,** 33-42.

Choi, S.Y., Lee, J.H., Jeon, Y.S., Lee, H.R., Kim, E.J., Ansaruzzaman, M., Bhuiyan, N.A., Endtz, H.P., Niyogi, S.K., Sarkar, B.L., Nair, G.B., Nguyen, B.M., Hien, N.T., Czerkinsky, C., Clemens, J.D., Chun, J., and Kim, D.W. (2010). Multilocus variable-number tandem repeat analysis of *Vibrio cholerae* O1 El Tor strains harbouring classical toxin B. *J Med Microbiol* 59**,** 763-769.

Choi, S.Y., Rashed, S.M., Hasan, N.A., Alam, M., Islam, T., Sadique, A., Johura, F.T., Eppinger, M., Ravel, J., Huq, A., Cravioto, A., and Colwell, R.R. (2016). Phylogenetic Diversity of *Vibrio cholerae* Associated with Endemic Cholera in Mexico from 1991 to 2008. *MBio* 7.

Chow, K.H., Ng, T.K., Yuen, K.Y., and Yam, W.C. (2001). Detection of RTX toxin gene in *Vibrio cholerae* by PCR. *J Clin Microbiol* 39**,** 2594-2597.

Chowdhury, F., Mather, A.E., Begum, Y.A., Asaduzzaman, M., Baby, N., Sharmin, S., Biswas, R., Uddin, M.I., Larocque, R.C., Harris, J.B., Calderwood, S.B., Ryan, E.T., Clemens, J.D., Thomson, N.R., and Qadri, F. (2015). *Vibrio cholerae* Serogroup O139: Isolation from Cholera Patients and Asymptomatic Household Family Members in Bangladesh between 2013 and 2014. *PLoS Negl Trop Dis* 9**,** e0004183.

Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J., Taviani, E., Jeon, Y.S., Kim, D.W., Brettin, T.S., Bruce, D.C., Challacombe, J.F., Detter, J.C., Han, C.S., Munk, A.C., Chertkov, O., Meincke, L., Saunders, E., Walters, R.A., Huq, A., Nair, G.B., and Colwell, R.R. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 106**,** 15442-15447.

CLSI (2010a). Methods for Antimicrobial Dilution and Disk Susceptibility Testing of Infrequently Isolated or Fastidious Bacteria. . *Approved Guideline 2nd edn, document M45–A2 (ISBN 1–56238–732–4). Wayne, PA: Clinical and Laboratory Standards Institute.*

CLSI (2010b). Performance standards for antimicrobial susceptibility testing; twentieth informational supplement. CLSI document M100-S20. CLSI, Wayne, PA.

Colwell, R., Brayton, P.R., Grimes, D.J., Roszak, D.B., Huq, S.A., and Palmer, L.M. (1985a). Viable but non-culturable *Vibrio cholerae* and related environmental pathogens in the environment: implications for release of genetically engineered microorganisms. *Nature Biotechnology* 3**,** 817 - 820.

Colwell, R., and Spira, W.M. (1992). The ecology of *Vibrio cholerae* and cholera. *in Cholera, eds Barua D., Greenouh W. B. I., editors. (New York: Plenum Press Inc.),* **,** p. 107-127.

Colwell, R.R. (1996). Global climate and infectious disease: the cholera paradigm. *Science* 274**,** 2025-2031.

Colwell, R.R. (2000). Viable but nonculturable bacteria: a survival strategy. *J Infect Chemother* 6**,** 121-125.

Colwell, R.R., Brayton, P.R., Grimes, D.J., Roszak, D.B., Huq, S.A., and Palmer, L.M. (1985b). Viable but Non-Culturable *Vibrio cholerae* and Related Pathogens in the Environment: Implications for Release of Genetically Engineered Microorganisms. *Bio/Technology* 3**,** 817-820.

Dalsgaard, A., Albert, M.J., Taylor, D.N., Shimada, T., Meza, R., Serichantalergs, O., and Echeverria, P. (1995). Characterization of *Vibrio cholerae* non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru. *J Clin Microbiol* 33**,** 2715-2722.

Dalsgaard, A., Forslund, A., Bodhidatta, L., Serichantalergs, O., Pitarangsi, C., Pang, L., Shimada, T., and Echeverria, P. (1999). A high proportion of *Vibrio cholerae* strains isolated from children with diarrhoea in Bangkok, Thailand are multiple antibiotic resistant and belong to heterogenous non-O1, non-O139 O-serotypes. *Epidemiol Infect* 122**,** 217-226.

Dalsgaard, A., Forslund, A., Sandvang, D., Arntzen, L., and Keddy, K. (2001). *Vibrio cholerae* O1 outbreak isolates in Mozambique and South Africa in 1998 are multiple-drug resistant, contain the SXT element and the *aadA2* gene located on class 1 integrons. *J Antimicrob Chemother* 48**,** 827-838.

Danin-Poleg, Y., Cohen, L.A., Gancz, H., Broza, Y.Y., Goldshmidt, H., Malul, E., Valinsky, L., Lerner, L., Broza, M., and Kashi, Y. (2007). *Vibrio cholerae* strain typing and phylogeny study based on simple sequence repeats. *J Clin Microbiol* 45**,** 736-746.

Danner, M.C., Robertson, A., Behrends, V., and Reiss, J. (2019). Antibiotic pollution in surface fresh waters: Occurrence and effects. *Sci Total Environ* 664**,** 793-804.

Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5**,** e11147.

Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* 74**,** 417-433.

Devault, A.M., Golding, G.B., Waglechner, N., Enk, J.M., Kuch, M., Tien, J.H., Shi, M., Fisman, D.N., Dhody, A.N., Forrest, S., Bos, K.I., Earn, D.J., Holmes, E.C., and Poinar, H.N. (2014). Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N Engl J Med* 370**,** 334-340.

Diaz-Quinonez, J.A. (2017). Genome Sequence Analysis of *Vibrio cholerae* clinical isolates from 2013 in Mexico reveals the presence of the strain responsible for the 2010 Haiti outbreak. *Gac Med Mex* 153**,** S91-S101.

Dixit, S.M., Johura, F.T., Manandhar, S., Sadique, A., Rajbhandari, R.M., Mannan, S.B., Rashid, M.U., Islam, S., Karmacharya, D., Watanabe, H., Sack, R.B., Cravioto, A., and Alam, M. (2014). Cholera outbreaks (2012) in three districts of Nepal reveal clonal transmission of multi-drug resistant *Vibrio cholerae* O1. *BMC Infect Dis* 14**,** 392.

Domman, D., Chowdhury, F., Khan, A.I., Dorman, M.J., Mutreja, A., Uddin, M.I., Paul, A., Begum, Y.A., Charles, R.C., Calderwood, S.B., Bhuiyan, T.R., Harris, J.B., Larocque, R.C., Ryan, E.T., Qadri, F., and Thomson, N.R. (2018). Defining endemic cholera at three levels of spatiotemporal resolution within Bangladesh. *Nature Genetics* 50**,** 951-955.

Domman, D., Quilici, M.L., Dorman, M.J., Njamkepo, E., Mutreja, A., Mather, A.E., Delgado, G., Morales-Espinosa, R., Grimont, P.a.D., Lizarraga-Partida, M.L., Bouchier, C., Aanensen, D.M., Kuri-Morales, P., Tarr, C.L., Dougan, G., Parkhill, J., Campos, J., Cravioto, A., Weill, F.X., and Thomson, N.R. (2017). Integrated view of *Vibrio cholerae* in the Americas. *Science* 358**,** 789-793.

Dutilh, B.E., Thompson, C.C., Vicente, A.C., Marin, M.A., Lee, C., Silva, G.G., Schmieder, R., Andrade, B.G., Chimetto, L., Cuevas, D., Garza, D.R., Okeke, I.N., Aboderin, A.O., Spangler, J., Ross, T., Dinsdale, E.A., Thompson, F.L., Harkins, T.T., and Edwards, R.A. (2014). Comparative genomics of 274 *Vibrio cholerae* genomes reveals mobile functions structuring three niche dimensions. *BMC Genomics* 15**,** 654.

Dutta, D., Chowdhury, G., Pazhani, G.P., Guin, S., Dutta, S., Ghosh, S., Rajendran, K., Nandy, R.K., Mukhopadhyay, A.K., Bhattacharya, M.K., Mitra, U., Takeda, Y., Nair, G.B., and Ramamurthy, T. (2013). *Vibrio cholerae* non-O1, non-O139 serogroups and cholera-like diarrhea, Kolkata, India. *Emerg Infect Dis* 19**,** 464-467.

Dziejman, M., Balon, E., Boyd, D., Fraser, C.M., Heidelberg, J.F., and Mekalanos, J.J. (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.

Dziejman, M., Serruto, D., Tam, V.C., Sturtevant, D., Diraphat, P., Faruque, S.M., Rahman, M.H., Heidelberg, J.F., Decker, J., Li, L., Montgomery, K.T., Grills, G., Kucherlapati, R., and Mekalanos, J.J. (2005). Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proc Natl Acad Sci U S A* 102**,** 3465-3470.

Eisenberg, M.C., Robertson, S.L., and Tien, J.H. (2013). Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. *Journal of theoretical biology* 324C**,** 84-102.

Emch, M., Feldacker, C., Islam, M.S., and Ali, M. (2008). Seasonality of cholera from 1974 to 2005: a review of global patterns. *Int J Health Geogr* 7**,** 31.

Eppinger, M., Pearson, T., Koenig, S.S.K., Pearson, O., Hicks, N., Agrawal, S., Sanjar, F., Galens, K., Daugherty, S., Crabtree, J., Hendriksen, R.S., Price, L.B., Upadhyay, B.P., Shakya, G., Fraser, C.M., Ravel, J., and Keim, P.S. (2014). Genomic Epidemiology of the Haitian Cholera Outbreak: a Single Introduction Followed by Rapid, Extensive, and Continued Spread Characterized the Onset of the Epidemic. *Mbio* 5.

Epstein, P.R. (1993). Algal blooms in the spread and persistence of cholera. *Biosystems* 31**,** 209-221.

Epstein, P.R., Ford, T.E., and Colwell, R.R. (1993). Marine ecosystems. *The Lancet* 342**,** 1216-1219.

Evins, G.M., Cameron, D.N., Wells, J.G., Greene, K.D., Popovic, T., Giono-Cerezo, S., Wachsmuth, I.K., and Tauxe, R.V. (1995). The emerging diversity of the electrophoretic types of *Vibrio cholerae* in the Western Hemisphere. *J Infect Dis* 172**,** 173-179.

Faast, R., Ogierman, M.A., Stroeher, U.H., and Manning, P.A. (1989). Nucleotide sequence of the structural gene, *tcpA*, for a major pilin subunit of *Vibrio cholerae*. *Gene* 85**,** 227-231.

Faruque, A.S., Alam, K., Malek, M.A., Khan, M.G., Ahmed, S., Saha, D., Khan, W.A., Nair, G.B., Salam, M.A., Luby, S.P., and Sack, D.A. (2007). Emergence of multidrug-resistant strain of *Vibrio cholerae* O1 in Bangladesh and reversal of their susceptibility to tetracycline after two years. *J Health Popul Nutr* 25**,** 241-243.

Faruque, S.M., Ahmed, K.M., Abdul Alim, A.R., Qadri, F., Siddique, A.K., and Albert, M.J. (1997). Emergence of a new clone of toxigenic *Vibrio cholerae* O1 biotype El Tor displacing *V. cholerae* O139 Bengal in Bangladesh. *J Clin Microbiol* 35**,** 624-630.

Faruque, S.M., Chowdhury, N., Kamruzzaman, M., Ahmad, Q.S., Faruque, A.S., Salam, M.A., Ramamurthy, T., Nair, G.B., Weintraub, A., and Sack, D.A. (2003a). Reemergence of epidemic *Vibrio cholerae* O139, Bangladesh. *Emerg Infect Dis* 9**,** 1116-1122.

Faruque, S.M., and Mekalanos, J.J. (2012). Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*. *Virulence* 3.

Faruque, S.M., Naser, I.B., Islam, M.J., Faruque, A.S., Ghosh, A.N., Nair, G.B., Sack, D.A., and Mekalanos, J.J. (2005a). Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc Natl Acad Sci U S A* 102**,** 1702-1707.

Faruque, S.M., Naser, I.B., Islam, M.J., Faruque, A.S.G., Ghosh, A.N., Nair, G.B., Sack, D.A., and Mekalanos, J.J. (2005b). Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc Natl Acad Sci U S A* 102**,** 1702-1707.

Faruque, S.M., Rahman, M.M., Asadulghani, Nasirul Islam, K.M., and Mekalanos, J.J. (1999). Lysogenic conversion of environmental *Vibrio mimicus* strains by CTXPhi. *Infect Immun* 67**,** 5723-5729.

Faruque, S.M., Sack, D.A., Sack, R.B., Colwell, R.R., Takeda, Y., and Nair, G.B. (2003b). Emergence and evolution of *Vibrio cholerae* O139. *Proc Natl Acad Sci U S A* 100**,** 1304-1309.

Federspiel, F., and Ali, M. (2018). The cholera outbreak in Yemen: lessons learned and way forward. *BMC public health* 18**,** 1338-1338.

Fiedler, C.J., Schönher, C., Proksch, P., Kerschbaumer, D.J., Mayr, E., Zunabovic-Pichler, M., Domig, K.J., and Perfler, R. (2018). Assessment of Microbial Community Dynamics in River Bank Filtrate Using High-Throughput Sequencing and Flow Cytometry. *Frontiers in Microbiology* 9.

Finch, M.J., Morris, J.G., Jr., Kaviti, J., Kagwanja, W., and Levine, M.M. (1988). Epidemiology of antimicrobial resistant cholera in Kenya and East Africa. *Am J Trop Med Hyg* 39**,** 484-490.

Folster, J.P., Katz, L., Mccullough, A., Parsons, M.B., Knipe, K., Sammons, S.A., Boncy, J., Tarr, C.L., and Whichard, J.M. (2014). Multidrug-resistant IncA/C plasmid in *Vibrio cholerae* from Haiti. *Emerg Infect Dis* 20**,** 1951-1953.

Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology* 13**,** 133.

Garg, P., Aydanian, A., Smith, D., J. Glenn M, J., Nair, G.B., and Stine, O.C. (2003). Molecular epidemiology of O139 *Vibrio cholerae*: mutation, lateral gene transfer, and founder flush. *Emerg Infect Dis* 9**,** 810-814.

Garg, P., Sinha, S., Chakraborty, R., Bhattacharya, S.K., Nair, G.B., Ramamurthy, T., and Takeda, Y. (2001). Emergence of fluoroquinolone-resistant strains of *Vibrio cholerae* O1 biotype El Tor among hospitalized patients with cholera in Calcutta, India. *Antimicrob Agents Chemother* 45**,** 1605-1606.

Ghosh, P., Naha, A., Basak, S., Ghosh, S., Ramamurthy, T., Koley, H., Nandy, R.K., Shinoda, S., Watanabe, H., and Mukhopadhyay, A.K. (2014a). Haitian variant *tcpA* in *Vibrio cholerae* O1 El Tor strains in Kolkata, India. *J Clin Microbiol* 52**,** 1020-1021.

Ghosh, P., Naha, A., Pazhani, G.P., Ramamurthy, T., and Mukhopadhyay, A.K. (2014b). Genetic traits of *Vibrio cholerae* O1 Haitian isolates that are absent in contemporary strains from Kolkata, India. *PLoS One* 9**,** e112973.

Ghosh, R., Nair, G.B., Tang, L., Morris, J.G., Sharma, N.C., Ballal, M., Garg, P., Ramamurthy, T., and Stine, O.C. (2008). Epidemiological study of *Vibrio cholerae* using variable number of tandem repeats. *FEMS Microbiol Lett* 288**,** 196-201.

Gil, A.I., Louis, V.R., Rivera, I.N., Lipp, E., Huq, A., Lanata, C.F., Taylor, D.N., Russek-Cohen, E., Choopun, N., Sack, R.B., and Colwell, R.R. (2004). Occurrence and distribution of *Vibrio cholerae* in the coastal environment of Peru. *Environ Microbiol* 6**,** 699-706.

Glass, R.I., Huq, I., Alim, A.R., and Yunus, M. (1980). Emergence of multiply antibiotic-resistant *Vibrio cholerae* in Bangladesh. *J Infect Dis* 142**,** 939-942.

Glass, R.I., Huq, M.I., Lee, J.V., Threlfall, E.J., Khan, M.R., Alim, A.R., Rowe, B., and Gross, R.J. (1983). Plasmid-borne multiple drug resistance in *Vibrio*

cholerae serogroup O1, biotype El Tor: evidence for a point-source outbreak in Bangladesh. *J Infect Dis* 147**,** 204-209.

Greenough, W.B., 3rd, Gordon, R.S., Jr., Rosenberg, I.S., Davies, B.I., and Benenson, A.S. (1964). Tetracycline in the Treatment of Cholera. *Lancet* 1**,** 355-357.

Guillaume, Y., Ternier, R., Vissieres, K., Casseus, A., Chery, M.J., and Ivers, L.C. (2018). Responding to Cholera in Haiti: Implications for the National Plan to Eliminate Cholera by 2022. *J Infect Dis* 218**,** S167-S170.

Guthmann, J.P. (1995). Epidemic cholera in Latin America: spread and routes of transmission. *J Trop Med Hyg* 98**,** 419-427.

Haque, R., Mondal, D., Duggal, P., Kabir, M., Roy, S., Farr, B.M., Sack, R.B., and Petri, W.A., Jr. (2006). Entamoeba histolytica infection in children and protection from subsequent amebiasis. *Infect Immun* 74**,** 904-909.

Haque, R., Mondal, D., Kirkpatrick, B.D., Akther, S., Farr, B.M., Sack, R.B., and Petri, W.A., Jr. (2003). Epidemiologic and clinical characteristics of acute diarrhea with emphasis on Entamoeba histolytica infections in preschool children in an urban slum of Dhaka, Bangladesh. *Am J Trop Med Hyg* 69**,** 398-405.

Harris, J.B., Larocque, R.C., Qadri, F., Ryan, E.T., and Calderwood, S.B. (2012). Cholera. *Lancet* 379**,** 2466-2476.

Hasan, N.A., Choi, S.Y., Eppinger, M., Clark, P.W., Chen, A., Alam, M., Haley, B.J., Taviani, E., Hine, E., Su, Q., Tallon, L.J., Prosper, J.B., Furth, K., Hoq, M.M., Li, H., Fraser-Liggett, C.M., Cravioto, A., Huq, A., Ravel, J., Cebula, T.A., and Colwell, R.R. (2012). Genomic diversity of 2010 Haitian cholera outbreak strains. *Proc Natl Acad Sci U S A* 109**,** E2010-E2017.

Hasan, N.A., Young, B.A., Minard-Smith, A.T., Saeed, K., Li, H., Heizer, E.M., Mcmillan, N.J., Isom, R., Abdullah, A.S., Bornman, D.M., Faith, S.A., Choi, S.Y., Dickens, M.L., Cebula, T.A., and Colwell, R.R. (2014). Microbial community profiling of human saliva using shotgun metagenomic sequencing. *PLoS One* 9**,** e97699.

Hashizume, M., Chaves, L.F., Faruque, A.S., Yunus, M., Streatfield, K., and Moji, K. (2013). A differential effect of Indian ocean dipole and El Nino on cholera dynamics in Bangladesh. *PLoS One* 8**,** e60001.

Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S.M., Woodgate, R., and Waldor, M.K. (2001). Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob Agents Chemother* 45**,** 2991-3000.

Hoshino, K., Yamasaki, S., Mukhopadhyay, A.K., Chakraborty, S., Basu, A., Bhattacharya, S.K., Nair, G.B., Shimada, T., and Takeda, Y. (1998). Development and evaluation of a multiplex PCR assay for rapid detection of toxigenic *Vibrio cholerae* O1 and O139. *FEMS Immunol Med Microbiol* 20**,** 201-207.

Hourigan, S.K., Subramanian, P., Hasan, N.A., Ta, A., Klein, E., Chettout, N., Huddleston, K., Deopujari, V., Levy, S., Baveja, R., Clemency, N.C., Baker, R.L., Niederhuber, J.E., and Colwell, R.R. (2018). Comparison of Infant Gut and Skin Microbiota, Resistome and Virulome Between Neonatal Intensive Care Unit (NICU) Environments. *Front Microbiol* 9**,** 1361.

Huq, A., Colwell, R.R., Chowdhury, M.A., Xu, B., Moniruzzaman, S.M., Islam, M.S., Yunus, M., and Albert, M.J. (1995). Coexistence of *Vibrio cholerae* O1 and O139 Bengal in plankton in Bangladesh. *Lancet* 345**,** 1249.

Huq, A., Grim, C., Colwell, R.R., and Nair, G.B. (2006). Detection, isolation, and identification of *Vibrio cholerae* from the environment. *Curr Protoc Microbiol* Chapter 6**,** Unit6A 5.

Huq, A., Haley, B.J., Taviani, E., Chen, A., Hasan, N.A., and Colwell, R.R. (2012). Detection, isolation, and identification of *Vibrio cholerae* from the environment. *Curr Protoc Microbiol* Chapter 6**,** Unit6A 5.

Huq, A., Sack, R.B., Nizam, A., Longini, I.M., Nair, G.B., Ali, A., Morris, J.G., Jr., Khan, M.N., Siddique, A.K., Yunus, M., Albert, M.J., Sack, D.A., and Colwell, R.R. (2005). Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl Environ Microbiol* 71**,** 4645-4654.

Huq, A., Small, E.B., West, P.A., Huq, M.I., Rahman, R., and Colwell, R.R. (1983). Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl Environ Microbiol* 45**,** 275-283.

Huq, A., West, P.A., Small, E.B., Huq, M.I., and Colwell, R.R. (1984). Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio*

cholerae serovar O1 associated with live copepods in laboratory microcosms. *Appl Environ Microbiol* 48**,** 420-424.

Islam, M.S., Drasar, B.S., and Bradley, D.J. (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.

Islam, M.S., Mahmud, Z.H., Ansaruzzaman, M., Faruque, S.M., Talukder, K.A., Qadri, F., Alam, M., Islam, S., Bardhan, P.K., Mazumder, R.N., Khan, A.I., Ahmed, S., Iqbal, A., Chitsatso, O., Mudzori, J., Patel, S., Midzi, S.M., Charimari, L., Endtz, H.P., and Cravioto, A. (2011). Phenotypic, genotypic, and antibiotic sensitivity patterns of strains isolated from the cholera epidemic in Zimbabwe. *J Clin Microbiol* 49**,** 2325-2327.

Islam, M.S., Midzi, S.M., Charimari, L., Cravioto, A., and Endtz, H.P. (2009). Susceptibility to fluoroquinolones of *Vibrio cholerae* O1 isolated from diarrheal patients in Zimbabwe. *JAMA* 302**,** 2321-2322.

Jain, M., Goel, A.K., Bhattacharya, P., Ghatole, M., and Kamboj, D.V. (2011). Multidrug resistant *Vibrio cholerae* O1 El Tor carrying classical *ctxB* allele involved in a cholera outbreak in South Western India. *Acta Trop* 117**,** 152-156.

Jesudason, M.V., and Saaya, R. (1997). Resistance of *Vibrio cholerae* O1 to nalidixic acid. *Indian J Med Res* 105**,** 153-154.

Jonson, G., Holmgren, J., and Svennerholm, A.M. (1992). Analysis of expression of toxin-coregulated pili in classical and El Tor *Vibrio cholerae* O1 in vitro and in vivo. *Infect Immun* 60**,** 4278-4284.

Jubair, M., Morris, J.G., Jr., and Ali, A. (2012). Survival of *Vibrio cholerae* in nutrient-poor environments is associated with a novel "persister" phenotype. *PLoS One* 7**,** e45187.

Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A., Alam, M., Sack, R.B., and Colwell, R. (2013). Environmental factors influencing epidemic cholera. *Am J Trop Med Hyg* 89**,** 597-607.

Jutla, A.S., Akanda, A.S., Griffiths, J.K., Colwell, R., and Islam, S. (2011). Warming oceans, phytoplankton, and river discharge: implications for cholera outbreaks. *Am J Trop Med Hyg* 85**,** 303-308.

Jutla, A.S., Akanda, A.S., and Islam, S. (2010). Tracking Cholera in Coastal Regions using Satellite Observations. *J Am Water Resour Assoc* 46**,** 651-662.

Kaas, R.S., Leekitcharoenphon, P., Aarestrup, F.M., and Lund, O. (2014). Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One* 9**,** e104984.

Kaevska, M., Videnska, P., Sedlar, K., and Slana, I. (2016). Seasonal changes in microbial community composition in river water studied using 454-pyrosequencing. *Springerplus* 5**,** 409.

Kaper, J.B., Morris, J.G., Jr., and Levine, M.M. (1995). Cholera. *Clin Microbiol Rev* 8**,** 48-86.

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

Karaolis, D.K., Lan, R., and Reeves, P.R. (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.

Keasler, S.P., and Hall, R.H. (1993). Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. *Lancet* 341**,** 1661.

Kendall, E.A., Chowdhury, F., Begum, Y., Khan, A.I., Li, S., Thierer, J.H., Bailey, J., Kreisel, K., Tacket, C.O., Larocque, R.C., Harris, J.B., Ryan, E.T., Qadri, F., Calderwood, S.B., and Stine, O.C. (2010). Relatedness of *Vibrio cholerae* O1/O139 isolates from patients and their household contacts, determined by multilocus variable-number tandem-repeat analysis. *J Bacteriol* 192**,** 4367-4376.

Khan, W.A., Saha, D., Ahmed, S., Salam, M.A., and Bennish, M.L. (2015). Efficacy of Ciprofloxacin for Treatment of Cholera Associated with Diminished Susceptibility to Ciprofloxacin to *Vibrio cholerae* O1. *PLoS One* 10**,** e0134921.

Kiiru, J., Mutreja, A., Mohamed, A.A., Kimani, R.W., Mwituria, J., Sanaya, R.O., Muyodi, J., Revathi, G., Parkhill, J., Thomson, N., Dougan, G., and Kariuki, S. (2013). A study on the geophylogeny of clinical and environmental *Vibrio cholerae* in Kenya. *PLoS One* 8**,** e74829.

Kim, E.J., Lee, C.H., Nair, G.B., and Kim, D.W. (2015). Whole-genome sequence comparisons reveal the evolution of *Vibrio cholerae* O1. *Trends Microbiol* 23**,** 479-489.

Kim, E.J., Lee, D., Moon, S.H., Lee, C.H., Kim, S.J., Lee, J.H., Kim, J.O., Song, M., Das, B., Clemens, J.D., Pape, J.W., Nair, G.B., and Kim, D.W. (2014). Molecular insights into the evolutionary pathway of *Vibrio cholerae* O1 atypical El Tor variants. *PLoS Pathog* 10**,** e1004384.

Kimsey, H.H., Nair, G.B., Ghosh, A., and Waldor, M.K. (1998). Diverse CTXphis and evolution of new pathogenic *Vibrio cholerae*. *Lancet* 352**,** 457-458.

Kitaoka, M., Miyata, S.T., Unterweger, D., and Pukatzki, S. (2011). Antibiotic resistance mechanisms of *Vibrio cholerae*. *J Med Microbiol* 60**,** 397-407.

Klinzing, D.C., Choi, S.Y., Hasan, N.A., Matias, R.R., Tayag, E., Geronimo, J., Skowronski, E., Rashed, S.M., Kawashima, K., Rosenzweig, C.N., Gibbons, H.S., Torres, B.C., Liles, V., Alfon, A.C., Juan, M.L., Natividad, F.F., Cebula, T.A., and Colwell, R.R. (2015). Hybrid *Vibrio cholerae* El Tor lacking SXT identified as the cause of a cholera outbreak in the Philippines. *MBio* 6.

Koch, R. (1884). An Address on Cholera and its Bacillus. *British medical journal* 2**,** 403-407.

Koskella, B., and Vos, M. (2015). Adaptation in Natural Microbial Populations. *Annual Review of Ecology, Evolution, and Systematics* 46**,** 503 - 522.

Kuleshov, K.V., Vodop'ianov, S.O., Dedkov, V.G., Markelov, M.L., Deviatkin, A.A., Kruglikov, V.D., Vodop'ianov, A.S., Pisanov, R.V., Mazrukho, A.B., Titova, S.V., Maleev, V.V., and Shipulin, G.A. (2016). Travel-Associated *Vibrio cholerae* O1 El Tor, Russia. *Emerging infectious diseases* 22**,** 2006-2008.

Kumar, P., Jain, M., Goel, A.K., Bhadauria, S., Sharma, S.K., Kamboj, D.V., Singh, L., Ramamurthy, T., and Nair, G.B. (2009). A large cholera outbreak due to a new cholera toxin variant of the *Vibrio cholerae* O1 El Tor biotype in Orissa, Eastern India. *J Med Microbiol* 58**,** 234-238.

Kumar, P., Mishra, D.K., Deshmukh, D.G., Jain, M., Zade, A.M., Ingole, K.V., and Yadava, P.K. (2014). Haitian variant *ctxB* producing *Vibrio cholerae* O1 with reduced susceptibility to ciprofloxacin is persistent in Yavatmal, Maharashtra, India, after causing a cholera outbreak. *Clin Microbiol Infect* 20**,** O292-293.

Kumar, P., Thulaseedharan, A., Chowdhury, G., Ramamurthy, T., and Thomas, S. (2011). Characterization of novel alleles of toxin co-regulated pilus A gene (*tcpA*) from environmental isolates of *Vibrio cholerae*. *Curr Microbiol* 62**,** 758-763.

Lam, C., Octavia, S., Reeves, P., Wang, L., and Lan, R. (2010). Evolution of seventh cholera pandemic and origin of 1991 epidemic, Latin America. *Emerg Infect Dis* 16**,** 1130-1132.

Lam, C., Octavia, S., Reeves, P.R., and Lan, R. (2012). Multi-locus variable number tandem repeat analysis of 7th pandemic *Vibrio cholerae*. *BMC Microbiology* 12**,** doi: 10.1186/1471-2180-1112-1182.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A.K., Wertheim, H.F., Sumpradit, N., Vlieghe, E., Hara, G.L., Gould, I.M., Goossens, H., Greko, C., So, A.D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A.Q., Qamar, F.N., Mir, F., Kariuki, S., Bhutta, Z.A., Coates, A., Bergstrom, R., Wright, G.D., Brown, E.D., and Cars, O. (2013). Antibiotic resistance-the need for global solutions. *Lancet Infect Dis* 13**,** 1057-1098.

Lee, J.H., Choi, S.Y., Jeon, Y.S., Lee, H.R., Kim, E.J., Nguyen, B.M., Hien, N.T., Ansaruzzaman, M., Islam, M.S., Bhuiyan, N.A., Niyogi, S.K., Sarkar, B.L., Nair, G.B., Kim, D.S., Lopez, A.L., Czerkinsky, C., Clemens, J.D., Chun, J., and Kim, D.W. (2009). Classification of hybrid and altered *Vibrio cholerae* strains by CTX prophage and RS1 element structure. *J Microbiol* 47**,** 783-788.

Lenski, R.E. (2017). What is adaptation by natural selection? Perspectives of an experimental microbiologist. *PLoS Genet* 13**,** e1006668.

Levine, M.M., Black, R.E., Clements, M.L., Cisneros, L., Saah, A., Nalin, D.R., Gill, D.M., Craig, J.P., Young, C.R., and Ristaino, P. (1982). The pathogenicity of nonenterotoxigenic *Vibrio cholerae* serogroup O1 biotype El Tor isolated from sewage water in Brazil. *J Infect Dis* 145**,** 296-299.

Lindenbaum, J., Greenough, W.B., and Islam, M.R. (1967). Antibiotic therapy of cholera. *Bull World Health Organ* 36**,** 871-883.

Lipp, E.K., Huq, A., and Colwell, R.R. (2002). Effects of global climate on infectious disease: the cholera model. *Clin Microbiol Rev* 15**,** 757-770.

Lizarraga-Partida, M.L., Mendez-Gomez, E., Rivas-Montano, A.M., Vargas-Hernandez, E., Portillo-Lopez, A., Gonzalez-Ramirez, A.R., Huq, A., and

Colwell, R.R. (2009). Association of *Vibrio cholerae* with plankton in coastal areas of Mexico. *Environ Microbiol* 11**,** 201-208.

Lizarraga-Partida, M.L., and Quilici, M.L. (2009). Molecular analyses of *Vibrio cholerae* O1 clinical strains, including new nontoxigenic variants isolated in Mexico during the Cholera epidemic years between 1991 and 2000. *J Clin Microbiol* 47**,** 1364-1371.

Lombardo, M.J., Michalski, J., Martinez-Wilson, H., Morin, C., Hilton, T., Osorio, C.G., Nataro, J.P., Tacket, C.O., Camilli, A., and Kaper, J.B. (2007). An in vivo expression technology screen for *Vibrio cholerae* genes expressed in human volunteers. *Proc Natl Acad Sci U S A* 104**,** 18229-18234.

Longini, I.M., Jr., Yunus, M., Zaman, K., Siddique, A.K., Sack, R.B., and Nizam, A. (2002). Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *J Infect Dis* 186**,** 246-251.

Manning, P.A., Clark, C.A., and Focareta, T. (1999). Gene capture in *Vibrio cholerae*. *Trends Microbiol* 7**,** 93-95.

Marin, M.A., Thompson, C.C., Freitas, F.S., Fonseca, E.L., Aboderin, A.O., Zailani, S.B., Quartey, N.K., Okeke, I.N., and Vicente, A.C. (2013). Cholera outbreaks in Nigeria are associated with multidrug resistant atypical El Tor and non-O1/non-O139 *Vibrio cholerae*. *PLoS Negl Trop Dis* 7**,** e2049.

Martin, Y.P., and Bianchi, M.A. (1980). Structure, diversity, and catabolic potentialities of aerobic heterotrophic bacterial populations associated with continuous cultures of natural marine phytoplankton. *Microb Ecol* 5**,** 265-279.

Mccarthy, S.A., and Khambaty, F.M. (1994). International dissemination of epidemic *Vibrio cholerae* by cargo ship ballast and other nonpotable waters. *Appl Environ Microbiol* 60**,** 2597-2601.

Merikanto, I., Laakso, J.T., and Kaitala, V. (2018). Outside-host phage therapy as a biological control against environmental infectious diseases. *Theoretical Biology and Medical Modelling* 15**,** 7.

Merrell, D.S., Butler, S. M., Qadri, F., Dolganov, N. A., Alam, A. , Cohen, M. B., Calderwood, S. B., Schoolnik, G. K., Camilli, A. (2002). Host-induced epidemic spread of the cholera bacterium. *Nature* 417**,** 642–645.

Merrell, D.S., and Camilli, A. (2000). Detection and analysis of gene expression during infection by in vivo expression technology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 355**,** 587-599.

Mhalu, F.S., Mmari, P.W., and Ijumba, J. (1979). Rapid emergence of El Tor *Vibrio cholerae* resistant to antimicrobial agents during first six months of fourth cholera epidemic in Tanzania. *Lancet* 1**,** 345-347.

Mlejnkova, H., and Sovova, K. (2010). Impact of pollution and seasonal changes on microbial community structure in surface water. *Water Sci Technol* 61**,** 2787-2795.

Mohapatra, S.S., Mantri, C.K., Fazil, H.U.T., and Singh, D.V. (2011). *Vibrio cholerae* O1 biotype El Tor strains isolated in 1992 from Varanasi, India harboured El Tor CTXΦ and classical *ctxB* on the chromosome-I and classical CTXΦ and classical *ctxB* on the chromosome-II. *Environ Microbiol Rep* 3**,** 783-790.

Montgomery, M., Jones, M.W., Kabole, I., Johnston, R., and Gordon, B. (2018). No end to cholera without basic water, sanitation and hygiene. *Bulletin of the World Health Organization* 96**,** 371-371A.

Moore, S., Thomson, N., Mutreja, A., and Piarroux, R. (2014a). Widespread epidemic cholera caused by a restricted subset of *Vibrio cholerae* clones. *Clin Microbiol Infect* 20**,** 373-379.

Moore, S.M., Azman, A.S., Zaitchik, B.F., Mintz, E.D., Brunkard, J., Legros, D., Hill, A., Mckay, H., Luquero, F.J., Olson, D., and Lessler, J. (2017). El Nino and the shifting geography of cholera in Africa. *Proc Natl Acad Sci U S A* 114**,** 4436-4441.

Moore, S.M., Shannon, K.L., Zelaya, C.E., Azman, A.S., and Lessler, J. (2014b). Epidemic risk from cholera introductions into Mexico. *PLoS Curr* 6.

Morais, L.L., Garza, D.R., Loureiro, E.C., Vale, E.R., Santos, D.S., Correa, V.C., Sousa, N.R., Gurjao, T.C., Santos, E.C., Vieira, V.V., Da Fonseca, E.L., and Vicente, A.C. (2013). Population and genetic study of *Vibrio cholerae* from the amazon environment confirms that the WASA-1 prophage is the main marker of the epidemic strain that circulated in the region. *PLoS One* 8**,** e81372.

Morita, M., Ohnishi, M., Arakawa, E., Bhuiyan, N.A., Nusrin, S., Alam, M., Siddique, A.K., Qadri, F., Izumiya, H., Nair, G.B., and Watanabe, H. (2008).

Development and validation of a mismatch amplification mutation PCR assay to monitor the dissemination of an emerging variant of *Vibrio cholerae* O1 biotype El Tor. *Microbiol Immunol* 52**,** 314-317.

Morita, M., Ohnishi, M., Arakawa, E., Yamamoto, S., Nair, G.B., Matsushita, S., Yokoyama, K., Kai, A., Seto, K., Watanabe, H., and Izumiya, H. (2010). Emergence and genetic diversity of El Tor *Vibrio cholerae* O1 that possess classical biotype *ctxB* among travel-associated cases of cholera in Japan. *J Med Microbiol* 59**,** 708-712.

Morris, J.G., Jr. (2011). Cholera-modern pandemic disease of ancient lineage. *Emerg Infect Dis* 17**,** 2099-2104.

Morris, J.G., Sztein, M.B., Rice, E.W., Nataro, J.P., Losonsky, G.A., Panigrahi, P., Tacket, C.O., and Johnson, J.A. (1996). *Vibrio cholerae* O1 Can Assume a Chlorine-Resistant Rugose Survival Form that Is Virulent for Humans. *J Infect Dis* 6**,** 1364–1368.

Morrison, J.M., Baker, K.D., Zamor, R.M., Nikolai, S., Elshahed, M.S., and Youssef, N.H. (2017). Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLoS One* 12**,** e0177488.

Mourino-Perez, R.R. (1998). Oceanography and the seventh cholera pandemic. *Epidemiology* 9**,** 355-357.

Mukandavire, Z., Liao, S., Wang, J., Gaff, H., Smith, D.L., and Morris, J.G., Jr. (2011). Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe. *Proceedings of the National Academy of Sciences of the United States of America* 108**,** 8767-8772.

Mukandavire, Z., Smith, D.L., and Morris, J.G., Jr. (2013). Cholera in Haiti: reproductive numbers and vaccination coverage estimates. *Scientific reports* 3**,** 997.

Mukherjee, A.K., Chowdhury, P., Bhattacharya, M.K., Ghosh, M., Rajendran, K., and Ganguly, S. (2009). Hospital-based surveillance of enteric parasites in Kolkata. *BMC Res Notes* 2**,** 110.

Mukhopadhyay, A.K., Basu, I., Bhattacharya, S.K., Bhattacharya, M.K., and Nair, G.B. (1998). Emergence of fluoroquinolone resistance in strains of *Vibrio*

cholerae isolated from hospitalized patients with acute diarrhea in Calcutta, India. *Antimicrobial agents and chemotherapy* 42**,** 206-207.

Mukhopadhyay, A.K., Chakraborty, S., Takeda, Y., Nair, G.B., and Berg, D.E. (2001). Characterization of VPI pathogenicity island and CTXphi prophage in environmental strains of *Vibrio cholerae*. *J Bacteriol* 183**,** 4737-4746.

Mukhopadhyay, A.K., Takeda, Y., and Balakrish Nair, G. (2014). Cholera outbreaks in the El Tor biotype era and the impact of the new El Tor variants. *Curr Top Microbiol Immunol* 379**,** 17-47.

Murphy, R.A., and Boyd, E.F. (2008). Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. *J Bacteriol* 190**,** 636-647.

Mutreja, A., Kim, D.W., Thomson, N.R., Connor, T.R., Lee, J.H., Kariuki, S., Croucher, N.J., Choi, S.Y., Harris, S.R., Lebens, M., Niyogi, S.K., Kim, E.J., Ramamurthy, T., Chun, J., Wood, J.L., Clemens, J.D., Czerkinsky, C., Nair, G.B., Holmgren, J., Parkhill, J., and Dougan, G. (2011). Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 477**,** 462-465.

Na-Ubol, M., Srimanote, P., Chongsa-Nguan, M., Indrawattana, N., Sookrung, N., Tapchaisri, P., Yamazaki, S., Bodhidatta, L., Eampokalap, B., Kurazono, H., Hayashi, H., Nair, G.B., Takeda, Y., and Chaicumpa, W. (2011). Hybrid & El Tor variant biotypes of *Vibrio cholerae* O1 in Thailand. *Indian J Med Res* 133**,** 387-394.

Naha, A., Pazhani, G.P., Ganguly, M., Ghosh, S., Ramamurthy, T., Nandy, R.K., Nair, B.G., Takeda, Y., and Mukhopadhyay, A.K. (2012). Development and Evaluation of a PCR Assay for Tracking the Emergence and Dissemination of Haitian Variant *ctxB* in *Vibrio cholerae* O1 Strains Isolated from Kolkata, India. *J Clin Microbiol* 50**,** 1733-1736.

Nair, G.B., Faruque, S.M., Bhuiyan, N.A., Kamruzzaman, M., Siddique, A.K., and Sack, D.A. (2002). New variants of *Vibrio cholerae* O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. *J Clin Microbiol* 40**,** 3296-3299.

Nair, G.B., Qadri, F., Holmgren, J., Svennerholm, A.M., Safa, A., Bhuiyan, N.A., Ahmad, Q.S., Faruque, S.M., Faruque, A.S., Takeda, Y., and Sack, D.A.

(2006). Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. *J Clin Microbiol* 44**,** 4211-4213.

Nandi, B., Nandy, R.K., Mukhopadhyay, S., Nair, G.B., Shimada, T., and Ghose, A.C. (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.

Nature Publishing Group (2011). Microbiology by numbers. *Nature Reviews Microbiology* 9**,** 628.

Nelson, E.J., Chowdhury, A., Flynn, J., Schild, S., Bourassa, L., Shao, Y., Larocque, R.C., Calderwood, S.B., Qadri, F., and Camilli, A. (2008). Transmission of *Vibrio cholerae* is antagonized by lytic phage and entry into the aquatic environment. *PLoS Pathog* 4**,** e1000187.

Nelson, E.J., Harris, J.B., Morris, J.G., Jr., Calderwood, S.B., and Camilli, A. (2009). Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat Rev Microbiol* 7**,** 693-702.

Nesper, J., Kraiss, A., Schild, S., Blass, J., Klose, K.E., Bockemuhl, J., and Reidl, J. (2002). Comparative and genetic analyses of the putative *Vibrio cholerae* lipopolysaccharide core oligosaccharide biosynthesis (*wav*) gene cluster. *Infect Immun* 70**,** 2419-2433.

Newton, R.J., Jones, S.E., Eiler, A., Mcmahon, K.D., and Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75**,** 14-49.

Nguyen, B.M., Lee, J.H., Cuong, N.T., Choi, S.Y., Hien, N.T., Anh, D.D., Lee, H.R., Ansaruzzaman, M., Endtz, H.P., Chun, J., Lopez, A.L., Czerkinsky, C., Clemens, J.D., and Kim, D.W. (2009). Cholera outbreaks caused by an altered *Vibrio cholerae* O1 El Tor biotype strain producing classical cholera toxin B in Vietnam in 2007 to 2008. *J Clin Microbiol* 47**,** 1568-1571.

Nusrin, S., Gil, A.I., Bhuiyan, N.A., Safa, A., Asakura, M., Lanata, C.F., Hall, E., Miranda, H., Huapaya, B., Vargas, G.C., Luna, M.A., Sack, D.A., Yamasaki, S., and Nair, G.B. (2009). Peruvian *Vibrio cholerae* O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. *J Med Microbiol* 58**,** 342-354.

O'shea, Y.A., Reen, F.J., Quirke, A.M., and Boyd, E.F. (2004). Evolutionary genetic analysis of the emergence of epidemic *Vibrio cholerae* isolates on the basis of comparative nucleotide sequence analysis and multilocus virulence gene profiles. *J Clin Microbiol* 42**,** 4657-4671.

Octavia, S., Salim, A., Kurniawan, J., Lam, C., Leung, Q., Ahsan, S., Reeves, P.R., Nair, G.B., and Lan, R. (2013). Population Structure and Evolution of Non-O1/Non-O139 *Vibrio cholerae* by Multilocus Sequence Typing. *PLoS One* 8**,** e65342.

Oliver, J.D. (2005). The viable but nonculturable state in bacteria. *J Microbiol* 43 Spec No**,** 93-100.

Olsvik, O. (1992). The cholera epidemic in Latin America. *Tidsskr Nor Laegeforen* 112**,** 1843-1846.

Olsvik, O., Wahlberg, J., Petterson, B., Uhlen, M., Popovic, T., Wachsmuth, I.K., and Fields, P.I. (1993). Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *J Clin Microbiol* 31**,** 22-25.

Ottemann, K.M., and Mekalanos, J.J. (1995). Analysis of *Vibrio cholerae* ToxR function by construction of novel fusion proteins. *Mol Microbiol* 15**,** 719-731.

Pacini, F. (1855). Microscopic Observations and Pathological Deductions on Asiatic Cholera. *The British and Foreign Medico-Chirurgical Review* 16**,** 144-145.

Pascual, M., Rodo, X., Ellner, S.P., Colwell, R., and Bouma, M.J. (2000a). Cholera dynamics and El Nino-Southern Oscillation. *Science* 289**,** 1766-1769.

Pascual, M., Rodó, X., Ellner, S.P., Colwell, R., and Bouma, M.J. (2000b). Cholera Dynamics and El Niño-Southern Oscillation. *Science* 289**,** 1766.

Pfau, J.D., and Taylor, R.K. (1996). Genetic footprint on the ToxR-binding site in the promoter for cholera toxin. *Mol Microbiol* 20**,** 213-222.

Phuc Nguyen, M.C., Woerther, P.L., Bouvet, M., Andremont, A., Leclercq, R., and Canu, A. (2009). Escherichia coli as reservoir for macrolide resistance genes. *Emerg Infect Dis* 15**,** 1648-1650.

Pollitzer, R. (1954). Cholera studies. 1. History of the disease. *Bulletin of the World Health Organization* 10**,** 421-461.

Quilici, M.L., Massenet, D., Gake, B., Bwalki, B., and Olson, D.M. (2010). *Vibrio cholerae* O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. *Emerg Infect Dis* 16**,** 1804-1805.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology* 35**,** 833.

Rahaman, M.H., Islam, T., Colwell, R.R., and Alam, M. (2015). Molecular tools in understanding the evolution of *Vibrio cholerae*. *Frontiers in microbiology* 6**,** 1040-1040.

Rahman, Z., Rahman, M.A., Rashid, M.-U., Monira, S., Johura, F.-T., Mustafiz, M., Bhuyian, S.I., Zohura, F., Parvin, T., Hasan, K., Saif-Ur-Rahman, K.M., Islam, N.N., Sack, D.A., George, C.M., and Alam, M. (2018). *Vibrio cholerae* Transmits Through Water Among the Household Contacts of Cholera Patients in Cholera Endemic Coastal Villages of Bangladesh, 2015–2016 (CHoBI7 Trial). *Frontiers in Public Health* 6.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S.K., Nair, G.B., Shimada, T., Takeda, T., Karasawa, T., Kurazano, H., Pal, A., and Et Al. (1993). Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* 341**,** 703-704.

Rashed, S.M., Azman, A.S., Alam, M., Li, S., Sack, D.A., Morris, J.G., Jr., Longini, I., Siddique, A.K., Iqbal, A., Huq, A., Colwell, R.R., Sack, R.B., and Stine, O.C. (2014). Genetic Variation of *Vibrio cholerae* during Outbreaks, Bangladesh, 2010-2011. *Emerg Infect Dis* 20**,** 54-60.

Rashed, S.M., Hasan, N.A., Alam, M., Sadique, A., Sultana, M., Hoq, M.M., Sack, R.B., Colwell, R.R., and Huq, A. (2017). *Vibrio cholerae* O1 with Reduced Susceptibility to Ciprofloxacin and Azithromycin Isolated from a Rural Coastal Area of Bangladesh. *Front Microbiol* 8**,** 252.

Rashed, S.M., Iqbal, A., Mannan, S.B., Islam, T., Rashid, M.U., Johura, F.T., Watanabe, H., Hasan, N.A., Huq, A., Stine, O.C., Sack, R.B., Colwell, R.R., and Alam, M. (2013). *Vibrio cholerae* O1 El Tor and O139 Bengal Strains Carrying *ctxB*(ET), Bangladesh. *Emerg Infect Dis* 19**,** 1713-1715.

Rashed, S.M., Mannan, S.B., Johura, F.T., Islam, M.T., Sadique, A., Watanabe, H., Sack, R.B., Huq, A., Colwell, R.R., Cravioto, A., and Alam, M. (2012).

Genetic characteristics of drug-resistant *Vibrio cholerae* O1 causing endemic cholera in Dhaka, 2006-2011. *J Med Microbiol* 61**,** 1736-1745.

Rashid, M.U., Almeida, M., Azman, A.S., Lindsay, B.R., Sack, D.A., Colwell, R.R., Huq, A., Morris, J.G., Jr., Alam, M., and Stine, O.C. (2016). Comparison of inferred relatedness based on multilocus variable-number tandem-repeat analysis and whole genome sequencing of *Vibrio cholerae* O1. *FEMS Microbiol Lett* 363.

Rashid, M.U., Rashed, S.M., Islam, T., Johura, F.T., Watanabe, H., Ohnishi, M., and Alam, M. (2015). CtxB1 outcompetes CtxB7 in *Vibrio cholerae* O1, Bangladesh. *J Med Microbiol*.

Raychoudhuri, A., Mukhopadhyay, A.K., Ramamurthy, T., Nandy, R.K., Takeda, Y., and Nair, B.G. (2008). Biotyping of *Vibrio cholerae* O1: Time to redefine the scheme. *Indian J Med Res* 128**,** 695-698.

Raychoudhuri, A., Patra, T., Ghosh, K., Ramamurthy, T., Nandy, R.K., Takeda, Y., Balakrish-Nair, G., and Mukhopadhyay, A.K. (2009). Classical *ctxB* in *Vibrio cholerae* O1, Kolkata, India. *Emerg Infect Dis* 15**,** 131-132.

Reimer, A.R., Van Domselaar, G., Stroika, S., Walker, M., Kent, H., Tarr, C., Talkington, D., Rowe, L., Olsen-Rasmussen, M., Frace, M., Sammons, S., Dahourou, G.A., Boncy, J., Smith, A.M., Mabon, P., Petkau, A., Graham, M., Gilmour, M.W., Gerner-Smidt, P., and Force, V.C.O.G.T. (2011). Comparative genomics of *Vibrio cholerae* from Haiti, Asia, and Africa. *Emerg Infect Dis* 17**,** 2113-2121.

Rivera, I.N., Chun, J., Huq, A., Sack, R.B., and Colwell, R.R. (2001). Genotypes associated with virulence in environmental isolates of *Vibrio cholerae*. *Appl Environ Microbiol* 67**,** 2421-2429.

Robins, W.P., and Mekalanos, J.J. (2014). Genomic science in understanding cholera outbreaks and evolution of *Vibrio cholerae* as a human pathogen. *Curr Top Microbiol Immunol* 379**,** 211-229.

Rodrigue, D.C., Popovic, T., and Wachsmuth, I.K. (1994). Non-toxigenic *Vibrio cholerae* O1 infections in the United States. *in Vibrio cholerae and Cholera: Molecular to Global Perspectives,* Wachsmuth, I. K., Blake, P. A., and Olsvik, Ø., eds. (Am. Soc. Microbiol., Washington, DC)**,** pp. 69 - 76.

Roy, M.A., Arnaud, J.M., Jasmin, P.M., Hamner, S., Hasan, N.A., Colwell, R.R., and Ford, T.E. (2018). A Metagenomic Approach to Evaluating Surface Water Quality in Haiti. *Int J Environ Res Public Health* 15.

Sack, D.A., Sack, R.B., and Chaignat, C.L. (2006). Getting serious about cholera. *N Engl J Med* 355**,** 649-651.

Sack, D.A., Sack, R.B., Nair, G.B., and Siddique, A.K. (2004). Cholera. *Lancet* 363**,** 223-233.

Safa, A., Bhuyian, N.A., Nusrin, S., Ansaruzzaman, M., Alam, M., Hamabata, T., Takeda, Y., Sack, D.A., and Nair, G.B. (2006). Genetic characteristics of Matlab variants of *Vibrio cholerae* O1 that are hybrids between classical and El Tor biotypes. *J Med Microbiol* 55**,** 1563-1569.

Safa, A., Nair, G.B., and Kong, R.Y. (2010). Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol* 18**,** 46-54.

Safa, A., Sultana, J., Dac Cam, P., Mwansa, J.C., and Kong, R.Y. (2008). *Vibrio cholerae* O1 hybrid El Tor strains, Asia and Africa. *Emerg Infect Dis* 14**,** 987-988.

Saha, D., Karim, M.M., Khan, W.A., Ahmed, S., Salam, M.A., and Bennish, M.L. (2006). Single-dose azithromycin for the treatment of cholera in adults. *N Engl J Med* 354**,** 2452-2462.

Saha, P.K., Koley, H., Mukhopadhyay, A.K., Bhattacharya, S.K., Nair, G.B., Ramakrishnan, B.S., Krishnan, S., Takeda, T., and Takeda, Y. (1996). Nontoxigenic *Vibrio cholerae* 01 serotype Inaba biotype El Tor associated with a cluster of cases of cholera in southern India. *J Clin Microbiol* 34**,** 1114-1117.

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

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6**,** e17288.

Schwartz, K., Hammerl, J.A., Göllner, C., and Strauch, E. (2019). Environmental and Clinical Strains of *Vibrio cholerae* Non-O1, Non-O139 From Germany Possess Similar Virulence Gene Profiles. *Frontiers in Microbiology* 10.

Seas, C., Miranda, J., Gil, A.I., Leon-Barua, R., Patz, J., Huq, A., Colwell, R.R., and Sack, R.B. (2000). New insights on the emergence of cholera in Latin America during 1991: the Peruvian experience. *Am J Trop Med Hyg* 62**,** 513-517.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30**,** 2068-2069.

Sepulveda, J., Valdespino, J.L., and Garcia-Garcia, L. (2006). Cholera in Mexico: the paradoxical benefits of the last pandemic. *Int J Infect Dis* 10**,** 4-13.

Shakya, G., Kim, D.W., Clemens, J.D., Malla, S., Upadhyaya, B.P., Dumre, S.P., Shrestha, S.D., Adhikari, S., Sharma, S., Rijal, N., Shrestha, S.K., Mason, C., and Kansakar, P. (2012). Phenotypic and genetic characterization of *Vibrio cholerae* O1 clinical isolates collected through national antimicrobial resistance surveillance network in Nepal. *World J Microbiol Biotechnol* 28**,** 2671-2678.

Shi, L., Miyoshi, S., Hiura, M., Tomochika, K., Shimada, T., and Shinoda, S. (1998). Detection of genes encoding cholera toxin (CT), zonula occludens toxin (ZOT), accessory cholera enterotoxin (ACE) and heat-stable enterotoxin (ST) in *Vibrio mimicus* clinical strains. *Microbiol Immunol* 42**,** 823-828.

Siddique, A.K., Baqui, A.H., Eusof, A., Haider, K., Hossain, M.A., Bashir, I., and Zaman, K. (1991). Survival of classic cholera in Bangladesh. *Lancet* 337**,** 1125-1127.

Siddique, A.K., and Cash, R. (2014). Cholera outbreaks in the classical biotype era. *Curr Top Microbiol Immunol* 379**,** 1-16.

Siddique, A.K., Nair, G.B., Alam, M., Sack, D.A., Huq, A., Nizam, A., Longini, I.M., Jr., Qadri, F., Faruque, S.M., Colwell, R.R., Ahmed, S., Iqbal, A., Bhuiyan, N.A., and Sack, R.B. (2010). El Tor cholera with severe disease: a new threat to Asia and beyond. *Epidemiol Infect* 138**,** 347-352.

Silva, A.J., and Benitez, J.A. (2016). *Vibrio cholerae* Biofilms and Cholera Pathogenesis. *PLoS Negl Trop Dis* 10**,** e0004330.

Sjolund-Karlsson, M., Reimer, A., Folster, J.P., Walker, M., Dahourou, G.A., Batra, D.G., Martin, I., Joyce, K., Parsons, M.B., Boncy, J., Whichard, J.M., and Gilmour, M.W. (2011). Drug-resistance mechanisms in *Vibrio cholerae* O1 outbreak strain, Haiti, 2010. *Emerg Infect Dis* 17**,** 2151-2154.

Smirnova, N.I., Krasnov, Y.M., Agafonova, E.Y., Shchelkanova, E.Y., Alkhova, Z.V., and Kutyrev, V.V. (2017). Whole-Genome Sequencing of *Vibrio cholerae* O1 El Tor Strains Isolated in Ukraine (2011) and Russia (2014). *Genome Announc* 5.

Snow, J. (1856). Cholera and the Water Supply in the South Districts of London in 1854. *J Public Health Sanit Rev* 2**,** 239-257.

Spira, W.M., Huq, A., Ahmed, Q.S., and Saeed, Y.A. (1981). Uptake of *Vibrio cholerae* biotype eltor from contaminated water by water hyacinth (eichornia crassipes). *Appl Environ Microbiol* 42**,** 550-553.

Steffen, R., Acar, J., Walker, E., and Zuckerman, J. (2003). Cholera: assessing the risk to travellers and identifying methods of protection. *Travel Medicine and Infectious Disease* 1**,** 80-88.

Stevenson, B.A., Hunter, D.W.F., and Rhodes, P.L. (2014). Temporal and seasonal change in microbial community structure of an undisturbed, disturbed, and carbon-amended pasture soil. *Soil Biology and Biochemistry* 75**,** 175-185.

Stewart, E.J. (2012). Growing unculturable bacteria. *Journal of bacteriology* 194**,** 4151-4160.

Stine, O.C., Alam, M., Tang, L., Nair, G.B., Siddique, A.K., Faruque, S.M., Huq, A., Colwell, R., Sack, R.B., and Morris, J.G., Jr. (2008). Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerg Infect Dis* 14**,** 831-833.

Stine, O.C., and Morris, J.G., Jr. (2014). Circulation and transmission of clones of *Vibrio cholerae* during cholera outbreaks. *Curr Top Microbiol Immunol* 379**,** 181-193.

Sultana, M., Nusrin, S., Hasan, N.A., Sadique, A., Ahmed, K.U., Islam, A., Hossain, A., Longini, I., Nizam, A., Huq, A., Siddique, A.K., Sack, D.A., Sack, R.B., Colwell, R.R., and Alam, M. (2018). Biofilms Comprise a Component of the Annual Cycle of *Vibrio cholerae* in the Bay of Bengal Estuary. *MBio* 9.

Tamplin, M.L., Gauzens, A.L., Huq, A., Sack, D.A., and Colwell, R.R. (1990). Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. *Appl Environ Microbiol* 56**,** 1977-1980.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011a). MEGA5: molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 28**,** 2731-2739.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011b). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 28**,** 2731-2739.

Taneja, N., Samanta, P., Mishra, A., and Sharma, M. (2010). Emergence of tetracycline resistance in *Vibrio cholerae* O1 biotype El Tor serotype Ogawa from north India. *Indian J Pathol Microbiol* 53**,** 865-866.

Tappero, J.W., and Tauxe, R.V. (2011). Lessons learned during public health response to cholera epidemic in Haiti and the Dominican Republic. *Emerging infectious diseases* 17**,** 2087-2093.

Taviani, E., Grim, C.J., Choi, J., Chun, J., Haley, B., Hasan, N.A., Huq, A., and Colwell, R.R. (2010). Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis. *FEMS Microbiol Lett* 308**,** 130-137.

Tay, C.Y., Reeves, P.R., and Lan, R. (2008). Importation of the major pilin TcpA gene and frequent recombination drive the divergence of the Vibrio pathogenicity island in *Vibrio cholerae*. *FEMS Microbiol Lett* 289**,** 210-218.

Thompson, F.L., Iida, T., and Swings, J. (2004). Biodiversity of vibrios. *Microbiol Mol Biol Rev* 68**,** 403-431, table of contents.

Thungapathra, M., Amita, Sinha, K.K., Chaudhuri, S.R., Garg, P., Ramamurthy, T., Nair, G.B., and Ghosh, A. (2002). Occurrence of antibiotic resistance gene cassettes aac(6')-Ib, *dfrA5*, *dfrA12*, and *ereA2* in class I integrons in non-O1, non-O139 *Vibrio cholerae* strains in India. *Antimicrob Agents Chemother* 46**,** 2948-2955.

Tkachenko, V.V. (1982). [Lipopolysaccharides of *Vibrio cholerae* and enterobacteria]. *Zh Mikrobiol Epidemiol Immunobiol***,** 20-28.

Toyama, D., Kishi, L.T., Santos-Junior, C.D., Soares-Costa, A., De Oliveira, T.C., De Miranda, F.P., and Henrique-Silva, F. (2016). Metagenomics Analysis of Microorganisms in Freshwater Lakes of the Amazon Basin. *Genome Announc* 4.

Tran, H.D., Alam, M., Trung, N.V., Kinh, N.V., Nguyen, H.H., Pham, V.C., Ansaruzzaman, M., Rashed, S.M., Bhuiyan, N.A., Dao, T.T., Endtz, H.P., and Wertheim, H.F. (2012). Multi-drug resistant *Vibrio cholerae* O1 variant El Tor isolated in northern Vietnam between 2007 and 2010. *J Med Microbiol* 61**,** 431-437.

Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 15**,** 524.

Valia, R., Taviani, E., Spagnoletti, M., Ceccarelli, D., Cappuccinelli, P., and Colombo, M.M. (2013). *Vibrio cholerae* O1 epidemic variants in Angola: a retrospective study between 1992 and 2006. *Front Microbiol* 4**,** 354.

Vezzulli, L., Grande, C., Reid, P.C., Hélaouët, P., Edwards, M., Höfle, M.G., Brettar, I., Colwell, R.R., and Pruzzo, C. (2016). Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic. *Proceedings of the National Academy of Sciences of the United States of America* 113**,** E5062-E5071.

Wachsmuth, I.K., Evins, G.M., Fields, P.I., Olsvik, O., Popovic, T., Bopp, C.A., Wells, J.G., Carrillo, C., and Blake, P.A. (1993). The molecular epidemiology of cholera in Latin America. *J Infect Dis* 167**,** 621-626.

Waldor, M.K., and Mekalanos, J.J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272**,** 1910-1914.

Waldor, M.K., Rubin, E.J., Pearson, G.D., Kimsey, H., and Mekalanos, J.J. (1997). Regulation, replication, and integration functions of the *Vibrio cholerae* CTXphi are encoded by region RS2. *Mol Microbiol* 24**,** 917-926.

Waldor, M.K., Tschape, H., and Mekalanos, J.J. (1996). A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *J Bacteriol* 178**,** 4157-4165.

Wallace, B. (1981). *Basic Population Genetics.* New York: Columbia University Press.

Wang, R., Liu, H., Zhao, X., Li, J., and Wan, K. (2018). IncA/C plasmids conferring high azithromycin resistance in *Vibrio cholerae*. *Int J Antimicrob Agents* 51**,** 140-144.

Wang, R., Lou, J., Liu, J., Zhang, L., Li, J., and Kan, B. (2012). Antibiotic resistance of *Vibrio cholerae* O1 El Tor strains from the seventh pandemic in China, 1961-2010. *Int J Antimicrob Agents* 40**,** 361-364.

Weill, F.X., Domman, D., Njamkepo, E., Almesbahi, A.A., Naji, M., Nasher, S.S., Rakesh, A., Assiri, A.M., Sharma, N.C., Kariuki, S., Pourshafie, M.R., Rauzier, J., Abubakar, A., Carter, J.Y., Wamala, J.F., Seguin, C., Bouchier, C., Malliavin, T., Bakhshi, B., Abulmaali, H.H.N., Kumar, D., Njoroge, S.M., Malik, M.R., Kiiru, J., Luquero, F.J., Azman, A.S., Ramamurthy, T., Thomson, N.R., and Quilici, M.L. (2019). Genomic insights into the 2016-2017 cholera epidemic in Yemen. *Nature* 565**,** 230-233.

Weill, F.X., Domman, D., Njamkepo, E., Tarr, C., Rauzier, J., Fawal, N., Keddy, K.H., Salje, H., Moore, S., Mukhopadhyay, A.K., Bercion, R., Luquero, F.J., Ngandjio, A., Dosso, M., Monakhova, E., Garin, B., Bouchier, C., Pazzani, C., Mutreja, A., Grunow, R., Sidikou, F., Bonte, L., Breurec, S., Damian, M., Njanpop-Lafourcade, B.M., Sapriel, G., Page, A.L., Hamze, M., Henkens, M., Chowdhury, G., Mengel, M., Koeck, J.L., Fournier, J.M., Dougan, G., Grimont, P.a.D., Parkhill, J., Holt, K.E., Piarroux, R., Ramamurthy, T., Quilici, M.L., and Thomson, N.R. (2017). Genomic history of the seventh pandemic of cholera in Africa. *Science* 358**,** 785-789.

Westrich, J.R., Ebling, A.M., Landing, W.M., Joyner, J.L., Kemp, K.M., Griffin, D.W., and Lipp, E.K. (2016). Saharan dust nutrients promote *Vibrio* bloom formation in marine surface waters. *Proc Natl Acad Sci U S A* 113**,** 5964-5969.

Westrich, J.R., Griffin, D.W., Westphal, D.L., and Lipp, E.K. (2018). *Vibrio* Population Dynamics in Mid-Atlantic Surface Waters during Saharan Dust Events. *Frontiers in Marine Science* 5.

White, P.B. (1938). The rugose variant of vibrios. *J Pathol Bacteriol* 46**,** 1-6.

Yen, M., and Camilli, A. (2017). Mechanisms of the evolutionary arms race between *Vibrio cholerae* and Vibriophage clinical isolates. *Int Microbiol* 20**,** 116-120.

Zerbino, D.R. (2010). Using the Velvet de novo assembler for short-read sequencing technologies. *Curr Protoc Bioinformatics* Chapter 11**,** Unit 11.15.

Zhang, S.Y., Tsementzi, D., Hatt, J.K., Bivins, A., Khelurkar, N., Brown, J., Tripathi, S.N., and Konstantinidis, K.T. (2019). Intensive allochthonous inputs along

the Ganges River and their effect on microbial community composition and dynamics. *Environ Microbiol* 21, 182-196.

Zhang, Y., Zheng, Z., Chan, E.W., Dong, N., Xia, X., and Chen, S. (2018). Molecular Characterization of *qnrVC* Genes and Their Novel Alleles in *Vibrio* spp. Isolated from Food Products in China. *Antimicrob Agents Chemother* 62.

List of Publications

Rashed, S.M., Hasan, N.A., Alam, M., Sadique, A., Sultana, M., Hoq, M.M., Sack, R.B., Colwell, R.R., and Huq, A. (2017). *Vibrio cholerae* O1 with Reduced Susceptibility to Ciprofloxacin and Azithromycin Isolated from a Rural Coastal Area of Bangladesh. Front Microbiol 8, 252.

Choi, S.Y., Rashed, S.M., Hasan, N.A., Alam, M., Islam, T., Sadique, A., Johura, F.T., Eppinger, M., Ravel, J., Huq, A., Cravioto, A., and Colwell, R.R. (2016). Phylogenetic Diversity of *Vibrio cholerae* Associated with Endemic Cholera in Mexico from 1991 to 2008. MBio 7, e02160.

Alam, M., Rashed, S.M., Mannan, S.B., Islam, T., Lizarraga-Partida, M.L., Delgado, G., Morales-Espinosa, R., Mendez, J.L., Navarro, A., Watanabe, H., Ohnishi, M., Hasan, N.A., Huq, A., Sack, R.B., Colwell, R.R., and Cravioto, A. (2014). Occurrence in Mexico, 1998-2008, of *Vibrio cholerae* CTX+ El Tor carrying an additional truncated CTX prophage. Proc Natl Acad Sci U S A 111, 9917-9922.

Rashed, S.M., Azman, A.S., Alam, M., Li, S., Sack, D.A., Morris, J.G., Jr., Longini, I., Siddique, A.K., Iqbal, A., Huq, A., Colwell, R.R., Sack, R.B., and Stine, O.C. (2014). Genetic Variation of *Vibrio cholerae* during Outbreaks, Bangladesh, 2010-2011. Emerg Infect Dis 20, 54-60.

Rashed, S.M., Kyle, B., Choi, S.Y., Hasan, N.A., Ceccarelli, D., Alam, M., Sack, D.A., Hoq, M.M., Huq, A., and Colwell, R.R. Phylodynamics of *Vibrio cholerae* O1 isolated in Rural Coastal area of Bangladesh. *Manuscript in preparation.*

Rashed, S.M., Hasan, N.A., Kyle, B., Alam, M., Sack, D.A., Hoq, M.M., Huq, A., and Colwell, R.R. Megagenomic analysis of surface water samples collected in Rural Coastal area of Bangladesh. *Manuscript in preparation.*