Putting *Vibrio cholerae* in context: the ecology and evolution of the Cholerae clade

By

Mohammad Tarequl Islam

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Microbiology and Biotechnology

Department of Biological Sciences
University of Alberta

ABSTRACT

Phylogenomic structuring of the *Vibrio cholerae* clade is an important area of study in the context of the alarming spike of human morbidities due to environmental pathogens around the world. Even after decades of research, the structuring of natural populations of *V. cholerae* and its sister species is poorly understood. Ecological and epidemiological importance of the clade shouts for a sound understanding of the phylogenetic relationships and environmental dynamics within the clade in their natural settings. In this thesis, my goal is to answer few broad questions on the biology of the *V. cholerae* clade: i) How different is the evolutionary trajectory of pathogenic lineages from environmental populations ii) To what extent environmental and biogeographic factors impact the intra-species diversity of the model species: *V. cholerae* iii) Is our understanding of the phylogeny and ecology of the *V. cholerae* clade consistent with the inclusion of the recently discovered sister species of *V. cholerae* from both environmental and clinical settings. To address these questions, I have employed culture-based techniques, phenotypic characterization, high throughput amplicon sequencing, whole genome sequencing, quantitative PCR (qPCR) as well as phylogenetic and taxonomic inference tools. I have found potential signatures in *V. cholerae* population structure in cholera endemic and non-endemic regions. I have also shown the impact of environmental factors on the intra-species diversity of *V. cholerae* population naturally occurring in a cholera endemic area. Furthermore, I am describing here the genetic, phylogenetic and biological traits, as well as interaction of a novel sister species with *V. cholerae*: *V. paracholerae* sp. nov. Moreover, I delineate the phylogenetic and molecular spectrum of a novel sister species of *V. cholerae*: *V. tarriae,* isolated from human and environmental samples only from cholera free locations in the USA. The studies presented herein consolidates the understanding of the natural diversity and ecology of the clade containing *V. cholerae* and its close relatives at species and sub-species level, which could be a significant stride in our knowledge on this group of microorganisms.

PREFACE

Some of the research conducted for this thesis are collaborative works, and individual author contributions are listed below for each chapter. Chapters 2, 3, and 4 have been published in peer-reviewed journals and are included in this thesisin practically unaltered forms (except for minor changes and formatting).

1. A version of Chapter 2 has been published as:
   "**Islam MT**, Alam M, Boucher Y. Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *International Microbiology*. 2017 Sep;20(3):106-115. doi: 10.2436/20.1501.01.291. "
   - M.T.I and Y.B. designed the study; M.T.I performed evolutionary molecular data analyses; M.T.I wrote the manuscript; Y.B. and MA. reviewed the manuscript.

2. A version of Chapter 3 has been published as:
   "**Islam MT**, Nasreen T, Kirchberger P, Liang KYH, Orata FD, Johura FT, Hussain NAS, Im MS,Tarr CL, Alam M, Boucher YF. Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp. nov.) with a history of association with humans. *Applied and Environmental Microbiology*. 2021. AEM0042221. doi:10.1128/AEM.00422-21."
   - M.T.I., T.N., M.A. and Y.F.B. designed the experiments. M.T.I., T.N. and F.J. performed data collection. M.T.I., P.K., K.Y.H.L. and F.O. performed data analysis. M.T.I., K.Y.H.L. and F.O performed bioinformatics analysis. M.I. performed genome sequencing and supplied meta data of isolates from CDC. M.A. and C.L.T. provided isolates used in this study. M.T.I. wrote the manuscript. Y.F.B. and C.L.T. provided critical reviews on the manuscript.

3. A version of Chapter 4 has been accepted for publication in *Microbial Ecology* as:
   "TaniaNasreen, **Mohammad Tarequl Islam***, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C.

Kirchberger, Eric Hill, Marzia Sultana, Rebecca J. Case, Munirul Alam and Yann F. Boucher. Dynamic subspecies population structure of *Vibrio* cholerae in Dhaka, Bangladesh. * Joint first author

- YB, TN, MTI and MA designed the project. TN, MTI and YB wrote the manuscript. MTI, FTJ, MS and MA helped perform sample collection and sample processing during field trips in Dhaka, Bangladesh. TN performed the qPCR and MTI did the amplicon sequencing. TN and MTI did the data analysis. KYHL helped in bioinformatics analysis. MTI, YB, FTJ, MS, RJC and MA reviewed the manuscript. MA and YB supervised the project.

4. A version of Chapter 5 is in preparation for publication in *Frontiers in Microbiology* as: "**Mohammad Tarequl Islam**, Kevin Liang, Monica S. Im, Jonathan Winkjer , Shelby Busby , Cheryl L. Tarr, and Yann F. Boucher *Vibrio tarriae*: A lactose-fermenting relative of *Vibrio cholerae* associated with opportunistic infections in humans."

    - M.T.I., and Y.F.B. designed the experiments. M.T.I., K.Y.H.L and F.O performed performed data analysis. M.T.I. wrote the manuscript. Y.F.B. and C.L.T. provided critical reviews on the manuscript.

Beyond the research conducted for this thesis, I was also involved in collaborative works that led to several publications. Appendix A provides a complete list of authored and coauthored works during my doctorate program.

ACKNOWLEDGEMENT

All praise and gratitude belong to the God who has created us and given us the opportunity to study this amazing world, every corner of which is rich in fascinating stories.

Firstly, I would like to thank my supervisor, Dr. Yann Boucher, for accepting me as a graduate student in your lab. You have been a fantastic supervisor and mentor. I have learnt so much from you over the years, not only about research and academic arena but also about the world outside our bubble. I would also like to thank Dr. Rebecca Case for the valuable insights during our lab meetings and various points of my PhD research. I thank my supervisory committee, Dr. Brian Lanoil, Dr. Felix Sperling, Dr. Norman Neumann and Dr. Alexander Kirschner for all your feedback and suggestions for the improvements of my thesis.

I am truly thankful to my lab mates: Dr. Paul Kirchberger, Dr. Fabini Orata, Dr. Tania Nasreen, Nora Hussain, and Kevin Liang for their significant contributions in my thesis. I am immensely grateful for all the collaborative work that was instrumental for generating data and completing my goals throughout this thesis, particularly, Dr. Munirul Alam (International Centre for diarrheal disease research) and Dr. Cheryll Tarr (Centers for Disease Control and Prevention) for their guidance and feedback on the project designs and manuscripts.

A big appreciation goes to my Edmonton family and students from the Bangladeshi community with whom I shared the ups and downs of our migrant lives with fun, affection, and compassion. I hope our paths will cross again in the near future.

Finally, thanks to my family for your encouragement in various forms. It has been a long journey so far. Thank you for staying beside me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction –Diversity, Ecology and environmental adaptation of vibrios

CHAPTER 1

1.1. *Vibrio*: a diverse group of environmental bacteria

The bacterial family Vibrionaceae is a diverse assemblage of aquatic organisms which represents a significant portion of oceans biomass and an important player of the earth's biogeochemical cycle (Hunt et al., 2008; Farmer et al., 2006). Members of this family are cosmopolitan in their distribution in marine environments, being present in salt marshes, deep-sea sediments, throughout the water column and associated with marine flora and fauna (Takemura et al., 2014). They are the most significant consumers of chitin (β-1,4-linked N-acetylglucosamine), the second most abundant polymer in the ocean (Pruzzo et al., 2008). Recycling of chitin by Vibrionaceae makes it an extremely important player to ocean health and also a major contributor to their capacity to flourish in marine environments (Hunt et al., 2008). The genus *Vibrio* within the Vibrionaceae family includes some of the best studied models for marine heterotrophic bacterial groups (Farmer III et al., 2015). The abundance, community structure and metabolic activities of this group have potential impacts on the local and global ecosystems. On top of that, this genus includes some environmental pathogens causing moderate to severe, life-threatening diseases to human and animals (Baker-Austin et al., 2018). *Vibrio* species are also important model organisms for the study of bacterial population biology, microbial adaptations, biofilm formation, natural competence, symbiosis, and quorum sensing pathways (Thompson et al., 2004; Boyd et al., 2015). Hence, understanding the ecological interactions in the natural habitat is instrumental to estimate and prevent the potential environmental and health impacts of this group of microorganisms on human lives.

1.2. Ecological adaptation of vibrios

Vibrios are thought to have evolved in the marine environment. However, most studied vibrios have shown signs of adaptation to diverse habitats: marine, estuarine and freshwater ecosystems. Underscoring their adaptability, species within the genus *Vibrio* can be found surviving in sediment, attached to biofilms on surfaces, free-swimming in the water-column or attached to or live associated with other organisms (Thompson et al., 2004; Takemura et al., 2014). Environmental conditions influence the habitat specialization as well as transient utilization of alternate niches (Takemura et al., 2014; Franco et al., 2020). An important strategy of vibrios to survive in harsh conditions is to attain a state of "viable but non culturable" (VBNC) which has been reported for many vibrios in the wild (Fakruddin et al., 2013). Another crucial aspect of the lifestyle of vibrios in the environment is the formation of biofilm which provides protection and means of survival in low nutrient or adverse conditions (Lutz et al., 2013; Sultana et al., 2018). Inresponse to environmental cues, change of physiology can occur, for instance alteration from free-swimming cells to "swarmer cells" that prosper in more viscous environments (Farmer III et al., 2015). The ecological diversity of the genus has been thought to be propelled by both adaptive changes in different habitats and horizontal gene transfer (HGT), the process of gaining genetic material from sources other than by the vertical transmission of DNA from parent to offspring (Thompson et al., 2004; Takemura et al., 2014). Vibrios are known for high level of genome plasticity which might have contributed to the adaptation to diverse habitats (Thompson et al., 2004; Le Roux and Blokesch, 2018). Horizontal gene transfer (HGT) can enable a subset of bacteria to have some competitive advantage over others in the population and exploit some alternative host and niches. For example, most pathogenic vibrios are thought to gain their

virulence factors via HGT that enable them to exploit a potential alternative habitat like human or animal body (Islam et al., 2017; Le Roux and Blokesch 2018).

1.3. Environmental abundance and distribution:

Vibrios are cosmopolitan in their distribution but tightly regulated by the environment. Environmental factors are major contributors in shaping the prevalence, physiological states, population structures, host-associations, virulence to other animals of this group of microorganisms. Salinity and temperature are the two major factors which stood out in previous studies as the defining factors for vibrio abundance (Lipp et al., 2002; Jutla et al., 2013). Vibrios are geographically spread all over the world but are more abundant in warmer waters (Lipp et al., 2002; Vezzulli et al., 2010). Temperature influences ecological interactions of vibrios and hence the abundance, activity and ability to impact other animals living in the environment (Lutz et al., 2013). The temporal fluctuation spectrum of abundance varies in tropical and temperate conditions. In temperate regions, temperature becomes the strongest driver of abundances of vibrios as the population dynamics appears to be directly coupled to water temperature (Blackwell and Oliver, 2008). Recent reports suggests that global climate change inducing water temperature increases is promoting proliferation of vibrios, particularly in temperate aquatic regions, and the effect is manifested by the rise of vibrio-associated illness cases in human (Vezzulli et al., 2010; Baker-Austin et al., 2018). However, during a large part of the year the water-column in temperate areas is too cold for many vibrios to proliferate, and thus use alternate strategies to survive in harsh environments and propagate in a suitable period of the year

Lipp et al., 2002; Colwell, 2004). Other environmental parameters have not been found to correlate with vibrio abundance consistently across the studies. Dissolved oxygen (DO), nitrogen, phosphate, pH, turbidity, dissolved organic carbon (DOC), total dissolved solids (TDS) measurements have been found occasionally linked to the vibrio abundance (Kirschner et al., 2008; Stamatakis, 2014; Takemura et al., 2014; Vezzulli et al., 2016). Together temperature and salinity has been found to be responsible for ~50% of the variation in the abundance of vibrios (Froelich and Daines, 2020). Temporal changes of these parameters have been found to have significant effect on the abundances of several vibrios in varying degrees depending on their tolerance limit of temperature and salinity (Huq et al., 2005; Kirschner et al., 2008). Warmer temperature helps the growth of these bacteria in the water and therefore global warming increases the probability of vibrio blooms, which can pose additional health risk to the human population (Huq et al., 2005). Temperature also has a role in virulence gene expression and there is evidence that warmer temperature induces switching on of virulence regulatory genes in several vibrios (Kimes et al., 2012; Weber et al., 2014). These bacteria usually have short generation time and, thus are capable of rapidly increasing the population number in response to drastic events like heatwaves, flooding, and nutrient enrichments (Lipp et al., 2002; Colwell, 2004).

1.4. Human-associated vibrios:

The study of vibrios gained significant importance largely because of the association of certain species with human and animal diseases. Out of the >100 described *vibrio* spp., at least 12 were isolated from clinical cases, hence thought to cause infections in humans (Table 1.1)

(Thompson et al., 2004; Takemura et al., 2014). These pathogenic vibrios can be routinely isolated from the coastal and brackish water around the world and can cause disease in human via ingestion of natural untreated water or contaminated aquatic organisms (Baker-Austin et al., 2018; Froelich and Daines, 2020). In the context of global warming, emergence of pathogenic vibrios in the environment is alarming and demands extra attention. The genus contains two species which are associated with pandemics in the global human population. Cholera, the historic pandemic disease which has shaken human civilization at least seven times in recorded history is caused by *Vibrio cholerae* (Colwell, 2004; Farmer III et al., 2015). *V. cholerae*, the type species of the genus *Vibrio*, was first described by Filippo Pacini in 1854. *V. cholerae* is still associated with the 7th pandemic of cholera and numerous epidemics around the world (will be discussed in more details in chapter 2). Due to the overwhelming impact of cholera, human diseases caused by the species of this genus have been divided in two major groups: cholera and non-cholera infections. Non-cholera *Vibrio* spp., such as *V. parahaemolyticus* and *V. vulnificus*, cause 'vibriosis', a group of infections with varying clinical manifestations depending on the pathogenic species, route of infection and host susceptibility (Baker-Austin et al., 2018). *V. parahaemolyticus* is one of the most important food-borne pathogens and the leading cause of human acute gastroenteritis, primarily following the consumption of raw, undercooked or contaminated seafood (Takemura et al., 2014). A subset of *V. parahaemolyticus* is called pandemic type for being globally spread and involved in human disease (Thompson et al., 2004; Takemura et al., 2014). The disease outcome from *V. parahaemolyticus* infection can be self- limiting to severe depending on the immunity level of the affected person. As an aftermath, the organism has also become a major threat to the shellfish and oyster industries (Chiang a nd Chuang, 2003) around the world.

Another well-known human pathogen from the genus is *V. vulnificus*. It can cause gastroenteritis when ingested and life-threatening septicemia when associated with wound infections (Farmer III et al., 2015; Baker-Austin et al., 2018). The bacteria is also known as "flesh eating bacteria" as wound infections by *V. vulnificus* can lead to bullous cellulitis and more serious outcomes like amputations of limbs and fatal septic shock due to multi organ failure. *V. damselae* was originally described as the causative agent of ulcerative disease in damselfish and found to cause wound infections in humans (Farmer III et al., 2015). In rare occasion, serious complication might arise as a form of fulminant septicemia, which mimics the one caused by *V. vulnificus* (Shin et al., 1996; Farmer III et al., 2015). *V. mimicus* was discovered during a DNA–DNA hybridization study of "biochemically atypical" strains that were originally characterized as "*Vibrio cholerae* sucrose +" and were highly related to, but distinct from, *V. cholerae* by DNA–DNA hybridization (Davis et al., 1981; Farmer III et al., 2015). It was later isolated from human clinical specimens and found to cause human gastroenteritis characterized by diarrhea, nausea, vomiting, abdominal cramps, and fever (Hasan et al., 2010).

V. fluvialis has been described as a novel species in 1981 (Lee et al., 1981) and found to be associated with diarrhea, gastroenteritis and in rare occasions, extraintestinal infections in human (Ramamurthy et al., 2014). *V. furnissii* is very closely related to *V. fluvialis* in terms of phenotypic and metabolic traits, and shows similar epidemiological patterns (Schirmeister et al., 2014). The two sister species could only be differentiated using molecular techniques (Schirmeister et al., 2014). The disease caused by the two species typically occurs after ingesting contaminated raw or undercooked seafood or after contact with warm marine environments (Jiang and Fu, 2001) associated to human infections in diverse geographic locations and might

have a human associated life cycle. *V. cincinnatiensis* is a halophilic species but was first isolated by Brayton et al. (Brayton et al., 1986) from blood and cerebral spinal fluid of a 70-year-old man with no known contact with seafood or saltwater; biochemical testing and the 5S rRNA sequence led to the description as new species (Farmer III et al., 2015). Since then, the species has been rarely isolated from human clinical specimens and its status as human enteric or wound pathogen is unclear (Farmer III et al., 2015; Jäckel et al., 2020). *V. hollisae* was first isolated from human feces and has been found to be associated with sporadic cases of diarrhea in humans (Hickman et al., 1982); and has also been isolated from blood (Farmer III et al., 2015). However, the causal role and pathogenic mechanisms of this organism to human disease is not well documented yet (Farmer III et al., 2015). *V. metschnikovii* was first isolated from a fowl that had died from a "cholera-like" disease and later frequently isolated from fresh, brackish, and marine waters (Wallet et al., 2005; Farmer III et al., 2015). It is widely distributed in the environment and found to cause opportunistic human infection leading to complications like septicemia and pneumonia (Wallet et al., 2005). *V. harveyi* is a marine bacterium which is a serious pathogen for aquatic animals and found to cause occasional human wound infections (Farmer III et al., 2015). *V. alginolyticus* is a halophilic organism which is found abundantly in the marine water. It rarely causes gastrointestinal tract infections and usually does superficial wound, ear (otitis media, otitis externa, myringitis), and conjunctival infections in humans that get exposed to the marine water (Citil et al., 2015). Beside these, the work in this thesis proposes two novel to cholera caused by toxigenic *V. cholerae*. genomic and clinical data presented here suggest that strains belonging to this clade had been associated to human infections in diverse geographic locations and might have a human associated life cycle.

1.5. Epidemiology of vibrio-associated infections:

An alarming trend of increasing number of infections due to vibrios have been observed around the globe in recent years. Warming of ocean waters are thought to be a major contributor to this rise. Epidemiological data suggests that the dynamics of vibrio infections are tightly linked with environmental conditions and seasonal fluctuations. For example, in USA the "vibrio season" is between May and October, when the risk of getting infected is high. However, global warming can change this scenario and prolong the high-risk season.

Table 1.1. Human associated vibrios

Species	Source of Infection	Major Route of Infection	Major Clinical manifestation
Vibrio cholerae	Fresh water, Sea water, Food	Oral	Cholera and gastroenteritis
Vibrio parahaemolyticus	Sea Food	Oral	Gastroenteritis
Vibrio vulnificus	Sea water, Sea Food	Oral, wound exposure	Gastroenteritis and wound infections
Vibrio alginolyticus	Sea Water	Wound exposure	Most commonly ear and wound infections
Vibrio fluvialis	Sea Water	Oral, wound exposure	Gastroenteritis

Vibrio hollisaea	Sea Food, sea water	Oral	Gastroenteritis and wound infections
Vibrio mimicus	Sea Water	Oral, wound exposure	Gastroenteritis
Vibrio metschnikovii	Sea Water	Oral	Gastroenteritis and sepsis
Vibrio damsela	Sea Water	Wound exposure	Wound infection, septicemia
Vibrio furnissii	Estuarine water	Oral	Gastroenteritis
Vibrio cincinnatiensis	Sea water, estuarine water	Unclear	Unclear

The ideal growth conditions for major pathogenic vibrios in the sea in terms of temperature and salinity are > 15 ℃ and moderate salinity (<25%). Thus, the expansion of the period having this condition potentially makes more days of the year risky for infections caused by aquatic exposure and via consumption of seafood. Additionally, sudden heatwaves can drastically increase the likelihood of infections, as the number of organisms and virulence increase with the rise of temperature. There has been reports of significant increase of vibrio-associated infections in different parts of the world due to heatwaves (Vezzulli et al., 2016; Baker-Austin et al., 2017). Floodwater intrusion can change the environmental conditions (i.e. salinity, nurtient content etc.) often making it more favorable to the

vibrio growth and potential blooms. After major floods and hurricanes, rapid changes to the vibrio community and associated infection rate have been reported (Liang and Messenger, 2018; Neogi et al., 2018). Climate change can also be related to the rise of sea-levels which can lead to the spread of these pathogens in broader geographic areas. El Nino, a cyclical natural weather phenomenon, has been found to impact both the population of vibrios in the sea and associated disease scenarios (Colwell, 2004; Logar-Henderson et al., 2019). The CDC estimated that the average annual incidence of all vibrio infections had increased by 41% between 1996 and 2005 in the USA. Interestingly, of all the major bacterial foodborne pathogens, vibrios are the only group that are currently increasing in incidence in the USA, indicating an emerging disease trend (Baker-Austin et al., 2017). In the global scenario, infections caused by vibrios are being reported from different localities in several continents which did not have any known history of those diseases (Baker-Austin et al., 2017). Climate warming appears to be playing a significant role in the expansion of the disease-causing impact of pathogenic vibrios (Baker-Austin et al., 2017). Laboratory observation supports the idea that pathogenic vibrios grow (Vezzulli et al., 2010; Mansour and Esseku, 2017) better in low-salinity warm water. Climate change can have direct impacts on marine ecosystems and higher life form like planktons. Planktons represent mobile nutrient-rich micro environments which can selectively enrich aquatic hetero trophic bacteria like vibrios (Lipp et al., 2002). Their ability to degrade chitin efficiently allows vibrios to utilize the chitinous exoskeletons of some plankton taxa as a source of carbon and nitrogen which can provide them with a competitive advantage in planktonic microhabitats (Turner et al., 2009). Plankton colonized by pathogenic Vibrio species can act as an indicator of their seasonal abundance and the frequency of vibrio infection (Turner et al., 2009). A study of long-term plankton datasets obtained using the continuous plankton recorder (CPR) from the 1960s onwards, identified

a significant increase in vibrio abundance in the North Sea during the 1980s onwards, which corresponded both temporally and spatially with an increase in sea surface temperature in the area (Vezzulli et al., 2012). In Canada, there was a significant increase in *Vibrio* cases between 2008 and 2015 which was correlated with the rise of sea surface temperature (Galanis et al., 2020). A recent outbreak of cholera from herring egg in Vancouver Island of Canada further draws attention to the alarming scenario of the increased likelihood of vibrio infection around the globe. In European countries as well, number of reported cases of vibrio infections are increasing and potentially pathogenic Vibrios are being isolated from environmental and clinical samples in increasing numbers (Kirschner et al., 2008; Roux et al., 2015). Prediction of future scenarios based on climate modelling suggest that these emerging pathogens are likely to continue to contribute to human diseases and pose a sustained public health threat (Escobar et al., 2015; Baker-Austin et al., 2017). However, other factors could also contribute to the observed increasing rates of vibriosis: an increase in farming and consumption of shellfish, an increase in population density and tourism in coastal regions, and improvements in surveillance, diagnosis, testing and reporting (Baker-Austin et al., 2017, 2018). Extreme climatic events have also been linked to an increase in reported infections (Baker-Austin et al., 2017). A sharp increase in vibrio wound infections was reported following Hurricane Katrina in USA in August 2005. *V. vulnificus* infections reports increased significantly in New Caledonia in 2008 following heavy precipitation in the area (Baker-Austin fertile conditions for this bacterium to thrive and may well have been a contributory factor for these infections. In coastal Bangladesh, frequent natural disasters have been found to impact vibrio community and likelihood of disease outbreaks in coastal villages (Jutla et al., 2013; Neogi et al., 2018).

1.6. Population biology of vibrios:

Populations are the units of diversity to study ecology of organisms, their ecological interactions, and the ecosystems. Population biology has enabled us to understand the diversity and use the knowledge to analyze ecological structuring of higher animals and plants. But for microorganisms, populations have been a very challenging concept to define (Rocha, 2018). The absence of a biologically meaningful species concept for microorganisms has left us using arbitrary units of diversity to analyze microbial populations so far (VanInsberghe et al., 2020). As a result, many of the fundamental concepts of ecology cannot be applied to approach very important questions in microbial ecology. In sexually reproducing organisms, species are generally defined as groups of interbreeding populations, and reproductive barrier created by various factors leads to the emergence of a new species (speciation) from the common ancestor (Coyne, 1992). This kind of barrier is not applicable to microorganisms, as horizontal gene transfer is very common even within phylogenetically distant bacterial groups (Polz et al., 2013). There is also rarity of phenotypic properties to make ecologically meaningful units of bacteria and archaea (Georgiades and Raoult, 2011). Thus, efforts have been made to define microbial populations into ecologically coherent units based on genetic, phylogenetic and ecological properties alongside distinguishing phenotypic traits (Rosselló-Mora and Amann, 2001). This approach, known as phylo-phenetic concept, proposes that a bacterial species is "a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity with respect to many independent characteristics, and is diagnosable by a discriminative phenotypic property" (Rosselló-Mora and Amann, 2001; Riley and Lizotte-Waniewski, 2009). Application of this approach has led to significant refinement of the

taxonomy of *Vibrio*, as well as discovery of novel species within the genus under this practical framework. The estimation and observation of diversity of vibrios has also seen a new era after advanced molecular tools became the mainstay to define bacterial species. The advent of high-throughput sequencing technologies revealed the staggering discrepancies between the number of species descriptions and the actual number of species living on earth (Goodwin et al., 2016; Sant'anna et al., 2020). Knowledge on *Vibrio* diversity also has started to be redefined with the use of matrices taking advantage of genome sequencing data. After 2000, 111 species were either redefined or newly been added as distinct species in the genus *Vibrio* (https://lpsn.dsmz.de/genus/vibrio). Previously, bacterial species would be defined based on phenotypic characteristics only which had led to many discrepancies in taxonomic designation of many species. Rita Colwell first proposed the polyphasic taxonomy for defining bacterial species to be used for classifying Vibrio species and identify new species (Colwell, 1970). This method/concept implies the integration of information coming from different levels and all available sources like phenotypic, genetic, ecological etc. Since then, DNA-DNA hybridization (DDH) data has been used as a mainstay to infer the vibrio species boundaries in the polyphasic taxonomy. DDH is still the gold standard test for defining new species (Auch et al., 2010). However, the conventional DDH technique suffers from some limitations in coping up with the expansion of environmental and biodiversity research. The technique is time consuming and tedious, needs the inclusion of reference strains in each new experiment (Owen and Pitcher, 1985). In addition, it involves non-portable special facilities available in a limited number of international laboratories which makes it a difficult technique to be

used widely around the world. Carl Woese's revolutionary breakthrough discovery involving the usage of 16S ribosomal RNA (rRNA) gene sequences offered a reproducible method for taxonomic classification of bacteria (Woese and Fox, 1977). Even though 16S sequence-based species criteria is not sufficient alone, because of some inherent limitations (Sant'anna et al., 2020), it's inclusion was important for a comprehensive approach of defining a novel species. With the inclusion of more genomic and phylogenetic data, a modified definition of a bacterial species became as 'a group of strains (including the type strain), having > 70% DDH similarity, < 5°C 'Tm, < 5% mol G +C difference of total genomic DNA, > 97% 16S rRNA identity' (Moore et al., 1987; Stackebrandt and Goebel, 1994). Current consensus in taxonomy is when an isolate of a given organism shows preliminary promise to differ from other organisms which is already assigned with a taxonomic designation, a comprehensive description of the organism is pursued. To check if the potentially novel species is sufficiently distinct from the closely related taxa, a polyphasic approach including all available information is used (Thompson et al., 2009). This approach encompasses phenotypic analysis, which examines the morphological, metabolic, physiological, and chemical characteristics; genotypic analysis which considers characteristics of the genome and phylogenetic analysis that seeks to place framework (Sant'anna et al., 2020).

1.7. The "Cholerae clade":

There are twelve genera within the family Vibrionaceae and *Vibrio* is the most prominent genus containing more than 110 recognized species (Boyd et al., 2015). The genus name *Vibrio* was coined by Pacini in 1854 during his studies on cholera, and it is one of the oldest names for a bacterial genus (Barcat, 2014). Members of the genus *Vibrio* are diverse in their ecology and pathogenic attributes. Vibrios generally have a requirement for salt, the

concentration of which varies for different species; sodium ion stimulates growth for all species in the genus and is an absolute requirement for most (Farmer III et al., 2015). This property provides a means of separating pathogenic vibrios into the non-halophilic species which grow on nutrient agar, and the halophilic species that require a salt supplement in their growth media. The taxonomic structure of the genus is being fine-tuned to accommodate the availability of new molecular and ecological evidence. Biological information on the close relatives of a dangerous environmental pathogen like *V. cholerae* is of great significance, because of their potential as emerging pathogens themselves and their interaction with *V. cholerae* in its natural habitats. It is also useful to define ecologically meaningful clades for future diversity analysis, tracking of potential pathogens and large-scale ecological studies. It's been proposed that the species *V. cholerae* and *V. mimicus*, for their non-halophilic nature, would be the basis of a modified more limited genus Vibrio; whereas the *Vibrio* species not closely related to *V. cholerae* will form the basis of many new genera (Farmer III et al., 2015). Recent discovery and description of a few new species closely related to *V. cholerae* and *V. mimicus* would be included in the potential new genera, including those described later in this thesis. So far, they are included in a un-official but practical group called "The Cholerae clade" (Boyd et al., 2015; Gregory and Boyd, 2021) with few other related species bearing similar properties. The Cholerae clade, named after the most famous and prominent member of the clade, *V. cholerae*, consists of species that are distinguishably more related to *V. cholerae* than any other lineages. This clade has been of significant interest, as *V. cholerae* is one of the most dangerous pathogens to human in history and these closely related species have been found to interact with *V. cholerae* in nature (Hasan et al., 2010; Orata et al., 2015).

Table 1.2. Comparison of significant traits among the potential members of the Cholerae clade

Species	Growth in 0% Nacl	Compared to *V. cholerae* type strain			GC %	Taxonomic status	Reference of the genome
		dDDH(%)	ANI(%)	16S Identity(%)			
V. cholerae	Yes	100	100	100	48.29	Type species	(Heidelberg et al., 2000)
V. paracholerae	Yes	67.6	95.53	99.61	48.14	Official species	(Islam et al.)
V. tarriae	Yes	63.8	95.13	99.94	47.21	Proposed species	This study
V. parilis	Unknown	37.8	86	99.94	46	Unofficial species	(Haley et al., 2010)
V. metoecus	Yes	38.8	86.87	99.94	46.87	Official species	(Kirchberger et al., 2014b)
V. mimicus	Yes	36	85.35	99.05	46.38	Official species	(Hasan et al., 2010)
V. fluvialis	No	25.6	76.04	94.01	49.93	Official species	(Schirmeister et al., 2014)
V. furnissiae	No	25.3	76.1	94.01	50.6	Official species	(Schirmeister et al., 2014)

V. hollisae	No	23	71.2	91.89	49.51	Official species	(Hickman et al., 1982)
V. vulnificus	No	25.4	74.39	95.18	46.72	Official species	(Chiang and Chuang, 2003)

1.8. Thesis objectives and outline

The overarching objective of this research is to define the phylogenetic structure and describe the intraspecies and interspecies diversity of the clade containing *V. cholerae* and closely related bacteria in the context of the global cholera pandemics. Chapter two reviews the key components of the ecology of *V. cholerae* and estimates the time of the major evolutionary events for the emergence of pandemic *V. cholerae* from environmental origins in the context of the global *V. cholerae* population. Chapter three describes subspecies level analysis of the populations of *V. cholerae* naturally occurring in cholera endemic Dhaka, Bangladesh and compares that with a cholera-free area (Oyster Pond, Falmouth, USA) using amplicon sequencing, qPCR and whole genome sequencing techniques. A divergent clade abundant in cholera endemic Dhaka is studied using comparative genomics, phylogeny and an array of phenotypic tests leading to the discovery of a novel sister species to *V. cholerae*: *V. paracholerae* (chapter 3). Chapter four analyzes the influence of environmental factors in the spatio-temporal dynamics of subspecies of *V. cholerae*. Lastly, chapter five presents the phenotypic, genotypic and phylogenetic description of a clinically relevant novel species, *V. tarriae*.

1.1.References

Arnosti, C. (2014). Patterns of Microbially Driven Carbon Cycling in the Ocean: Links between Extracellular Enzymes and Microbial Communities. *Adv. Oceanogr.* doi:10.1155/2014/706082.

Auch, A. F., von Jan, M., Klenk, H. P., and Goker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2, 117–134. doi:10.4056/sigs.531120.

Baker-Austin, C., Oliver, J. D., Alam, M., Ali, A., Waldor, M. K., Qadri, F., et al. (2018). *Vibrio* spp. infections. *Nat. Rev. Dis. Prim.* doi:10.1038/s41572-018-0005-8.

Baker-Austin, C., Trinanes, J., Gonzalez-Escalona, N., and Martinez-Urtaza, J. (2017). Non-Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends Microbiol.* 25, 76–84. doi:10.1016/j.tim.2016.09.008.

Barcat, J. A. (2014). [Filippo Pacini and cholera, 1854]. *Medicina (B. Aires).* 74, 77–79.

Blackwell, K. D., and Oliver, J. D. (2008). The ecology of *Vibrio vulnificus*, *Vibrio cholerae*, and *Vibrio parahaemolyticus* in North Carolina Estuaries. *J. Microbiol.* doi:10.1007/s12275-007-0216-2.

Boyd, E. F., Carpenter, M. R., Chowdhury, N., Cohen, A. L., Haines-Menges, B. L., Kalburge, S. S., et al. (2015). Post-Genomic Analysis of Members of the Family Vibrionaceae. *Microbiol. Spectr.* 3, 10.1128/microbiolspec.VE-0009–2014. doi:10.1128/microbiolspec.VE-0009-2014.

Brayton, P. R., Bode, R. B., Colwell, R. R., MacDonell, M. T., Hall, H. L., Grimes, D. J., et al. (1986). *Vibrio cincinnatiensis* sp. nov., a new human pathogen. *J. Clin. Microbiol.*

doi:10.1128/jcm.23.1.104-108.1986.

Chiang, S.-R., and Chuang, Y.-C. (2003). *Vibrio vulnificus* infection: clinical manifestations, pathogenesis, and antimicrobial therapy. *J. Microbiol. Immunol. Infect.* 36, 81–88.

Citil, B. E., Derin, S., Sankur, F., Sahan, M., and Citil, M. U. (2015). *Vibrio alginolyticus* Associated Chronic Myringitis Acquired in Mediterranean Waters of Turkey. *Case Rep. Infect. Dis.* 2015, 187212. doi:10.1155/2015/187212.

Colwell, R. R. (1970). Polyphasic taxonomy of the genus vibrio: numerical taxonomy of *Vibrio cholerae, Vibrio parahaemolyticus*, and related *Vibrio* species. *J. Bacteriol.* doi:10.1128/jb.104.1.410-433.1970.

Colwell, R. R. (2004). Infectious disease and environment: Cholera as a paradigm for waterborne disease. *Int. Microbiol.* doi:10.2436/im.v7i4.9485.

Coyne, J. A. (1992). Genetics and speciation. *Nature* 355, 511–515. doi:10.1038/355511a0.

Davis, B. R., Fanning, G. R., Madden, J. M., Steigerwalt, A. G., Bradford, H. B., Smith, H. L., et al. (1981). Characterization of biochemically atypical *Vibrio cholerae* strains and designation of a new pathogenic species, *Vibrio mimicus*. *J. Clin. Microbiol.* doi:10.1128/jcm.14.6.631-639.1981.

Dorman, M. J., Kane, L., Domman, D., Turnbull, J. D., Cormie, C., Fazal, M. A., et al. (2019). The history, genome and biology of NCTC 30: A non-pandemic *Vibrio cholerae* isolate from World War One. *Proc. R. Soc. B Biol. Sci.* doi:10.1098/rspb.2018.2025.

Escobar, L. E., Ryan, S. J., Stewart-Ibarra, A. M., Finkelstein, J. L., King, C. A., Qiao, H., et al. (2015). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Trop.* 149, 202–211. doi:10.1016/j.actatropica.2015.05.028.

Fakruddin, M., Mannan, K. S. Bin, and Andrews, S. (2013). Viable but Nonculturable Bacteria:

Food Safety and Public Health Perspective. *ISRN Microbiol.* 2013, 703813. doi:10.1155/2013/703813.

Farmer III, J. J., Michael Janda, J., Brenner, F. W., Cameron, D. N., and Birkhead, K. M. (2015). "*Vibrio*," in *Bergey's Manual of Systematics of Archaea and Bacteria* (American Cancer Society), 1–79. doi:https://doi.org/10.1002/9781118960608.gbm01078.

Farmer, J. J. (2006). "The Family Vibrionaceae," in *The Prokaryotes: A Handbook on the Biology of Bacteria Volume 6: Proteobacteria: Gamma Subclass*, eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (New York, NY: Springer New York), 495–507. doi:10.1007/0-387-30746-X_17.

Franco, A., Rückert, C., Blom, J., Busche, T., Reichert, J., Schubert, P., et al. (2020). High diversity of *Vibrio* spp. associated with different ecological niches in a marine aquaria system and description of *Vibrio aquimaris* sp. nov. *Syst. Appl. Microbiol.* doi:10.1016/j.syapm.2020.126123.

Froelich, B. A., and Daines, D. A. (2020). In hot water: effects of climate change on Vibrio–human interactions. *Environ. Microbiol.* 22, 4101–4111. doi:https://doi.org/10.1111/1462-2920.14967.

Galanis, E., Otterstatter, M., and Taylor, M. (2020). Measuring the impact of sea surface temperature on the human incidence of *Vibrio* sp. infection in British Columbia, Canada, 1992–2017. *Environ. Heal.* 19, 58. doi:10.1186/s12940-020-00605-x.

Gardner, A. D., and Venkatraman, K. V. (1935). The antigens of the cholera group of vibrios. *J. Hyg. (Lond).* doi:10.1017/S0022172400032265.

Georgiades, K., and Raoult, D. (2011). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1, 151. doi:10.3389/fmicb.2010.00151.

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* doi:10.1038/nrg.2016.49.

Gregory, G. J., and Boyd, E. F. (2021). Stressed out: Bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from Vibrionaceae. *Comput. Struct. Biotechnol. J.* 19, 1014–1027. doi:https://doi.org/10.1016/j.csbj.2021.01.030.

Haley, B. J., Grim, C. J., Hasan, N. A., Choi, S. Y., Chun, J., Brettin, T. S., et al. (2010). Comparative genomic analysis reveals evidence of two novel *Vibrio* species closely related to V. cholerae. *BMC Microbiol.* doi:10.1186/1471-2180-10-154.

Hasan, N. A., Grim, C. J., Haley, B. J., Chun, J., Alam, M., Taviani, E., et al. (2010). Comparative genomics of clinical and environmental *Vibrio mimicus*. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1013825107.

Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae* . *Nature* 406, 477–483. doi:10.1038/35020000.

Hickman, F. W., Farmer, J. J., Hollis, D. G., Fanning, G. R., Steigerwalt, A. G., Weaver, R. E., et al. (1982). Identification of *Vibrio hollisae* sp. nov. from patients with diarrhea. *J. Clin. Microbiol.* 15, 395–401. doi:10.1128/jcm.15.3.395-401.1982.

Hunt, D. E., Gevers, D., Vahora, N. M., and Polz, M. F. (2008). Conservation of the chitin utilization pathway in the Vibrionaceae. *Appl. Environ. Microbiol.* 74, 44–51. doi:10.1128/AEM.01412-07.

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., et al. (2005). Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* doi:10.1128/AEM.71.8.4645-4654.2005.

Islam, M. T., Alam, M., and Boucher, Y. (2017). Emergence, ecology and dispersal of the

pandemic generating *Vibrio cholerae* lineage. *Int Microbiol* 20, 106–115. doi:10.2436/20.1501.01.291.

Islam, M. T., Nasreen, T., Kirchberger, P., Liang, K. Y. H., Orata, F. D., Johura, F.-T., et al. Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp. nov.) with a history of association with humans. *Appl. Environ. Microbiol.* 0, AEM.00422-21. doi:10.1128/AEM.00422-21.

Jäckel, C., Hammerl, J. A., Arslan, H.-H.-T., Göllner, C., Vom Ort, N., Taureck, K., et al. (2020). Phenotypic and Genotypic Characterization of Veterinary *Vibrio cincinnatiensis* Isolates. *Microorganisms* 8, 739. doi:10.3390/microorganisms8050739.

Jiang, S. C., and Fu, W. (2001). Seasonal abundance and distribution of *Vibrio cholerae* in coastal waters quantified by a 16S-23S intergenic spacer probe. *Microb. Ecol.* doi:10.1007/s00248-001-0029-5.

Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A., et al. (2013). Environmental factors influencing epidemic cholera. *Am. J. Trop. Med. Hyg.* doi:10.4269/ajtmh.12-0721.

Kimes, N. E., Grim, C. J., Johnson, W. R., Hasan, N. A., Tall, B. D., Kothary, M. H., et al. (2012). Temperature regulation of virulence factors in the pathogen *Vibrio coralliilyticus*. *ISME J.* 6, 835–846. doi:10.1038/ismej.2011.154.

Kirchberger, P. C., Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., et al. (2014a). *Vibrio metoecus* sp. nov., a close relative of *Vibrio cholerae* isolated from coastal brackish ponds and clinical specimens. *Int. J. Syst. Evol. Microbiol.* 64, 3208–3214. doi:10.1099/ijs.0.060145-0.

Kirchberger, P. C., Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., et al.

(2014b). *Vibrio metoecus* sp. nov., a close relative of *Vibrio cholerae* isolated from coastal brackish ponds and clinical specimens. *Int J Syst Evol Microbiol* 64, 3208–3214. doi:10.1099/ijs.0.060145-0.

Kirschner, A. K. T., Schlesinger, J., Farnleitner, A. H., Hornek, R., Süß, B., Golda, B., et al. (2008). Rapid Growth of Planktonic *Vibrio cholerae* Non-O1/Non-O139 Strains in a Large Alkaline Lake in Austria: Dependence on Temperature and Dissolved Organic Carbon Quality. *Appl. Environ. Microbiol.* 74, 2004–2015. doi:10.1128/AEM.01739-07.

Le Roux, F., and Blokesch, M. (2018). Eco-evolutionary Dynamics Linked to Horizontal Gene Transfer in Vibrios. *Annu. Rev. Microbiol.* 72, 89–110. doi:10.1146/annurev-micro-090817-062148.

LEE, J. V, SHREAD, P., FURNISS, A. L., and BRYANT, T. N. (1981). Taxonomy and Description of *Vibrio fluvialis* sp. nov. (Synonym Group F Vibrios, Group EF6). *J. Appl. Bacteriol.* 50, 73–94. doi:https://doi.org/10.1111/j.1365-2672.1981.tb00873.x.

Liang, S. Y., and Messenger, N. (2018). Infectious Diseases After Hydrologic Disasters. *Emerg. Med. Clin. North Am.* 36, 835–851. doi:10.1016/j.emc.2018.07.002.

Lipp, E. K., Huq, A., and Colwell, R. R. (2002). Effects of global climate on infectious disease: The cholera model. *Clin. Microbiol. Rev.* doi:10.1128/CMR.15.4.757-770.2002.

Logar-Henderson, C., Ling, R., Tuite, A. R., and Fisman, D. N. (2019). Effects of large-scale oceanic phenomena on non-cholera vibriosis incidence in the United States: implications for climate change. *Epidemiol. Infect.* 147, e243–e243. doi:10.1017/S0950268819001316.

Lutz, C., Erken, M., Noorian, P., Sun, S., and McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. *Front. Microbiol.* doi:10.3389/fmicb.2013.00375.

Mansour, G., and Esseku, H. (2017). Situation analysis of the urban sanitation sector in Ghana. *WSUP - Water Sanit. Urban Poor*.

Moore, W. E. C., Stackebrandt, E., Kandler, O., Colwell, R. R., Krichevsky, M. I., Truper, H. G., et al. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/00207713-37-4-463.

Neogi, S. B., Lara, R., Alam, M., Harder, J., Yamasaki, S., and Colwell, R. R. (2018). Environmental and hydroclimatic factors influencing *Vibrio* populations in the estuarine zone of the Bengal delta. *Environ. Monit. Assess.* 190, 565. doi:10.1007/s10661-018-6925-7.

Orata, F. D., Kirchberger, P. C., Méheust, R., Jed Barlow, E., Tarr, C. L., and Boucher, Y. (2015). The dynamics of genetic interactions between *Vibrio metoecus* and *Vibrio cholerae*, two close relatives co-occurring in the environment. *Genome Biol. Evol.* doi:10.1093/gbe/evv193.

Owen, R., and Pitcher, D. (1985). Current methods for estimating DNA base composition and levels of DNA-DNA hybridization. *Soc. Appl. Bacteriol. Tech. Ser.*, 67–93.

Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* doi:10.1016/j.tig.2012.12.006.

Pruzzo, C., Vezzulli, L., and Colwell, R. R. (2008). Global impact of Vibrio cholerae interactions with chitin. *Environ. Microbiol.* 10, 1400–1410. doi:https://doi.org/10.1111/j.1462-2920.2007.01559.x.

Ramamurthy, T., Chowdhury, G., Pazhani, G., and Shinoda, S. (2014). Vibrio fluvialis: an emerging human pathogen. *Front. Microbiol.* 5, 91. doi:10.3389/fmicb.2014.00091.

Riley, M. A., and Lizotte-Waniewski, M. (2009). Population genomics and the bacterial species

concept. *Methods Mol. Biol.* 532, 367–377. doi:10.1007/978-1-60327-853-9_21.

Rocha, E. P. C. (2018). Neutral Theory, Microbial Practice: Challenges in Bacterial Population Genetics. *Mol. Biol. Evol.* 35, 1338–1347. doi:10.1093/molbev/msy078.

Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67. doi:10.1111/j.1574-6976.2001.tb00571.x.

Roux, F. Le, Wegner, K. M., Baker-Austin, C., Vezzulli, L., Osorio, C. R., Amaro, C., et al. (2015). The emergence of Vibrio pathogens in Europe: ecology, evolution, and pathogenesis (Paris, 11–12th March 2015). *Front. Microbiol.* 6, 830. doi:10.3389/fmicb.2015.00830.

Sant'anna, F. H., Reiter, K. C., Almeida, P. de F., and Passaglia, L. M. P. (2020). Systematic review of descriptions of novel bacterial species: Evaluation of the twenty-first century taxonomy through text mining. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/ijsem.0.004070.

Sawabe, T., Kita-Tsukamoto, K., and Thompson, F. L. (2007). Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. *J. Bacteriol.* doi:10.1128/JB.00693-07.

Schirmeister, F., Wieczorek, A., Dieckmann, R., Taureck, K., and Strauch, E. (2014). Evaluation of molecular methods to discriminate the closely related species Vibrio fluvialis and Vibrio furnissii. *Int. J. Med. Microbiol.* 304, 851–857. doi:https://doi.org/10.1016/j.ijmm.2014.09.001.

Shin, J. H., Shin, M. G., Suh, S. P., Ryang, D. W., Rew, J. S., and Nolte, F. S. (1996). Primary Vibrio damsela Septicemia. *Clin. Infect. Dis.* 22, 856–857. doi:10.1093/clinids/22.5.856.

Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* doi:10.1099/00207713-44-4-846.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. doi:10.1093/bioinformatics/btu033.

Sultana, M., Nusrin, S., Hasan, N. A., Sadique, A., Ahmed, K. U., Islam, A., et al. (2018). Biofilms Comprise a Component of the Annual Cycle of *Vibrio cholerae* in the Bay of Bengal Estuary. *MBio* 9. doi:10.1128/mBio.00483-18.

Takemura, A. F., Chien, D. M., and Polz, M. F. (2014). Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* doi:10.3389/fmicb.2014.00038.

Thompson, C. C., Vicente, A. C. P., Souza, R. C., Vasconcelos, A. T. R., Vesth, T., Alves, N., et al. (2009). Genomic taxonomy of vibrios. *BMC Evol. Biol.* doi:10.1186/1471-2148-9-258.

Thompson, F. L., Iida, T., and Swings, J. (2004). Biodiversity of Vibrios. *Microbiol. Mol. Biol. Rev.* doi:10.1128/mmbr.68.3.403-431.2004.

Turner, J. W., Good, B., Cole, D., and Lipp, E. K. (2009). Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME J.* 3, 1082–1092. doi:10.1038/ismej.2009.50.

VanInsberghe, D., Arevalo, P., Chien, D., and Polz, M. F. (2020). How can microbial population genomics inform community ecology? *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* 375, 20190253. doi:10.1098/rstb.2019.0253.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P. C., Colwell, R. R., Höfle, M. G., et al. (2012). Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *ISME J.* 6, 21–30. doi:10.1038/ismej.2011.89.

Vezzulli, L., Grande, C., Reid, P. C., Hélaouët, P., Edwards, M., Höfle, M. G., et al. (2016). Climate influence on Vibrio and associated human diseases during the past half-century in

the coastal North Atlantic. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1609157113.

Vezzulli, L., Pruzzo, C., Huq, A., and Colwell, R. R. (2010). Environmental reservoirs of *Vibrio cholerae* and their role in cholera. *Environ. Microbiol. Rep.* doi:10.1111/j.1758- 2229.2009.00128.x.

Wallet, F., Tachon, M., Nseir, S., Courcol, R. J., and Roussel-Delvallez, M. (2005). *Vibrio metschnikovii* pneumonia. *Emerg. Infect. Dis.* 11, 1641–1642. doi:10.3201/eid1110.050177.

Weber, G. G., Kortmann, J., Narberhaus, F., and Klose, K. E. (2014). RNA thermometer controls temperature-dependent virulence factor expression in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14241–14246. doi:10.1073/pnas.1411570111.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* 74, 5088–5090. doi:10.1073/pnas.74.11.5088.

CHAPTER 2: Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage

A version of this chapter has been published as:

"Islam MT, Alam M, Boucher Y. Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *International Microbiology*. 2017 Sep;20(3):106-115. doi:10.2436/20.1501.01.291."

M.T.I and Y.B. designed the study; M.T.I performed evolutionary molecular data analyses; M.T.I wrote the manuscript; Y.B. and MA. reviewed the manuscript.

CHAPTER 2

2.1. Abstract

Cholera continues to remain a major health threat globally as causative agent for the pandemic form of the disease, *Vibrio cholerae* emerges in evolved forms to survive and spread into new habitats. Understanding the ecology, evolution and epidemiology of pandemic *V. cholerae* is central to study complex diseases like cholera. In this review, we present recent advancements of our knowledge on emergence and spread of pandemic generating lineage of *V. cholerae* on the light of established eco-evolutionary observations. We investigated specific ecological interactions shaping seasonal cholera, role of innate and acquired genetic and host-mediated factors in pandemic causing ability and evolutionary drivers for the emergence of pandemic generating lineage of *V. cholerae.* On the basis of the current understanding, we outline future threats and required improved biogeographical and genomic study schemes to combat this global problem.

2.2. Introduction

Microorganisms causing human disease often have complex dynamics of transmission, persistence and dispersion within their natural reservoirs (Jones et al., 2008). *Vibrio cholerae*, the causative agent of cholera, is a unique model system to study the effect of such environment-human interactions in shaping a deadly infectious disease from aquatic origin (Lipp et al., 2002). Cholera has been endemic for centuries in many countries in South Asia and Africa, where it occurs almost every year, infecting people encountering the pathogen through consumption of untreated water (Kaper et al., 1995). According to World Health Organization (WHO) reports, there are roughly 1.3 to 4.0 million cases, and 21 000 to 143 000 deaths worldwide due to cholera (WHO Cholera Fact Sheet, 2017). Cholera can also become

pandemic and so far, seven recorded pandemics of cholera have shaken the world since 1817, killing millions of people worldwide and warranting extensive study of the disease (Kaper et al., 1995; Faruque et al., 1998). Unlike some other pandemic infectious diseases, cholera is unlikely to be eradicated. This is because its causative agent, *V. cholerae,* is an autochthonous member of marine and estuarine ecosystems around the world, and there is no clear transmission vector to serve as a means to control its human association cycle (Lipp et al., 2002; Colwell, 2004b). Two main hypotheses have been proposed to explain the global spread of pandemic cholera. The first is that human travel has spread the bacteria from endemic countries to other parts of the world (Shapiro et al., 2016). An alternative hypothesis is that ocean currents and maritime transfer of the pathogenic bacteria are responsible for the spread of cholera, with climatic events (i.e El Nino, global warming) playing a major role in shaping cholera pandemics (Colwell, 2004b; Vezzulli et al., 2012). Regardless of how the causative agent of cholera spreads across the world, occurrence of the disease is significantly influenced by living standard of people in cholera prone regions as well as the climatic conditions. For example, in countries with compromised water supply and sanitation infrastructure, excessive rainfall and flooding can lead to massive cholera epidemics if pathogenic variants of *V. cholerae* are present in the environment (Lipp et al., 2002; Vezzulli et al., 2016). Environmental links are evident in the dynamics of the disease, as cholera incidence in endemic areas usually show seasonal patterns i.e. number of cholera cases increases and reaches a peak in specific months every year (Stine et al., 2008; Jutla et al., 2013). In the environment, *V. cholerae* exists as a diverse species. However, a very small portion of the heterogenous population of *V. cholerae* strains found in nature is capable of causing human disease (Kaper et al., 1995). The outer membrane

lipopolysaccharides of *V. cholerae* have a region named O-antigen, the synthesis of which is encoded by a diverse combination of genes, giving rise to the remarkable diversity of more than 200 serogroups (starting from O1) of *V. cholerae* observed in nature (Chun et al., 2009a). Although isolated infections and outbreaks have been caused by various *V. cholerae* genotypes, pandemics are only caused by strains from a single lineage, most of which display the O1 antigen on their surface (Figure 2.1). Understanding cholera pandemics is therefore dependent on determining ecological characteristics of this specific lineage, not necessarily the entire *V. cholerae* species. This lineage has usually been referred to as *V. cholerae* O1/ O139, which is misleading as it includes several strains displaying other serogroups (such as O37), and that many unrelated harmless strains display the O1 and O139 antigens (Chun et al., 2009a). Here we will refer to the monophyletic phylogenetic group containing all genotypes responsible for cholera pandemics as the Pandemic Generating (PG) lineage to avoid confusion. The factors affecting environmental persistence, survival during inter-epidemic periods, emergence and spread of pathogenic genotypes from this lineage are still poorly understood. This is because population level analysis over a wide range of geographical locations and variety of potential niches requires large-scale sampling and sequencing of a bacteria forming only a small proportion (usually <1%) of natural populations. Most of our current knowledge on *V. cholerae* ecology has been obtained from cultivation-based studies. Despite their limited sampling size, these studies have created a solid foundation to develop culture-independent population-level approaches that will enable investigations of pandemic *V. cholerae* local and global dynamics.

2.3. Seasonality of cholera is mediated by the ecological interactions of naturally occurring *V. cholerae*

V. cholerae has been detected in aquatic habitats from the tropics to temperate waters world-wide (Lutz et al., 2013), underscoring it's highly adaptable and persistent nature over a broad range of environmental conditions (Takemura et al., 2014). However, cholera incidence patterns can vary greatly among geographic locations. It is firmly endemic in some South Asian countries where it appears in distinct seasonal patterns (Emch et al., 2008). Other regions, such as parts of South America and Africa, have historically had only sporadic epidemics of cholera (Emch et al., 2008). In Bangladesh, cholera maintains an annual cycle with two infection peaks; before monsoon and just after monsoon (Stine et al., 2008; Alam et al., 2011). This marked seasonality of cholera appears to be closely linked with the changes in flora and fauna populations in the coastal environment where pathogenic *V. cholerae* exists, mediated by micro- and macro-level environmental factors such as water temperature, salinity, organic matter concentrations, abundance of planktonic surface and water consumption (Lipp et al., 2002; Colwell, 2004b) and ultimately influenced by larger-scale climatic variables (Emch et al., 2008). In the aquatic environment, *V. cholerae* are known to be associated with phytoplankton, zooplankton, chitinous animals, aquatic plants, protozoa, bivalves, fish and water birds (Lutz et al., 2013). These associations could serve as environmental reservoirs where the pathogen can live over time, with the potential to be disseminated and cause cholera outbreaks in nearby human populations (Azarian et al., 2016; Vezzulli et al., 2016). Many of these studies looked at the entire *V. cholerae* species, not specifically the PG lineage. Although the ecological parameters leading to high abundance of PG *V. cholerae* are not well understood, it is clear that during the annual cholera epidemic periods, the conditions in the coastal ecosystems are ideal for the multiplication and transmission of *these bacteria*. As a result, these water sources contain high enough concentration of PG *V. cholerae* to cause human disease upon consumption in sufficient amount

(64, 38). Those months display blooms of phytoplankton, providing food for zooplanktons, both potential resources for growth of *V. cholerae* in water (Huq et al., 2005; Constantin de Magny et al., 2008; Turner et al., 2009). Studies have found significant correlations with seasonal bloom in aquatic microorganisms and cholera incidence rates in nearby populations. During inter-epidemic periods, PG *V. cholerae* remains mostly undetectable by routine microbiological culture-dependent assays based on growth of the targeted bacteria on selective culture media (Alam et al., 2007). However, using culture-independent techniques, including direct fluroscence antibody (DFA) assay and PCR, presence of live PG *V. cholerae* in the water can readily be detected year-round, indicating a survival strategy making them either rare, unable to grow under laboratory conditions, or concentrated in specific reservoirs (host, sediments, particles, etc) (Alam et al., 2007). Indeed, *V. cholerae* possesses the ability to switch into a viable but non-culturable (VBNC) or dormant state in response to nutrient deprivation or other stresses (Colwell, 2000; Alam et al., 2007). In its VBNC state, the *V. cholerae* cells become coccoid (as opposed to their normal curved rod shape) and do not respond readily to typical microbiological medium hence cannot be detected by culture-based surveillance. These non-culturable cells have been found to retain pathogenic ability upon passage through animal intestine (Colwell, 2000). Such a dormant state could serve as a survival strategy in inter-epidemic months, as resuscitation could occur once conditions are favorable again during epidemic months (Colwell, 2000).

In natural aquatic habitats, a possible reservoir of *V. cholerae* is chitin-containing organisms. Chitin can serve as a source of energy, carbon, and nitrogen and a substrate for biofilm formation for this microorganism (Yildiz and Visick, 2009; Lutz et al., 2013). Biofilms provide a microenvironment favoring survival and persistence, displaying increased resistance to various

stresses, facilitating success in both of *V. cholerae* ecological niches, namely aquatic habitats and the human body (Yildiz and Visick, 2009). Association with plankton as biofilms could lead to the persistence of the pathogen during inter-epidemic periods (Vezzulli et al., 2016), possibly an important part of the seasonal cycle of cholera in endemic areas. The infectious dose required to cause human disease is quite high (~ 10^4 to 10^{11}) for cholera, the bacterium needing to pass the acidic stomach to reach small intestine where the cholera toxin is effective (Nelson et al., 2009). The association of the bacterium with biotic or abiotic surfaces could lower the infectious dose, as indicated by the observation that ingestion of *V. cholerae* along with food products decreases the required number to cause infection (Kaper et al., 1995; Almagro-Moreno and Taylor, 2013). Biofilm formation can also simply increase the likelihood of ingesting a larger dose than free living *V. cholerae,* thus increasing chance of successful human infection (Almagro-Moreno and Taylor, 2013). Moreover, biofilm-derived *V. cholerae* was found to show hyper- infectious phenotypes, leading to a reduction of the infectious dose by orders of magnitude lower in contrast to the ingestion of planktonic cells (Tamayo et al., 2010).

These characteristics have particular significance in maintaining seasonal cycle of the bacteria in the aquatic environment, especially in cholera endemic areas. Most of the evidence for *V. cholerae* association with hosts in the environment remains anecdotal and no systematic study has been done so far. Moreover, studies of its presence in aquatic niches are rarely specific to the PG lineage, but look at the species as a whole, which is likely to display significant ecological variations below the species level. A culture-based study of a *V. cholerae* population in two connected water bodies in a single coastal location in northeastern USA, encompassing extensive sampling and phylogenetic analysis, revealed subspecies level divergence within competing

genotypes (Kirchberger et al., 2016). Clonal complexes, which are groups of closely related *V. cholerae* strains as defined by multi-locus sequence typing, showed distinct spatial distributions across adjacent water bodies and water column size fractions (free-living, small and large particle associated), indicating likely subspecies level ecological differentiations. If this finding is confirmed in other ecosystems, it is possible that pathogenic *V. cholerae* such as the PG lineage differ from non-pathogenic strains in their ecological preferences. One possibility is that PG lineage strains could show significant association with zooplankton, helping to explain the seasonal cholera epidemics correlating with planktonic blooms. Such blooms could increase the number of bacteria in the water and trigger the epidemic cycle. It would also help explain why filtration of water with Sari cloth, which can trap larger particles and their associated microbes, can reduce the incidence of cholera (Huq et al., 2010).

2.4. The role of human hosts in the *V. cholerae* life cycle

In inter-epidemic periods, strains with pathogenic potential are rarely isolated in environmental surveillance (Alam et al., 2007). It seems that during the initiation of seasonal cholera epidemics, there is an enrichment period for pathogenic *V. cholerae* in the combined niche of aquatic habitats and human body (Almagro-Moreno and Taylor, 2013). During inter-epidemic periods, most of the *V. cholerae* cells in nature have been found to be in the VBNC state (Colwell, 2000; Alam et al., 2007). Passages in animal models have been shown to allow resuscitation of *V. cholerae* from the VBNC state, suggesting a potential advantage for cells able to survive passage through a human/animal host (Colwell, 2000; Almagro-Moreno and Taylor, 2013). Furthermore, production of cholera toxin, which causes massive diarrhea, aids rapid spread/dissemination of the bacteria into the nearby water in very high concentrations (Nelson et al., 2009). This rapid

increase is likely to be important to outnumber competing microbes, predators and lytic phages in natural reservoirs. Secreted bacteria from the infected human can also be hyper-infectious, requiring less bacteria to cause subsequent infection, fostering the epidemic cycle (Merrell et al., 2002; Nelson et al., 2009). Positive effect of human association was also evident in regulation of the type VI secretion system (T6SS), which has roles in competitive fitness of the bacteria in both human host and environment. Pandemic type *V. cholerae* strains were found to activate T6SS only inside the small intestine of animal host not in vitro conditions indicating active host role in infection (Bachmann et al., 2015). Human association can also contribute to the survival and persistence of *V. cholerae* in the environment (Levade et al., 2017). It was found that the transfer rate of CTXφ genetic element in *V. cholerae* was higher within the mice gastrointestinal tracts than under laboratory conditions (Waldor and Mekalanos, 1996). Baharoglu et al have shown that the SOS response in *V. cholerae*, triggered by pH changes, oxidative stress or exposure to DNA damaging antibiotics, can increase rate of gene cassette insertion in integrons (a gene capture element in bacteria) with a potential of acquiring new advantageous genes (Baharoglu et al., 2010). The human intestine is an environment allowing for extensive interaction of diverse microorganisms at high densities and could serve to induce acquisition of virulence and adaptive genes in *V. cholerae.* These observations make human gastrointestinal tract as a possible niche/hotspot for the exchange of crucial virulence or other advantageous gene clusters, which might have a significant role in the evolution of pathogenicity within *V. cholerae* populations.

Human hosts also serve as a vehicle to transfer the bacterium into new places through various forms of transportation. Asymptomatic carriers are of special interest in this case (Nelson et al.,

2009), as they could play an important role in transporting pathogenic *V. cholerae* to a new habitat (King et al., 2008). Asymptomatic carriage was proposed to have role in initiating the recent cholera epidemic in Haiti (58), where pandemic *V. cholerae* were introduced by UN troops originating from a cholera endemic area (Katz et al., 2013). Asymptomatic infections are mild enough to go undetected and estimates of the ratio of asymptomatic to symptomatic infections have ranged from 3:1 to 100:1 (King et al., 2008). Asymptomatic carriers usually posses certain level of immunity to the disease, can be physically healthy individuals and thus travel anywhere unnoticed and shed approximately 10^3 bacteria per gram of stool (Nelson et al., 2009). Shed pathogenic *V. cholerae* can grow in numbers in the nearby environment if conditions are ideal; a scenario that would be consistent with the Haiti epidemic (Jutla et al., 2013). In areas where people do not have immunity to cholera at sufficient levels or at all (such as Haiti), disease can spread rapidly. Understanding the role of asymptomatic infections in the dynamics of endemic and epidemic cholera requires detailed investigation over a wide geographical range.

2.5. Genetic factors influencing the dual stage life cycle of pandemic *V. cholerae*

Pandemic *V. cholerae* have both environmental and human stages in their life cycle in cholera endemic areas (Nelson et al., 2009). To maintain a successful seasonal epidemic cycle, these *V. cholerae* strains need to adapt with two different competitive niches. The ability of *V. cholerae* to survive in many different environmental niches is largely due to their inherent and/or acquired resistance to environmental shifts (Lutz et al., 2013). As a species, *V. cholerae* has certain attributes which makes it predisposed to survive in the human body. This includes the ability to

live in freshwater, grow well at human body temperature, utilize human intestinal biopolymers with aquatic analogs, form biofilms, resist acidic passage in the stomach and evade the host immune system (Almagro-Moreno and Taylor, 2013; Boucher et al., 2015). In addition to these inherent abiliuties of the species, pandemic *Vibrio cholerae* harbor several genetic elements directly contributing to virulence. These include the two major virulence factors cholera toxin (CT) and the toxin co-regulated pilus (TCP), as well as other genes thought to be associated with the infection process in humans, such as repeat in toxin (RTX), mannose sensitive hemolysis agglutination pillin (*mshA*, pillin gene *pilE*, hemolysisn (*hlyA*), and sialic acid degradation gene (*nanH*)(Kaper et al., 1995; Faruque et al., 1998). However, the exact role of all of these genes in the infection process is not clear. The gene set crucial for providing competitive advantage in environmental survival and human to environment transition to the pandemic *V. cholerae* are also not fully understood. However, regulation of virulence and fitness related genes are critical for the long-term viability of the bacteria in human. A recent study using transposon mutagenesis combined with massively parallel sequencing (Tn-seq) revealed 133 genes including 76 genes previously unknown for having any role in human infection, as contributing to survival of pandemic *V. cholerae O1* in infant rabbit model (Kamp et al., 2013). When dissemination from host into the environment were studied, 165 genes were found to be important for survival into the pond water including genes having known or hypothetical roles in energy production and conservation, cell wall and outer membrane biogenesis, electron transport, flagellar biosynthesis, transcriptional regulation and transportation (Kamp et al., 2013). Fu et al, identified 400 genes potentially critical for the fitness of *V. cholerae* in the infant rabbit intestine (Fu et al., 2013).

Among these, genes for encoding outer membrane porin ompU were found important for the fitness of the bacterium inside the host and genes for ycogen utilization and storage were shown to be crucial for dissemination, survival and persistence of host released *V. cholerae* into the environment (Kamp et al., 2013; Conner et al., 2016). Shapiro et al also found evidence for allelic differentiations in *ompU*, linked to virulence and environmental survival; where there seems to be a trade-off between human adapted and environment specific roles of the gene among the *V. cholerae* populations. One genetic system, which is believed to be significant in survival and fitness of pandemic *V. cholerae* in both human and environmental stages, is the type VI secretion system (T6SS) (Pukatzki and Provenzano, 2013; Unterweger et al., 2014). *V. cholerae* has been shown to contain three gene clusters, each harboring different combinations of effector –immunity proteins. In each effector-immunity module type, an effector gene encodes a protein which can kill other surrounding bacteria, and their corresponding immunity gene encode a protein protecting the bacteria from their matching effector (Unterweger et al., 2014; Kirchberger et al., 2017). Killing of non-compatible cells by T6SS might serve as source of readily available food and DNA from other bacteria, with potentially beneficial or protective functions. In environments where *V. cholerae* deals with low nutrient conditions, type VI mediated killing could generate supplementary source of nutrients to maintain their physiological activities (Unterweger et al., 2014). In the human intestine however, type VI mediated killing ability could give a selective advantage to a particular strain competing with the commensal host flora. It was found that mucin in human intestine can activate T6SS system in pandemic *V. cholerae*, whereas bile acid further modulate its activity (Bachmann et al., 2015).

After excretion with diarrheal stool, *V. cholerae* with activated T6SS are potentially better equipped to fight against bacterial and eukaryotic predators in the aquatic environment (Miyata et al., 2010). Thus, T6SS might give significant advantage to the pandemic lineage inside the human host as well in the environment by outcompeting others for the colonization of a desired niche and providing energetic benefits from lysed cells upon entry into and before exiting the human host (Pukatzki and Provenzano, 2013).

All known PG lineage *V. cholerae* strains consistently possess at least three genomic islands which are not shared with other *V. cholerae*: CTXφ, Vibrio Pathogenicity Island I (VPI1) and Vibrio Pathogenicity Island 2 (VPI2) (Faruque et al., 1998; Chun et al., 2009b) (Figure 2.1). Maintenance of these genetic elements and coexistence into the same genome can be crucial for the disease-causing ability. They are part of a virulence gene repertoire that has been acquired progressively over centuries by horizontal gene transfer (Figure 2.1) (Chun et al., 2009b). Acquisition of these factors can give advantage to PG lineage *V. cholerae* over benign environmental strains in surviving and exploiting the human gut as an ecological niche (Figure 2.1). The toxin corregulated pilus (TCP), one of the two main virulence factors of pandemic *V. cholerae*, is encoded within the horizontally acquired VPI1 (Kaper et al., 1995) and serves as the essential colonization factor and receptor for the CTXφ phage, which carries the second major virulence factor, cholera toxin (CT). Phylogenetic analysis suggests that VPI1 was acquired long before CTXφ by the ancestors of modern pandemic *V. cholerae,* making them capable of integrating CTXφ in their genome to become cholera causing agents (Figure 2.1). Beside TCP, VPI1 also encodes metallo protease TagA, which can breakdown mucin glycoproteins, and cell-several genes for sialic acid transport and catabolism, some of which were found to provide

surface glycans, making them available as a source of nutrients for the bacterium (Pukatzki and Provenzano, 2013). Saccharides, mucins and the glycocalix on the surface of human gut epithelial cells provide energy sources necessary for the growth and multiplication of the bacterium during the early stages of infection (Pukatzki and Provenzano, 2013). VPI-2 encodes competitive advantage against competing bacteria in the gut to *V. cholerae* in mouse model (Cho et al.). Two genomic islands exclusively present in 7th pandemic strains, Vibrio Seventh Pandemic Island 1 and 2 (VSP1 and VSP2) encode yet to be fully described but potentially important functions for the pathogenesis and survival of the lineage. For example, VSP1 encodes a transcription factor required for efficient colonization of human epithelial cells (Merrell et al., 2002). All these genetic elements likely provided advantages in aquatic populations before giving a fitness benefit inside the human host, which is a secondary niche for *V. cholerae* (Boucher et al., 2015). Toxin corregulated pilus encoded in VPI-1 was shown to be crucial for bacterial interactions required for biofilm differentiation on chitinaceous surfaces and thus likely to have important role in ecological fitness (Reguera and Kolter, 2005). As TCP also serves as a receptor for CTXφ, its expression during the formation of biofilms also fosters CTXφ transduction and thus represents an ecological setting outside the host in which selection for a host colonization factor may take place (Reguera and Kolter, 2005). We have already mentioned that T6SS can serve as a weapon in defense against predation by eukaryotic grazers or other competing bacteria in the aquatic environment (Pukatzki and Provenzano, 2013). Several studies have also found factors involved in pathogenesis to be expressed or required in association of the bacterium with algae, i.e an increase in toxin production was observed in *V. cholerae* when in association with the green alga *Rhizoclonium fontanum* (Islam, 1990).

These findings support the idea that these pathogenicity factors have environmental functions and are useful for bacterial survival and persistence outside of the human host (Reguera and Kolter, 2005). Transcriptional profiling of V. cholerae secreted in stool from cholera patients revealed that genes involved in nutrient acquisition and motility were highly expressed whereas genes for chemotaxis were expressed at lower levels (Matz and Kjelleberg, 2005; Kamp et al., 2013). It appeared that *V. cholerae* differentially regulate gene expression inside the human body and during passage to the environment, i.e. turns off expression of particular virulence genes as part of a program for dissemination to the environment. These changes in gene expression are thought to be linked to efficient exit from the host, re-entry to the aquatic environment and maintenance of an hyper-infectious state which enhances subsequent water borne spread of the cholera by lowering the infectious dose significantly (Merrell et al., 2002). When *V. cholerae* is shed into water, it is likely to encounter a drastic change in physiological conditions, i.e drop in osmolarity, temperature and nutrient availability. *V. cholerae* transitioned into pond water was found to repress genes for protein synthesis and energy metabolism and induction of phosphate and nitrogen scavenging genes indicating a adaptive program in response to the low nutrient condition (Nelson et al., 2009; Kamp et al., 2013). The glycogen utilization and storage program was found to have a central role in this adaption to transition between to vastly different niches (Kamp et al., 2013).

Although it is assumed that PG *V. cholerae* survive inter-epidemic periods in aquatic reservoirs, they could potentially also reside in the human gut during that time. If ingesting PG *V. cholerae* at low concentrations from the environment, human carriers may not show disease symptoms

but could still be colonized as asymptomatic carriers. These carriers could shed pathogenic clones into the nearby water bodies (Nelson., et al., 2009) eventually facilitating the initiation of a seasonal epidemic (Alam et al., 2006). Full-blown cholera can be considered helpful for bacterial dispersion in the environment in large numbers, cholera stool containing between 1010 and 1012 bacteria per litre (Nelson., et al., 2009). This enables a particular type of bacteria to outnumber other types in the environment and with a continuous annual cycle, they get selective advantage over other non- pathogenic type to sustain in an environment-human-environment life cycle. This kind of competitive advantage is not uncommon, which can explain how over time pathogenic *V. cholerae* settles endemicity within a population associated with a natural reservoir. But the ability to become a successful pandemic agent capable of causing on human infection and environmental survival cycle requires a constellation of virulence, regulatory and survival genes to be acquired and maintained for a long time. Environmental *V. cholerae* had to gradually change over a long period of time with continued selective pressure for such a genetic combination to evolve. The basic genetic backbone which made the evolution of pandemic variants possible apparently evolved a single time in the ancestor of the PG lineage (Figure 2.1). Actual pandemic variants evolved twice independently within this lineage, giving rise to two major pandemic biotypes (Chun et al., 2009b; Boucher et al., 2015).

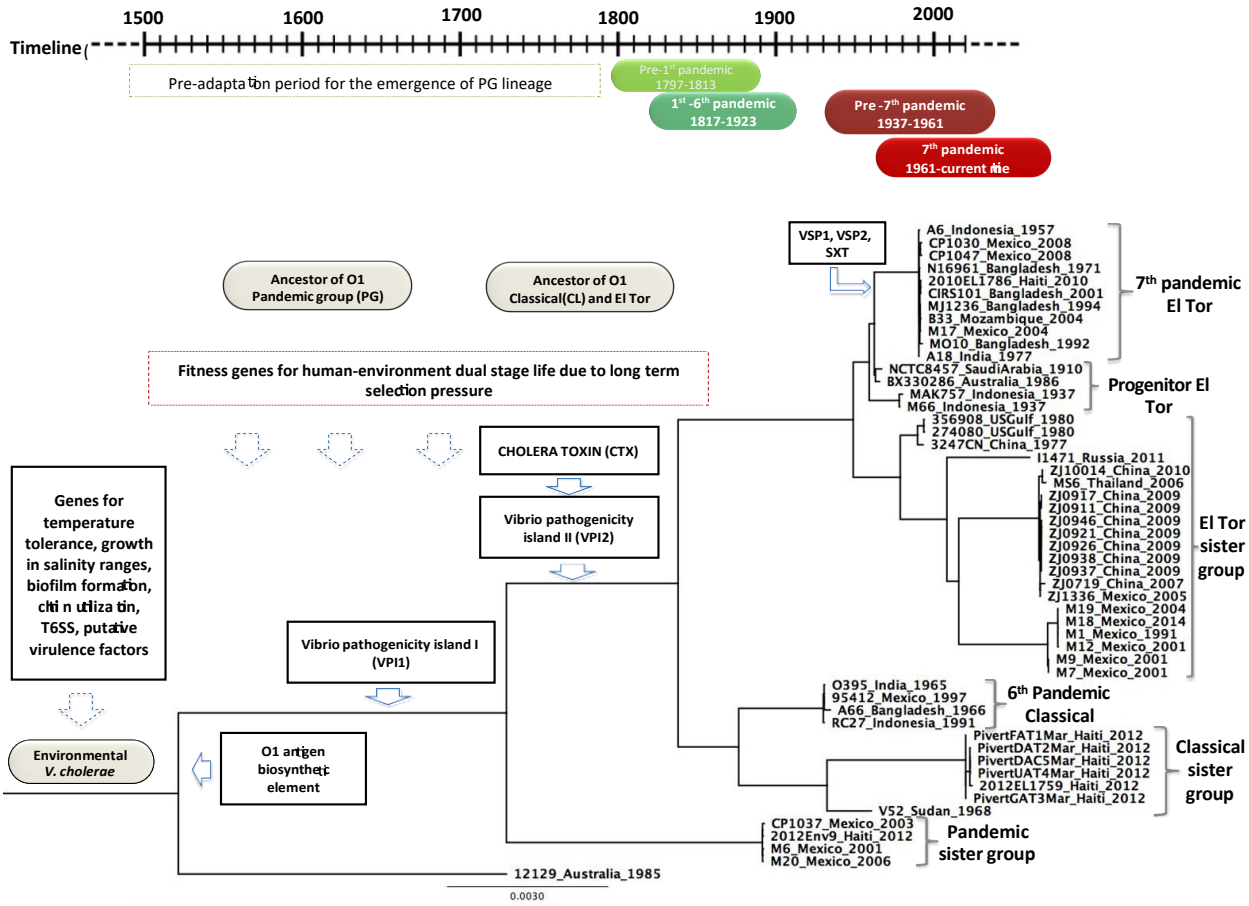


Figure. 2.1. Emergence and evolution of pandemic *V. cholerae* in a phylogenetic context

The maximum likelihood phylogenomic tree was constructed from the alignment of locally collinear blocks (2,784,396 bp) using GTR gamma substitution model with 100 bootstrap replicates (all nodes had >98% bootstrap support). Environmental non-toxigenic O1 strain 12129 was used as the outgroup to root the tree. The source and isolation year of the strains used are indicated next to the strain names. The timeline shows important events on the path of evolution toward pandemic *V. cholerae* from environmental origin. Important genetic elements including virulence factors are denoted in solid rectangles whereas putative genetic events are denoted in dotted rectangles. Stadium boxes denote for the emergence or appearance of lineages or pandemics.

2.6. The rise and spread of a deadly pathogen

Lethal variants capable of causing pandemic cholera emerged twice independently from two branches of the pandemic generating lineage (PG), the classical biotype (possibly in Asia

between 1500 and 1800) and the El Tor biotype (Indonesia, between 1930 and 1960) (Chun et al., 2009b; Alam et al., 2012; Hu et al., 2016). How, where and when did this PG lineage with the capacity to generate extremely virulent variants emerge from heterogeneous environmental *V. cholerae* population?

Despite being a widespread aquatic bacterium, *V. cholerae* as a species has several characteristics which makes it predisposed for the survival in the human gut (Almagro-Moreno and Taylor, 2013; Boucher et al., 2015). These traits seemingly provide a basic genetic background that fortuitously makes survival in a human host more likely but are not sufficient for *V. cholerae* to become a human pathogen, which requires virulence factors and other fitness genes to be added to this background and enhance their potential to cause human disease on a stable basis (Kaper et al., 1995; Faruque et al., 1998; Chun et al., 2009b). However, no lineage specific genomic region or genes were found in genome wide analysis, which is exclusively, present in PGs but absent in environmental groups (EGs) (63). Thus, no particular gene or gene families could be linked to the emergence and evolution of the PG group. Hence, virulence adaptive polymorphism was proposed to play a vital role in the process which implied that the environmental ancestor of the PGs had a particular genomic back-bone containing alleles of core genes that served as 'preadaptation' and enhanced its potential to give rise to the pandemic clones (Shapiro et al., 2016). A proposed conceptual model states that a variety of virulence related genes circulate in a diverse, recombining environmental gene pool, which is maintained in the population through various biotic and abiotic selective pressures. Upon encountering a new ecological opportunity, such as human consumption or transient

colonization of other animal hosts, proliferation and gradual expansion of the clones encoding an advantageous combination of vital genes for virulence and pandemicity is selected. These pre-adapted lineages can then serve as progenitors to acquire crucial virulence factors either in the environment or inside human body to mediate the emergence of pandemic *V. cholerae* (Boucher et al., 2016).

Pandemic causing *V. cholerae* appears to have the optimized genetic systems to maintain a dual stage life cycle as opposed to most of their benign environmental counterparts. Epidemiological and genetic data suggests that only O1 serogroup PG lineage has been successful in evolving and maintaining these adaptive traits (Chun at al., 2009). Sporadic cholera cases are caused by *V. cholerae* strains outside this lineage throughout the world (Kaper et al., 1995; Faruque et al., 1998) but none of them could be established as a long-term etiological agent to cause consistent seasonal cholera episodes. This complex genetic capability is unlikely to be created only in a single lineage without a consistent selection pressure over an extended period of time. This kind of evolutionary drive is most likely to have happened in Ganges delta, which has been endemic for cholera in at least the last three centuries and represents a unique ecosystem for *V. cholerae* (Boucher et al., 2016). It has been proposed that extensive contact between *V. cholerae* living in the coastal brackish waters and dense human population drinking from that water over centuries has created the circumstances for the emergence of pandemic lineage. Fecal-oral circulation of the bacterium in the local environmental reservoirs could have led to the selection and enrichment of variants capable of thriving both in the human gut and the environment (16). This hypothesis implies that long-

term association with human host is the driving factor for the emergence of *V. cholerae* with pandemic capabilities. Recent phylogenomic data suggests that currently ongoing seventh pandemic of cholera might have originated from Bay of Bengal and from there spread to other parts of the world in several waves (Mutreja et al., 2013). Hu et al hypothesized that the Middle East and Indonesia played essential roles in the evolution of seventh pandemic strains (Hu et al., 2016). However, this hypothesis is based on the analysis of very few strains and remains highly speculative.

Even though this pandemic generating lineage, also termed phylocore genome (PG), is a distinct monophyletic group from an extremely diverse environmental pool, it can be divided into two main phylogenetic branches; PG1 and PG2 (Chun et al., 2009b). The PG1 branch contains strains of the El Tor biotype and the PG2 branch those of the Classical biotype. These biotypes differ from each other by certain phenotypic and molecular traits. Strains of Classical biotype clade (PG2) are known to be responsible for the sixth and presumably the earlier pandemics, whereas strains from the El Tor clade (PG1) are the causative agent of the currently ongoing seventh pandemic of cholera starting in 1961 (Faruque et al., 1998). Classical biotype strains have not been isolated since the early nineties even from Southeast Asia, where they were last found and have thus been considered as completely outcompeted by the El Tor biotype both from clinical and environmental settings. Expansion of the 7th pandemic has given rise to new variants of the prototype *V. cholerae* O1 El Tor regularly during the (Safa et al., 2010). These variants include strains harboring classical biotype features within an El Tor genetic backbone and/or other divergent genetic features including mutations in major virulence factors and hence are named atypical El Tor (Safa et al., 2010).

After the initial wave of the current pandemic spread prototype El Tor strains across the world, two additional waves spread the variants of El Tor strains, each wave mostly displacing bacteria from the preceding one and has been a feature of global cholera epidemiology (Mutreja et al., 2013). The 7th pandemic of cholera struck South America in 1991 via Peruvian coast and reached in Mexico the same year (Alam et al., 2010). In 2010, one of the most devastating cholera epidemics in history occurred in Haiti, killing thousands of people (Katz et al., 2013). Cholera has now set residencies in the local environment of Mexico and Haiti, even though both the countries did not have any recorded cases in 100 years before the recent epidemics happened. Cholera epidemics leading to endemicity of PG *V. cholerae* O1 in the affected area have prompted extensive environmental sampling of natural waters in Mexico and Haiti. These have revealed remarkable diversity of pandemic-related *V. cholerae* for countries, which did not have any known history of cholera until recently (Azarian et al., 2014; Choi et al., 2016). In Mexico, a recent series of retrospective studies have reported the discovery of classical, prototype El Tor, atypical El Tor, and non-toxigenic O1 strains with some unusual genetic features in *V. cholerae* strains isolated from 1983 to 2008 (Alam et al., 2010, 2012; Choi et al., 2016). Along with this surprising diversity, there was presence of strains grouping at the base of the PG lineage that led to Classical and El Tor biotype strains, hence candidates for being considered as previously undetected descendants of the ancestor of the two biotypes (Boucher, 2016; Choi et al., 2016)(Figure 2.1). In Haiti as well, where cholera cases could clearly be attributed to the atypical El Tor strains introduced from Nepal (Katz et al., 2013), presence of this divergent lineage (termed pandemic sister group) in the water was evident (Azarian et al.,

2014; Boucher, 2016). Azarian et al. estimated the time for the divergence of this lineage from the common ancestor of pandemic V. cholerae around 1548 C.(Azarian et al., 2016), long before the report of the first pandemic in 1817. These observations are consistent with the historic records suggesting that descendants of the *V. cholerae* PG lineage common ancestor have been globally distributed for centuries and that this dissemination happened long before the first recorded pandemic (Barua, 1972; Boucher et al., 2015). Presence of *V. cholerae* strains belonging to the PG lineage but clearly distinct from the Classical and El Tor strains have been isolated sporadically from around the world over the last few decades, including some non-endemic regions i.e., US Gulf coast, Australia, Russia, Thailand and China (Chun et al., 2009a; Boucher, 2016; Hu et al., 2016) (Figure 2.1) and have been reported very recently from Haiti and Mexico (Katz et al., 2013; Azarian et al., 2014; Boucher, 2016). Presence of these non-pandemic members of the PG lineage in wide geographical locations underscores that genomic database of *V. cholerae* today is extremely biased by clinical isolates and large-scale environmental sampling in wide geographical areas are needed to get a better picture of the diversity and global distribution of the PG lineage. Even though these close relatives of pandemic causing strains in most case lack the main virulence factor CT, they harbour TCP, which can act as the receptor for CT. The rest of their genetic backbone is also very similar to pandemic strains (Azarian et al., 2016; Choi et al., 2016). Therefore, the possibility for the emergence of novel *V. cholerae* with pandemic potential from this globally spread lineage cannot be discounted.

2.7. Combating future cholera scenario

Pathogenic bacteria with environmental reservoirs like PG *V. cholerae*, which has to survive in both host and environmental conditions, need to maintain a delicate balance to the two very contrasting life styles. The drastic transition from environment to human and vice versa requires adaptations for both human body and the aquatic environment. The currently ongoing 7th pandemic is the longest in duration and largest in geographical span. During the course of this pandemic, cholera has struck countries in virtually every continent except Antarctica and has even become endemic in countries other than Asia and Africa, surviving in those geographic settings successfully and causing regular cholera outbreaks (6, 11, 8). El Tor biotype strains are known to have better survival in the environment than Classical biotype strains (Faruque et al., 1998), whereas the Classical type toxin is found to cause more severe cholera then the El Tor type (Kaper et al., 1995). Currently found variants of prototype El Tor strains possess Classical type toxin in the El Tor genetic backbone, which is likely to make them more potent pandemic causing agents. 7^{th} pandemic isolates contain two genomic islands, VSP-I and VSP-II, which were not found consistently in other lineages. Even though exact function of these elements is not well understood, VSP1 encodes a transcription factor, which was shown to be required for efficient intestinal colonization (Merrell et al., 2002). Their consistent and exclusive presence in current 7^{th} pandemic isolates implies that they might well have significant roles in environmental fitness and pathogenic capabilities (Almagro-Moreno and Taylor, 2013). From 1992 and onwards, most El Tor strains have been found to harbor a integrative conjugative element called SXT, which is known to serve as hotspot for acquisition of genes including resistance to certain antimicrobials and environmental persistence (Mutreja et al., 2013). Acquisition of antimicrobial resistance can becrucial for the success of the 51 modern El Tor strains as a long-lasting

disease- causing agent (Hu et al., 2016). The high fitness of the currently circulating strains might have selected for traits constraining their evolution. Most of the 7th pandemic clinical V. cholerae strains isolated since 2000, including strains causing epidemic cholera in Haiti, were found to harbor an integrative These observations suggest that the 7th pandemic of cholera is likely to continue in the near future. New atypical variants of the El Tor biotype are likely to emerge and could trigger new waves of the pandemic. In 1992, *V. cholerae* O139 emerged in the Ganges delta region and caused severe cholera outbreaks in various parts of Asia and even suggested as a possible "Eighth pandemic of cholera" by some investigators (Kaper et al., 1995; Faruque et al., 1998). Even though serogroup O139 became rare since 2005, it is still being isolated sporadically from environmental and clinical samples (Alam et al., 2006). Of concern is also the possibility of a novel pandemic biotype, separate from El Tor or Classical but still belonging to the PG lineage, would emerge. As the PG lineage as already genetrated two pandemic biotypes independently, a third one is a real threat, especially given ongoing global warming and rapidly changing climatic conditions. In ideal transmission and dissemination settings, these novel biotypes or variant of current pandemic biotypes can adapt to the environment and spread to non-cholera regions via human or environmental carriers to cause cholera outbreaks on a global scale.

Ecological niche modeling taking current and future climatic condition in consideration has predicted a latitudinal increase in potential areas of *V. cholerae* distribution in the future (Escobar et al., 2015). Effective methodologies to predict cholera outbreaks one to several months in advance would make controlling cholera outbreaks much easier. It is

presumed that *V. cholerae* was originally a marine bacterium that could persist in estuarine, coastal waters over a broad range of environmental 52 conditions (Lipp et al., 2002; Colwell, 2004a). In cholera endemic areas, water current, flooding and human activity might carry the bacteria inland, where it can adapt and survive in varying degrees to infect human population drinking contaminated water (Akanda et al.). Over the last decades, studies have identified potential environmental variables associated with *V. cholerae* occurrence. Ocean chlorophyll has been found to have the most consistent association with number of cholera cases in nearby populations and thus is thought to be a potential indicator for future cholera prediction (Constantin de Magny et al., 2008; Emch et al., 2008). However, prediction models for a complex and dynamic environmental disease like cholera would require more in-depth understanding of the ecology and biogeography of this pathogen, especially of its pandemic-generating lineage.

2.8. Concluding remarks:

In 2010, cholera killed more than 8000 people in Haiti, country that did not have any recent history of cholera (Orata et al., 2014). War torn Yemen is currently facing the devastation of cholera, one of the worst outbreaks on record, with nearly 2000 deaths and around 50,0000 suspected cholera cases till August 2017 (http://www.who.int/mediacentre/news/releases/2017/cholera-yemen-mark/en/). Haiti and Yemen episodes underscore the massive threat cholera poses even in this modern time, showing the need for more effective approaches to prevention and control of this deadly disease. Thus, a global scale coordination of bio-geographical and genomic studies is warranted to improve prevention and management of future cholera epidemics.

2.9. References:

Akanda, A. S., Jutla, A. S., Gute, D. M., Sack, R. B., Alam, M., Huq, A., et al. Population Vulnerability to Biannual Cholera Outbreaks and Associated Macro-Scale Drivers in the Bengal Delta. *Am. Soc. Trop. Med. Hyg.* 89, 950–959. doi:10.4269/ajtmh.12-0492.

Alam, M., Hasan, N. A., Sadique, A., Bhuiyan, N. A., Ahmed, K. U., Nusrin, S., et al. (2006). Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl Env. Microbiol* 72, 4096–4104. doi:10.1128/AEM.00066-06.

Alam, M., Islam, A., Bhuiyan, N., Rahim, N., Hossain, A., Khan, G. Y., et al. (2011). Clonal transmission, dual peak, and off-season cholera in Bangladesh. *Infect. Ecol. Epidemiol.* doi:10.3402/iee.v1i0.7273.

Alam, M., Islam, M. T., Rashed, S. M., Johura, F., Bhuiyan, N. A., Delgado, G., et al. (2012). *Vibrio cholerae* classical biotype strains reveal distinct signatures in Mexico. *J. Clin. Microbiol.* 50, 2212–2216. doi:10.1128/JCM.00189-12.

Alam, M., Nusrin, S., Islam, A., Bhuiyan, N. A., Rahim, N., Delgado, G., et al. (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. doi:10.1128/JCM.00866-10.

Alam, M., Sultana, M., Balakrish Nair, G., Siddique, A. K., Hasan, N. A., Bradley Sack, R., et al. (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.* doi:10.1073/pnas.0705599104.

Almagro-Moreno, S., and Taylor, R. K. (2013). Cholera: Environmental Reservoirs and Impact on Disease Transmission. *Microbiol. Spectr.* 1. doi:10.1128/microbiolspec.OH-0003-2012.

Azarian, T., Ali, A., Johnson, J. A., Jubair, M., Cella, E., Ciccozzi, M., et al. (2016). Non-toxigenic environmental *Vibrio cholerae* O1 strain from Haiti provides evidence of pre-pandemic cholera in Hispaniola. *Sci. Rep.* 6, 36115. doi:10.1038/srep36115.

Azarian, T., Ali, A., Johnson, J. A., Mohr, D., Prosperi, M., Veras, N. M., et al. (2014). Phylodynamic Analysis of Clinical and Environmental *Vibrio cholerae* Isolates from HaitiReveals Diversification Driven by Positive Selection. *MBio* 5, e01824-14. doi:10.1128/mBio.01824-14.

Bachmann, V., Kostiuk, B., Unterweger, D., Diaz-Satizabal, L., Ogg, S., and Pukatzki, S. (2015). Bile Salts Modulate the Mucin-Activated Type VI Secretion System of Pandemic *Vibrio cholerae*. *PLoS Negl. Trop. Dis.* 9, e0004031–e0004031. doi:10.1371/journal.pntd.0004031.

Baharoglu, Z., Bikard, D., and Mazel, D. (2010). Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet.* 6, e1001165–e1001165. doi:10.1371/journal.pgen.1001165.

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

Boucher, Y. (2016). Sustained Local Diversity of *Vibrio cholerae* O1 Biotypes in a PreviouslyCholera-Free Country. *MBio* 7, e00570-16. doi:10.1128/mBio.00570-16.

Boucher, Y., Orata, F. D., and Alam, M. (2015). The out-of-the-delta hypothesis: Dense human populations in low-lying river deltas served as agents for the evolution of a deadly

pathogen. *Front. Microbiol.* doi:10.3389/fmicb.2015.01120.

Choi, S. Y., Rashed, S. M., Hasan, N. A., Alam, M., Islam, T., Sadique, A., et al. (2016). Phylogenetic diversity of *Vibrio cholerae* associated with endemic cholera in Mexico from1991 to 2008. *MBio*. doi:10.1128/mBio.02160-15.

Chun, J., Grim, C. J., Hasan, N. A., Je, H. L., Seon, Y. C., Haley, B. J., et al. (2009a). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0907787106.

Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., et al. (2009b). 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. doi:10.1073/pnas.0907787106.

Colwell, R. R. (2000). Viable but nonculturable bacteria: A survival strategy. *J. Infect. Chemother.* doi:10.1007/PL00012151.

Colwell, R. R. (2004a). Infectious disease and environment: cholera as a paradigm for waterborne disease. *Int Microbiol* 7, 285–289.

Colwell, R. R. (2004b). Infectious disease and environment: Cholera as a paradigm for waterborne disease. *Int. Microbiol.* doi:10.2436/im.v7i4.9485.

Conner, J. G., Teschler, J. K., Jones, C. J., and Yildiz, F. H. (2016). Staying Alive: *Vibrio cholerae*'s Cycle of Environmental Survival, Transmission, and Dissemination. *Microbiol. Spectr.* 4. doi:10.1128/microbiolspec.VMBF-0015-2015.

Constantin de Magny, G., Murtugudde, R., Sapiano, M. R. P., Nizam, A., Brown, C. W., Busalacchi, A. J., et al. (2008). Environmental signatures associated with cholera epidemics. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17676–17681. doi:10.1073/pnas.0809654105.

Dalia, A. B., Seed, K. D., Calderwood, S. B., and Camilli, A. (2015). A globally distributed mobile genetic element inhibits natural transformation of *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* 112, 10485–10490. doi:10.1073/pnas.1509097112.

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. doi:10.1186/1476-072X-7-31.

Escobar, L. E., Ryan, S. J., Stewart-Ibarra, A. M., Finkelstein, J. L., King, C. A., Qiao, H., et al. (2015). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Trop.* 149, 202–211. doi:10.1016/j.actatropica.2015.05.028.

Faruque, S. M., Albert, M. J., and Mekalanos, J. J. (1998). Epidemiology, Genetics, and Ecology of Toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* doi:10.1128/mmbr.62.4.1301-1314.1998.

Fu, Y., Waldor, M. K., and Mekalanos, J. J. (2013). Tn-Seq analysis of *Vibrio cholerae* intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe* 14, 652–663. doi:10.1016/j.chom.2013.11.001.

Hu, D., Liu, B., Feng, L., Ding, P., Guo, X., Wang, M., et al. (2016). Origins of the current seventh cholera pandemic. *Proc. Natl. Acad. Sci. U. S. A.* 113, E7730–E7739. doi:10.1073/pnas.1608732113.

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., et al. (2005). Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* doi:10.1128/AEM.71.8.4645-4654.2005.

Huq, A., Yunus, M., Sohel, S. S., Bhuiya, A., Emch, M., Luby, S. P., et al. (2010). Simple sari cloth filtration of water is sustainable and continues to protect villagers from cholera in Matlab, Bangladesh. *MBio* 1. doi:10.1128/mBio.00034-10.

Islam, M. S. (1990). Effect of various biophysicochemical conditions on toxigenicity of *Vibrio cholerae* 01 during survival with a green alga, Rhizoclonium fontanum, in an artificial aquatic environment. *Can. J. Microbiol.* 36, 464–468. doi:10.1139/m90-081.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993. doi:10.1038/nature06536.

Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A., et al. (2013). Environmental factors influencing epidemic cholera. *Am. J. Trop. Med. Hyg.* doi:10.4269/ajtmh.12-0721.

Kamp, H. D., Patimalla-Dipali, B., Lazinski, D. W., Wallace-Gadsden, F., and Camilli, A. (2013). Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *PLoS Pathog.* 9, e1003800. doi:10.1371/journal.ppat.1003800.

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

Katz, L. S., Petkau, A., Beaulaurier, J., Tyler, S., Antonova, E. S., Turnsek, M. A., et al. (2013). Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. *MBio* 4, e00398-13. doi:10.1128/mBio.00398-13.

King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature* 454, 877–880. doi:10.1038/nature07084.

Kirchberger, P. C., Orata, F. D., Barlow, E. J., Kauffman, K. M., Case, R. J., Polz, M. F., et al. (2016). A small number of phylogenetically distinct clonal complexes dominate a coastal *Vibrio cholerae* population. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01177-16.

Kirchberger, P. C., Unterweger, D., Provenzano, D., Pukatzki, S., and Boucher, Y. (2017). Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. *Sci. Rep.* doi:10.1038/srep45133.

Levade, I., Terrat, Y., Leducq, J.-B., Weil, A. A., Mayo-Smith, L. M., Chowdhury, F., et al. (2017). *Vibrio cholerae* genomic diversity within and between patients. *Microb. genomics* 3. doi:10.1099/mgen.0.000142.

Lipp, E. K., Huq, A., and Colwell, R. R. (2002). Effects of global climate on infectious disease: The cholera model. *Clin. Microbiol. Rev.* doi:10.1128/CMR.15.4.757-770.2002.

Lutz, C., Erken, M., Noorian, P., Sun, S., and McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. *Front. Microbiol.* doi:10.3389/fmicb.2013.00375.

Matz, C., and Kjelleberg, S. (2005). Off the hook - How bacteria survive protozoan grazing. *Trends Microbiol.* doi:10.1016/j.tim.2005.05.009.

Merrell, D. S., Butler, S. M., Qadri, F., Dolganov, N. A., Alam, A., Cohen, M. B., et al. (2002). Host-induced epidemic spread of the cholera bacterium. *Nature* 417, 642–645. doi:10.1038/nature00778.

Miyata, S. T., Kitaoka, M., Wieteska, L., Frech, C., Chen, N., and Pukatzki, S. (2010). The *Vibrio cholerae* Type VI Secretion System: Evaluating its Role in the Human Disease Cholera. *Front. Microbiol.* 1, 117. doi:10.3389/fmicb.2010.00117.

Mutreja, A., Kim, D. W., Thomson, N., Connor, T. R., Hee, J., Lebens, M., et al. (2013). Evidence for multiple waves of global transmission within the seventh cholera pandemic. *Nature*. doi:10.1038/nature10392.Evidence.

Nelson, E. J., Harris, J. B., Morris Jr, J. G., Calderwood, S. B., and Camilli, A. (2009). Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat. Rev. Microbiol.* 7, 693–702. doi:10.1038/nrmicro2204.

Orata, F. D., Keim, P. S., and Boucher, Y. (2014). The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy. *PLOS Pathog.* 10, 1–5. doi:10.1371/journal.ppat.1003967.

Pukatzki, S., and Provenzano, D. (2013). *Vibrio cholerae* as a predator: lessons from evolutionary principles. *Front. Microbiol.* 4, 384. doi:10.3389/fmicb.2013.00384.

Reguera, G., and Kolter, R. (2005). Virulence and the Environment: a Novel Role for *Vibrio cholerae* Toxin-Coregulated Pili in Biofilm Formation on Chitin. *J. Bacteriol.* 187, 3551–3555. doi:10.1128/JB.187.10.3551-3555.2005.

Safa, A., Nair, G. B., and Kong, R. Y. C. (2010). Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol.* 18, 46–54. doi:10.1016/j.tim.2009.10.003.

Shapiro, B. J., Levade, I., Kovacikova, G., Taylor, R. K., and Almagro-Moreno, S. (2016). Origins of pandemic *Vibrio cholerae* from environmental gene pools. *Nat Microbiol* 2, 16240. doi:10.1038/nmicrobiol.2016.240.

Stine, O. C., Alam, M., Tang, L., Nair, G. B., Siddique, A. K., Faruque, S. M., et al. (2008). Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerg. Infect. Dis.* doi:10.3201/eid1405.071116.

Takemura, A. F., Chien, D. M., and Polz, M. F. (2014). Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* doi:10.3389/fmicb.2014.00038.

Tamayo, R., Patimalla, B., and Camilli, A. (2010). Growth in a Biofilm Induces a Hyperinfectious Phenotype in *Vibrio cholerae*. *Infect. Immun.* 78, 3560–3569. doi:10.1128/IAI.00048-10.

Turner, J. W., Good, B., Cole, D., and Lipp, E. K. (2009). Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME J.* 3, 1082–1092. doi:10.1038/ismej.2009.50.

Unterweger, D., Miyata, S. T., Bachmann, V., Brooks, T. M., Mullins, T., Kostiuk, B., et al.

(2014). The *Vibrio cholerae* type VI secretion system employs diverse effector modules for intraspecific competition. *Nat Commun* 5, 3549. doi:10.1038/ncomms4549.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P. C., Colwell, R. R., Höfle, M. G., et al. (2012). Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *ISME J.* 6, 21–30. doi:10.1038/ismej.2011.89.

Vezzulli, L., Grande, C., Reid, P. C., Hélaouët, P., Edwards, M., Höfle, M. G., et al. (2016). Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1609157113.

Waldor, M. K., and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1914. doi:10.1126/science.272.5270.1910.

Yildiz, F. H., and Visick, K. L. (2009). Vibrio biofilms: so much the same yet so different. *Trends Microbiol.* 17, 109–118.

CHAPTER 3: Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp.nov.) with a history of association with humans

A version of this chapter has been published as:

"**Islam MT**, Nasreen T, Kirchberger P, Liang KYH, Orata FD, Johura FT, Hussain NAS, Im MS, Tarr CL, Alam M, Boucher YF. Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species (*Vibrio paracholerae* sp. nov.) with a history of association with humans. *Applied and Environmental Microbiology*. 2021 Jun 16: AEM0042221. doi:10.1128/AEM.00422-21."

M.T.I., T.N., M.A. and Y.F.B. designed the experiments. M.T.I., T.N. and F.J. performed data collection. M.T.I., P.K., K.Y.H.L. and F.O. performed data analysis. M.T.I., K.Y.H.L. and F.O performed bioinformatics analysis. M.I. performed genome sequencing and supplied meta data of isolates from CDC. M.A. and C.L.T. provided isolates used in this study. M.T.I. wrote the manuscript. Y.F.B. and C.L.T provided critical reviews on the manuscript.

CHAPTER 3

3.1. Abstract

Most efforts to understand the biology of *Vibrio cholerae* have focused on a single group, the pandemic-generating lineage harboring the strains responsible for all known cholera pandemics. Consequently, little is known about the diversity of this species in its native aquatic environment. To understand the differences in the *V. cholerae* populations inhabiting regions with a history of cholera cases and those lacking such a history, a comparative analysis of population composition was performed. Little overlap was found in lineage compositions between those in Dhaka (cholera endemic) located in the Ganges delta, and of Falmouth (no known history of cholera), a small coastal town on the United States east coast. The most striking difference was the presence of a group of related lineages at high abundance in Dhaka which was completely absent from Falmouth. Phylogenomic analysis revealed that these lineages form a cluster at the base of the phylogeny for the *V. cholerae* species, sufficiently differentiated genetically and phenotypically to form a novel species. A retrospective search revealed that strains from this species have been anecdotally found from around the world and were isolated as early as 1916 from a British soldier in Egypt suffering from choleraic diarrhoea. In 1935 Gardner and Venkatraman unofficially referred to a member of this group as *Vibrio paracholerae*. In recognition of this earlier designation, we propose the name *Vibrio paracholerae* sp. nov. for this bacterium. Genomic analysis suggests a link with human populations for this novel species and substantial interaction with its better-known sister species.

3.2. Introduction

Vibrio cholerae is the causative agent of cholera, the disease which has shaken human civilization from the last few centuries and continues to be a public health threat, especially to the developing world (Ali et al., 2012; Kaper et al., 1995). Its pathogenesis and epidemiology have been extensively studied, but the aquatic part of its life cycle is still not fully understood. Strikingly few close relatives have been found for this species in recent years, most being initially classified as *V. cholerae*-like bacteria. One of them was occasional human pathogen *Vibrio mimicus,* which was proposed as a new species in 1981 based on phenotypic characteristics (Davis et al., 1981). Later, genome-based studies established the molecular basis of its importance as a pathogen, close association and exchange of important virulence genes with *V. cholerae* (Hasan et al., 2010; Neogi et al., 2019). Two other closely related novel species, *Vibrio* sp. (unofficially named *Vibrio parilis*) and *Vibrio metoecus,* were more recently isolated alongside *V. cholerae* from coastal waters (Haley et al., 2010; P C Kirchberger et al., 2014) and found to exchange genetic material with their well-known sister species in aquatic environments (Haley et al., 2010; Orata et al., 2015). Biological information on the close relatives of a dangerous environmental pathogen like *V. cholerae* is of significance, because of their potential as emerging pathogens themselves and their interaction with *V. cholerae* in its natural habitats. Even though this diverse species is ubiquitous in tropical and temperate coastal waters world-wide, cholera is only caused by a specific lineage of *V. cholerae* in which the O1 antigen is ancestral, the pandemic-generating (PG) *V. cholerae* lineage. It is not clear whether aquatic *V. cholerae* maintains a significantly different population structure in cholera endemic and non-endemic areas, and if this structure is influenced by co-occurring species. This is a crucial gap in our understanding of the factors defining cholera endemicity and driving local and global biogeographic dispersal patterns of *V. cholerae.* It has recently become possible to investigate the details of the population structure of *V. cholerae* and its close relatives, using a

molecular marker based the single copy housekeeping gene (*viuB*, vibriobactin utilization protein subunit B), which provides subspecies level resolution (P C Kirchberger et al., 2020). This method was used to study a cholera-free region on the east coast of the USA, the Oyster Pond ecosystem (Falmouth, Massachusetts), where differences in abundance of individual alleles in particular locations/habitats indicated potential adaptation to ecological conditions at the subspecies level (P C Kirchberger et al., 2016, 2020). A similar study was performed in *V. cholerae* populations in an inland location (Dhaka) in cholera-endemic Bangladesh (P C Kirchberger et al., 2020; Mohammad Tarequl Islam, Tania Nasreen, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C. Kirchberger, Marzia Sultana, Rebecca J. Case, n.d.). Here, to understand the role played by subspecies population structure in disease, we compared the *V. cholerae* population from inland Bangladesh with that from the east coast of the USA. This revealed that distribution and abundance of major lineages of *V. cholerae* differ significantly in the two distinct ecosystems. Both globally distributed as well as locally adapted lineages of *V. cholerae* are found in the two environments studied. One of the most striking differences was the presence of several related lineages in Dhaka forming a divergent clade at the base of the *V. cholerae* species in a phylogenomic analysis, which were completely absent in the coastal USA location. Genomic characterization of these lineages reveals that they form a novel species closely related to but distinct from *V. cholerae*. A revision of recent and decades old historical isolates related to this novel species indicates that it has been found in similar environments to pandemic *V. cholerae* for decades and is associated with human infections ranging from septicemia to choleraic diarrhea.

3.3. Materials and Methods

3.3.1. Sample collection and processing

Environmental water samples were collected every two weeks between June 2015 and March 2016 from seven points along the water bodies surrounding Dhaka city, which is located in the central part of

Bangladesh (23.8103° N, 90.4125° E). One-time water samples were collected from two natural coastal water bodies in Mathbaria (22.2920° N, 89.9580° E) and Kuakata (21.8210° N, 90.1214° E), which are geographically adjacent to the coast of the Bay of Bengal and approximately 200 km and 250 km southwest of Dhaka, respectively. One liter of water was collected from each sites in sterile Nalgene bottles placed in an insulated plastic box, and transported at ambient air temperature from the site of collection to the central laboratory of the International Center for Diarrheal Disease Research, Bangladesh (ICDDR,B), in Dhaka. Oyster pond sampling was performed at the same spots and approximates same time of the day in the pond and the nearby lagoon connected to the ocean in monthly intervals from June to October as described in Kirchberger *et al* (P C Kirchberger et al., 2016). 50 liters of water were filtered through 0.22µm sterivex filters (Mo Bio Laboratories Inc., Carlsbad, CA, USA) for the collection of biomass. Genomic DNA was extracted from the biomass using the protocol described by Wright et al. (Wright et al., 2009).

3.3.2. Isolation and identification of isolates

Bacterial isolates were recovered as described elsewhere (Alam et al., 2006). Briefly, water samples were enriched in APW (Difco Laboratories, Detroit, Mich.) at 37°C for 6 to 8 h before plating. About 5 µl of enriched APW broth was streaked, using an inoculating loop, onto both thiosulfate-citrate-bile-salts-sucrose (TCBS) and TTGA and incubated at 37°C for 18 to 24 h. Colonies with the characteristic appearance of *V. cholerae* were confirmed by standard biochemical and serological tests (and, in the case of the latter, by testing with polyvalent and monoclonal antibodies specific for *V. cholerae* O1 or O139) and, finally, by PCR.

3.3.3 Phenotypic tests

For the comparison of phenotypic characteristics, Biolog phenotypic microarray plates PM1, PM2A, PM14A, PM16A, PM18C were used (Bochner et al., 2001). Overnight cultured bacterial colonies were inoculated into Biolog IF-0a Base medium to reach 85 % turbidity followed by 1:200 dilution aliquoted into IF-10b medium supplemented with Dye Mix A as indicated by the manufacturer instructions. The mixture was then added into wells of Biolog PM1 and PM2A plate containing various carbon sources and PM14A, PM16A and PM18C plates containing substrates of various antimicrobials and heavy metal salts. The incubation and monitoring of the growth of inocula were done for 96 h in the presence of sole carbon source or the heavy metals, growth causes reduction of the dye, resulting in purple colour formation.

3.3.4 Quantitative PCR (qPCR)

Estimation of *Vibrio cholerae* number was done using qPCR following the protocol described elsewhere (Nasreen et al., 2020). Briefly, qPCR protocol described in Nasreen et al was followed (Nasreen et al., 2020). Target probe for *viuB*, 5'-/56-FAM/TCATTTGGC/ZEN/CAGAGCATAAACCGGT/ 3IABkFQ/-3', forward primer 5'-TCGGTATTGTCTAACGGTAT-3', and reverse primer 5'-CGATTCGTGAGGGTGATA-3' was used. The volume of the PCR reaction was 10 µl containing 5 µl of 2× Dynamite qPCR master mix (MBSU, University of Alberta, Edmonton, Canada), 1 µl of each of 500 nM primer-250 nM probe mix, 1 µl of molecular grade water and 2 µl of DNA template. Real-time quantitative PCR was performed under the following conditions: initial primer activation at 95 °C for 2 min followed by 40 cycles of 95 °C for 15 s, 60 °C for 1 min in Illumina Eco Real-Time PCR system.

3.3.5. Amplicon sequencing

Amplicon sequencing of *viuB* gene was performed following the method described elsewhere (P C Kirchberger et al., 2020). To amplify 293 bp of the *viuB* region from DNA extracted from water samples, a touchdown PCR was performed using 0.5 µL each of 10 pmol forward and reverse primers (for *viuB*: *viuB*2f 5'-CCGTTAGACAATACCGAGCAC-3' and *viuB*5r 5'-TTAGGATCGCGCACTAACCAC-3'), 0.4 µL of 10 mM dNTP mix (ThermoFisher), 0.4 of µL Phire Hot Start II DNA Polymerase (ThermoFisher), 0.5 µL of molecular biology grade bovine serum albumin (20 mg/mL, New England Biolabs), 5 µL of 5× Phire Buffer, and 2 µL of template DNA. The PCR reaction was performed as follows: initial denaturation at 98°C for 4 min; followed by 10 cycles of denaturation at 98°C for 10 sec, annealing at 60°C for 6 sec (reduced by 1°C per cycle), and extension 72°C for 1 sec; followed by 23 cycles of denaturation at 98°C for 10 sec, annealing at 50°C for 6 sec (reduced by 1°C per cycle), and extension at 72°C for 1 sec; and a final extension at 72°C for 1 min. In preparation for sequencing, dual-indexed sequences were tagged using indices developed by Kozich *et al.* (Kozich et al., 2013) as follows: 2 µL of preceding *viuB* PCR amplification reaction were used as template for a tagging PCR reaction; initial denaturation at 98°C for 30 sec; followed by two cycles of denaturation at 98°C for 10 sec, annealing at 55°C for 6 sec, and extension at 72°C for 1 sec; and final extension at 72°C for 1 min. Eight tagging reactions were performed for each sample and products were pooled and ran on a 2% agarose gel in 1× Tris-Acetate-EDTA buffer. The appropriate bands (428 bp) were cut out of the gel. PCR products were then purified using Wizard SV Gel and PCR Clean-Up System (Promega) according to the instructions by the manufacturer. Concentration of clean PCR products was then measured using a Qubit Fluorometer (ThermoFisher) with a Qubit dsDNA HS Assay Kit (ThermoFisher) and pooled together in equal concentrations (>10 ng/µL). The pooled samples were then concentrated using a Wizard SV Gel and PCR Clean-Up System (Promega). Quality control of the pooled and concentrated sample was done using an Agilent 2100 Bioanalyzer. Sequencing was performed using Illumina MiSeq

technology with a v3 (600 cycles) reagent kit.

3.3.6. Amplicon sequence analysis

De-multiplexed raw reads from the sequencing run were processed in R (R Development Core Team, 2011) using the DADA2 pipeline 1.4.0 (Callahan et al., 2016). First 10 bp of forward and reverse reads were trimmed and reads with a maximum expected error rate >1 was discarded. Chimera detection implemented in DADA2 was then performed on pooled samples. To account for the possibility of real chimeras between protein coding genes from closely related organisms (due to recombination or homoplastic mutations), chimeras were compared with a reference dataset of *viuB* alleles found in 782 sequenced *V. cholerae* genomes (obtained from GenBank). Only *viuB* alleles composed of more than 1,000 reads, found in multiple samples (with an average of 100,000 reads per sample) were considered for further analysis. Samples were rarefied to the level of the sample with the lowest reads using mothur 1.39.5 (Schloss et al., 2009), and further analysis was performed in R, with statistical tests and distance calculations performed using the VEGAN 2.4-6 package (Oksanen et al., 2008). Bray-Curtis similarity was calculated based on relative read abundance of each allele in different samples in Primer-E Software Suite and used for similarity percentage (SIMPER) and non-metric multi-dimensional scaling (NMDS) analysis.

3.3.7. Whole-genome sequencing and core genome phylogeny

The genomes of 23 strains from Dhaka belonging to various *viuB* genotypes were chosen for whole-genome sequencing as described in Orata *et al.* (Orata et al., 2015). Sequencing libraries were prepared from the genomic DNA using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA) and sequenced using Illumina MiSeq sequencing platforms (2 × 250-bp paired-end reads). Quality

control and *de novo* assembly of the reads were done using default parameters in CLC Genomics workbench 7 (Qiagen). Whole-genome alignment was performed using Mugsy v1.2.3 (Angiuoli & Salzberg, 2011) with default parameters, and a maximum likelihood tree was built from this alignment using RaxML v8 (Stamatakis, 2014) under the GTR+GAMMA model with 100 bootstrap replicates. Additional *V. cholerae* genomes were downloaded from GenBank. The maximum likelihood phylogenomic tree was constructed from the alignment of locally collinear blocks (2,094,734 bp) using GTR gamma substitution model with 100 bootstrap replicates.

3.3.8. Comparative genomic analysis

The genome sequences were annotated with RAST 2.0 (Aziz et al., 2008). Genomic distances were calculated in Geneious (Kearse et al., 2012). Core and accessory genes were determined with BPGA finding orthologous protein-coding genes clustered into families based on a 30% amino acid sequence identity (Chaudhari et al., 2016). Group specific genes were clustered using a custom-made Python program. BLAST atlas of the genomes and genomic islands were carried out using gview server (https://server.gview.ca/). Effector and immunity genes in Type VI secretion system loci were typed as previously described in Kirchberger et al., 2017 (Paul C. Kirchberger et al., 2017).

3.4. Results and Discussion

3.4.1. Pandemic related strains increase total *V. cholerae* abundance in Dhaka and reduce local diversity

One of the main differences between the *V. cholerae* populations from Oyster Pond (Falmouth, USA) and Dhaka (Bangladesh) is, unsurprisingly, the abundance of the pandemic generating (PG) lineage,

which includes strains responsible for the current 7th pandemic. Water samples were previously collected biweekly from seven different sites in the water bodies surrounding Dhaka city for nine continuous months (from June, 2015 to March, 2016), as well as a population from Oyster Pond over the summers of 2008 and 2009 in Cape Cod on the USA east coast (P C Kirchberger et al., 2016). Here we compare the *V. cholerae* populations from these two areas to gain insights on the differences between a region that is non-endemic for cholera and experiences strong seasonal variation, with a tropical area endemic for the disease. High-throughput sequencing of *viuB* marker gene amplicons was used to analyse the subspecies composition of *V. cholerae* in these two populations. Amplicons of this gene were annotated following a previously established scheme (P C Kirchberger et al., 2020), in which diversity within the *V. cholerae* species is measured based on relative abundance and distribution of *viuB* alleles. Each allele represents a *V. cholerae* lineage, the diversity of which is roughly equivalent to that of a clonal complex as traditionally defined by Multi-locus Sequence Typing (P C Kirchberger et al., 2020). A single *viuB* allele (*viuB*-73) is found to be uniquely associated with the pandemic generating (PG) lineage. Abundance and distribution of *viuB* alleles in samples collected from the two locations were estimated from *viuB* amplicon sequencing data normalized by quantitative data of *viuB* gene copy numbers determined by qPCR (Nasreen et al., 2020). Total abundance of *V. cholerae* in the two locations varied significantly (Kruskal-Wallis test, $p<0.1$), being almost twice as high on average in Dhaka (2.30×10^5 gene copies/litre) than in Oyster Pond (1.25×10^5 gene copies/litre) (**Fig. 1A**). However, when PG *V. cholerae* O1 (*viuB*-73) were excluded (quantified independently of other lineages using qPCR of the *rfb*O1 gene), average abundance was very similar in the two locations (Kruskal-Wallis, $p<0.01$). The PG lineage was the predominant genotype in Dhaka, with an average abundance of $1.4X10^5$ *rfb*O1 gene copies/litre, whereas it was just a minor member of the population in Oyster Pond, with an average abundance of 1.5×10^4 gene copies/litre (**Fig. 1A**). qPCR analysis confirmed that PG *V. cholerae* O1

present in the Oyster Pond population were non-toxigenic (CTX negative), as opposed to the vast majority of PG *V. cholerae* O1 in Dhaka being toxigenic (CTX positive) (Nasreen et al., 2020). Similarity percentage (SIMPER) analysis based on Bray Curtis dissimilarity suggests that the allele most responsible for the overall dissimilarities between Dhaka and Oyster Pond is indeed *viuB*-73. This allele was predominant throughout the nine month sampling period in Dhaka (Mohammad Tarequl Islam, Tania Nasreen, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C. Kirchberger, Marzia Sultana, Rebecca J. Case, n.d.), constituting around 60% of the total *V. cholerae* population on average whereas its presence was stochastic in Oyster Pond, with around 5% of the total population (P C Kirchberger et al., 2020). Population structure indices (Shannon diversity and Pielou's evenness) were significantly lower in Dhaka than in Oyster Pond (Kruskal-Wallis test, $P<0.1$) (**Fig. 1B**). This indicates a more stable and diverse *V. cholerae* community structure in the coastal location and a less diverse community dominated by fewer alleles in inland Bangladesh (Dhaka), likely because of the dominance of *viuB*-73 in that environment. Dhaka's aquatic reservoirs therefore seem to harbour a *V. cholerae* community highly dominated by the PG lineage that is most likely to be affected substantially by human activity. It is one of the most densely populated megacities in the world and has long history of suffering from recurring cholera (Rafique et al., 2016). Sustenance of the cholera causing genotype (PG) in the environment could be the driving factor to shape the overall population of *V. cholerae* in Dhaka. The reduction of intra-species diversity by PG *V. cholerae* in cholera endemic Dhaka could be attributed to the potential selective advantage of colonizing human gut (Shapiro et al., 2016), which would result in a constant output to water reservoirs. Type six secretion-mediated killing could also lead to the reduction of diversity, giving advantage to PG *V. cholerae* in a resource limited competitive environment, where PG is a superior competitor to other *V. cholerae* lineages under certain conditions. For example, in competition assays PG *V. cholerae* have been shown to outcompete others lineages in higher temperature (37°C), whereas at lower temperature

(25°C) other environmental lineages could outcompetehtem (Nora Hussein, Paul Kirchberger, n.d.). Environmental conditions, i.e. the lower salinity seen in Dhaka (**Table S1**) could also give advantage to PG strains over others, as they have been shown to be more prevalent in lower salinity environments relative to other lineages (P C Kirchberger et al., 2020).

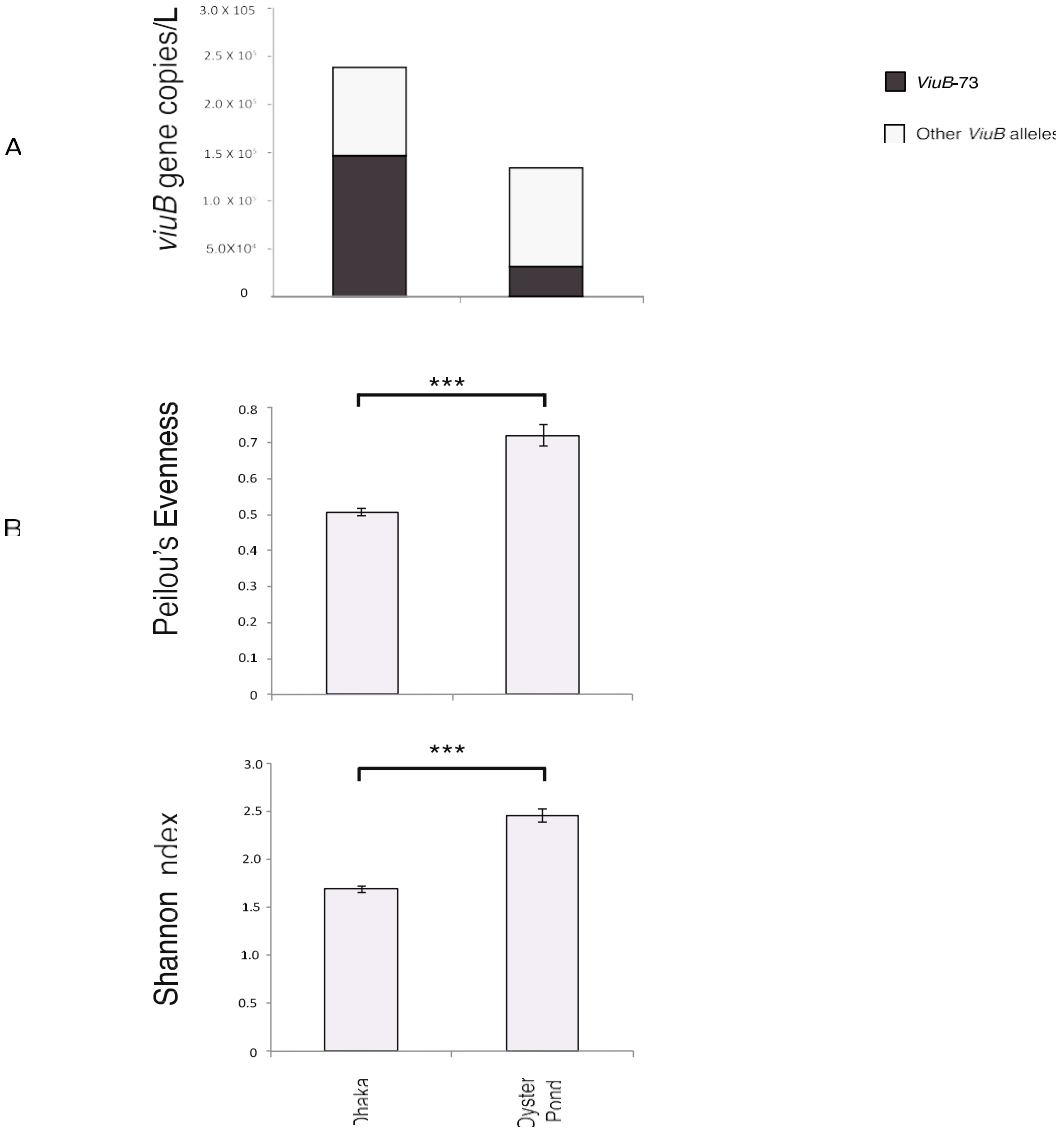


Fig. 3.1. Abundance and Diversity of *Vibrio cholerae* populations in two geographic locations: Dhaka and Oyster Pond. A: Absolute average abundance of *V. cholerae* in two locations. *viuB* gene copy numbers were quantified from qPCR data; average of *viuB* gene copies for all the samples in two locations were calculated and used as the proxy for the *V. cholerae* abundance. Total height of the bar represents total *V. cholerae* (*viuB*), black segment represents *viuB*-73 and clear

segment represents other *viuB* alleles. B: Evenness and diversity of the two *V. cholerae* populations measured by Peilou's evenness and Shannon diversity indices based on analysis of *viuB* alleles. Statistical significance was measured by Kruskal-Wallis test; *** : statistically significant differences (Kruskal-Wallis $p<0.1$).

In Oyster Pond, *V. cholerae* were present in substantially larger abundance in a coastal pond and lagoon compared to nearby ocean waters, indicating a likely ecological barrier (Paul C. Kirchberger et al., 2020). Overall, it is plausible that the combination of human population density and environmental factors creates the conditions for the sustenance of a natural *V. cholerae* population dominated by the PG lineage in Dhaka.

3.4.2 A novel divergent lineage is endemic to inland Bangladesh

Besides the PG lineage, the population composition of *V. cholerae* sampled over 6 to 9 months was strikingly different in Dhaka and Oyster Pond. This was determined by using the abundance and distribution data of individual *viuB* alleles from the two locations. Non-metric multi-dimensional scaling (NMDS) was performed to compare the two communities and statistical significance of community structure dissimilarity was evaluated using the analysis of similarity (ANOSIM) with a Bray-Curtis distance matrix. In the NMDS plot, samples from Dhaka and Oyster Pond clustered separately and community structure dissimilarity was statistically significant (ANOSIM $R=0.75$, P value <1%) (**Fig. 2**).

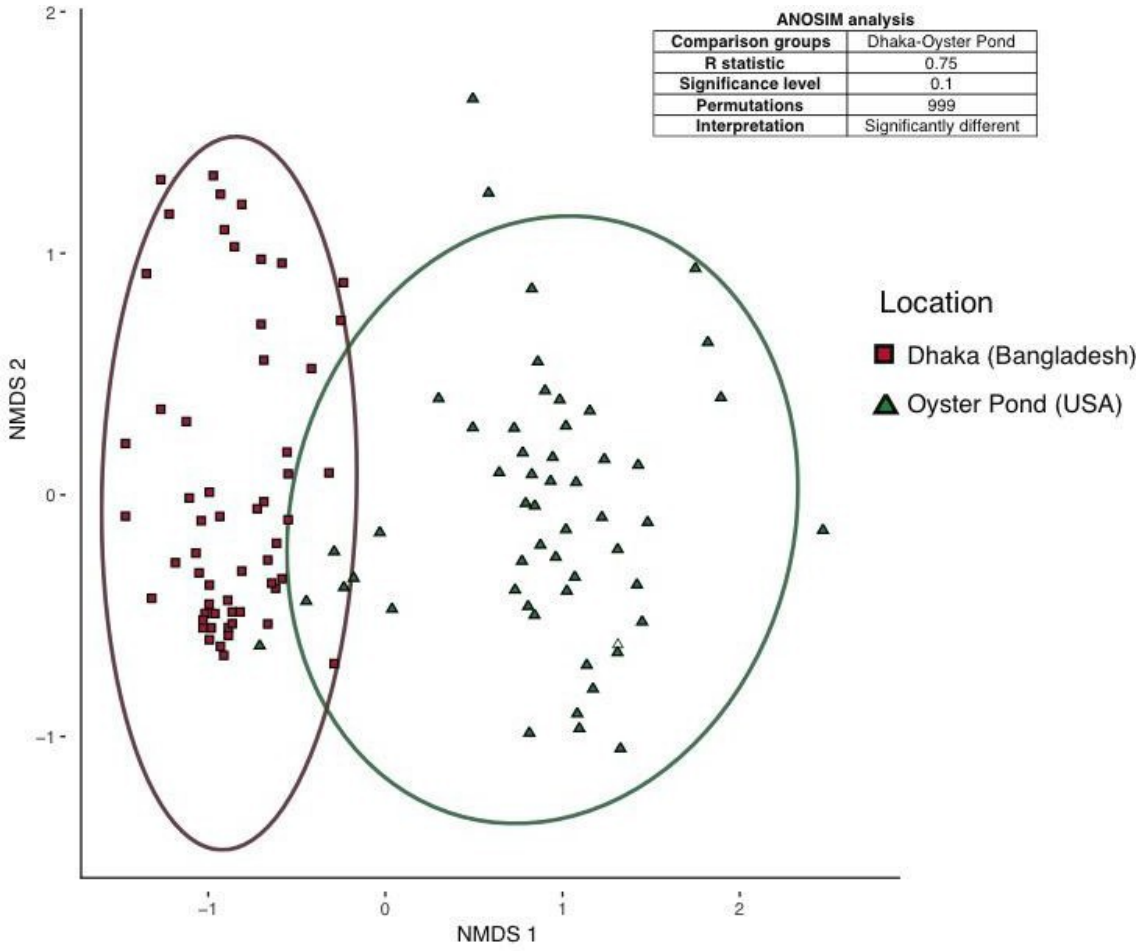
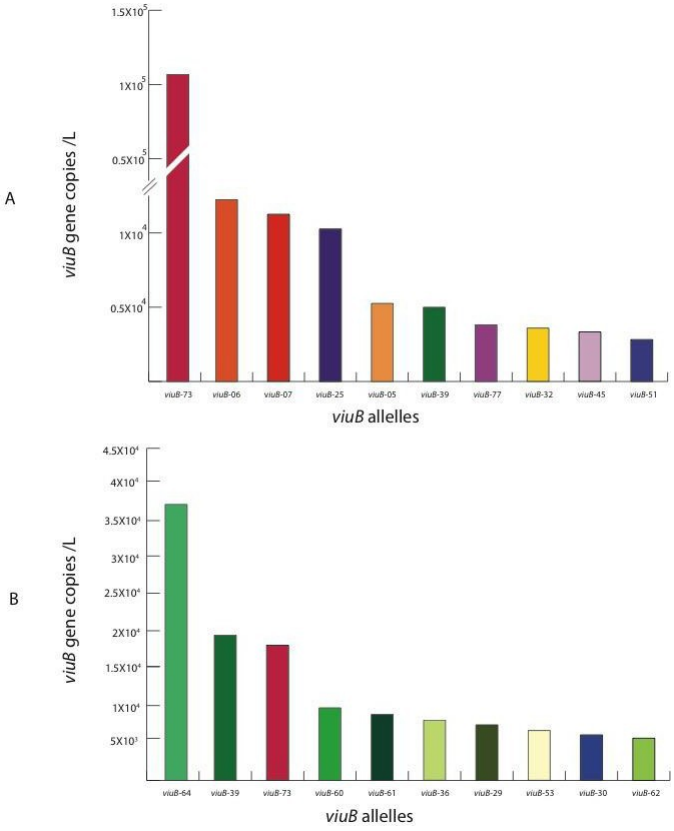


Fig. 3.2. Non-metric multi-dimensional scaling (NMDS) plot comparing beta diversity of *Vibrio cholerae* populations from two aquatic environments. Population compositions were compared using Bray–Curtis dissimilarity matrix with ellipses representing 95% confidence intervals. Dataset was composed of *viuB* gene amplicon sequences normalized by qPCR copy numbers. NMDS plot (stress 0.16) shows distinct clustering of samples from the two locations shown along the first two axes labeled as NMDS1 and NMDS 2. Analyses of similarity (ANOSIM) results are displayed in the box inside the plot describing dissimilarity between pairs of samples from the two locations.

Only two major alleles were shared between these locations from a total of 13 *viuB* alleles in Dhaka and 15 alleles in Oyster pond (each individual allele constituting at least 1% of the *V. cholerae* population). The most abundant alleles in Dhaka after *viuB*-73 were *viuB*-06, *viuB*-07, *viuB*-25 and *viuB*-05 (**Fig. 3**). Of these four, three are exclusively found in Dhaka (*viuB*-05, *viuB*-06 and *viuB*-07) and are of particular interest. Together, they composed ~15% of the average Dhaka *V. cholerae* population and have been found to display higher abundance in sites surrounded by a high human population density and levels of pollution (19).



Fig, 3.3. Abundance of the most prevalent *viuB* alleles at two locations: A. Dhaka (Bangladesh); B. Oyster Pond (USA). Total *viuB* gene copy numbers were obtained by qPCR. Relative abundance of each allele was determined by amplicon sequencing. Specific colours were used for individual alleles to be consistent with the scheme described elsewhere (P C Kirchberger et al., 2020). The ten most abundant alleles for each location were selected for comparison between the two locations.

Table 3.1: Demographic information of the *Vibrio paracholerae* sp. nov. and *Vibrio cholerae* isolates used in this study.

Isolates	Origin	Year of isolation	Source	Phylogenetic group	Genome accession number
N16961	Bangladesh	1971	Clin	PG *Vibrio cholerae* O1	GCA_000006745.1
2010EL1786	Haiti	2010	Clin	PG *Vibrio cholerae* O1	GCA_000166455.2
EDC 721	Bangladesh	2015	Env	PG *Vibrio cholerae* O1	WYCI00000000
MJ1236	Bangladesh	1994	Clin	PG *Vibrio cholerae* O1	GCA_000022585.1
MO10	India	1992	Clin	PG *Vibrio cholerae* O1	GCA_000152425.1
2740-80	USA	1980	Env	PG *Vibrio cholerae* O1	GCA_001683415.1
BX3330286	Australia	1985	Env	PG *Vibrio cholerae* O1	GCA_000174335.1
O395	India	1965	Clin	PG *Vibrio cholerae* O1	GCA_000021625.1
95412	Mexico	1997	Env	PG *Vibrio cholerae* O1	GCA_000348105.2
V52	Sudan	1968	Clin	PG *Vibrio cholerae* O37	GCA_000167935.2
2012EL1759	Haiti	2012	Env	PG *Vibrio cholerae* O1	JNEW01000000
2012Env9	Haiti	2012	Env	Non-PG *Vibrio cholerae* O1	GCA_000788715.2
MZO-03	Bangladesh	2001	Clin	Non-PG *Vibrio cholerae* non-O1/O139	GCA_000168935.3
12129	Australia	1985	Env		ACFQ00000000

				Non-PG *Vibrio cholerae* O1	
1587	Peru	2000	Clin	Non-PG *Vibrio cholerae* non-O1/O139	GCA_000168895.2
CP1035	Mexico	2004	Clin	Non-PG *Vibrio cholerae* non-O1/O139	AJRM00000000
EDC 689	Bangladesh	2015	Env	Non-PG *Vibrio cholerae* non-O1/O139	WYCR00000000
EDC 715	Bangladesh	2016	Env	Non-PG *Vibrio cholerae* non-O1/O139	WYCJ00000000
EDC 800	Bangladesh	2016	Env	Non-PG *Vibrio cholerae* non-O1/O139	WYCA00000000
EM1676A	Bangladesh	2011	Env	Non-PG *Vibrio cholerae* non-O1/O139	GCA_000348345.2
2012ENV-92	Haiti	2012	Env	Non-PG *Vibrio cholerae* non-O1/O139	GCA_000788755.1

HE 48	Haiti	2010	Env	Non-PG *Vibrio cholerae* non-O1/O139	GCA_000220785.2
VCC19	Brazil	1994	Env	*Vibrio paracholerae*	GCA_000438805.2
877163	Bangladesh	2002	Env	*Vibrio paracholerae*	GCA_001402745.1
EDC 792	Bangladesh	2016	Env	*Vibrio paracholerae*	WYCC00000000
EDC 690	Bangladesh	2015	Env	*Vibrio paracholerae*	WUWI00000000
EDC 716	Bangladesh	2015	Env	*Vibrio paracholerae*	WYBZ00000000
EDC 717	Bangladesh	2015	Env	*Vibrio paracholerae*	WYBY00000000
HE09	Haiti	2010	Env	*Vibrio paracholerae*	GCA_000221405.1
HE16	Haiti	2010	Env	*Vibrio paracholerae*	GCA_000303085.1
CISM300506	Mozambique	2008	Clin	*Vibrio paracholerae*	GCA_002097755.1
CISM1163068	Mozambique	2012	Clin	*Vibrio paracholerae*	GCA_002097815.1
49093 DA89	Thailand	1992	Clin	*Vibrio paracholerae*	GCA_000737015.1
SIO	USA	2003	Waste water	*Vibrio paracholerae*	GCA_001857455.1
2017V-1144	USA	2017	Stool	*Vibrio paracholerae*	GCA_003312015.1
2014V-1107	USA	2014	Stool	*Vibrio paracholerae*	GCA_003311945.1
2017V-1105	USA	2017	Wound	*Vibrio paracholerae*	GCA_003311975.1
2016V-1114	USA	2016	Stool	*Vibrio paracholerae*	GCA_003312085.1
2016V-1111	USA	2016	Stool	*Vibrio paracholerae*	GCA_003311965.1
2017V-1176	USA	2017	Animal feed	*Vibrio paracholerae*	GCA_003312095.1

2016V-1091	USA	2016	Stool	*Vibrio paracholerae*	GCA_003312065.1
2017V-1110	USA	2017	Wound	*Vibrio paracholerae*	GCA_003312005.1
87395	UK	UK	UK	*Vibrio paracholerae*	GCA_000348085.2
07-2425	UK	UK	UK	*Vibrio paracholerae*	GCA_003311905.1
NCTC30	Egypt	1916	Clin	*Vibrio paracholerae*	LS997868.1

PG, Pandemic generating; Clin, clinical; Env, environmental; UK, unknown.

To have more information on the lineages found in Dhaka, 23 *V. cholerae* strains isolated from the city during the study period were selected for whole genome sequencing: nine *V. cholerae* O1 harbouring the *viuB*-73 allele and fourteen *V. cholerae* non-O1/O139 isolates displaying a diversity of *viuB* alleles. Four strains possessed *viuB* alleles 05, 06, 07 and 08 (EDC690, EDC716, EDC717 and EDC792) and were found to be part of a very long branch occupying a basal position in a global core genome phylogeny compared to the rest of the *V. cholerae* strains (hence termed Long Branch clade or LB) (**Fig. 4**). This phylogenetic group has not been described in any other studies, although other strains from public databases, isolated from different parts of the world, also belong to it (**Table 1**). Nine isolates were recovered from human clinical specimens across the United States and reported to the Centre for Disease Control (CDC) as part of the surveillance conducted under the Cholera and Other *Vibrio* Illness Surveillance (COVIS) program (Islam et al., 2018). Two more isolates originate from stool samples of diarrheal patients in Mozambique in 2008 (Garrine et al., 2017) and one isolate was recovered from a diarrheal patient from Thailand in 1993 and described as *V. cholerae* serogroup O155 (Bishop-Lilly et al., 2014). Seven additional isolates have been found to belong in the clade for a total of twenty-three as of August 2019 (**Table 1**).

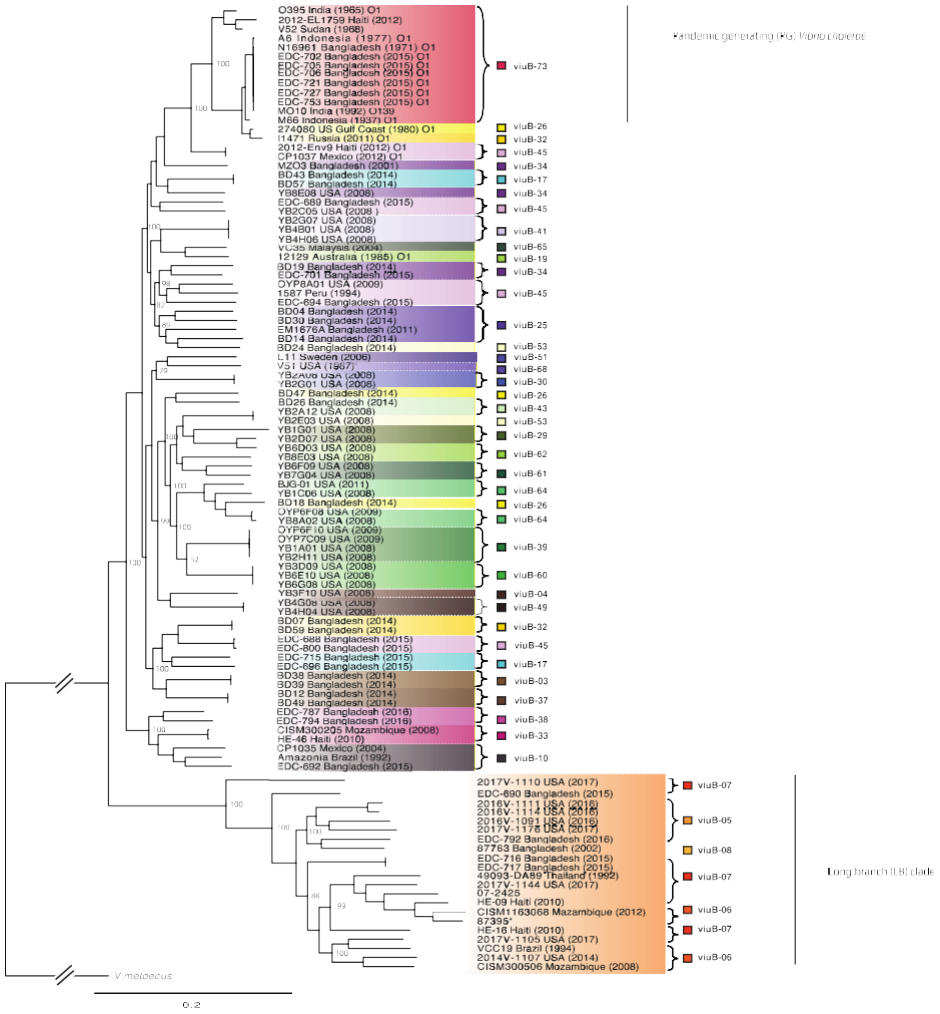


Fig. 3.4. Whole-genome phylogeny of *Vibrio cholerae* strains found in Dhaka and Oyster Pond populations. The phylogenetic tree was inferred using Parsnp v1.2 (Treangen et al., 2014) based on the reference genome of *V. cholerae* O1 El Tor N16961 and includes representative strains from other environments**.** Leaves of the tree were colored according to the *viuB* allele found in that genome. Statistical support of relevant nodes was estimated by bootstrap analysis (1000 replicates, indicated as a percentage). The scale bar represents nucleotide substitutions per site**.**

Table 3.2: Phenotypic traits differentiating *Vibrio paracholerae* sp. nov. from its closest relatives *Vibrio cholerae* and *Vibrio metoecus*.

Phenotypic test	*Vibrio paracholerae* sp. nov.				*Vibrio cholerae*				*Vibrio metoecus*			
	1	2	3	4	5	6	7	8	9	10	11	12
***α*-cyclodextrin**	+	+	+	+	-	-	-	-	+	+	-	+
Pectin	+	+	+	+	-	-	-	-	+	+	+	+
Mono methyl succinate	-	-	-	-	+	+	+	+	-	-	-	-
D-Mannose	-	-	+	-	+	+	+	+	+	+	+	+
L-aspartic acid	-	-	+	-	+	+	+	+	+	+	+	+
Citric acid	-	-	+	-	+	+	+	+	+	+	+	+
***α*-keto glutaric acid**	-	-	+	+	+	+	+	+	-	-	+	-
N-acetyl-D-galactosamine	-	-	-	-	-	-	-	-	+	+	+	+
D-glucuronic acid	-	-	-	-	-	-	-	-	+	+	+	+
Acetoin production	+	+	+	+	+	+	+	+	-	-	-	-

Tested strains: 1, EDC 792; 2, EDC 690; 3, 2016V-1111; 4, 2016V-1091; 5, N16961; 6, V52; 7, YB3B05; 8, YB8E08; 9, 082459; 10, OP6B; 11, OP4B; 12, OP3H. +, Growth/positive test result; -, no growth/negative test result; ND, not determined. †Results for *V. cholerae* and *V. metoecus* strains were obtained from Kirchberger *et al* (Paul C Kirchberger et al., 2014).

3.4.3. A sister species to *Vibrio cholerae*?

Comparative genome analysis suggests that LB isolates could represent a new species, which would be the closest relative of *V. cholerae* described to date. Based on the genome sequences, G+C content of the strains belonging to LB clade was 46-48.1%, falling within the known range of the genus *Vibrio*. Overall, representative strain (EDC-792) of the clade shared 2883 genes with type strains from both *V. cholerae* (N16961) and *V. metoecus* (OP3H) and 2996 genes with *V. cholerae* alone (Fig. S1). The genomes of twenty-two LB isolates were compared with a set of *V. cholerae* strains containing the same number (n=22) of representatives from both pandemic and non-pandemic lineages (**Table 1**). This comparison revealed that genetic distance between LB strains and *V. cholerae* fall below or at the threshold of the species cut-off values. Indeed, Digital DNA-DNA hybridization (dDDH) values ranged from 82-100% within the LB clade and 69-70% with *V. cholerae,* whereas Average Nucleotide Identity (ANI) values ranged from 97-100% within the group and 95-96% with *V. cholerae* strains, respectively. DDH values are considered to be the gold standard for species designation and a value of ≤70% presents as an indication that the tested organism belongs to a different species than the type strain(s) used as reference (Auch et al., 2010). ANI has been proposed as an alternative genomic statistics to DDH and the cut off values of 95-96% has been used for species delineation (Ciufo et al., 2018). In this case, all the strains from the LB clade had a dDDH value of ≤69% and ANI value of ≤96% when compared with the *V. cholerae* type strain N16961. In large scale taxonomic studies, dDDH has been shown to overpower ANI values for the purpose of species designation (Jin et al., 2020; Meier-Kolthoff et al., 2014), suggesting that dDDH values should be given more importance when other evidence (phylogenetic, phenotypic and ecological) support a novel species designation. Thus, according to the current species definition (Richter & Rossello-Mora, 2009), the LB clade meets the genotypic criteria to qualify as a candidate for a novel species designation. It also meets the phylogenetic criteria, as it

biochemical and growth characteristics, they clearly differed for some phenotypic characteristics (**Table 2**). All four LB strains tested could utilize ***α*-**cyclodextrin as a sole carbon source, whereas none of the tested *V. cholerae* strains could. Cyclodextrin utilization requires a specific category of amylases, which has not been reported in *V. cholerae* so far (Fiedler et al., 1996). *In silico* analysis revealed that LB strains possess a gene cluster (genes 03367 to 03379 in the NCTC30 genome, NZ_LS997867) containing homologs of genes encoding cyclomaltodextrin glucanotransferase (*amyM*), ABC transporter MalK (*malK*), glycosidase MalE (*malE*), glucosamine N-acetyltransferase, cyclodextin specific porin (*cycA*), cyclodextrin binding protein (*cycB*), cyclodextrin transport system permease (*malF*), cyclodextrin transport system permease (*malD/malG*) and neopullulanase (*nplT*) (**Table 3**). Only two (9%) *V. cholerae* strains in our dataset (n=22) and a similar percentage in the NCBI database possessed this cluster whereas 100% of the LB clade strains (n=22) harboured it. This gene cluster might be associated with the cyclodextrin degradation phenotype, as reported previously (Davis and Park, 1962). LB isolates also differed clearly from *V. cholerae* in the utilization of pectin and mono methyl succinate (Table 2). In contrast with *V. cholerae*, 75% (3 out of 4) of LB isolates tested were found to be lacking the ability to utilize D-mannose, L-aspartic acid and citric acid (**Table 2**). D-mannose was found to be readily utilized by both *V. cholerae* and *V. metoecus* tested in this study and previous literature reported that ~80% of *V. cholerae* are capable of utilizing this sugar (Davis et al., 1981). The gene cluster encompassing *manP* to *manA* (VC1820 to VC1827 in N16961 genome, AE008352.1), including the well-known mannose-6 phosphate isomerase (*manA*) gene required for this process (Sun & Altenbuchner, 2010), was present in all the tested *V. cholerae* (n=22) and *V. metoecus* (n=22) strains, whereas it was found in only ~40% (9/22) of LB strains **(Table 3**). Notably, three isolates that were unable to utilize mannose lack the gene cluster *manP-manA*, whereas the only isolate (2016V-1111) of the four tested isolates from the LB clade which could utilize mannose (**Table** 2) possesses the gene cluster in its genome. We could not find the genetic

basis for the other phenotypic differences between LB isolates and *V. cholerae*. LB strains are similar to *V. cholerae* in N-Acetyl-D-Galactosamine D- glucuronic acid utilization tests and acetoin production, which differentiates both of the groups from *V. metoecus* (P C Kirchberger et al., 2014). Resistance to 96 drugs or metals were also tested at different concentrations, and *V. cholerae* and LB isolates showed similar profiles in most, although three chemicals elicited differential responses by the two species. LB isolates were resistant to cadmium chloride, sodium selenite and dichlofluanid in contrast to the sensitivity of the *V. cholerae* strains towards those chemicals (**Table S2**). Thus, phylogenetic, genotypic, phenotypic and ecological data supports the designation of a novel species for the LB clade. Very recently, genome sequencing efforts of a historical collection of isolates from cholera or cholera-like diseases have identified a strain isolated during the first World War (in 1916) from a soldier convalescent in Egypt as a divergent *V. cholerae* (Dorman et al., 2019). This NCTC30 strain actually belongs to the LB clade found in this study (**Fig. 5**). Interestingly, NCTC30 was initially designated as "*Vibrio paracholerae*" and the disease caused was described as choleraic and termed as 'paracholera' (Gardner and Venkatraman, 1935). To honour its history, we propose the name *Vibrio paracholerae* sp. nov. (EDC-792^{T}) for this novel species. Inclusion of this novel species would revise the phylogeny of the *V. cholerae* species complex, making *V. paracholerae* sp. nov. the sister species to *V. cholerae* and *V. metoecus* as the sister species to *V. mimicus* (**Fig. 5**). This improved description of the species complex containing pandemic *V. cholerae* will provide a framework for understanding the emergence and evolution of human pathogenic Vibrios from their environmental common ancestor.

Table 3.3. Major genetic traits differentiating *Vibrio paracholerae* sp. nov. from its closest relatives: *Vibrio cholerae* and *Vibrio metoecus*

Genomic island/Gene cluster	Genomic position in reference genomes		Present in % strains			Putative function	Reference
	N16961 (AE003852.1) locus	NCTC30 (LS997868.1) locus	VC (n=22)	VP (n=22)	VM (n=22)		
RND efflux pump gene cluster	Absent	818-823	0	90	0	Resistance to antimicrobials and heavy metals	This study
GI-66	Absent	1923-1927	0	100	0	Iron regulation	35
Cyclo-maltodextrin operon	Absent	3367-3375	9	100	68	Cyclodext-rin utilization	27
VCA1102-1111 (N16961)	VCA_1102-VCA_1111	Absent	100	0	100	Fatty acid biosynthesis, Heme billiverdin,	This study

						thermost-able hemolysin	
Tor operon	VC_1692-VC_1694, VC_1719-VC_1720	Absent	100	0	100	Virulence gene regulation	30
Glutathione regulated Potassium pump	VC_2606-VC_2607	Absent	100	0	100	Potassium regulation	This study
Beta lactamase	Absent	3210	0	60	0	Resistance to β-lactams	25

VC, *Vibrio cholerae*; VP, *Vibrio paracholerae* sp. nov; VM, *Vibrio metoecus.* Reference genomes N16961 (*V. cholerae*) and NCTC30 (*V. paracholerae* sp. nov.) were used for determining locus positions of the gene clusters.

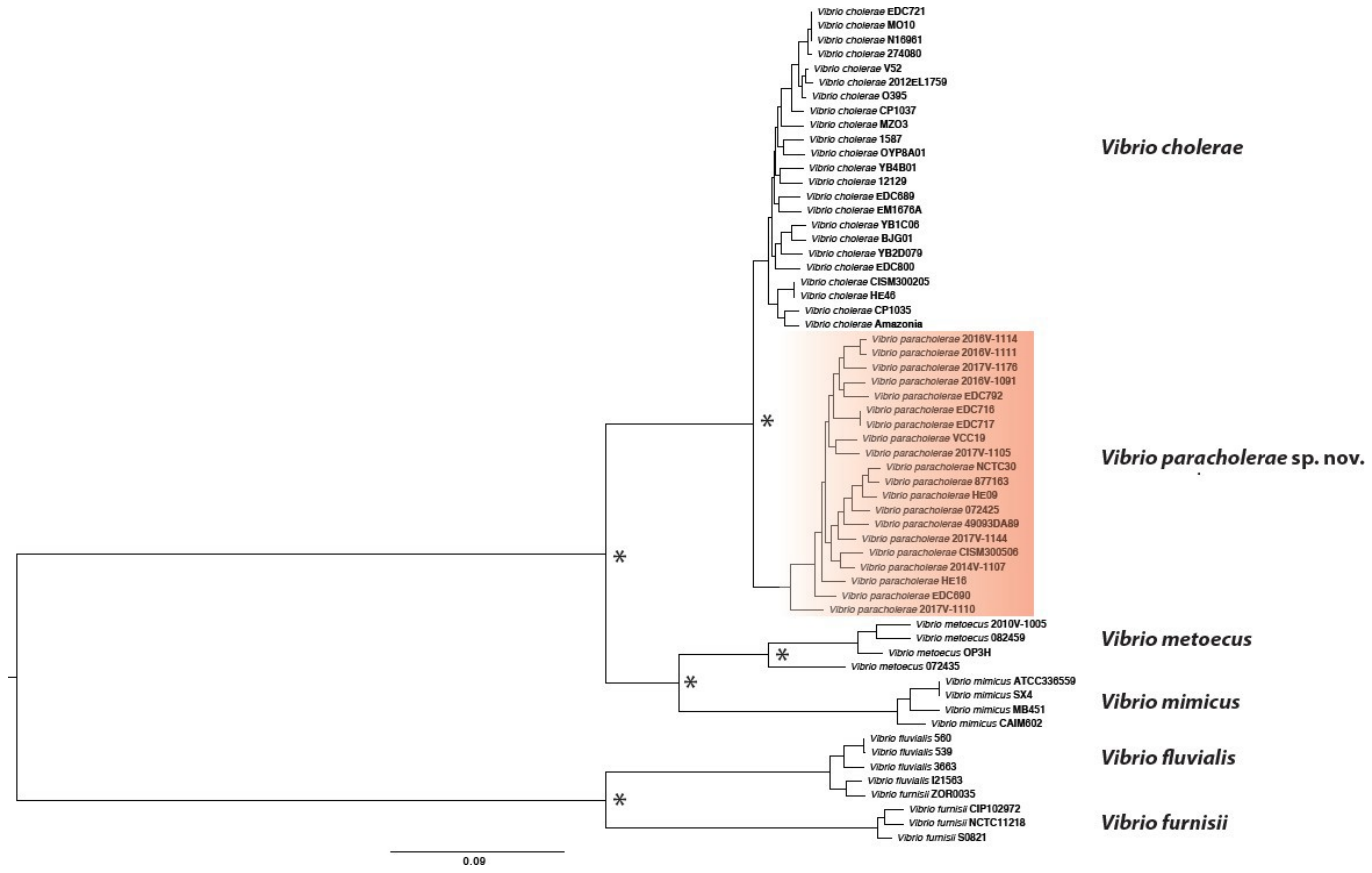


Fig. 3.5. Whole-genome phylogenetic tree of *Vibrio paracholerae* sp. nov. along with its closest sister species. The maximum likelihood phylogenetic tree was constructed from the core genome alignment of ≈2.1M bp using GTR gamma substitution model. Corresponding nodes with relevant Bootstrap support over 70% from the 100 replicates were indicated with *. The scale bar represents nucleotide substitutions per site.

3.4.4. A potential threat to humans?

To be a successful disease-causing agent to humans, a bacterial pathogen of aquatic origin needs to have the ability to survive in the environment and colonize the human body. In cholera endemic Dhaka, *V. paracholerae* sp. nov. has been found to exist abundantly in local water reservoirs. In one particular site, the number even surpassed that of PG *V. cholerae*, which was otherwise the most predominant lineage found in Dhaka(Islam et al., 2021). That site (Kamrangir char) happened to be *V. paracholerae* sp. nov. strains lack CTX, VPI-1, and VPI-2 (Table S3; Fig. S3), three major elements known to be essential for *V. cholerae* to cause cholera (Kaper et al., 1992). They also lack

lack a cluster of genes (VC1692, VC1694, VC1719, and VC1720 in the N16961 genome) encoding proteins for the "Tor operon" required for trimethylamine N-oxide respiration in *V. cholerae.* The genes in this operon have been shown to be crucial for cholera toxin production, cytotoxicity, and intestinal colonization of *V. cholerae* in an infant mouse model (Lee et al., 2012). The Tor operon was found in 100% of *V. cholerae* (n = 22) and *V. metoecus* (n = 4) strains in our data set, which indicates that it was likely lost in the *V. paracholerae* sp. nov. phylogenetic branch. All the *V. paracholerae* sp. nov. strains in our data set possessed the RTX toxin gene cluster (Table S3), a virulence factor for *V. cholerae* known to have a role in interactions with eukaryotes (Lin et al., 1999). Interestingly, four *V. paracholerae* sp. nov. strains (22%) (including NCTC30) possess type three secretion system (T3SS) genes (Table S3), an established virulence factor for nonpandemic *V. cholerae* (33). The historical strain NCTC30, which was isolated from a choleraic patient, contains a T3SS genetically more similar to a *V. parahaemolyticus* T3SS than that of one found in a subset of *V. cholerae* (acquired within a VPI-2 insertion site locus of PG *V. cholerae*) (Dziejman et al., 2005). The region of *V. paracholerae* sp. nov. strains containing this T3SS is 97% identical at the nucleotide sequence level to the T3SS island found in *V. parahaemolyticus*, highlighting a recent acquisition from this human-pathogenic species (Fig. S4). However, NCTC30 does not have the mobile integrative conjugative element (ICE) SXT/RC39, whereas four *V. paracholerae* sp. nov. strains contain this element in their genomes (Table S3). SXT/ RC39-related ICEs are found in most PG *V. cholerae* strains since 2001 and are believed to be involved in improved fitness in 7th pandemic El Tor *V. cholerae* (Mutreja et al., 2012). ICEs found in *V. paracholerae* sp. nov. strains showed 95.7% to 96.4% nucleotide sequence identity in the conserved homologous regions to the first reported SXT element in

V. cholerae SXTMO10, originally detected in V. cholerae O139 strain MO10. However, ICEs found in *V. paracholerae* sp. nov. strains were more related to ICEVchMex1, which was first reported to be present in an environmental *V. cholerae* non-O1/O139 strain (Burrus et al., 2006) based on the nucleotide sequence of the integrase (int) gene.

V. parahaemolyticus T3SS than that of one found in a subset of *V. cholerae* (acquired within VPI-2 insertion site locus of PG *V. cholerae*) (Dorman et al., 2019). The region of *V. paracholerae* sp. nov. strains containing this T3SS is 97% identical at the nucleotide sequence level to the T3SS island found in *V. parahaemolyticus, highlighting* a recent acquisition from this human pathogenic species (**Fig. S4**). However, NCTC30 does not have the mobile integrative conjugative element (ICE) SXT/ RC39, whereas four *V. paracholerae* sp. nov. strains contain this element in their genomes (Table S3). SXT/RC39 related ICE's are found in most PG *V. cholerae* strains since 2001 and is believed to be involved in improved fitness in 7th pandemic El Tor *V. cholerae* (Chun et al., 2009; Mutreja et al., 2013). ICE found in *V. paracholerae* sp. nov. strains showed 95.7-96.4% nucleotide sequence identity in the conserved homologous regions to the first reported SXT element in *V. cholerae*: SXT^{MO10}, originally detected in *V. cholerae* O139 strain MO10 (Burrus et al., 2006). However, ICE found in *V. paracholerae* sp. nov. strains were more related to ICEVchMex1, first reported to be present in a environmental *V. cholerae* non-O1/O139 strain (Burrus et al., 2006), based on the nucleotide sequence of the integrase (*int*) gene (Fig. S7). Additionally, 50% of the *V. parcholerae* sp. nov. strains also contained genes for the recently discovered cholix toxin, thought to be an important virulence factor for *V. cholerae* (Lugo & Merrill, 2015). Recently, a novel island has been found in the VPI-1 insertion site, designated as GI*Vch*S12, which contains CRISPR-Cas and type VI secretion system (T6SS) modules (Labbate et al., 2016). CRISPR-Cas is a defense mechanism that bacteria can utilize against unwanted lateral gene transfer by recognizing foreign DNA and

cleaving it. CRISPR-cas modules are usually unevenly distributed across taxa but can potentially provide fitness advantage to the bacteria in a mixed population (Labbate et al., 2016; Westra et al., 2019). Seven out of twenty-two strains (~31%) strains contain a version of GIVchS12 island in the region to the VPI-1 insertion site of PG *V. cholerae* (Fig. S5). T6SS operon of *Vibrio cholerae* contain diverse arrays of toxic effector and cognate immunity genes, which are thought to play an important role in the environmental lifestyle and adaptation to the human host (Hussain et al., 2021). The functional T6SS genetic element corresponds to three loci across the *V. cholerae* genome termed aux-1, aux-2 and large cluster (Unterweger et al., 2014). Like *V. cholerae* and its close relatives, *V. mimicus* and *V. metoecus* (Kirchberger et al., 2014), *V. paracholerae* sp. nov. possesses all three chromosomally encoded type VI secretion system (T6SS) loci and shares its effector & immunity gene pool with these species (Table S5). However, for a similar phylogenetic diversity, the diversity of the T6SS effector-immunity domain is much less in *V. paracholerae* sp. nov. when compared to that of *V. cholerae* observed (Hussain et al., 2021; Paul C. Kirchberger et al., 2017). This surprising lack of diversity could represent an ecological adaption in *V. paracholerae*, which could be driving a fitness advantage in specific niches from this restricted set of T6SS effector and immunity proteins.

Apart from the known virulence genes usually found in *V. cholerae*, gene content analysis revealed a few species-specific genetic traits in *V. paracholerae* sp. nov. Two genes were present in all 22 *V. paracholerae* sp. nov. strains, with no homolog found in any *V. cholerae* strains. These two genes encode a lysR family transcriptional regulator (WP_001924807.1) and HAD_IB family hydrolase (WP_071179638.1). Both of these genes are part of a previously reported genomic island (GI-66) found in *Vibrio albensis* (Choi et al., 2016). This GI contains iron-related regulatory

cleaving it. CRISPR-cas modules are usually unevenly distributed across taxa but can potentiallyprovide fitness advantage to the bacteria in a mixed population (Labbate et al., 2016; Westra et al., 2019). Seven out of twenty-two strains (~31%) strains contain a version of GIVchS12 island in the region to the VPI-1 insertion site of PG *V. cholerae* (Fig. S5). T6SS operon of *Vibrio cholerae* contain diverse arrays of toxic effector and cognate immunity genes, which are thought to play an important role in the environmental lifestyle and adaptation to the human host (Hussain et al., 2021). The functional T6SS genetic element corresponds to three loci across the *V. cholerae* genome termed aux-1, aux-2 and large cluster (Unterweger et al., 2014). Like *V. cholerae* and its close relatives, *V. mimicus* and *V. metoecus* (Kirchberger et al., 2014), *V. paracholerae* sp. nov. possesses all three chromosomally encoded type VI secretion system (T6SS) loci and shares its effector & immunity gene pool with these species (Table S5). However, for a similar phylogenetic diversity, the diversity of the T6SS effector-immunity domain is much less in *V. paracholerae* sp. nov. when compared to that of *V. cholerae* observed (Hussain et al., 2021; Paul C. Kirchberger et al., 2017). This surprising lack of diversity could represent an ecological adaption in *V. paracholerae*, which could be driving a fitness advantage in specific niches from this restricted set of T6SS effector and immunity proteins.

Apart from the known virulence genes usually found in *V. cholerae*, gene content analysis revealed a few species-specific genetic traits in *V. paracholerae* sp. nov. Two genes were present in all 22 *V. paracholerae* sp. nov. strains, with no homolog found in any *V. cholerae* strains. These two genes encode a lysR family transcriptional regulator (WP_001924807.1) and HAD_IB family hydrolase (WP_071179638.1). Both of these genes are part of a previously reported genomic island (GI-66) found in *Vibrio albensis* (Choi et al., 2016). This GI contains iron-related regulatory

genes that can be significant in regulation of iron scavenging in this group of organisms. Iron acquisition is thought to be an important aspect for regulation of virulence as well as host selectivity/specificity (Payne et al., 2016). Most (90%) of *V. paracholerae* sp. nov. strains also harboured a novel RND efflux pump (Table 3), thought to be critical for intrinsic and induced antimicrobial resistance, virulence gene expression, colonization in animal host and environmental regulation of stress response (Bina et al., 2008). Efflux pumps have been proposed to be important for expelling bile out of the cell, and the resulting bile resistance would be key to overcoming this challenge inside the human gut (Trastoy et al., 2018). RND efflux pumps have specifically been found to confer increased bile resistance in other gram-negative bacteria (Routh et al., 2010). The novel RND gene cluster is absent in both *V. cholerae* and *V. metoecus* but homologs have been found in the halophilic bacteria *V. cincinnatiensis* (Brayton et al., 1986) and a bile associated isolate of *V. fluvialis* (Zheng et al., 2017). Other than efflux pumps, ToxR and TolC have been also proposed to be crucial for bile resistance, and like *V. cholerae* strains, all the *V. paracholerae* sp. nov. strains possess both genes. All these factors make *V. paracholerae* sp. nov. a potential candidate for a species associated to the human gut microbial population and underscores the importance of studying their biology in greater detail.

3.4.5. Interaction of *Vibrio paracholerae* sp. nov. with pandemic *Vibrio cholerae* could have impacted the ecology and evolution of both species

Horizontal gene transfer (HGT) among species sharing an ecological niche can have a major impact on their evolution (Polz et al., 2013). As *V. paracholerae* sp. nov. (VP) co-exists with *V. cholerae* (VC) in natural ecosystems (at least in Dhaka), it is expected that HGT could take place between these two groups. To assess the propensity of interspecies HGT, potential gene transfer events between two

groups (VC and VP) were inferred based on phylogenetic congruence of individual genes. Maximum likelihood (ML) trees were constructed for each of the core and accessory gene families present in at least two strains from each group. A gene transfer was hypothesized if a member of a group clustered with members of the other group in a clade, and the gene tree could not be partitioned into perfect clades, which must consist of all members from the same group and only of that group (Orata et al., 2015; Schliep et al., 2011). In our groups of 22 VC and 22 VP strains, 216 HGT events were hypothesized involving 82 gene families from VC to VP, but only 62 events from VP to VC involving 33 gene families. All of the core genes transferred from VP to VC were acquired by strains outside of the PG group. In the case of accessory genes, we could infer 82 potential transfer events from VC to VP and 54 events from VP to VC. Only four events involved strains belonging to the PG clade. Thus, gene transfer directionality was biased from VC to VP, VP being the recipient of HGT in most cases. Lower rate of HGT towards *V. cholerae* was previously reported in case of the co-occurring *V. metoecus,* which has a lower abundance in the environment (Orata et al., 2015). This gene transfer bias could be attributed to the dominance of *V. cholerae* in cholera endemic region, as it is generally more abundant than *V. paracholerae* sp. nov. and therefore, more likely to be a DNA donor (Nasreen et al., 2020). Among the accessory genes transferred from *V. cholerae* to *V. paracholerae* sp. nov., there were proteins related to O antigen synthesis, Type six secretion system, iron regulation, chaperone and multi-drug resistance and putative metabolic functions. There are examples of a single gene or even a small set of nucleotides within a gene acquired via HGT impacting the ecology and pathogenicity of *Vibrio* species (Levade et al., 2017; Polz et al., 2013). Thus, the HGT events in *V. paracholerae* sp. nov. underscore the possibility for species co-existing with PG *V. cholerae* to acquire virulence and fitness-related genes to become pathogenic to human and/or novel ecological traits. Gene transfer events have led to the rise of virulent *V. cholerae* before, the most striking example

being the rise of *V. cholerae* O139. The latter emerged in Bangladesh and India in 1992 and has been hypothesized to have originated via genetic recombination of the O-antigen region from a serogroup O22 strain to a serogroup O1 El Tor strain (Faruque et al., 2003).

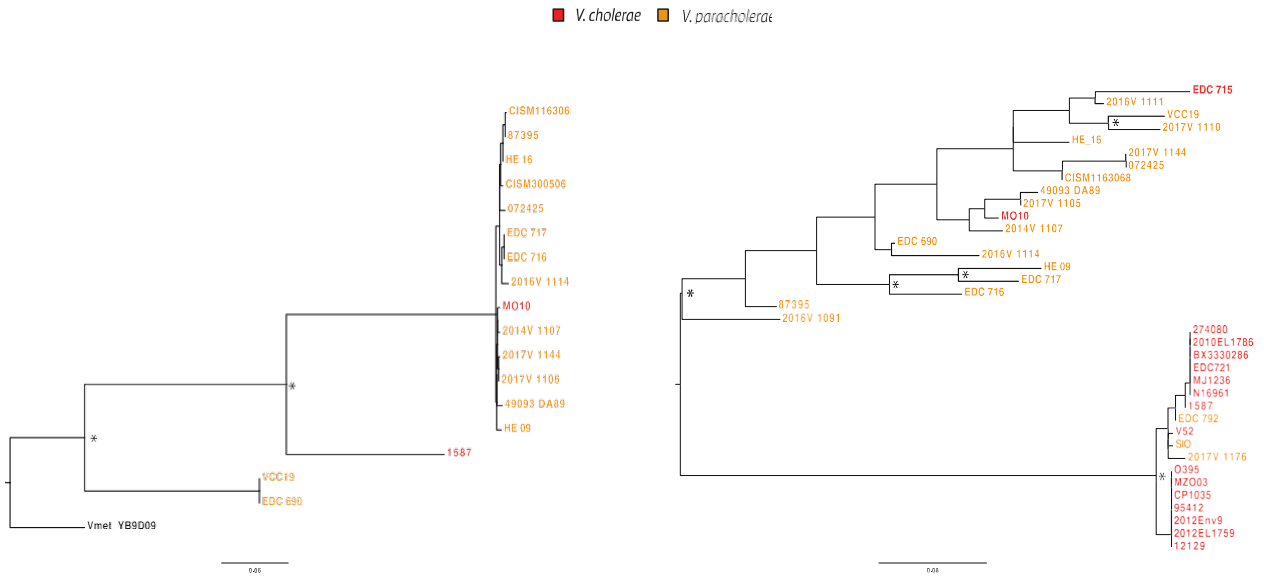


Fig. 3.6: Phylogenetic tree of O-antigen cluster genes found in *Vibrio paracholerae* and *Vibrio cholerae*. Maximum likelihood trees were constructed using A) 705 bp nucleotide alignment of the gene encoding UDP-glucose 4-epimerase and B) 564 bp alignment of the gene encoding UDP-N-acetylgalactosaminyltransferase. Nodes with relevant Bootstrap support over 70 of 100 replicates are indicated with *****. The scale bar represents nucleotide substitutions per site.

After its emergence, *V. cholerae* O139 remained an important cause of widespread cholera epidemics in that region until 2004, along with *V. cholerae* O1 El Tor (Faruque et al., 2003). Interestingly, among the accessory genes inferred as subject to HGT from VP to VC, genes encoding UDP-glucose 4-epimerase (EC 5.1.3.2) and UDP-N-acetylgalactosaminyltransferase appear to be a transfer in an ancestor of O139 strain MO10 from the *V. paracholerae* sp. nov. clade (Fig. 6). Both of these genes are involved in O-antigen biosynthesis and could be of significance in the emergence and evolution of *V. cholerae* O139 hypothesized as the source of the O139 regions not found is *V. cholerae* O1 El Tor. However,

as a human pathogen and pandemic agent. Even though it will require further investigation to find out how and to what extent *V. paracholerae* sp. nov. as a species contributed to the emergence and evolution of *V. cholerae* O139, these transfer events could be considered as examples of how interaction of this close relative with *V. cholerae* could impact the epidemiology of cholera. O antigen region is a highly variable region in terms of gene contents. Previously, O22 strains have been there are parts of the O antigen region in *V. cholerae* reference O139 strain MO10 (designated as region IV by Yamasaki *et al*.), for which the origin could not be attributed to either O22 or *V. cholerae* O1 (Yamasaki et al., 1999). Recently, closely related homologs of genes in region IV of *V. cholerae* O139 Strains were found in *V. fujianensis* (Huang et al., 2020). This indicates that the origin of MO10' O antigen region is most likely a result of multiple recombination events from various origins, forming the mosaic O-antigen encoding region in *V. cholerae* O139. *V. paracholerae* sp. nov. strains contain at least nine genes found in the *V. cholerae* O139 reference strain MO10 which are not found in *V. cholerae* O1 (Fig. S6). These observations suggest *V. paracholerae* sp. nov. could have been an important player of the gene transfer dynamics that gave rise to the second pathogenic serogroup of *V. cholerae*.

3.5. Description of the novel species *Vibrio paracholerae* sp. nov.

Vibrio paracholerae (pa.ra.chol.er.ae. Gr. prep. para alongside; L. gen. f. n. cholerae bilious disease; referring to the isolation of the type strain alongside *Vibrio cholerae*, the causative agent of cholera. A Gram stain negative, oxidase positive, curved rod-shaped bacterium, roughly 1.25-2 mm in length and 0.4 mm in width. Exhibits motility by means of a single polar flagellum. Growth is observed at 30 °C with salt concentrations in the range of 0–6.0 % NaCl; no growth occurs in the presence of 10 % NaCl. The ability to utilize α-cyclodextrin and pectin as well as the lack of ability to utilize mono methyl succinate differentiates the species from its closest relative *V. cholerae* (99% identity of 16S *rRNA*

gene). Forms *V. cholerae*-like yellow circular colonies on TCBS agar and circular colonies of creamy-white colour on TSB agar. Positive for carbon utilization from D-glucose, D-fructose, sucrose, maltose, D- galactose, maltotriose, D-mannitol, D-trehalose, D-glucose-6-phosphate, N-acetyl-D-glucosamine, glycerol, succinic acid, L-Lactic Acid, L-glutamic acid, fumaric acid, acetic acid, L-proline, D-alanine, L-asparagine, 2-deoxy adenosine, adenosine, inosine and L-serine, dextrin, gelatin, glycogen, D-glucosamine and D-lactic acid methyl ester. Variation of response observed between strains in utilization of D-mannose, D-cellobiose, D-fructose-6-phosphate, D-psicose, α-D-lactose, L-aspartic acid, D- glucuronic Acid, D-gluconic acid, D,L-α-glycerol phosphate, D,L-malic acid, D-ribose, tween 20, thymidine, α-keto-glutaric acid, tween 40, tween 80, α-keto-butyric acid, uridine, L-glutamine, α-hydroxy-glutaric acid-γ-lactone, ß-methyl-D-glucoside, citric acid, L-threonine, L-alanine, L-alanyl-glycine, glycine , histidine, methyl pyruvate, D-malic acid, glycyl-L-proline and pyruvic acid. The proposed type strain EDC-792 was isolated from environmental water in Dhaka, Bangladesh in 2016; the strain displays all the properties given above for the species. The proposed type strain (EDC-792) shows 95.6% ANI and 69% dDDH, respectively to the type strain of the closest sister species *Vibrio cholerae* (N16961). The oldest isolate of the species NCTC30 described by Dorman *et al*. (Dorman et al., 2019) is likely to have some phenotype changing mutations that urged for a new type strain with typical properties present in all the known strains of the species.

3.6. Conclusions

Culture-independent analysis below the species level in inland cholera endemic and coastal non-endemic locations in distinct geographic settings identified differences in the population structures present in these environments. It revealed that human influences are likely to be a major factor shaping communities of that species in cholera endemic areas. In urban tropical Dhaka, found in inland

Bangladesh, PG *V. cholerae* was abundant and continuously present, but accompanied by members of a related but phylogenetically distinct clade, which could represent a novel species. The abundance of this putative species, '*Vibrio paracholerae* sp. nov.', in Dhaka and its absence from Oyster Pond on the USA east coast, indicates that it is not a ubiquitous member of aquatic communities. In addition to those identified from the COVIS program in the USA and from Mozambique and Thailand, a number of strains of Vibrio spp. have been very recently isolated from clinical cases in China and Korea which would belong to this species according to the genome sequence similarities they share with strains analyzed here (**Table S4**). An indirect association of their abundance with human population density indicates that they could be adapted to the human gut in cholera endemic areas (Mohammad Tarequl Islam, Tania Nasreen, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C. Kirchberger, Marzia Sultana, Rebecca J. Case, n.d.). They could therefore occasionally become pathogenic by acquiring pathogenicity gene clusters or cause opportunistic infections in vulnerable individuals. Its history, biology, genetic traits and coexistence with a pathogenic sister species makes it a risk as an emerging human pathogen. Its potential contribution to the evolution of new pathogenic variants of *V. cholerae* (such as PG lineage O139) and likely influence on their population structure highlights the importance of studying this novel species in the context of a globally distributed infectious disease.

3.7. Reference

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., & Colwell, R. R. (2006). Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl Environ Microbiol*, *72*(6), 4096–4104. https://doi.org/10.1128/AEM.00066-06

Ali, M., Lopez, A. L., Ae You, Y., Eun Kim, Y., Sah, B., Maskery, B., & Clemens, J. (2012). The global burden of cholera. *Bulletin of the World Health Organization*. https://doi.org/10.2471/blt.11.093427

Bina, X. R., Provenzano, D., Nguyen, N., & Bina, J. E. (2008). *Vibrio cholerae* RND family effluxsystems are required for antimicrobial resistance, optimal virulence factor production, and colonization of the infant mouse small intestine. *Infection and Immunity*. https://doi.org/10.1128/IAI.01620-07

Bishop-Lilly, K. A., Johnson, S. L., Verratti, K., Luu, T., Khiani, A., Awosika, J., Mokashi, V. P., Chain, P. S., & Sozhamannan, S. (2014). Genome sequencing of 15 clinical Vibrio isolates, including 13 non-o1/non-o139 serogroup strains. *Genome Announc*, *2*(5). https://doi.org/10.1128/genomeA.00893-14

Bochner, B. R., Gadzinski, P., & Panomitros, E. (2001). Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Research*. https://doi.org/10.1101/gr.186501

Boyd, E. F., Carpenter, M. R., Chowdhury, N., Cohen, A. L., Haines-Menges, B. L., Kalburge, S. S., Kingston, J. J., Lubin, J. B., Ongagna-Yhombi, S. Y., & Whitaker, W. B. (2015). Post-Genomic Analysis of Members of the Family Vibrionaceae. *Microbiology Spectrum*, *3*(5), 10.1128/microbiolspec.VE-0009–2014. https://doi.org/10.1128/microbiolspec.VE-0009-2014

Brayton, P. R., Bode, R. B., Colwell, R. R., MacDonell, M. T., Hall, H. L., Grimes, D. J., West, P. A., & Bryant, T. N. (1986). Vibrio cincinnatiensis sp. nov., a new human pathogen. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.23.1.104-108.1986

Burrus, V., Quezada-Calvillo, R., Marrero, J., & Waldor, M. K. (2006). SXT-Related Integrating Conjugative Element in New World *Vibrio cholerae*. *Applied and Environmental Microbiology*, *72*(4), 3054–3057. https://doi.org/10.1128/AEM.72.4.3054-3057.2006

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*. https://doi.org/10.1038/nmeth.3869

Chaudhari, N. M., Gupta, V. K., & Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Scientific Reports*. https://doi.org/10.1038/srep24373

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., & Colwell, R. R. (2016). Phylogenetic diversity of *Vibrio cholerae* associated with endemic cholera in Mexico from 1991 to 2008. *MBio*. https://doi.org/10.1128/mBio.02160-15

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., Lee, J. H., Brettin, T. S., Bruce, D. C., Challacombe, J. F., Detter, J. C., Han, C. S., Munk, A. C., Chertkov, O., Meincke, L., Saunders, E., … 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*(36), 15442–15447. https://doi.org/10.1073/pnas.0907787106

Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., Brover, S., Schoch, C. L.,

Kimchi, A., & DiCuccio, M. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol*, *68*(7), 2386–2392. https://doi.org/10.1099/ijsem.0.002809

Davis, B. R., Fanning, G. R., Madden, J. M., Steigerwalt, A. G., Bradford, H. B., Smith, H. L., & Brenner, D. J. (1981). Characterization of biochemically atypical *Vibrio cholerae* strains and designation of a new pathogenic species, *Vibrio mimicus*. *Journal of Clinical Microbiology*.https://doi.org/10.1128/jcm.14.6.631-639.1981

DAVIS, G. H., & PARK, R. W. (1962). A taxonomic study of certain bacteria currently classified as Vibrio species. *Journal of General Microbiology*. https://doi.org/10.1099/00221287-27-1-101

Dorman, M. J., Kane, L., Domman, D., Turnbull, J. D., Cormie, C., Fazal, M. A., Goulding, D. A., Russell, J. E., Alexander, S., & Thomson, N. R. (2019). The history, genome and biology of NCTC 30: A non-pandemic *Vibrio cholerae* isolate from World War One. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2018.2025

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., & Mekalanos, J. J. (2005). Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0409918102

Faruque, S. M., Sack, D. A., Sack, R. B., Colwell, R. R., Takeda, Y., & Nair, G. B. (2003). Emergence and evolution of *Vibrio cholerae* O139. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0337468100

Fiedler, G., Pajatsch, M., & Böck, A. (1996). Genetics of a novel starch utilisation pathway present in Klebsiella oxytoca. *Journal of Molecular Biology*. https://doi.org/10.1006/jmbi.1996.0085

Garrine, M., Mandomando, I., Vubil, D., Nhampossa, T., Acacio, S., Li, S., Paulson, J. N., Almeida, M., Domman, D., Thomson, N. R., Alonso, P., & Stine, O. C. (2017). Minimal genetic change in *Vibrio cholerae* in Mozambique over time: Multilocus variable number tandem repeat analysis and whole genome sequencing. *PLoS Negl Trop Dis*, *11*(6), e0005671. https://doi.org/10.1371/journal.pntd.0005671

Haley, B. J., Grim, C. J., Hasan, N. A., Choi, S. Y., Chun, J., Brettin, T. S., Bruce, D. C., Challacombe, J. F., Detter, J. C., Han, C. S., Huq, A., & Colwell, R. R. (2010). Comparative genomic analysis reveals evidence of two novel Vibrio species closely related to V. cholerae. *BMC Microbiology*. https://doi.org/10.1186/1471-2180-10-154

Hasan, N. A., Grim, C. J., Haley, B. J., Chun, J., Alam, M., Taviani, E., Hoq, M., Munk, A. C., Saunders, E., Brettin, T. S., Bruce, D. C., Challacombe, J. F., Detter, J. C., Han, C. S., Xie, G., Nair, G. B., Huq, A., & Colwell, R. R. (2010). Comparative genomics of clinical and environmental *Vibrio mimicus*. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1013825107

Huang, Z., Yu, K., Fang, Y., Dai, H., Cai, H., Li, Z., Kan, B., Wei, Q., & Wang, D. (2020). Comparative Genomics and Transcriptomics Analyses Reveal a Unique Environmental Adaptability of Vibrio fujianensis. *Microorganisms*, *8*(4), 555. https://doi.org/10.3390/microorganisms8040555

Hussain, N. A. S., Kirchberger, P. C., Case, R. J., & Boucher, Y. F. (2021). Modular Molecular Weaponry Plays a Key Role in Competition Within an Environmental *Vibrio cholerae*

Population. *Frontiers in Microbiology*, *12*, 671092. https://doi.org/10.3389/fmicb.2021.671092

Islam, M. T., Liang, K., Im, M. S., Winkjer, J., Busby, S., Tarr, C. L., & Boucher, Y. (2018). Draft Genome Sequences of Nine Vibrio sp. Isolates from across the United States Closely Related to *Vibrio cholerae*. *Microbiol Resour Announc*, *7*(21). https://doi.org/10.1128/MRA.00965-18

Jin, Y., Zhou, J., Zhou, J., Hu, M., Zhang, Q., Kong, N., Ren, H., Liang, L., & Yue, J. (2020). Genome-based classification of Burkholderia cepacia complex provides new insight into its taxonomic status. *Biology Direct*, *15*(1), 6. https://doi.org/10.1186/s13062-020-0258-5

Kaper, J. B., Morris Jr., J. G., & Levine, M. M. (1995). Cholera. *Clin Microbiol Rev*, *8*(1), 48–86.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bts199

Kirchberger, P C, Orata, F. D., Barlow, E. J., Kauffman, K. M., Case, R. J., Polz, M. F., & Boucher, Y. (2016). A Small Number of Phylogenetically Distinct Clonal Complexes Dominate a Coastal *Vibrio cholerae* Population. *Appl Environ Microbiol*, *82*(18), 5576–5586. https://doi.org/10.1128/AEM.01177-16

Kirchberger, P C, Orata, F. D., Nasreen, T., Kauffman, K. M., Tarr, C. L., Case, R. J., Polz, M. F., & Boucher, Y. F. (2020). Culture-independent tracking of *Vibrio cholerae* lineages reveals complexspatiotemporal dynamics in a natural population. *Environ Microbiol*. https://doi.org/10.1111/1462-2920.14921

Kirchberger, P C, Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., Tarr, C. L., &

Boucher, Y. (2014). Vibrio metoecus sp. nov., a close relative of *Vibrio cholerae* isolated from coastal brackish ponds and clinical specimens. *Int J Syst Evol Microbiol*, *64*(Pt 9), 3208–3214. https://doi.org/10.1099/ijs.0.060145-0

Kirchberger, Paul C., Orata, F. D., Nasreen, T., Kauffman, K. M., Tarr, C. L., Case, R. J., Polz, M. F., & Boucher, Y. F. (2020). Culture-independent tracking of *Vibrio cholerae* lineages reveals complex spatiotemporal dynamics in a natural population. *Environmental Microbiology*. https://doi.org/10.1111/1462-2920.14921

Kirchberger, Paul C., Unterweger, D., Provenzano, D., Pukatzki, S., & Boucher, Y. (2017). Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. *Scientific Reports*. https://doi.org/10.1038/srep45133

Kirchberger, Paul C, Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., Tarr, C. L., & Boucher, Y. (2014). Vibrio metoecus sp. nov., a close relative of *Vibrio cholerae* isolated from coastal brackish ponds and clinical specimens. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt 9), 3208–3214. https://doi.org/10.1099/ijs.0.060145-0

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01043-13

Labbate, M., Orata, F. D., Petty, N. K., Jayatilleke, N. D., King, W. L., Kirchberger, P. C., Allen, C., Mann, G., Mutreja, A., Thomson, N. R., Boucher, Y., & Charles, I. G. (2016). A genomic island in *Vibrio cholerae* with VPI-1 site-specific recombination characteristics contains CRISPR-Cas and type VI secretion modules. *Scientific Reports*, *6*(1), 36891. https://doi.org/10.1038/srep36891

Lee, K. M., Park, Y., Bari, W., Yoon, M. Y., Go, J., Kim, S. C., Lee, H. Il, & Yoon, S. S. (2012). Activation of cholera toxin production by anaerobic respiration of trimethylamine N-oxide in *Vibrio cholerae*. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M112.394932

Levade, I., Terrat, Y., Leducq, J.-B., Weil, A. A., Mayo-Smith, L. M., Chowdhury, F., Khan, A. I., Boncy, J., Buteau, J., Ivers, L. C., Ryan, E. T., Charles, R. C., Calderwood, S. B., Qadri, F., Harris, J. B., LaRocque, R. C., & Shapiro, B. J. (2017). *Vibrio cholerae* genomic diversity within and between patients. *Microbial Genomics*, *3*(12). https://doi.org/10.1099/mgen.0.000142

Lin, W., Fullner, K. J., Clayton, R., Sexton, J. A., Rogers, M. B., Calia, K. E., Calderwood, S. B., Fraser, C., & Mekalanos, J. J. (1999). Identification of a *Vibrio cholerae* RTX toxin gene clusterthat is tightly linked to the cholera toxin prophage. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.96.3.1071

Lugo, M. R., & Merrill, A. R. (2015). The father, son and cholix toxin: The third member of the DT group mono-ADP-ribosyltransferase toxin family. In *Toxins*. https://doi.org/10.3390/toxins7082757

Meier-Kolthoff, J. P., Klenk, H.-P., & Göker, M. (2014). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt_2), 352–356. https://doi.org/https://doi.org/10.1099/ijs.0.056994-0

Mohammad Tarequl Islam, Tania Nasreen, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C. Kirchberger, Marzia Sultana, Rebecca J. Case, M. A. and Y. F. B. (n.d.). Dynamic Subspecies Population Structure of *Vibrio cholerae* in Dhaka, Bangladesh. *Preprint at Https://Www.Researchsquare.Com/Article/Rs-364004/V1*.

Mutreja, A., Kim, D. W., Thomson, N., Connor, T. R., Hee, J., Lebens, M., Niyogi, S. K., Kim, E. J., Ramamurthy, T., Chun, J., Parkhill, J., & Dougan, G. (2013). Evidence for multiple waves of global transmission within the seventh cholera pandemic. *Nature*. https://doi.org/10.1038/nature10392.Evidence

Nasreen, T., Hussain, N. A. S., Islam, M. T., Orata, F. D., Kirchberger, P. C., Case, R. J., Alam, M., Yanow, S. K., & Boucher, Y. F. (2020). Simultaneous quantification of vibrio metoecus and *Vibrio cholerae* with its o1 serogroup and toxigenic subpopulations in environmental reservoirs.*Pathogens*. https://doi.org/10.3390/pathogens9121053

Neogi, S. B., Chowdhury, N., Awasthi, S. P., Asakura, M., Okuno, K., Mahmud, Z. H., Islam, M. S., Hinenoya, A., Nair, G. B., & Yamasaki, S. (2019). Novel cholera toxin variant and ToxT regulon in environmental *Vibrio mimicus* isolates: Potential resources for the evolution of *Vibrio cholerae* hybrid strains. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01977-18

Nora Hussein, Paul Kirchberger, Y. F. B. (n.d.). Type six secretion mediated competetion in *Vibrio cholerae*. *Frontiers in Microbiology*.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P. M., Stevens, M. H. H., & & Wagner, H. (2008). The vegan package. *Community Ecology Package*.

Orata, F. D., Kirchberger, P. C., Méheust, R., Jed Barlow, E., Tarr, C. L., & Boucher, Y. (2015). The dynamics of genetic interactions between Vibrio metoecus and *Vibrio cholerae*, two close relatives co-occurring in the environment. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evv193

Orata, F. D., Xu, Y., Gladney, L. M., Rishishwar, L., Case, R. J., Boucher, Y., Jordan, I. K., & Tarr, C.

L. (2016). Characterization of clinical and environmental isolates of Vibrio cidicii sp. nov., a close relative of Vibrio navarrensis. *International Journal of Systematic and Evolutionary Microbiology*, *66*(10), 4148–4155. https://doi.org/10.1099/ijsem.0.001327

Payne, S. M., Mey, A. R., & Wyckoff, E. E. (2016). Vibrio Iron Transport: Evolutionary Adaptation to Life in Multiple Environments. *Microbiology and Molecular Biology Reviews*. https://doi.org/10.1128/mmbr.00046-15

Polz, M. F., Alm, E. J., & Hanage, W. P. (2013). Horizontal gene transfer and the evolution of bacterial and archaeal population structure. In *Trends in Genetics*. https://doi.org/10.1016/j.tig.2012.12.006

R Development Core Team, R. (2011). R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*. https://doi.org/10.1007/978-3-540-74686-7

Rafique, R., Rashid, M. U., Monira, S., Rahman, Z., Mahmud, M. T., Mustafiz, M., Saif-Ur-Rahman, K. M., Johura, F. T., Islam, S., Parvin, T., Bhuyian, M. S., Sharif, M. B., Rahman, S. R., Sack, D. A., Sack, R. B., George, C. M., & Alam, M. (2016). Transmission of Infectious *Vibrio cholerae* through Drinking Water among the Household Contacts of Cholera Patients (CHoBI7 Trial). *Front Microbiol*, *7*, 1635. https://doi.org/10.3389/fmicb.2016.01635

Richter, M., & Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A*, *106*(45), 19126–19131. https://doi.org/10.1073/pnas.0906412106

Routh, M. D., Zalucki, Y., Su, C. C., Long, F., Zhang, Q., Shafer, W. M., & Yu, E. W. (2010). Efflux pumps of the Resistance-nodulation-division family: A perspective of their structure, function,

and regulation in gram-negative bacteria. *Advances in Enzymology and Related Areas of Molecular Biology*. https://doi.org/10.1002/9780470920541.ch3

Schliep, K., Lopez, P., Lapointe, F. J., & Bapteste, É. (2011). Harvesting evolutionary signals in a forest of prokaryotic gene trees. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msq323

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01541-09

Shapiro, B. J., Levade, I., Kovacikova, G., Taylor, R. K., & Almagro-Moreno, S. (2016). Origins of pandemic *Vibrio cholerae* from environmental gene pools. *Nat Microbiol*, *2*, 16240. https://doi.org/10.1038/nmicrobiol.2016.240

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btu033

Sun, T., & Altenbuchner, J. (2010). Characterization of a mannose utilization system in bacillus subtilis. *Journal of Bacteriology*. https://doi.org/10.1128/JB.01673-09

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., Munn, C. B., & Swings, J. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Applied and Environmental Microbiology*, *71*(9), 5107–5115. https://doi.org/10.1128/AEM.71.9.5107-5115.2005

Trastoy, R., Manso, T., Fernández-García, L., Blasco, L., Ambroa, A., Pérez Del Molino, M. L., Bou, G., García-Contreras, R., Wood, T. K., & Tomás, M. (2018). Mechanisms of bacterial tolerance and persistence in the gastrointestinal and respiratory environments. In *Clinical Microbiology Reviews*. https://doi.org/10.1128/CMR.00023-18

Treangen, T. J., Ondov, B. D., Koren, S., & Phillippy, A. M. (2014). The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology*. https://doi.org/10.1186/s13059-014-0524-x

Unterweger, D., Miyata, S. T., Bachmann, V., Brooks, T. M., Mullins, T., Kostiuk, B., Provenzano, D., & Pukatzki, S. (2014). The *Vibrio cholerae* type VI secretion system employs diverse effector modules for intraspecific competition. *Nat Commun*, *5*, 3549. https://doi.org/10.1038/ncomms4549

Westra, E. R., van Houte, S., Gandon, S., & Whitaker, R. (2019). The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* (Vol. 374, Issue 1772, p. 20190101). https://doi.org/10.1098/rstb.2019.0101

Wright, J. J., Lee, S., Zaikova, E., Walsh, D. A., & Hallam, S. J. (2009). DNA extraction from 0.22 μM Sterivex filters and cesium chloride density gradient centrifugation. *Journal of Visualized Experiments*. https://doi.org/10.3791/1352

Yamasaki, S., Shimizu, T., Hoshino, K., Ho, S.-T., Shimada, T., Nair, G. B., & Takeda, Y. (1999). The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are closely related to those of *Vibrio cholerae* O22. *Gene*, *237*(2), 321–332. https://doi.org/https://doi.org/10.1016/S0378-1119(99)00344-3

Zheng, B., Jiang, X., Cheng, H., Guo, L., Zhang, J., Xu, H., Yu, X., Huang, C., Ji, J., Ying, C., Feng, Y., Xiao, Y., & Li, L. (2017).

Genome characterization of two bile-isolated Vibrio fluvialis strains: An insight into pathogenicity and bile salt adaption. *Scientific Reports*. https://doi.org/10.1038/s41598-017- 12304-8

CHAPTER 4: Dynamic subspecies population structure of *Vibrio cholerae* in Dhaka, Bangladesh

A version of this chapter has been accepted for publication as:

"TaniaNasreen*, **Mohammad Tarequl Islam***, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C. Kirchberger, Eric Hill, Marzia Sultana, Rebecca J. Case, Munirul Alam and Yann F. Boucher. Dynamic subspecies population structure of *Vibrio cholerae* in Dhaka, Bangladesh. (2015). Accepted in *Microbial Ecology*. * Joint first authors.

MTI, YB, TN, and MA designed the project. TN and MTI wrote the manuscript. MTI, FTJ, MS and MA performed sample collection and sample processing during field trips in Dhaka, Bangladesh. TN performed the qPCR and MTI did the amplicon sequencing. MTI and TN performed the data analysis. KYHL helped in bioinformatics analysis. MTI, YB, FTJ, MS, RJC and MA reviewed the manuscript. MA and YB supervised the project.

CHAPTER 4

4.1. Abstract

Cholera has been endemic to the Ganges delta for centuries. Although the causative agent, *Vibrio cholerae,* is autochthonous to coastal and brackish water, cholera occurs continually in Dhaka, the inland capital city of Bangladesh which is surrounded by fresh water. Despite the persistence of this problem, little is known about the environmental abundance and distribution of lineages of *V. cholerae*, the most important being the pandemic generating lineage (PG) consisting mostly of serogroup O1 strains. To understand spatial and temporal dynamics of PG and other lineages belonging to the *V. cholerae* species in surface water in and around Dhaka city, we used qPCR and high throughput amplicon sequencing. Seven different freshwater sites across Dhaka were investigated for six consecutive months and physiochemical parameters were measured *in situ*. Total abundance of *V. cholerae* was found to be relatively stable throughout the six months sampling period, with 2×10^5 to 4×10^5 genome copies/L at six sites and around 5×10^5 genome copies/L at the site located in the most densely populated part of Dhaka city. PG O1 *V. cholerae* was present in high abundance during the entire sampling period and composed between 24-92% of the total *V. cholerae* population, only showing occasional but sudden reductions in abundance. In instances where PG O1 lost its dominance, other lineages underwent a rapid expansion while the size of the total *V. cholerae* population remained almost unchanged. Intraspecies richness of *V. cholerae* was positively correlated to salinity, conductivity and total dissolved solids (TDS), while it was negatively correlated to dissolved oxygen (DO) concentration in water. Interestingly, negative correlation was observed specifically between PG O1 and salinity, even though the

changes in this variable were minor (0-0.8 ppt). Observations in this study suggest that at the subspecies level, population composition of naturally occurring *V. cholerae* can be influenced by fluctuations in environmental factors, which can lead to altered competition dynamics among the lineages.

4.2. Introduction

Vibrio cholerae is a normal inhabitant of aquatic environments such as rivers, estuaries and coastal waters and has been detected in diverse geographic locations worldwide(Colwell, 2004). Toxigenic strains of *V. cholerae* are capable of causing cholera, an acute life-threatening diarrheal disease (Lipp et al., 2002; Vezzulli et al., 2010), which is a major public health concern because of its high morbidity and mortality with an estimation of 1.3 to 4 million cases and 21,000 to 143,000 deaths worldwide each year (Ali et al., 2012). Seven cholera pandemics have struck human civilization so far; the first pandemic of cholera started in 1817 and was followed by six others in the next two hundred years, leaving a devastating human death toll (Faruque et al., 1998). In the environment, *V. cholerae* is a diverse species, with more than 200 serogroups being identified based on their surface polysaccharide O antigen (Chatterjee and Chaudhuri, 2003). However, only O1 serogroup dominates the pandemic generating (PG) lineage (Chun et al., 2009b; Islam et al., 2017b), which has been responsible for all seven cholera pandemics, and other serogroups mostly represent *V. cholerae* environmental strains which are generally non-toxigenic (Faruque et al., 1998; Chun et al., 2009a). Serogroup O1 strains of the PG lineage are further classified into two biotypes, the Classical biotype that was shown to cause the fifth and sixth pandemic and believed to be associated with the earlier pandemics, and the El Tor biotype which

is the causative agent for the seventh pandemic of cholera (Faruque et al., 1998; Islam et al., 2017b).

Cholera remains as an emerging and reemerging disease today (Faruque et al., 1998). The seventh pandemic of cholera started in 1961 and is still ongoing, known as the world's longest and most persistent pandemic (Ali et al., 2012; Islam et al., 2017b). Despite being a global disease, the world's worst cholera epidemics can be traced back to Ganges delta, hence that area has been thought to serve as a reservoir for the disease (Mutreja et al., 2013; Boucher et al., 2015b). High population density and improper sewage disposal along with uncontrolled industrialization constantly pollute the water sources in this region. Dhaka, the capital city of Bangladesh, is considered as a hyper-endemic region for cholera. The disease continually exists at a low incidence but exhibits biannual seasonal outbreaks in this area (Longini et al., 2002). Consequently, Dhaka has been a center for clinical research on cholera for many years. Recent advancement on cholera surveillance intended to increase prevention, preparedness, intervention and awareness of the disease has been chiefly established based on analysis of clinical cases and clinical samples (Deen et al., 2020). However, little attention has been paid to dynamics of the natural population of *V. cholerae*, which could play an important role in the epidemiological outcome of cholera. As *V. cholerae* is an autochthonous member of the aquatic environments, understanding its population structure and ecology requires a knowledge of the temporal and spatial variations in abundance of various genotypes in its natural habitat.

Most studies looking at the environmental aspects of the disease have been dependent on culture-based isolation of toxigenic *V. cholerae* (Jutla et al., 2013). Culture-based studies underestimate *V. cholerae* abundance and diversity, because of important limitations including enrichment biases and the viable but non-culturable (VBNC) state (Colwell, 2000; Alam et al., 2007, 2010). Real-

time qPCR analyses of environmental DNA targeting *V. cholerae* or O1 serogroup-specific genes are quantitative but do not provide information on the subspecies composition of the populations analyzed (Rashid et al., 2017). Additionally, culture-independent studies, including 16S rRNA sequencing, can be helpful for genus identification but do not usually provide resolution at the species or subspecies level (Chun et al., 1999). A fluorescent antibody staining method has been used to enumerate viable *V. cholerae* cells in water samples (Chowdhury et al., 1995), but only targets the O1 or O139 serogroups and cannot distinguish between O1 strains belonging to the pandemic generating (PG) lineage and strains from other lineages bearing that antigen.

To overcome these limitations, we have applied high-throughput sequencing of a species-specific, highly variable region of a gene, encoding the vibriobactin utilization protein, *viuB* (Kirchberger et al., 2020). This protein releases iron captured by the siderophore vibriobactin inside the cell (Wyckoff et al., 2007), which is a housekeeping function for *V. cholerae*. In combination with using qPCR(Nasreen et al., 2020), amplification and sequencing of the partial *viuB* gene allowed for the elucidation of the spatio-temporal abundance of the PG and other lineages belonging to the *V. cholerae* species in Dhaka's water reservoirs during six consecutive months from October 2015 to March 2016, roughly encompassing after-monsoon, fall, winter and spring seasons in Bangladesh. The study reveals the continual presence of the PG lineage and occasional reduction of its abundance in conjunction with increases in other lineages, providing insights on the influence of environmental factors in subspecies level population dynamics of *V. cholerae*.

4.3. Materials and Methods

4.3.1. Study site

Surface water samples were collected from seven different locations (site 1 to site 7) in Dhaka (23.8103º N, 90.4125º E), Bangladesh (Fig. 4.1), biweekly for six consecutive months from October 2015 to March 2016. Dhaka is the capital city of Bangladesh surrounded by a river system mostly composed of four rivers: Turag, Buriganga, Shitalakshya and Balu. A population of >21 million was recorded for this area in 2020, with a density of 23,234 people/ km^2 within a total area of 300 km^2 (Fig. 4.1) (World population review, 2020)). The physical distance between site 1 and site 7 is shorter (9.9 km) than the distance between site 1 and site 7 to the other five sites (approximately 21 to 25 km). The climate of Dhaka is categorized as tropical wet and dry with a distinct monsoon season. The average water temperature recorded is 26.1 ºC (19.1 ºC in Jan and 29.1 ºC in June). The visual inspection of the study areas indicated that local markets surrounded these sites, and human intrusion such as bathing, swimming, washing household utensils dishes as well as bathing domestic animals were frequent. Occasional direct defecation in the water at several study sites were also noticeable.

4.3.2 Sample collection and processing

Water samples (200 ml) were collected directly from the sources in sterile Nalgene bottles. Using 50 ml sterile polypropylene syringes 50 ml water sample was filtered through 0.22 μm Sterivex filters (Millipore). Total DNA extraction from the biomass on the filters was done through the following three consecutive steps: cell lysis and digestion, DNA extraction, and DNA concentrating and washing according to the protocol developed by Wright et al. (Wright et al., 2009). Briefly, the filters with trapped biomass are incubated with lysozyme (ThermoFisher) to lyse the cells and RNaseA (New England Biolabs) to remove RNA at 37°C for 1 hour. Afterwards, the filteres are treated with Proteinase K (New England Biolabs) and 20% SDS at 55°C for 2 hours.

Then the lysate were treated with equal volume of Phenol:Chloroform:Isoamyl Alcohol (IAA) to seprate the DNA layer from proteins and debris follwed by equal volume of Chloroform:IAA, to remove any remaining phenol from the DNA sample. Final DNA was eluted in TE buffer for downstream applications. To reduce impurities that can act as PCR inhibitors during amplification, all extracted DNA samples were further treated with One step PCR inhibitory removal kit (ZYMO Research) by following the user manual instructions with 90-180 µl of yield achieved from 100-200 µl of the eluted extracted DNA sample. Treated samples were kept at -20° C for further analysis.

4.3.3. Culture based identification

Bacterial isolates were isolated as described elsewhere (Alam et al., 2010). Briefly, water samples were enriched in APW (Difco) at 37°C for 6 to 8 h before plating. About 5 µl of enriched APW broth was streaked, using an inoculating loop, onto thiosulfate-citrate bile-salts-sucrose (TCBS) plates at 37°C for 18 to 24 h. Yellow colonies on TCBS agar were subcultured onto gelatin agar (Difco) plates. Gelatinase-positive colonies were subjected to PCR for amplification of the *V. cholerae* species-specific *ompW* gene and O1 serogroup specific *rfbO1* gene for the confirmation.

4.3.4. Physicochemical parameters

Surface water quality was measured in situ at the sampling sites. EXO2 multiparameter sonde (YSI, Xylam Brand, USA) allowed for simultaneous measurement of pH, dissolved oxygen (DO), conductivity, total dissolved solids (TDS), salinity and water temperature. Properly calibrated sensors attached to the instrument were placed in each site while sampling and data were recorded for analysis.

4.3.5. PCR amplification and amplicon sequencing

A touchdown PCR was performed to amplify a 293bp region of the *viuB* gene from DNA extracted from biomass. Master mix for PCR contained 5 µl of 5× Phire Buffer (ThermoFisher), 0.4 µl of 10 mM dNTP mix (ThermoFisher), 0.4 of µl Phire Hot Start II DNA Polymerase (ThermoFisher), 0.5 µl of Molecular Biology Grade Bovine Serum Albumin (20 mg/mL, New England Biolabs), 0.5 µl each of 10 pmol forward and reverse primers (for *viuB*: *viuB*2f 5'-CCGTTAGACAATACCGAGCAC-3' and *viuB*5r 5'-TTAGGATCGCGCACTAACCAC-3') (Kirchberger et al., 2020) and 2 µl of template DNA. The PCR reaction was performed as follows: initial denaturation at 98 °C for 4 min, followed by 10 cycles of denaturation at 98 °C for 10 sec, annealing at 60 °C for 6 sec (reduced by 1°C per cycle), and extension 72 °C for 1 sec; followed by 23 cycles of denaturation at 98 °C for 10 sec, annealing at 50 °C for 6 sec (reduced by 1°C per cycle), and extension at 72 °C for 1 sec; and a final extension at 72 °C for 1 min. Dual-indexed sequences using indices developed by Kozich et al. (Kozich et al., 2013) were used to prepare amplified *viuB* products for sequencing. Amplified *viuB* products (2 µl) were used as template for the tagging PCR reaction with the same reagents as above with a set of forward and reverse primers that contained appropriate Illumina-adapters, a sample-specific 8 nucleotide index sequence, a 10 nucleotide pad, 2 nucleotide linker, and the gene specific sequence described above, for a total of 70 and 65 bp (Kirchberger et al., 2020). This tagging PCR reaction was performed as follows: initial denaturation at 98 °C for 30 sec; followed by two cycles of denaturation at 98 °C for 10 sec, annealing at 55 °C for 6 sec, and extension at 72 °C for 1 sec; and final extension at 72 °C for 1 min. Use of gene-specific primers during amplification and subsequent tagging to create dual-indexed PCR products facilitated improved yield of amplicons and prevented biased amplification due to unexpected interaction of non-primer sequences with the template.

Additionally, eight tagging reactions were done for each sample to obtain an adequate concentration of amplicon DNA for further analysis. All eight reactions of the same sample were pooled together and run on a 2% agarose gel in 1× Tris-Acetate-EDTA buffer where two bands of very similar size, a smaller band (around 360 bp) representing only half-tagged PCR products, and a slightly bigger band (428 bp) of the fully tagged product were visualized. The 428bp bands were cut out of the gel. PCR products were purified using Wizard SV Gel and PCR Clean-Up System (Promega) according to the instructions by the manufacturer. The concentration of cleaned PCR products was then measured using a Qubit Fluorometer (ThermoFisher) with a Qubit dsDNA HS Assay Kit (ThermoFisher) and pooled together in equal concentrations (>10 ng/µl). The pooled samples were then concentrated using a Wizard SV Gel and PCR Clean-Up System (Promega) according to the instructions by the manufacturer. Quality control of the pooled and the concentrated sample was performed using an Agilent 2100 Bioanalyzer system. Sequencing was performed using a v3 (600 cycles) Illumina sequencing reagent kit and in Illumina MiSeq platform using the sequencing facility at the Molecular Biology Service Unit (MBSU) at University of Alberta, Canada

4.3.6. Sequence Analysis

Processing of amplicon sequence reads was performed following the procedure described by Kirchberger et al (Kirchberger et al., 2020). Briefly, de-multiplexed raw reads were processed in R (Ripley, 2001) using the DADA2 pipeline (Callahan et al., 2016). Forward and reverse reads were trimmed due to a drop-off in read quality in the first 10 bp as well as after 240 bp and 160 bp for forward and reverse reads, respectively. Assembled overlapping forward and reverse reads were therefore 272 bp in length, 11 bp shorter than the fully sequenced region. Reads with a maximum expected error rate >1% were also discarded based on DADA2 analysis. After this

procedure, 1072 unique sequences remained in the dataset. Chimera detection implemented in DADA2 was then performed on pooled samples, leaving a total of 460 unique sequences. Sequence reads were assigned to *viuB* alleles, 25 of which were composed of more than 1,000 reads (with an average of 100,000 reads per sample) and of with the 12 most abundant alleles representing >99% of all reads were considered for further analysis.

4.3.7. Real-time qPCR amplification

A real-time qPCR assay was performed to determine the abundance of total *V. cholerae* and toxigenic *V. cholerae* O1(Nasreen et al., 2020). For total *V. cholerae*, the *viuB* gene encoding vibriobactin utilization protein B was targeted using the probe, 5'-/56-FAM/TCA TTT GGC/ZEN/CAG AGC ATA AAC CGG T/3IABkFQ/-3',; forward and reverse primers, 5'-TCG GTA TTG TCT AAC GGT AT-3' and 5'-CGA TTC GTG AGG GTG ATA-3', respectively, for a 77-bp product. For *V. cholerae* O1 serogroup, *rfbO1* gene was targetted; Probe, 5'-

/5HEX/AGAAGTGTG/ZEN/

TGGGCCAGGTAAAGT/3IABkFQ/-3', forward 5'- GTAAAGCAGGATGGAAACATATTC-3' And reverse, 5'- TGGGCTTACAAACTCAAGTAAG-3' primers were used for a 113bp product(Nasreen et al., 2020). For the qPCR reaction, Dynamite qPCR Mastermix which is a proprietary mix, developed and distributed by the Molecular Biology Service Unit (MBSU) at University of Alberta, Canada was used. The volume of the PCR reaction was 10 µl containing 5 µl of 2× Dynamite qPCR master mix, 1 µl of each of 500 nM primer-250 nM probe mix, 1 µl of molecular grade water and 2 µl of DNA template. Real-time quantitative PCR was performed using the following conditions: initial primer activation at 95 °C for 2 min followed by 40 cycles of 95 °C for 15 s, 60 °C for 1 min in Illumina Eco Real-Time PCR system. The qPCR method has been

reported to have a analytical detection limit for the *viuB* and *rfbO1* genes as three gene copies per reaction with 1.5×10^6 copies/L of water as the lowest detectable number without filtration (Nasreen et al., 2020).

4.3.8. Statistical and multivariate analyses

Two-dimensional visualizations of the overall *V. cholerae* community structure was performed using NMDS (Non-metric multidimensional scaling) with Bray-Curtis distance at operational taxonomic unit (OTU) levels based on unique *viuB* alleles. 'bioenv' function in vegan package was used to find a non-parametric monotonic relationship between the dissimilarities in the samples matrix and for plotting the location of each site in a two dimensional space (Torondel et al., 2016). Abundance and environmental variables data were analyzed using R. Multiple subplots were generated for the environmental variables against *viuB* allele richness, and a correlation test was performed to determine any significant relationships between them. Hierarchical clustering analysis was done with the R function hclust to identify clusters of *viuB* alleles with the most similar distribution based on the abundance data.

4.3.9. Phylogenetic analysis and determination of genome similarity

For phylogenetic analysis, two representative strains (Table C.1) were chosen for each major *viuB* allele. *V. cholerae* strains isolated from Dhaka, genome of which has already been sequenced (Islam et al.) and reference genome sequences were obtained from NCBI database. Whole genome analysis was done using mugsy v1.2.3 with default parameters (Angiuoli and Salzberg, 2011). Mugsy outputs were analyzed using galaxy web server (Galaxy: a web-based genome analysis tool for experimentalists) (https://usegalaxy.org), and the phylogenetic tree was constructed using

Raxml v8.2.11 under the GTRGAMMA model with 100 bootstrap replicates (Stamatakis, 2014). Defaults were chosen for all other parameters. The phylogenetic tree was visualized using iTOL (Letunic and Bork, 2007). *in-silico* DNA-DNA Hybridization (dDDH), which is a proxy for traditional DDH value was used to calculate genomic similarities (Auch et al., 2010). All pairwise comparisons for dDDH were calculated using GGDC with default parameters (Meier-Kolthoff et al., 2013). Allelic differences were also used to determine genomic similarities. To determine allelic differences each genome was first annotated using RAST (Aziz et al., 2008). A set of 2443 genes common in most *V. cholerae* strains were then identified using Usearch (Edgar, 2013). Allele designations and identification were subsequently performed using automated scripts made available by BIGSdb (Jolley and Maiden, 2010). Finally, pairwise allelic differences for all isolates were calculated using an in-house script and only loci present in both isolates of a pair are considered. Sequences for all gene alleles are available on https://pubmlst.org/vcholerae under the cgMLST scheme.

4.4. Results and Discussion

4.4.1. The pandemic generating lineage is the most abundant *Vibrio cholerae* genotype in Dhaka's water system

Dhaka, one of the most densely populated cities in the world (>21 million residents as of June 2020), is located within the Ganges delta. It is an inland city, with a water system primarily consisting of freshwater rivers and canals. Cholera is endemic there and shows biannual peaks in reported cases during the spring and fall, i.e. before and after the monsoons (Alam et al., 2011). However, little is known about the abundance and distribution of *V. cholerae* in natural waterbodies and if it correlates with environmental and human factors. To track the abundance of

total *V. cholerae* and pandemic generating (PG) lineage of *V.* cholerae, we used culture-based detection as well as qPCR analysis. A 272 bp hypervariable stretch of the *viuB* marker gene, which is present in a single copy in *V. cholerae* (Kirchberger et al., 2020) was amplified and sequenced from fortnightly samples taken in seven different water reservoirs in and around Dhaka city from October 2015 to Mach 2016 (**Fig. 4.1**).

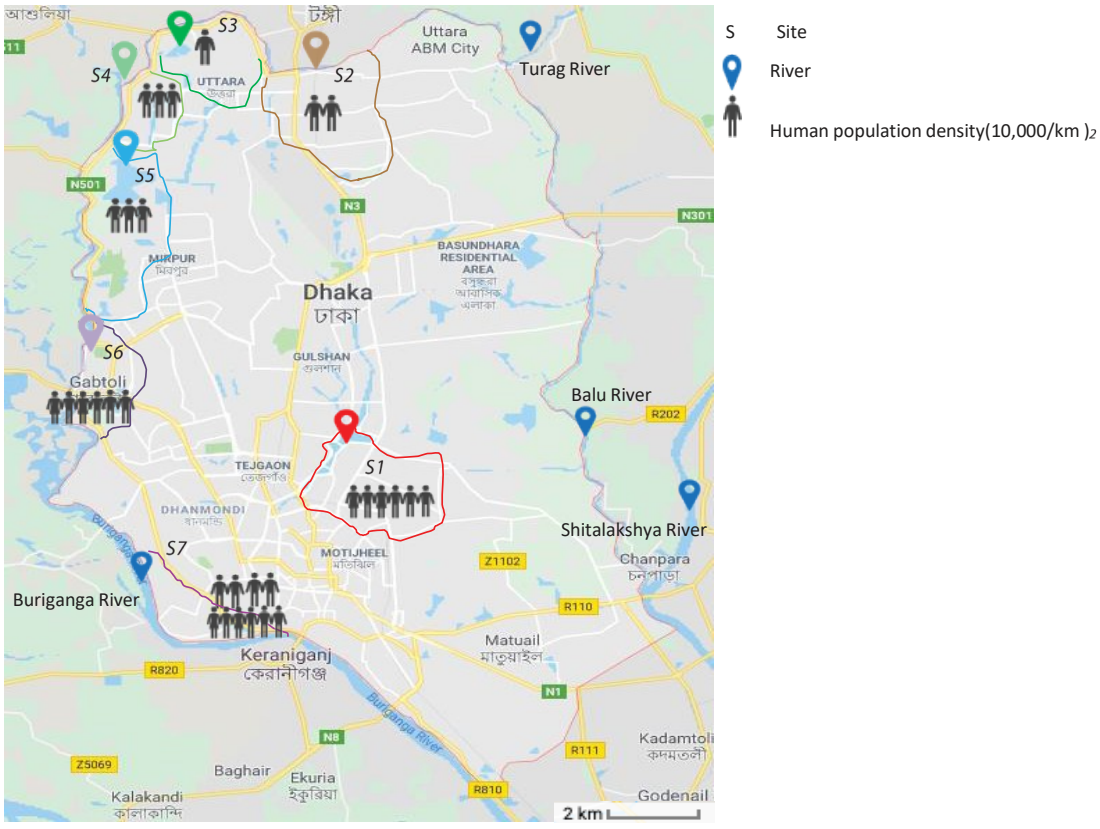


Fig. 4.1. Location of sampling sites in the study. Dhaka, the capital city of Bangladesh, is surrounded by the river system of four rivers including Turag river, Buriganga river, Shitalakshya river and Balu river, as indicated by blue pins on the map. Seven different sampling sites are indicated with 'S' (from S1 to S7) along with the approximate human population density corresponding to each site (marked by pins with distinct colors). Information on the humanpopulation density in this figure was adapted from Khatun et al., 2017 [54].

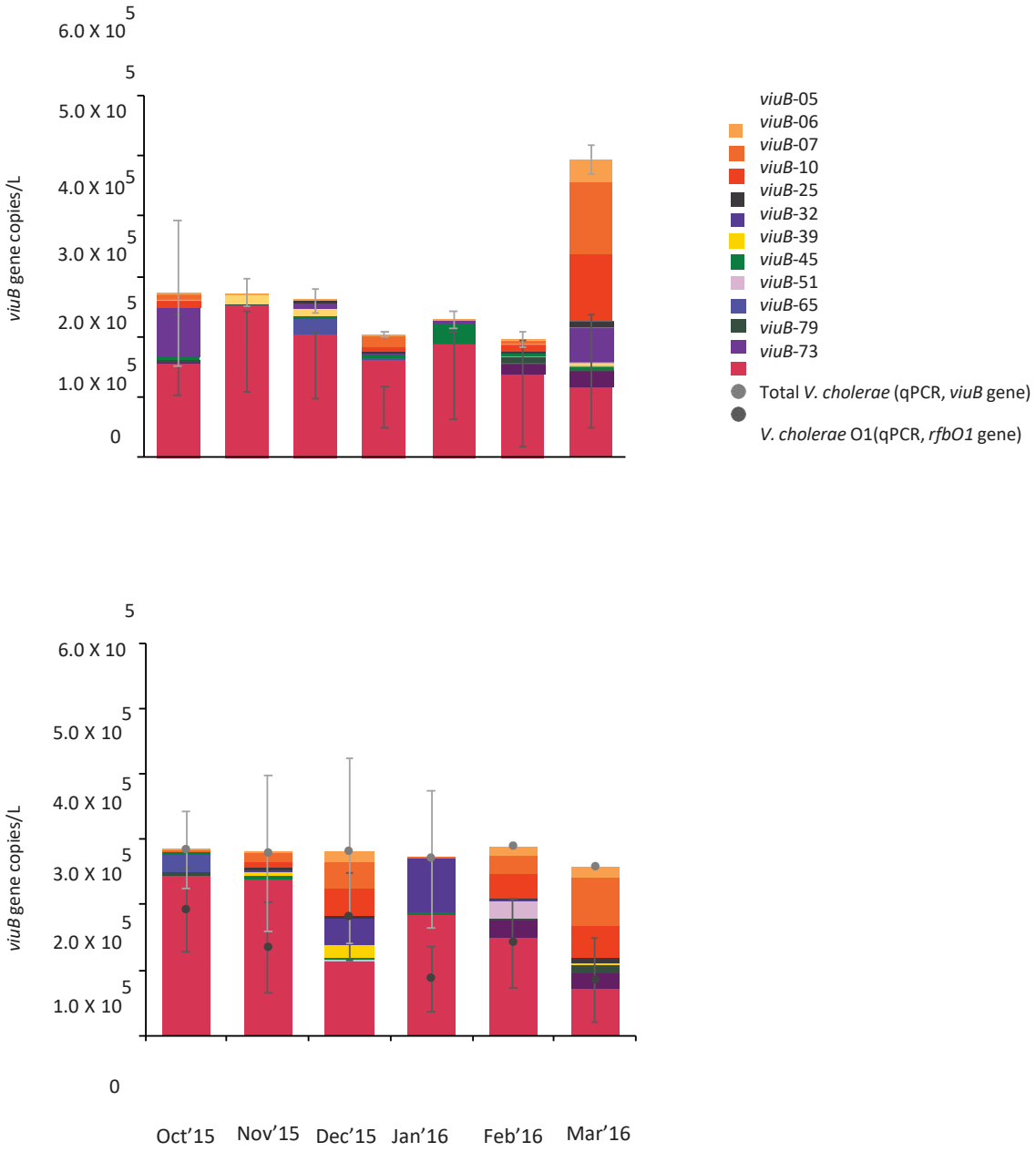


Fig. 4.2. Spatial and temporal distribution of *viuB* alleles found in Dhaka, Bangladesh. A) Spatial variation of *viuB* alleles relative abundance. Each of which corresponds to a different *V. cholerae* lineage (Fig. C.1). Error barsrepresent standard deviation from the average abundance at each site across the entire time series; B) Temporal variation of *viuB* alleles relative abundance. Error bars represent standard deviation from the average monthly abundance across all sites. Aquatic biomass was extracted from water samples collected from seven different sites around Dhaka, Bangladesh in six consecutive months (Oct 2015 to Mar 2016). Total *V. cholerae* abundance was determined by qPCR amplification of the *viuB* gene marker and used to normalize the number of the 12 most abundant*viuB* alleles sequences representing >99% of all reads. Light grey dots denote the average count of total *V. cholerae* genomes copies at each site enumerated by qPCR. Proportions of different *viuB* alleles are denoted with colors specificto each allele. The relative abundance of *viuB*-73, a proxy for PG *V. cholerae* (mostly composed by O1 serogroup strains in Dhaka), was confirmed by qPCR with the amplification of rfbO1 gene (specific for the O1 serogroup), whichis represented by dark grey dots.

Using *viu*B amplicon sequencing, 25 *viuB* alleles were found in total, differing from each other by two or more single nucleotide polymorphisms. Each unique allele represents a specific *V. cholerae* lineage, roughly the equivalent of a multilocus sequence typing (MLST) clonal complex (Kirchberger et al., 2016, 2020). Amongst them, 12 *viuB* alleles (each with > 20,000 sequence reads) were included for further analysis of the *V. cholerae* community composition, while the other 13 alleles representing < 1% of the total population were excluded following the strategy adopted previously (Kirchberger et al., 2020). Total amplicon sequencing reads were normalized based on qPCR data. The count of *viuB*-73 allele from amplicon sequencing corresponding to the PG lineage (Kirchberger et al., 2020) significantly correlated with the *rfBO1* copy number obtained by qPCR (Pearson correlation p value <0.05). *viuB*-73 therefore largely corresponds to the 7th pandemic El Tor *V. cholerae* in the Dhaka environment and was present in all seven sites throughout the six months of sampling (**Fig. 4.2**). That this allele dominates *V. cholerae* communities in the water bodies of Dhaka is not surprising, given the endemicity of cholera in the city and the contamination of water bodies with human waste. Interestingly, *viuB*-73 displayed high abundance based on *viuB* sequencing throughout the six months sampling period, as opposed to only occasional detection of strains corresponding to this genotype using culture-based techniques (Alam et al., 2007, 2011). Sustained presence of PG *V. cholerae* detected by both qPCR and amplicon sequencing underscores the importance of culture-independent methods of tracking toxigenic *V. cholerae* in nature. Although regular cholera episodes occur year-round, the bacterium is usually only culturable during the two seasonal peaks (Alam et al., 2011), presumably being

present in its viable but nonculturable (VBNC) state at other times (Colwell, 2000). In subspecies-level population analysis, *viu*B-73 remained the predominant allele in Dhaka, representing 24% to 92% of the total *V. cholerae* population. In a study of the natural *V. cholerae* population in cholera-free Oyster Pond on the US east coast, *viuB*-73 was found sporadically at lower relative abundance and other *viu*B alleles dominated the *V. cholerae* population (Kirchberger et al., 2020). Even though current study could not present the picture of a whole year due to our sampling limitations, it revealed that PG *V. cholerae* was present even in the so called 'non-epidemic' months of the year (December-February) and was the predominant genotype throughout the sampling period in the surface water of cholera endemic Dhaka. During these months, a majority of the *V. cholerae* population are thought to remain in VBNC form (Alam et al., 2007; Sultana et al., 2018), which could be the cause of their disappearance from culture-based surveillance. Another possibility is that there is a lower level of virulence gene expression during this period is because of the lower water temperature. Other environmental or human factors may also lead to the reduction of the number of infections during this period, even though toxigenic *V. cholerae* survives in the environment. Drastic change in the environment like flooding and excessive rain fall can contaminate the water supply and lead to the rise of clinical cases to pose a threat to the healthcare system.

4.4.2 Subspecies level diversity correlates with variation in environmental parameters

Studies conducted to look into populations of *V. cholerae* in natural habitats have shown that different environmental factors might influence their abundance (Lutz et al., 2013; Takemura et al., 2014). But how variation in these parameters influence subspecies level diversity of *V. cholerae* is not known. In this study, several environmental variables were measured *in situ* : pH,

dissolved oxygen (DO), conductivity, total dissolved solids (TDS), salinity and water temperature. This was to gain insight on their effect on the absolute abundance of the species and relative abundance of the different *viuB* alleles found in the water reservoirs in Dhaka. Salinity, pH and temperature showed little variation at the seven sites sampled, ranging from 0-0.8 ppt, 6.5-8.0 and 30-32°C, respectively (Table C.3). Conductivity (134.5 -1608 μs/cm) and TDS (67.2-804 mg/L) changed noticeably with sites and time of sampling (Table C.3). DO also changed across the time and space ranging from 0.1-5.05 mg/L (Table C.3). Potential links between environmental parameters and the diversity (richness) and population composition were assessed using correlation analyses. Diversity (richness) of *V. cholerae* alleles was significantly correlated with salinity, conductivity, TDS and DO (**Fig. 4.3, Fig. C.1**). Salinity, conductivity and TDS were found to be positively correlated with allele richness (Pearson correlation coefficient 0.44, 0.41 and 0.40 respectively; $P<0.01$), while DO (mg/L) showed a negative correlation (Pearson correlation -0.30; $P<0.05$) with that population characteristic. In previous studies, various physicochemical variables have been found to be associated with the abundance and persistence of *V. cholerae* in aquatic environments and with the risk of cholera outbreaks (Huq et al., 2005; Stine et al., 2008). Temperature and salinity have been observed to influence planktonic populations, which is a well-known habitat for *V. cholerae* in the aquatic ecosystem (Lutz et al., 2013). Adaptation to a wide range of salinity levels also facilitates *V. cholerae*'s survival in various aquatic environments (from coastal to inland water) (Huq et al., 2005; Takemura et al., 2014). Previous studies suggest that abundance of *V. cholerae* decreases with increasing salinity and that they are most abundant in salinities ranging from 0 to 10 ppt (Huq et al., 2005; Takemura et al., 2014).

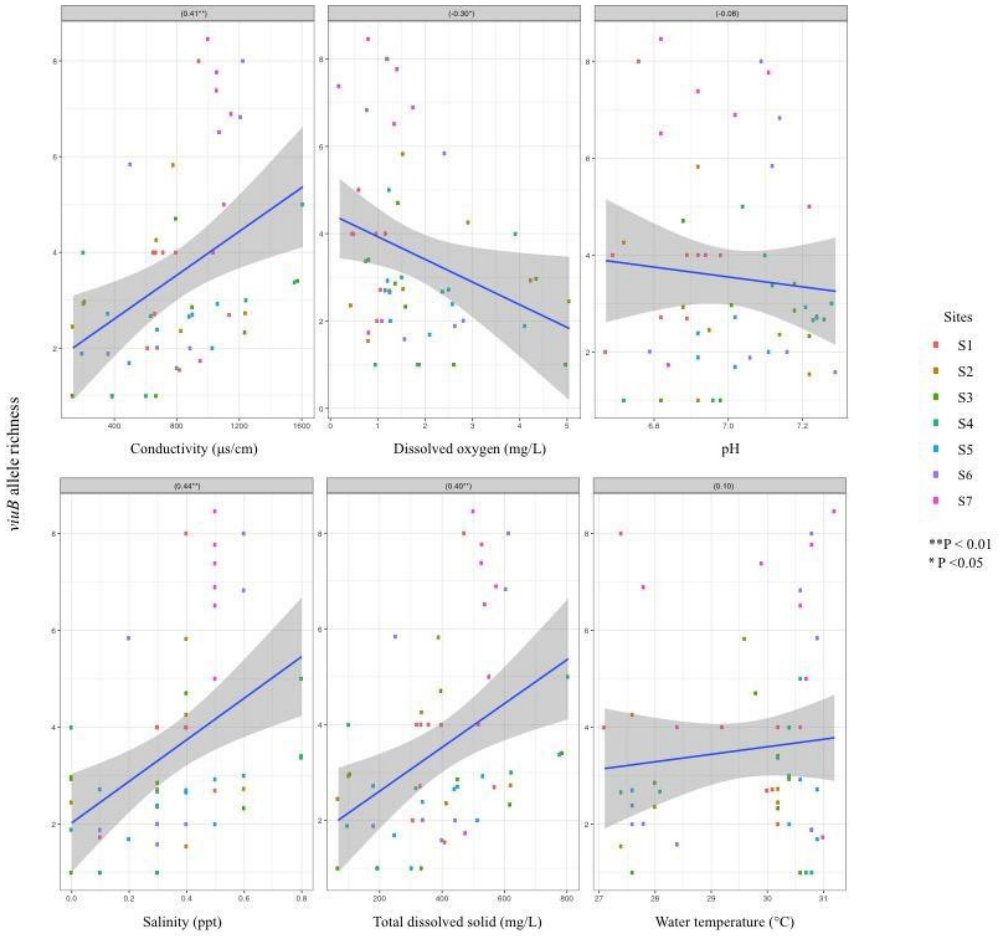


Fig. 4.3. Correlation of environmental variables and viuB allele richness. A correlation test was performed between viuB allele richness and each environmental variable measured in R (vegan library's bioenv function). Each scatter plot shows a different environmental variable (x axis) and the corresponding viuB allele richness (y axis). Pearson's correlations are given in parentheses, and significance is indicated (*P<0.05 and **P<0.01) above each panel. Sites are indicated by different colors.

Multiple linear mixed effect regression analysis was performed to find out which of the independent parameter impacted the richness of the subspecies population significantly. A repeated measure analysis was performed using a site-level random intercept term. When all the

six environmental variables were used in the model, it revealed that Salinity, conductivity, and TDS were highly correlated with each other. In the final model, only four parameters (pH, DO, salinity and WT) were therefore considered as independent predictors of the variance in *viuB* allele richness (**Table 4.1**). The only variable that was found to be significant in that regards (at significance level of 0.05) is salinity (p-value=<0.001) (**Table 4.1**), increasing richness by 19.1% for each increase of 0.1 ppt. As conductivity and TDS were strongly correlated with salinity (Pearson's rho= 0.93 for both), the effect of those two parameters on richness can be expected to be similar to salinity.

Table 4.1: Summary of results from multiple linear mixed effect regression analysis. Predicted changes of the richness of *V. cholerae* community are shown as percentage values in response to a unit change of the four independentpredictors. A repeated measures analysis was performed using a site-level random intercept term. All reported p- values were two-sided and considered statistically significant when less than 0.05.

Predictor	**% Change (95% Confidence Interval)**	**p-value**	**Interpretation**
pH	-0.971% (-10.092% – 8.603%)	0.823	For every 0.1 increase in pH, richness decreases by 0.971%
DO	14.903% (-5.420% – 37.125%)	0.094	For every 1.0 increase in dissolved oxygen, richness increases by 14.903%
Salinity	19.135% (8.938% – 29.978%)	<0.001	For every 0.1 increase in salinity, richness increases by 19.135%
WT	2.365% (-7.150% – 13.005%)	0.635	For every 1.0 increase in water temperature, richness increases by 2.365%

Redundancy analysis (RDA) was done using Monte Carlo permutation test with the aim of testing the significance of the constraint ranking model to reveal what environmental factor/s impacted the shift in *viuB* allele composition. In the RDA biplot, two axes explained 27.1% of the variation in allele composition (**Fig. 4.4**). Salinity was the strongest factor that was positively correlated

with allele richness.

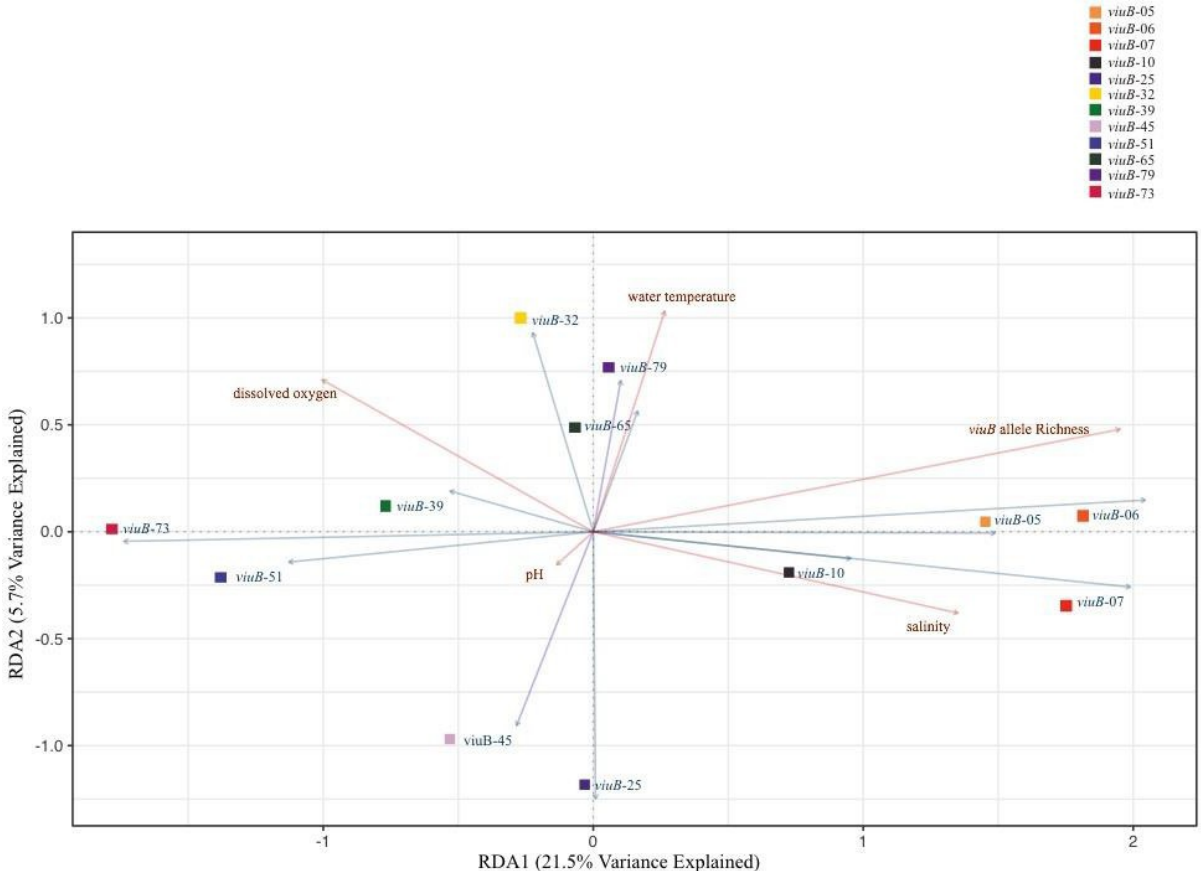


Fig. 4.4. Redundancy analysis (RDA) illustrating the relationships between *Vibrio cholerae* community at the subspecies level and environmental variables. Environmental factors and *viuB* allele richness were marked by red arrows relative to the first two canonical axes (RDA1 and RDA2). The angles in the biplot between viuB alleles and environmental variables, and between viuB alleles themselves or environmental variable themselves, reflect their correlations. An angle less than 90° suggests positive correlations, and an angle approaching 180° suggests a strong negative correlation between variables. The length of the arrow measures the degree of effect an environmental variable has on the community, a longer arrow indicates a greater effect on the community.

In a study of the *V. cholerae* population in cholera free Oyster Pond coastal ecosystem (Falmouth, MA, USA), the *viuB*-73 allele was rarely found in the ocean (Kirchberger et al., 2020), whereas it was present in the brackish pond and lagoon water connected to the ocean, suggesting that high salinity might represent an environmental barrier to the dispersal and range of cholera. The negative correlation between *viuB*-73 abundance and salinity observed in this study also suggest that PG *V. cholerae* (*viuB*-73) might be adapted to low salinity. The PG lineage represented by *viuB*-73 might have higher tolerance of rapid variations at low salinities, as it was consistently predominating in all but one site of Dhaka where salinity levels fluctuated between 0-0.8 ppt. The only site at which *viuB*-73 was not dominant was site 7 (**Fig. 4.5**), in which salinity was more stable and only varied from 0.4 ppt to 0.5 ppt, and never dropped below 0.4 ppt. This suggests that even at the subspecies level, small environmental fluctuations could have an influence on population composition of *V. cholerae*. In the low-lying Ganges delta, salinity intrusion is considered a major threat due to the changing climate. Reduced upstream discharge, sea level rise and other catastrophic events such as cyclones can lead to increase in the salinity of inland water bodies (Huq et al., 2005). A salinity increase of about 26% was recorded in coastal regions in Bangladesh over the last 35 years (Mahmuduzzaman et al., 2014). Such a shift in salinity could be affecting the composition of *V. cholerae* populations, possibly changing the distribution and abundance of various lineages, some of which could pose a threat to human health.

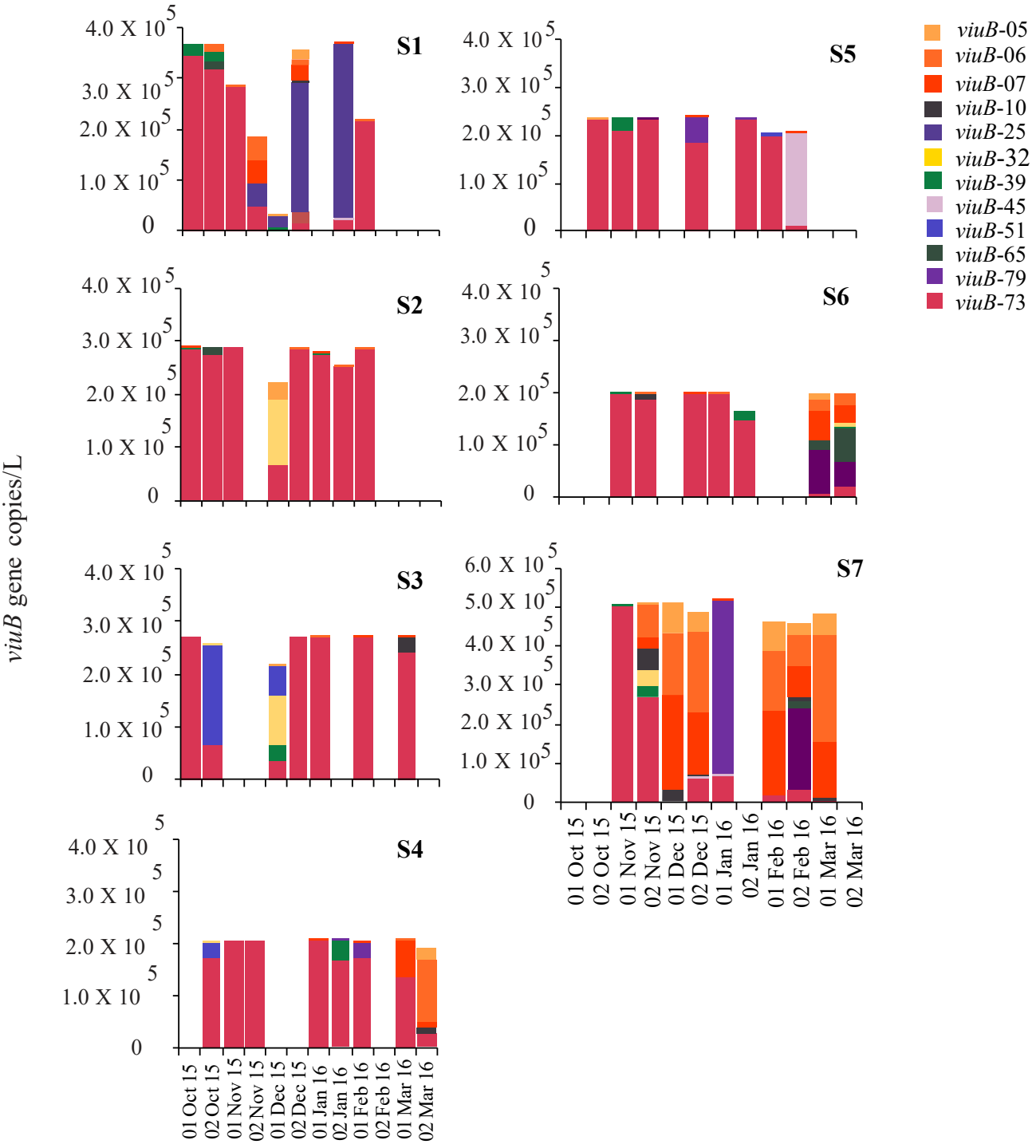


Fig. 4.5. Abundance of *V. cholerae* lineages represented by different *viuB* alleles present at various water reservoirs in Dhaka, Bangladesh. Water samples were collected from seven different sites (S1-S7) in six consecutive months (Oct 2015 to Mar 2016). The partial *V. cholerae* specific *viuB* gene was amplified and sequenced from DNA extracted from the biomass of these samples. Relative abundance of *viuB* alleles representing various lineages is

presented on a scale of absolute abundance of V. cholerae genome copies determined by qPCR of the *viuB* gene. Blankspaces indicate missing samples.

DO, a measure of free non-compound oxygen present in an aquatic system, is another variable influencing microbial communities (Aldunate et al., 2018). A study in Hood Canal (Washington, USA) demonstrated that there was a strong negative correlation between bacterial richness and DO (Spietz et al., 2015). So far, there are no studies describing the correlation between the intraspecies diversity of *V. cholerae* and DO, but it was demonstrated earlier that *V. cholerae* is most abundant in low DO environments (Blackwell and Oliver, 2008; León Robles et al., 2013). In our analysis, DO was found to impact the *V. cholerae* diversity negatively in the Dhaka environment (**Fig. 4.3, Fig. 4.4**). DO could also be one of the factors contributing to the differential adaptation at the subspecies level. Among the sampling locations, site 7 had the lowest average DO (0.90 mg/L). Average DO in site 1 was also low at 0.95 mg/L, but the other five sites (site 2-6) had noticeably higher average DO concentration ranging from 2.7 to 3.0 mg/L. Conductivity, salinity and TDS usually have a negative relationship with DO, which might have caused the lower DO observed in sites 1 and 7. Site 7 was the only one where *viuB* -73 was outcompeted in abundance by other *viuB* alleles (*viuB*-05 and *viuB*-06) (**Fig. 4.5**). *V. cholerae* being a facultative anaerobe, can adapt to the low-oxygen conditions by utilizing alternative energy-producing pathways (i.e. nitrate utilization) (Bueno et al., 2020). It is possible that some lineages are at an advantage in lower oxygen concentrations, giving them the ability to co-exist with other lineages that usually outcompete them, leading to increased diversity at lower DO.

Temperature is another factor affected by climate change that is known to have an impact on *V. cholerae* populations. While this species is found at a wide range of temperatures (10 to 30 °C), the highest abundance is observed at >20 °C (Takemura et al., 2014). However, no significant

correlation between temperature and *viuB* allele diversity (Pearson correlation coefficient 0.1) were observed in Dhaka (**Fig. 4.3**). Water temperatures remained within the range of 27.4°C to 30.8°C throughout the sampling period, a pattern that differs dramatically from regions where shifts in temperature are noticeable during summer and winter (Jiang and Fu, 2001). In a temperate region, attachment to particles and hosts has been shown to increase when temperature increases above 22 °C, contributing to changes in the lineage composition of a *V. cholerae* population when a seasonal change occurs in a temperate climate (Kirchberger et al., 2020). It is likely that at conditions with consistent high temperatures such as those found in a tropical region like Bangladesh, *V. cholerae* is not overly responsive to this parameter, either in terms of its growth rate or particle attachment behavior. Overall, this study unravels the role of important environmental parameters on the composition of the *V. cholerae* community in Dhaka. It must be noted that observations from this study are based on biweekly sampling at a single point per site. We have also lost some samples due to difficulties in transportation, which led to a lack of continuity in temporal data (**Fig. 4.5**). As a result, this study is likely to have missed small fluctuations in environmental determinants and population composition. A high-resolution time series analysis would pinpoint on how even small daily fluctuations in those parameters impact the *V. cholerae* community at a subspecies level and the possible ecological outcome of those changes in the population composition.

High human population density correlates with changes in the *V. cholerae* population of Dhaka reservoirs

Dhaka is one of the most densely populated areas in the world, with a density of 23,234 people per square kilometer within a total area of 300 square kilometers. This huge burden of human population has significant impact on the ecology and evolution of *V. cholerae* and consequently

epidemiology of cholera in this region (Boucher et al., 2015a; Islam et al., 2017a). Based on the demographic records of Dhaka city, the area surrounding sampling site 7 (Kamrangir char) is the most densely populated (100,000 people/km^2) among the seven sites studied, whereas population density at the other six sites ranged from 10,000/km^2 to 60,000/km^2 (**Fig. 4.1**) (Hafiza Khatun, Nishat Falgunee, 2017). Thus, human impacts on the water reservoir in this area is expected to be much higher than other sites, with a higher level of fecal contamination and industrial waste mostly from tannery industrial units. This area is mostly inhabited by a dense low-income population where people use shared hygiene facilities such as showers and toilets (Nelson et al., 2014). Open defecation has been reported from poorly maintained shared facilities. Additionally, high population density in this area frequently causes an overload of septic tanks, which results in the overflow of untreated effluent to the water reservoir (Mansour and Esseku, 2017). The open drainage system commonly causes mixing between sewage and fresh water, increasing the possibility of *V. cholerae* transmission between the water reservoir and local population. Amongst the seven different sites studied, most of them had TDS concentration of < 300mg/L on average, except sites 1 and 7, where TDS varied from 308 to 472 mg/L and from 476 to 575 mg/L, respectively (Table C.3). TDS value comes from the combination of the disassociated electrolytes and other compounds such as dissolved organic matter (Rusydi, 2018). Typically, natural bodies of water have dissolved solids due to the dissolution and weathering of rocks and soil. However, human activities also influence the concentration of TDS in water reservoirs, which is especially likely in Dhaka's inland water bodies because of urban runoff as well as wastewater discharge. This heavy influence of human population at Kamrangir char likely led to different population dynamics of *V. cholerae* at this site compared to other locations in the city. However, in our sampling sites, conductivity, salinity and TDS were found to be tightly correlated, indicating inorganic ions might be the major contributor of the variation in TDS values. Hence, major cause

of higher conductivity, salinity and TDS values observed in site 7 can also be indicators of increased chemical pollution there in comparison to other sites. Noticeable differences in the diversity and population composition in site 7 might suggest that human population density can have a direct or indirect impact on the *V. cholerae* population at the subspecies level.

Although absolute abundance of total *V. cholerae* remained stable at all seven sites throughout the six months sampling period, there are some for site 7, where total *V. cholerae* abundance was around 28% to 43% higher than any other site (**Fig. 4.2**). It is unclear if this higher abundance is due to a more constant input of *V. cholerae* from human waste, an indirect increase in numbers from nutrients linked to this waste, or the existence of a niche allowing the expansion of a particular lineage that does not compete with others (sympatry). Unlike sites 1 to 6, at which *viuB*-73 was more abundant than all other alleles combined (**Fig. 4.5**), three *viuB* alleles (*viuB*-05, *viuB*-06 and *viuB*-07) had greater combined abundance than *viuB*-73 at site 7 (**Fig. 4.5**). The trend is most pronounced during December 2015 to March 2016, coinciding with a decrease of water quality in the reservoir during December to April (Hasan et al., 2019). These lineages have so far not been found in environmental surveys outside of Dhaka. Interestingly, these three alleles (*viuB*-05, *viuB*-06 and *viuB*-07) have been found in strains representing a basal long branch clade in whole-genome phylogeny of known global *V. cholerae* strains (Islam et al.). Representative isolates of this clade were found to be indistinguishable from typical *V. cholerae* based on conventional phenotypic tests, but they are phylogenetically and genotypically divergent (Islam et al., 2018).

Allelic differences and *in-silico* DDH of whole genome sequences show that organisms represented by these three alleles (*viuB*-05, *viuB*-06 and *viuB*-07) are not more closely related to each other on average than most pairs of *V. cholerae* strains would be (**Fig. C.2**). Despite this lack of genetic similarity, their presence and absence are strongly correlated, as they are usually co-occurring (**Fig. 4.6, Fig. 4.4**).

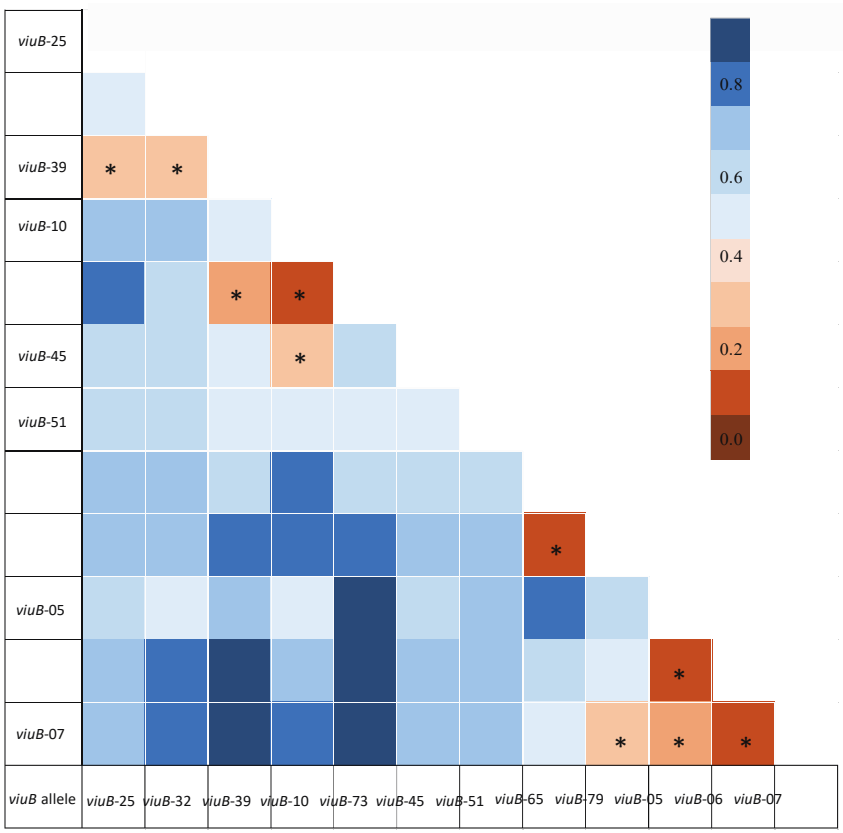


Fig. 4.6. Co-occurrence of V. cholerae lineages based of viuB sequence typing in Dhaka. The Hopkins statistics (H) was used to assess the clustering tendency of various viuB allele sequences in the dataset. The threshold value was 0.5, meaning that if H < 0.5, data showed significant clustering. Visualization of analyzed viuB amplicon sequence data is presented with blue and brown color gradients (H < 0.5, indicating that the data is highly clusterable which is shown with a brown color gradient and H > 0.5, indicating that the data is not clusterable which is described by blue color gradient). Significant clustering (H < 0.5) is observed between viuB alleles 05, 06 and 07. * indicates significant clustering.

Surprising dominance of the *viuB*-05, 06 and 07 lineages over *viuB*-73 exclusively at the Kamrangir char (site 7) location suggests that this group of lineages is specifically adapted to the environmental conditions found at this site. As the site also stood out from a human population density point of view, and that these lineages were most abundant in this site, it suggests a potential human link to the ecology of these lineages in Dhaka. Indeed, strains phylogenetically related with

these lineages have been isolated from human samples in different parts of the world [59]. Abundance of those lineages was positively correlated with TDS, conductivity and salinity, which can be considered as indicators of the impact of human intrusion in the water. Among the seven sampling sites, these three related parameters were highest on average at site 7. These three alleles (*viuB*-05, 06, 07) were most abundant at site 7 (54% of total *V. cholerae*), where *viuB*-73 had the lowest abundance among all sites (24%). This suggests that some form of competition could be taking place between different *V. cholerae* lineages and/or that the lineages respond differently to environmental factors present at this site. Recently, in a more robust study of genomic, phylogenetic and phenotypic characterization of isolates harboring these alleles has suggested for it to be a novel sister species of *V. cholerae* and named as *V. paracholerae.* Possible link of the human population to the abundance and distribution of this novel species makes this lineage of bacteria compelling candidates for future studies looking into the ecology of human adapted genotypes of *V. cholerae*.

4.4.3 Intraspecies interaction could influence relative abundance of PG *V. cholerae* O1

Even though overall *viuB*-73 was the predominant allele over the six months sampling period, spatial and temporal analysis indicate that direct or indirect intraspecies competition among *V. cholerae* genotypes could play an important role in the population dynamics observed in Dhaka. Whenever *viuB*-73 displayed a drop in abundance, another lineage carrying a *viuB*-05, *viuB*-06, *viuB*-07, *viuB*-25, *viuB*-45, *viuB*-51 or *viuB*-79 allele underwent a rapid expansion. Although the abundance of these other alleles was not directly quantified by qPCR, the absolute abundance measurement of the total *V. cholerae* and PG *V. cholerae* O1 by qPCR (Fig. 4.2) supported the observation coming from amplicon sequencing data. Furthermore, this shift in abundances is repeated multiple times within our dataset. For example, *viuB*-25 increased in relative abundance

when the *viuB*-73 abundance decreased at sites 1 and 7 (Fig. 4.5). Notably, a new *viuB* allele, *viuB*-79 (for which no cultured isolates have been found), also appeared when *viuB*-73 abundance was reduced at sites 6 and 7 (Fig. 4.5). All of these alleles showed an inverse correlation with *viuB*-73 relative abundance (Fig. 4.6). Hierarchical clustering analysis of *viuB* amplicon sequencing data (Fig. 4.6) shows that statistically significant clustering (Hopkins statistics [*H*] < 0.5) only occurred between *viuB* alleles representing strains harboring *viuB*-05, *viuB*-06 and *viuB*-07 alleles (*H* <0.5). *viuB*-73 showed significant negative correlation with most other *viuB* alleles (*H* >0.5) (Lawson and Jurs, 1990). The only *viuB* alleles positively correlated with *viuB*-73 were *viuB*-10 and *viuB*-39 (Fig. 4.6), the latter being a ubiquitous allele present at low abundance across sampling sites as well as other geographical locations (Fig. 4.5).

The cause of these shifts in the abundance of some alleles is unclear. It could be due to a differential response to environmental factors or trophic interactions. In the Oyster Pond (MA, USA) *V. cholerae* population, divergent responses of different lineages were observed in ocean, lagoon and pond, where ecosystem parameters varied substantially(Kirchberger et al., 2020). It is plausible that *V. cholerae* lineages responded differently to even the slight environmental variations observed at the sites sampled in Dhaka. Differential response of lineages to phage predation can also be a cause of shifts in the community composition. Effect of predation by bacteriophages on the *V. cholerae* population composition can be modulated by the environmental factors i.e. nutrient availability [61], which in turn can give advantage to certain lineages to outcompete others under certain environmental conditions.

Another factor influencing this differential response could be the ability of various strains to avoid predation. The abundance of *V. cholerae* is influenced by grazing from heterotrophic protists (Lutz et al., 2013). To overcome the grazing pressure *V. cholerae* executes different strategies such as

morphological shift, i.e. from smooth to rugose, resulting in the production of VPS (*Vibrio* polysaccharide) that helps to encase themselves in biofilm and resist predation (Matz and Kjelleberg, 2005). *V. cholerae* can also survive predation by becoming intracellular in a range of amoeba (Abd et al., 2004). They are also able to kill grazers using T6SS (MacIntyre et al., 2010). This system encodes a syringe-like structure that can pierce cellular envelopes of other bacteria and some eukaryotes, injecting effector proteins that can kill the recipient if it does not process the cognate immunity protein (Unterweger et al., 2014). This phenomenon could influence the population composition of *V. cholerae*, with different lineages having varying effectors and immunity proteins providing varying predatory success (Unterweger et al., 2014; Kirchberger et al., 2017). Another possibility is that incompatibility between subspecies likely plays a role in the diversity and dynamic of the *V. cholerae* populations in Dhaka. Most of these genotypes T6SS effector and immunity protein profiles suggest they are incompatible and can kill each other on contact (Hussain et al., 2021). Even though the PG lineage carrying *viuB*-73 could seemingly outcompete strains represented by all other alleles at sites 1-6, it was outcompeted by the three co-occurring alleles corresponding to the *viuB*-05, 06 and 07 group at site 7. This site differs from the other six sites in terms of environmental parameters and surrounding human population density. These observations suggest that environmental conditions can play an important role in shaping the intra-species competition of *V. cholerae* in their natural environment to impact diversity and subspecies level population dynamics of the *V. cholerae*.

4.5 Conclusion

For centuries, the Ganges delta has been a reservoir for pandemic and non-pandemic *V. cholerae* lineages. Cholera endemicity is usual in this area, and Dhaka is one of the most densely populated

megacities in the world, with cholera epidemics occurring biannually before and after the rainy season. This study revealed the consistent presence of PG *V. cholerae* lineage in the natural water bodies of Dhaka at a considerable proportion of the total *V. cholerae* population, suggesting that the toxigenic bacterium is circulating year-round in this urban aquatic environment, or is constantly shed by human carriers. Moreover, population analysis with subspecies level resolution revealed that other *V. cholerae* lineages were coexisting with PG *V. cholerae* O1 in that environment. Intraspecies niche specialization and potentially subspecies interactions with these other lineages could decrease PG *V. cholerae* O1 abundance occasionally, especially when environmental parameters, such as consistently higher TDS, salinity or lower dissolved oxygen concentration favored other lineages. Variability of these parameters, even on a small scale, correlated with changes in *V. cholerae* population composition and diversity, with more temporal stability of salinity and TDS at one particular site correlating with a dramatically different lineage composition. The same site also displayed the highest human population density among the seven sampling locations, indicating humans as a possible contributing source of organic and inorganic materials, or of *V. cholerae* themselves, altering *V. cholerae* population composition. Consistent human interaction and frequent leaking of sewage into the water being linked to a substantially different lineage composition would suggest that human gut may also serves as a potential reservoir for PG O1 and other lineages of *V. cholerae,* resulting in year-long persistence of *V. cholerae* belonging to these lineages through the transmission cycle between human and aquatic environments. A study of the microbiomes of individuals living near water reservoirs is essential to identify if humans play a direct or indirect role in the differences observed between the aquatic *V. cholerae* populations found at different Dhaka sites.

References

Abd, H., Weintraub, A., and Sandström, G. (2004). Interaction between *Vibrio cholerae* and Acanthamoeba castellanii. *Microb. Ecol. Health Dis.* doi:10.1080/08910600410029190.

Alam, M., Hasan, N. A., Sultana, M., Nair, G. B., Sadique, A., Faruque, A. S., et al. (2010). Diagnostic limitations to accurate diagnosis of cholera. *J Clin Microbiol* 48, 3918–3922. doi:10.1128/JCM.00616-10.

Alam, M., Islam, A., Bhuiyan, N., Rahim, N., Hossain, A., Khan, G. Y., et al. (2011). Clonal transmission, dual peak, and off-season cholera in Bangladesh. *Infect. Ecol. Epidemiol.* doi:10.3402/iee.v1i0.7273.

Alam, M., Sultana, M., Balakrish Nair, G., Siddique, A. K., Hasan, N. A., Bradley Sack, R., et al. (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.* doi:10.1073/pnas.0705599104.

Aldunate, M., De la Iglesia, R., Bertagnolli, A. D., and Ulloa, O. (2018). Oxygen modulates bacterial community composition in the coastal upwelling waters off central Chile. *Deep. Res. Part II Top. Stud. Oceanogr.* doi:10.1016/j.dsr2.2018.02.001.

Ali, M., Lopez, A. L., Ae You, Y., Eun Kim, Y., Sah, B., Maskery, B., et al. (2012). The global burden of cholera. *Bull. World Health Organ.* doi:10.2471/blt.11.093427.

Angiuoli, S. V., and Salzberg, S. L. (2011). Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics*. doi:10.1093/bioinformatics/btq665.

Auch, A. F., Klenk, H. P., and Göker, M. (2010). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genomic Sci.*

doi:10.4056/sigs.541628.

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*. doi:10.1186/1471-2164-9-75.

Blackwell, K. D., and Oliver, J. D. (2008). The ecology of Vibrio vulnificus, *Vibrio cholerae*,and Vibrio parahaemolyticus in North Carolina Estuaries. *J. Microbiol.* doi:10.1007/s12275-007-0216-2.

Boucher, Y., Orata, F. D., and Alam, M. (2015a). The out-of-the-delta hypothesis: dense human populations in low-lying river deltas served as agents for the evolution of a deadly pathogen. *Front Microbiol* 6, 1120. doi:10.3389/fmicb.2015.01120.

Boucher, Y., Orata, F. D., and Alam, M. (2015b). The out-of-the-delta hypothesis: Dense human populations in low-lying river deltas served as agents for the evolution of a deadly pathogen. *Front. Microbiol.* doi:10.3389/fmicb.2015.01120.

Bueno, E., Pinedo, V., and Cava, F. (2020). Adaptation of *Vibrio cholerae* to Hypoxic Environments. *Front. Microbiol.* doi:10.3389/fmicb.2020.00739.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. doi:10.1038/nmeth.3869.

Chatterjee, S. N., and Chaudhuri, K. (2003). Lipopolysaccharides of *Vibrio cholerae*: I. Physical and chemical characterization. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1639, 65–79. doi:10.1016/j.bbadis.2003.08.004.

Chowdhury, M. A. R., Xu, B., Montilla, R., Hasan, J. A. K., Huq, A., and Colwell, R. R. (1995).

A simplified immunofluorescence technique for detection of viable cells of *Vibrio cholerae* O1 and O139. *J. Microbiol. Methods*. doi:10.1016/0167-7012(95)00066-6.

Chun, J., Grim, C. J., Hasan, N. A., Je, H. L., Seon, Y. C., Haley, B. J., et al. (2009a). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0907787106.

Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., et al. (2009b). 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. doi:10.1073/pnas.0907787106.

Chun, J., Huq, A., and Colwell, R. R. (1999). Analysis of 16S-23S rRNA intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. *Appl. Environ. Microbiol.* doi:10.1128/aem.65.5.2202-2208.1999.

Colwell, R. R. (2000). Viable but nonculturable bacteria: A survival strategy. *J. Infect. Chemother.* doi:10.1007/PL00012151.

Colwell, R. R. (2004). Infectious disease and environment: Cholera as a paradigm for waterborne disease. *Int. Microbiol.* doi:10.2436/im.v7i4.9485.

Deen, J., Mengel, M. A., and Clemens, J. D. (2020). Epidemiology of cholera. *Vaccine*. doi:10.1016/j.vaccine.2019.07.078.

Edgar, C. R. C. (2013). USEARCH. *J. Natl. Cancer Inst.*

Faruque, S. M., Albert, M. J., and Mekalanos, J. J. (1998). Epidemiology, Genetics, and Ecology of Toxigenic*Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* doi:10.1128/mmbr.62.4.1301-1314.1998.

Hafiza Khatun, Nishat Falgunee, M. J. R. K. (2017). Analyzing urban population density gradient of Dhaka Metropolitan. *Geogr. OnlineTM Malaysian J. Soc. Sp.* 11, 1–13.

Hasan, M. K., Shahriar, A., and Jim, K. U. (2019). Water pollution in Bangladesh and its impact on public health. *Heliyon*. doi:10.1016/j.heliyon.2019.e02145.

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., et al. (2005). Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* doi:10.1128/AEM.71.8.4645-4654.2005.

Hussain, N. A. S., Kirchberger, P. C., Case, R. J., and Boucher, Y. F. (2021). Modular Molecular Weaponry Plays a Key Role in Competition Within an Environmental *Vibrio cholerae* Population. *Front. Microbiol.* 12, 671092. doi:10.3389/fmicb.2021.671092.

Islam, M. T., Alam, M., and Boucher, Y. (2017a). Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *Int Microbiol* 20, 106–115. doi:10.2436/20.1501.01.291.

Islam, M. T., Alam, M., and Boucher, Y. (2017b). Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *Int. Microbiol.* doi:10.2436/20.1501.01.291.

Islam, M. T., Liang, K., Im, M. S., Winkjer, J., Busby, S., Tarr, C. L., et al. (2018). Draft Genome Sequences of Nine Vibrio sp. Isolates from across the United States Closely Related to *Vibrio cholerae*. *Microbiol Resour Announc* 7. doi:10.1128/MRA.00965-18.

Islam, M. T., Nasreen, T., Kirchberger, P., Liang, K. Y. H., Orata, F. D., Johura, F.-T., et al. Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp. nov.) with a history of association with humans. *Appl. Environ. Microbiol.* 0, AEM.00422-21. doi:10.1128/AEM.00422-21.

Jiang, S. C., and Fu, W. (2001). Seasonal abundance and distribution of *Vibrio cholerae* incoastal waters quantified by a 16S-23S intergenic spacer probe. *Microb. Ecol.* doi:10.1007/s00248-001-0029-5.

Jolley, K. A., and Maiden, M. C. J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. doi:10.1186/1471-2105-11-595.

Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A., et al. (2013). Environmental factors influencing epidemic cholera. *Am. J. Trop. Med. Hyg.* doi:10.4269/ajtmh.12-0721.

Kirchberger, P. C., Orata, F. D., Barlow, E. J., Kauffman, K. M., Case, R. J., Polz, M. F., et al. (2016). A small number of phylogenetically distinct clonal complexes dominate a coastal *Vibrio cholerae* population. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01177-16.

Kirchberger, P. C., Orata, F. D., Nasreen, T., Kauffman, K. M., Tarr, C. L., Case, R. J., et al. (2020). Culture-independent tracking of *Vibrio cholerae* lineages reveals complex spatiotemporal dynamics in a natural population. *Environ. Microbiol.* doi:10.1111/1462-2920.14921.

Kirchberger, P. C., Unterweger, D., Provenzano, D., Pukatzki, S., and Boucher, Y. (2017). Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. *Sci. Rep.* doi:10.1038/srep45133.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01043-13.

Lawson, R. G., and Jurs, P. C. (1990). New Index for Clustering Tendency and Its Application to Chemical Problems. *J. Chem. Inf. Comput. Sci.* doi:10.1021/ci00065a010.

León Robles, A., Acedo Félix, E., Gomez-Gil, B., Quiñones Ramírez, E. I., Nevárez-Martínez, M., and Noriega-Orozco, L. (2013). Relationship of aquatic environmental factors with the abundance of *Vibrio cholerae*, Vibrio parahaemolyticus, *Vibrio mimicus* and Vibrio vulnificus in the coastal area of Guaymas, Sonora, Mexico. *J. Water Health*. doi:10.2166/wh.2013.160.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*. doi:10.1093/bioinformatics/btl529.

Lipp, E. K., Huq, A., and Colwell, R. R. (2002). Effects of global climate on infectious disease: The cholera model. *Clin. Microbiol. Rev.* doi:10.1128/CMR.15.4.757-770.2002.

Longini, I. M., 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.* doi:10.1086/341206.

Lutz, C., Erken, M., Noorian, P., Sun, S., and McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. *Front. Microbiol.* doi:10.3389/fmicb.2013.00375.

MacIntyre, D. L., Miyata, S. T., Kitaoka, M., and Pukatzki, S. (2010). The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1012931107.

Mahmuduzzaman, M., Ahmed, Z. U., Nuruzzaman, A. K. M., and Ahmed, F. R. S. (2014). Causes of Salinity Intrusion in Coastal Belt of Bangladesh. *Int. J. Plant Res.*

Mansour, G., and Esseku, H. (2017). Situation analysis of the urban sanitation sector in Ghana. *WSUP - Water Sanit. Urban Poor*.

Matz, C., and Kjelleberg, S. (2005). Off the hook - How bacteria survive protozoan grazing. *Trends Microbiol.* doi:10.1016/j.tim.2005.05.009.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. doi:10.1186/1471-2105-14-60.

Mutreja, A., Kim, D. W., Thomson, N., Connor, T. R., Hee, J., Lebens, M., et al. (2013). Evidence for multiple waves of global transmission within the seventh cholera pandemic. *Nature*. doi:10.1038/nature10392.Evidence.

Nasreen, T., Hussain, N. A. S., Islam, M. T., Orata, F. D., Kirchberger, P. C., Case, R. J., et al. (2020). Simultaneous quantification of vibrio metoecus and *Vibrio cholerae* with its o1 serogroup and toxigenic subpopulations in environmental reservoirs. *Pathogens*. doi:10.3390/pathogens9121053.

Nelson, K. B., Karver, J., Kullman, C., and Graham, J. P. (2014). User perceptions of shared sanitation among rural households in Indonesia and Bangladesh. *PLoS One*. doi:10.1371/journal.pone.0103886.

Rashid, R. Bin, Ferdous, J., Tulsiani, S., Mackie Jensen, P. K., and Begum, A. (2017). Development and validation of a novel real-time assay for the detection and quantification of *Vibrio cholerae*. *Front. Public Heal.* doi:10.3389/FPUBH.2017.00109.

Ripley, B. D. (2001). The R Project in Statistical Computing. *MSOR Connect.* doi:10.11120/msor.2001.01010023.

Rusydi, A. F. (2018). Correlation between conductivity and total dissolved solid in various type of water: A review. in *IOP Conference Series: Earth and Environmental Science (2018)* doi:10.1088/1755-1315/118/1/012019.

Spietz, R. L., Williams, C. M., Rocap, G., and Horner-Devine, M. C. (2015). A dissolved oxygen threshold for shifts in bacterial community structure in a seasonally hypoxic estuary. *PLoS One*. doi:10.1371/journal.pone.0135731.

Stamatakis, A. (2014). RAxML. *Manual/tutorial*.

Stine, O. C., Alam, M., Tang, L., Nair, G. B., Siddique, A. K., Faruque, S. M., et al. (2008). Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerg. Infect. Dis.* doi:10.3201/eid1405.071116.

Sultana, M., Nusrin, S., Hasan, N. A., Sadique, A., Ahmed, K. U., Islam, A., et al. (2018). Biofilms Comprise a Component of the Annual Cycle of *Vibrio cholerae* in the Bay of Bengal Estuary. *MBio* 9. doi:10.1128/mBio.00483-18.

Takemura, A. F., Chien, D. M., and Polz, M. F. (2014). Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* doi:10.3389/fmicb.2014.00038.

Torondel, B., Ensink, J. H. J., Gundogdu, O., Ijaz, U. Z., Parkhill, J., Abdelahi, F., et al. (2016). Assessment of the influence of intrinsic environmental and geographical factors on the bacterial ecology of pit latrines. *Microb. Biotechnol.* doi:10.1111/1751-7915.12334.

Unterweger, D., Miyata, S. T., Bachmann, V., Brooks, T. M., Mullins, T., Kostiuk, B., et al. (2014). The *Vibrio cholerae* type VI secretion system employs diverse effector modules forintraspecific competition. *Nat. Commun.* doi:10.1038/ncomms4549.

Vezzulli, L., Pruzzo, C., Huq, A., and Colwell, R. R. (2010). Environmental reservoirs of *Vibrio cholerae* and their role in cholera. *Environ. Microbiol. Rep.* doi:10.1111/j.1758-2229.2009.00128.x.

Wright, J. J., Lee, S., Zaikova, E., Walsh, D. A., and Hallam, S. J. (2009). DNA extraction from 0.22 μM Sterivex filters and cesium chloride density gradient centrifugation. *J. Vis. Exp.* doi:10.3791/1352.

Wyckoff, E. E., Mey, A. R., and Payne, S. M. (2007). Iron acquisition in *Vibrio cholerae*. in *BioMetals* doi:10.1007/s10534-006-9073-4.

CHAPTER 5: *Vibrio tarriae*: A lactose-fermenting relative of *Vibrio cholerae* associated with opportunistic infections in humans

A version of this chapter is in preparation for publication in *Frontiers of Microbiology* as:

"**Mohammad Tarequl Islam**, Kevin Liang, Monica S. Im, Jonathan Winkjer , Shelby Busby, Cheryl L. Tarr, and Yann F. Boucher. *Vibrio tarriae*: A lactose-fermenting relative of *Vibrio cholerae* associated with opportunistic infections in humans."

M.T.I., and Y.F.B. designed the experiments. M.T.I., K.Y.H.L and F.O performed data analysis. M.T.I. wrote the manuscript. Y.F.B. and C.L.T. provided criticalreviews on the manuscript.

CHAPTER 5

5.1. Abstract

A number of bacteria with close resemblance to *Vibrio cholerae* have been isolated over the years by Centre for Disease Control (CDC), which could not be assigned proper taxonomic designation based on preliminary identification methods. Nine such isolates have been found to share 16S rRNA gene identity exceeding 99 % with *V. cholerae,* yet their DNA-DNA hybridization (DDH) and average nucleotide identity (ANI) values with this well-known bacterium are below the species cut off, indicating they represent a potentially novel species. Phylogenetic analysis of their core genome places this group of isolates in their own monophyletic clade, within the so called "cholerae clade" but distinct from any other species. Extensive phenotypic characterization reveals traits such as the ability to utilize D-lactose, N-acetyl-D-galactosamine and pectin and the lack of ability to consume D-mannose, D- serine, citric acid, D-ribose, propionic acid, D-alanine, mono-methyl succinate and caproic acid, distinguishing the novel species from *V. cholerae*. Comparative genomic analysis reveals it displays a unique set of siderophore genes, suggesting that Iron acquisition strategies could be vital for the divergence of the novel species from their common ancestor with *V. cholerae*. Average Nucleotide Identity and DNA-DNA hybridization values obtained from the pairwise comparisons of whole-genome sequences of these isolates to *V. cholerae* ranges from 94.4-95.07 % and 60.4-62.1%, respectively. Based on the phenotypic, genetic and phylogenetic differences observed, we propose these isolates represent a novel species of the genus *Vibrio*, for which the name *Vibrio tarriae* sp. nov. is proposed. Strain 2521-89 (=DSM 112461=CCUG 75318),

isolated from lake water, is the type strain.

5.2. Introduction

From a clinical perspective, the "cholerae clade" represents one of the most important group of bacteria among the diverse assemblages within the genus *Vibrio*. The clade is named after *V. cholerae*, the type strain of the genus and the causative agent of cholera, a devastating pandemic disease (Islam et al., 2017a). *V. cholerae* is an extensively studied bacteria, largely because of its lethality. It is also a model system for environmental pathogens causing human disease (Huq et al., 1983). Despite this, our understanding of the natural diversity *V. cholerae* and its close relatives have been limited until recently (Islam et al.; Boyd et al., 2015). A few closely related species have recently been found, most being initially classified as *V. cholerae*-like bacteria in the last few decades (Haley et al., 2010; Hasan et al., 2010). Interestingly, only two serogroups (O1 and O139) of *V. cholerae* are known to be capable of causing cholera (Islam et al., 2017b). Diseases caused by *Vibrio* bacteria other than *Vibrio cholerae* O1/O139 are usually known as vibriosis, and the arbitrary group is known as "Non- cholera vibrios". Cholera is endemic in different parts of the world but not a major concern for countries with well-developed wastewater management and public health systems. However, vibriosis infections are an emerging problem even in countries with such facilities. In the USA, *Vibrio*-related diseases are on the rise, affecting an increasing number of people coming in contact with water during recreational or professional activities (Baker-Austin et al., 2018). Global climate change is thought to have contributed to this alarming increase in cases (Colwell, 2004; Baker-Austin et al., 2018), as vibrios are aquatic bacteria and their ecology is tightly correlated with the environmental conditions (Huq et al., 1983; Colwell, 2004). To track the emergence of novel

and potentially dangerous species, surveillance is paramount and has been carried out regularly in recent years. The Centers for Disease Control (CDC) have been carrying out collection and identification of potential *Vibrio* pathogens from human clinical specimens across the United States and reported them as part of the surveillance conducted under the Cholera and Other *Vibrio* Illness Surveillance (COVIS) program (Islam et al., 2018).

The pathogenic potential of species related to *V. cholerae* vary and so far, they are mostly known for being associated with opportunistic infections (Baker-Austin et al., 2018). But their interaction with *V. cholerae* and other pathogenic bacteria in nature and inside the human gut can be a source of threat. Closely related species have been found to exchange genetic material, including genes encoding virulence factors, with *V. cholerae*. For example, *V. mimicus* has been found to contain virulence genes including cholera toxin and was isolated from diarrheal patients (Hasan et al., 2010). *V. metoecus* have also been found to frequently exchange genes via horizontal gene transfer (HGT) with *V. cholerae* (Orata et al., 2015). The recently described closest known sister species of *V. cholerae*, *V. paracholerae,* had historically been associated with human infections and co-exists with *V. cholerae* in cholera endemic areas (Dorman et al., 2019; Islam et al., n.d.). Genetic interactions of these relatives with *V. cholerae* in a shared environment can be significant for the ecology of this clade and the epidemiology of the diseases caused. Addition of these species to the "cholera clade" is crucial for the study of diversity and evolution of their famous pathogenic sister.

In this study, we employed a polyphasic approach including phenotypic, genomic, and phylogenetic data to describe a novel species isolated from both human end environmental samples from across USA. The name *V. tarriae* sp. nov. is proposed for the novel species. This name was chosen to honor the decades of public service of Dr. Cheryl L. Tarr in fighting

vibriosis and cholera across the world and her contribution to our understanding of *Vibrio* taxonomy, physiology and epidemiology.

5.3. Materials and Methods:

5.3.1. Bacterial isolates:

The bacterial isolates were collected by center for disease control (CDC) as part of national surveillance system in USA for human infection with pathogenic species of the family Vibrionaceae, which cause vibriosis and cholera. Nine isolates were identified as *Vibrio* spp. by preliminary culture identification (Table 5.1); genome sequencing was performed, and genome comparison identified it as a close relative of *V. cholerae* (Liang et al., 2017, 2019). The isolates were transported to the University of Alberta following standard bacterial isolate transportation procedure described elsewhere (Alam et al., 2010).

5.3.2. Phenotypic tests

The isolates were sub-cultured by streaking on thiosulfate citrate bile salts sucrose (TCBS) agar (Becton Dickinson) and LB agar (Difco). The cultures were then incubated overnight at 30 °C. Single colonies from the LB agar cultures were tested using the Analytical Profile Index (API) 20 NE (bioMerieux), the Phenotype MicroArray 1 (PM1) and 2 (PM2) microplate (Biolog) according to the instructions of the manufacturers. Single colonies from the TSB agar cultures were used for the characterization of the isolates using Analytical Profile Index (API) 20 NE (bioMerieux), gram staining and viewing under a light microscope (Carl Zeiss) at 1000X magnification was done following the procedure described by Orata *et al*. (Orata et al., 2016). Permissive growth temperatures were determined in Brain heart infusion broth (BHI) and incubation at a range of 4–45°C, whereas permissive salinity concentrations were determined in

TSB at 30℃ in a range of 0–10 % NaCl.

5.3.3. Genetic and bioinformatic analysis:

For genotypic characterization, genome sequences were obtained from the GenBank database and genome accession numbers are listed in Table 5.1. The G+C content was determined from whole-genome sequences (Table 5.1), using Geneious 8.1.2 (Kearse et al., 2012). Pairwise average nucleotide identity (ANI) was calculated using JSpecies V1.2.1 (Richter and Rosselló-Móra, 2009). Pairwise percent DNA–DNA hybridization (dDDH) was also calculated in silico using the Genome-to-Genome Distance Calculator 2.0 (GGDC) (Meier-Kolthoff et al., 2013). The genome sequences were annotated with RAST 2.0 (Aziz et al., 2008). Core and accessory genes were determined with BPGA finding orthologous protein-coding genes clustered into families based on a 30% amino acid sequence identity (Chaudhari et al., 2016). The gene families unique to *V. tarriae* sp. nov. were determined using a custom-made Python program. Genome sequences were aligned using Mugsy (Angiuoli and Salzberg, 2011) and the alignments were concatenated, stripping columns with at least one gap, using Geneious 8.1.2. This resulted i n a single alignment with a total length of 972,240 bp, which was used to reconstruct a maximum-likelihood tree with RAxML 8.2.8 (Stamatakis, 2014) using the GTR (general time reversible) nucleotide substitution model with gamma distribution of rate categories and 100 bootstrap replicates. For multi-locus sequence analysis (MLSA), six housekeeping genes *adk*, *gyrB, pyrH, pgi, recA* and *rpoA* were selected (Thompson et al., 2005; Kirchberger et al., 2014).

From the DNA sequences, a concatenated alignment of 7,392 bp was obtained and used to reconstruct a maximum-likelihood tree.

5.4. Results and discussions

5.4.1. Phenotypic traits

Isolates of *V. tarriae* sp. nov. used in this study were obtained from the CDC (Table 5.1).

Phenotypic characterization was performed on four representative isolates of *V. tarriae* sp. nov., and compared with same number of isolates from *V. cholerae*, *V. paracholerae* and *V. metoecus* (Tables 5.2 and Table D.1). All the *V. tarriae* sp. nov. isolates studied exhibited growth in TSB without NaCl. Ability to grow in the media without additional salt is a characteristic of the so called "cholerae clade" which would include *V. cholerae, V. paracholerae, V. metoecus* and *V. mimicus*, differentiating it from the rest of the vibrios (Boyd et al., 2015; Farmer III et al.,2015). The isolation of strain 2521-89 of *V. tarriae* sp. nov. from lake water (Liang et al., 2017) suggests that the species is able to survive in freshwater environments. Furthermore, the isolates tested were also able to survive at 40 °C, which is an ability usually present in pathogenic vibrios that can survive inside the human body (Farmer III et al., 2015).

Vibrio tarriae sp. nov. resembles *V. cholerae* (160 of 190 tests or 84%; Table 2 and Table S1) and *V. paracholerae* (172 of 190 tests or 90%) in the majority of phenotypic characteristics tested. However, twelve phenotypic features distinguished *V. tarriae* sp. nov. from *V. cholerae* and at least two from *V. paracholerae* (summarized in Table 5.2). Notably, the new species tested positive for the utilization of N-acetyl galactose amine and pectin as the sole carbon and energy source, while *V. cholerae* was negative for these tests. In contrast to *V. cholerae*, it was unable to utilize D-mannose, Citric acid, propionic acid, monomethyl succinate, caproic acid, D- Serine, D-Glucose-6-phosphate, glycyl-L-proline and D-fructose-6-phosphate. It also differs from *V. paracholerae* by two phenotypic traits: the ability to utilize N-acetyl galactose amine and inability to utilize α-cyclodextrin.

Table 5.1: Information of *V. tarriae* sp. nov., strains along with reference strains of closely related species used inthis study.

Species	Strain	Year of isolation	Source of isolation	Genome accession number	Reference
	2521-89	1989	Water	NZ_CP022353.1	(Liang et al.,2019)

V. tarriae sp. nov.	2523-88	Unknown	Blood	QKKG01000001.1	(Liang et al.,2019)
	2015V-1076	2015	Blood	QKKH00000000.1	(Liang et al.,2019)
	2016V-1018	2016	Wound	QKKI00000000.1	(Liang et al.,2019)
	2016V-1062	2016	Stool	QKKJ00000000.1	(Liang et al.,2019)
	2017V-1038	2017	Stool	QKKK00000000.1	(Liang et al.,2019)
	2017V-1070	2017	Stool	QKKL00000000.1	(Liang et al.,2019)
	2017V-1085	2017	Blood	QKKM00000000.1	(Liang et al.,2019)
	2017V-1124	2017	Wound	QKKN00000000.1	(Liang et al.,2019)
V. cholerae	N16961	1971	Clinical	CP028827	(Heidelberg etal., 2000)
	V52	1968	Clinical	KQ410497.1	(Chun et al.,2009)
	YB3B05	2015	Water	LBGB01000047.1	(Kirchberger etal., 2020)
	YB8E08	2015	Water	LBGN01000044.1	(Kirchberger etal., 2020)
V. paracholerae	EDC 792	2015	Water	WYCC00000000	(Islam et al.)
	EDC 690	2015	Water	WUWI00000000	(Islam et al.)
	EDC 716	2015	Water	WYBZ00000000	(Islam et al.)
	2016V-1091	2016	Stool	QKKQ00000000.1	(Islam et al.)
V. metoecus	OP3H	2006	Water	JJMN00000000.1	(Kirchberger et al., 2014)
	07-2435	2007	Clinical	LCUE00000000	(Kirchberger et al., 2014)
	YB4D01	2009	Water	LBGO00000000	(Kirchberger et al., 2014)
	08-2459	2008	Clinical	LCUF00000000	(Kirchberger et al., 2014)

Species of the "cholerae clade" are usually known to be lactose negative, and the ability to ferment lactose is a distinguishing characteristic for *V. vulnificus* (Oliver et al., 1983). In our phenotypic tests, all four *V. tarriae* sp. nov. isolates and one *V. paracholerae* isolate were positive for growth on lactose, whereas all *V. cholerae* and *V. metoecus* isolates were negative for lactose utilization (Table 5.2).

5.4.2. Genomic and phylogenetic characterization of the novel species

The DNA G+C content of the V. tarriae sp. nov. isolates range from 47.1–47.2 mol%, which is consistent the known range for the genus Vibrio (38.0–51.0 mol%) (Farmer III et al., 2015). Their 16S rRNA gene sequences are >99% identical to *V. cholerae*, *V. paracholerae*, *V. metoecus* and *V. mimicus* strains in the NCBI database. DNA-DNA relatedness of the strains with these species were determined using two matrixes: average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) by pairwise comparisons of whole-genome sequences. The ANI between isolates within the species *V. tarriae* sp. nov. range from 98-99% (Fig. 5.1). In contrast, the ANI value for *V. tarriae* sp. nov. was 95% with V. cholerae and 93-94% with *V. paracholerae*. Since the results are close to the cut-off value of 96% ANI for two genomes to belong to the same species (Richter and Rosselló-Móra, 2009), we complemented our ANI results with dDDH. The GGDC package was used to calculate percent dDDH in silico to mimic wet lab-based DDH (Meier-Kolthoff et al., 2013). dDDH within the species *V. tarriae* sp. nov. ranges from 81.5-85%, whereas they range from 60.6 to 67.9 % with V. cholerae and 59.8-67% with *V. paracholerae* (Fig. 5.1).

MLSA further supports our proposal of a novel species. Single-copy, protein-coding core genes are used as alternatives to 16S rRNA gene sequences for the identification and phylogenetic analysis of various species of the genus *Vibrio*, since there is a lack of species-level resolution using 16S rRNA gene sequences (Gladney & Tarr, 2014; Thompson et al., 2005). The *V. tarriae* sp. nov. isolates form a monophyletic clade that is distinct from *V. cholerae* and other *Vibrio* species, with 100% bootstrap support using six housekeeping genes (*adk*, *gyrB, pyrH, pgi, recA* and *rpoA*)(Kirchberger et al., 2014; Orata et al., 2016) (Fig. 5.2). These genes have been used to for the taxonomic characterization of vibrios and to describe novel species within the genus. The average patristic distance calculated from this tree between the *V. tarriae* sp. nov. isolates and the *V. cholerae* and *V. paracholerae* isolates is 0.06 and 0.09, respectively, while lower average distances of 0.03 are obtained when comparing isolates within the species (Table D.1). To further demonstrate this distinction, a phylogeny was reconstructed based on core genome sequences (Fig. 5.2) This analysis also shows the distinct clustering of *V. tarriae* sp. nov. isolates from other *Vibrio* species. It placed the *V. tarriae* sp. nov. lineage into the context of a larger *Vibrio* phylogeny, showing that the novel species is distinct from all Vibrionaceae that have been characterized to date but fall into the "*V. cholerae* clade". It is more closely related to *V. cholerae* than *V. metoecus* or *V. mimicus* but forms an outgroup to the *V. cholerae* and *V. paracholerae* sister species (Fig. 5.2). Overall, phylogenetic analyses confirm the position of *V. tarriae* sp. nov. in the genus *Vibrio*, which forms a monophyletic clade distinct from other known *Vibrio* species supporting its identification as a novel species. This distinction is further confirmed by ANI and percent DDH below 96 % and 70 %, respectively, between species.

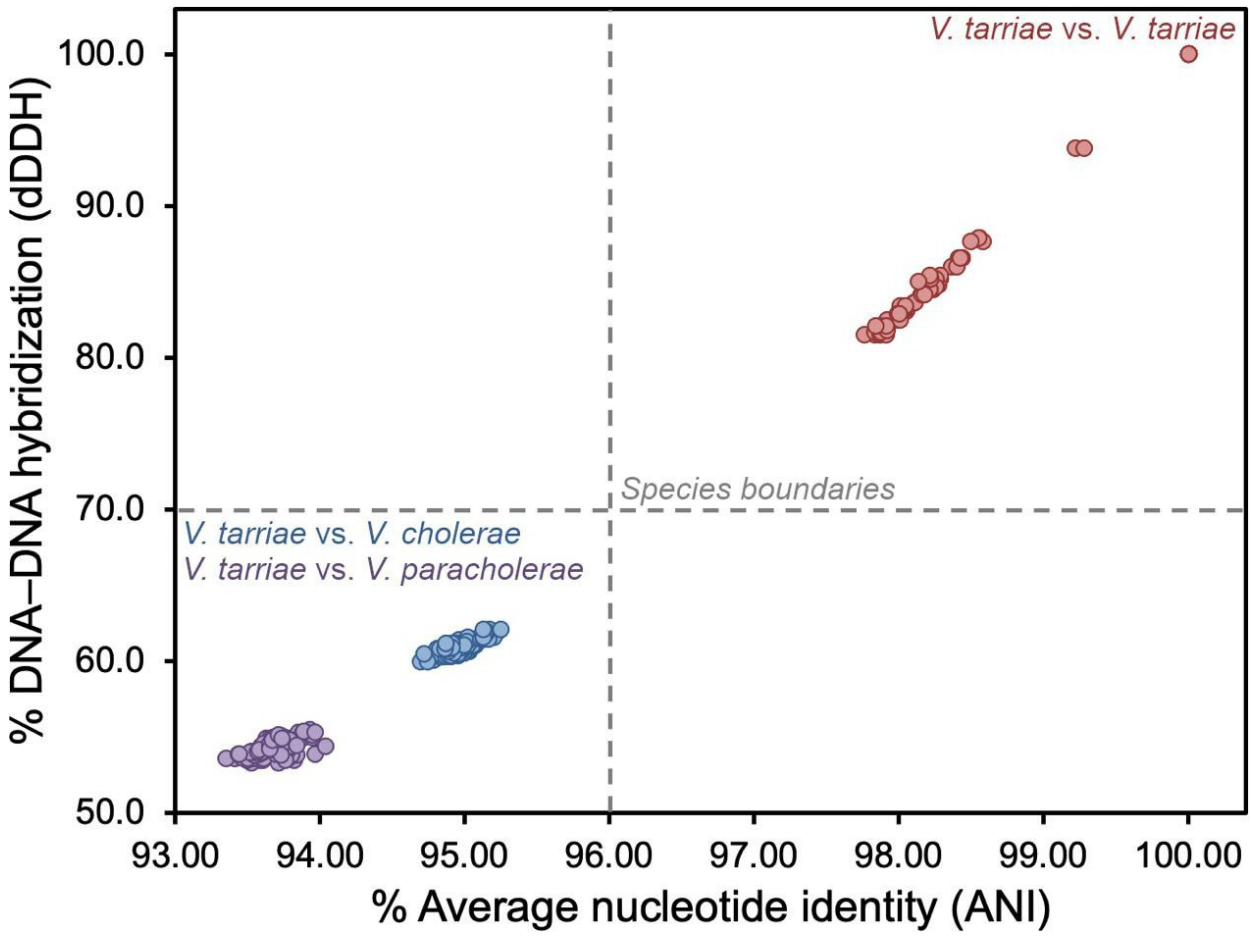


Fig. 5.1. Pairwise percent dDDH and ANI comparisons between isolates of *V. tarriae* sp. nov. and with *V. cholerae* and *V. paracholerae*. Species boundary cut-offs are shown by dotted lines.

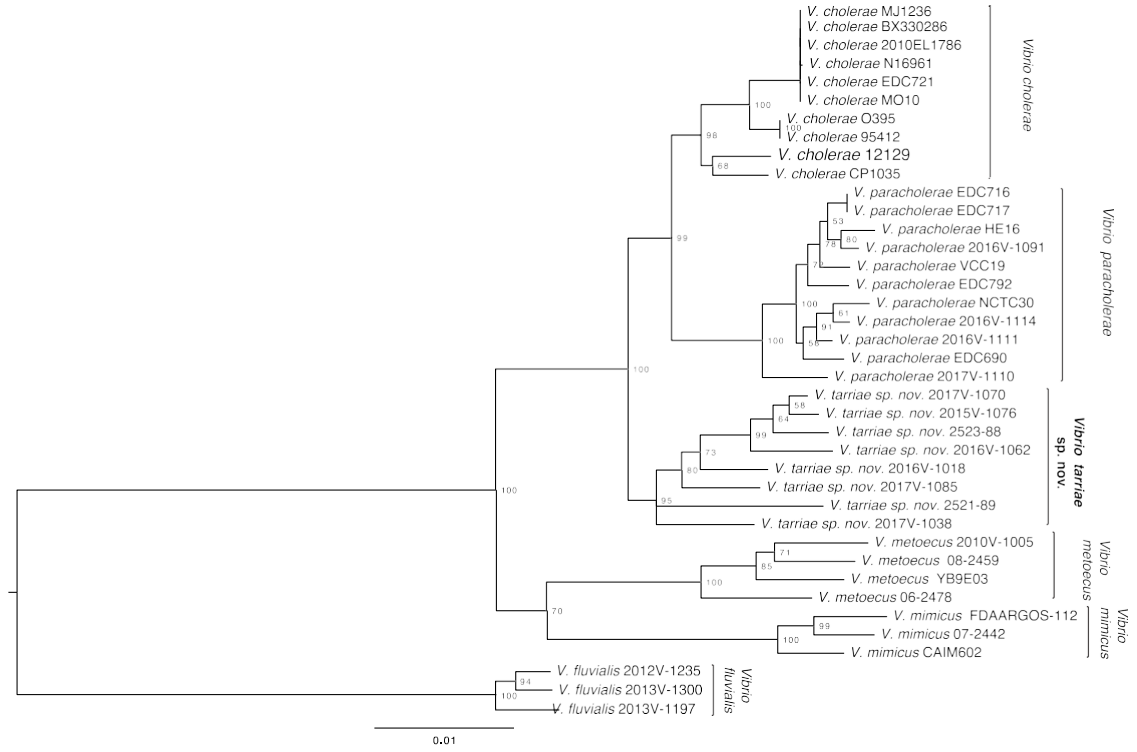


Fig. 5.2. Phylogenetic relationship of *V. tarriae* sp. nov. and its closest relatives based on a concatenated alignment of DNA sequences of six protein coding housekeeping genes: *adk*, *gyrB* , *pyrH*, *pgi*, *recA* and *rpoA*. Neighbour-Joining method was used for making the tree using Tamua-Nei substitution model with 100 bootstrap replicates. Bootstrap values are shown in the nodes of relevant branches.

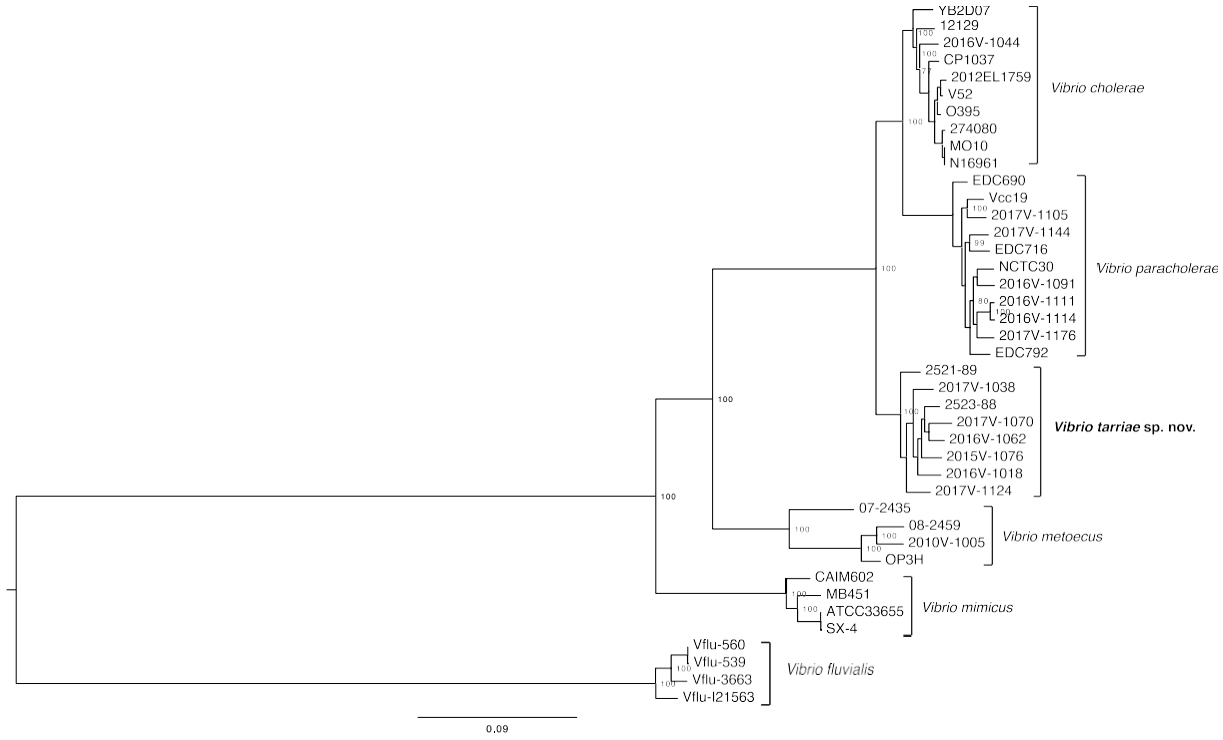


Fig. 5.3. Core genome phylogenetic tree of *V. tarriae* sp. nov. along with its closest sister species. The maximum likelihood phylogenetic tree was constructed from the core genome alignment of ≈2.1M bp using GTR gamma substitution model. Corresponding nodes with relevant Bootstrap support over 70% from the 100 replicates were indicated. The scale bar represents nucleotide substitutions per site.

5.4.3. Gene content analysis:

Gene content analysis revealed a few species-specific genetic traits in *V. tarriae* sp. nov., which differentiates it from *V. cholerae* and *V. paracholerae* (Table 5.3). There were at least 11 gene clusters / islands which were present in all nine *V. tarriae* sp. nov. strains but absent in all *V. cholerae* (n=22) and *V. paracholerae* (n=22) strains. Noticeably, the siderophore system has some conserved differences in *V. tarriae* sp. nov. compared with the other two most closely related species. Genes for aerobactin (CEQ48_09230-CEQ48_09275) common to *V. metoecus* and *V. mimicus* are al s o present in *V. tarriae* sp. nov. but absent in *V. cholerae* and *V. paracholerae.* On the other hand, *V. tarriae* sp. nov. strains lack genes for the vibriobactin siderophore system present in *V. cholerae* and *V. paracholerae.* Production of unique

siderophores can provide competitive advantage to a species from its neighbors in a shared habitat (Wyckoff et al., 2007). Also, possession of additional varieties of siderophore receptors can be useful in scavenging iron captured by siderophores secreted by other bacteria in an iron limiting condition (Thode et al., 2018).

There are some signatures of divergence from *V. cholerae* and *V. paracholerae* in the ion transport system as well. A gene cluster (CEQ48_14570-CEQ48_14585) of ~4.0 kb found in all *V. tarriae* sp. nov. strains is not present in any of its close relatives. The genes found in this cluster have homologs in *V. Vulnificus* with 79% nucleotide identity. Gene sequence similarity suggests that this biosynthetic gene cluster encodes putative aminobenzoyl-glutamate family (AbgT) transporter system which could have been acquired via horizontal transfer by *V. tarriae* sp. nov. A few other gene clusters associated with sodium transport system (CEQ48_11160-CEQ48_11162), peptidase dipeptidase system (CEQ48_10265-CEQ48_10295) and an MFS transporter (CEQ48_05115-CEQ48_05190) are present in *V. tarriae* sp. nov., *V. metoecus* and *V. mimicus* but absent in *V. cholerae* and *V. paracholerae.* These acquired transporters can be important mediators for the diversification of substrate utilization among these sister species. *V. tarriae* sp. nov. isolates possess gene cluster for cytochrome C nitrite reductase (CEQ48_12450-CEQ48_12457) which is missing in *V. cholerae*. Nitrite reductase have been thought to be required for growth in anoxic conditions using nitrate reduction (Bueno et al., 2018). Seven out of nine isolates of *V. tarriae* sp. nov. analyzed in this study were isolated from human clinical samples (Table 5.1), which suggests that the species is a opportunistic pathogens to human. However, the pathogenesis and prognosis of the disease caused could not be clearly defined. Genetic differences from it's well known close relative *V. cholerae* could be important clues to find their mechanism of virulence and pathogenesis as a potential human pathogen.

Table 5.3. Major genetic traits differentiating *Vibrio tarriae* sp. nov. from its closest relatives: *Vibrio cholerae, Vibrio paracholerae and Vibrio metoecus.* Reference genomes N16961 (*V. cholerae*) and 2521-89 (*V. paracholerae* sp. nov.) were used for determining locus positions of the gene clusters.

Gene cluster	Locus in reference genome (NZ_CP022353.1)	Locus in reference genome (AE003852.1)	*V. tarriae* sp. nov.	*V. cholerae*	*V. paracholerae*	*V. metoecus*	Predicted function
Siderophore aerobactin	CEQ48_09230-CEQ48_09275	Missing	+	-	-	+	Iron regulation
Peptidase-dipeptidase	CEQ48_10265-CEQ48_10295	Missing	+	-	-	+	Ion transport
Sodium transporter	CEQ48_11160-CEQ48_11162	Missing	+	-	-	+	Sodium transport
Cytochrome C nitrite reductase	CEQ48_12450-CEQ48_12457	Missing	+	-	-	+	Nitrite reduction
S-formylglutathione hydrolase	CEQ48_14570-CEQ48_14583	Missing	+	-	-	-	Ion transport
Chemotaxis protein	CEQ48_18370-CEQ48_18386	Missing	+	-	-	+	Chemotaxis
Metallophosphoesterase	CQ48_02995-CEQ48_03040	Missing	+	-	-	-	Ion transport
Chemotaxis protein	CEQ48_03175-CEQ48_03235	Missing	+	-	-	+	Chemotaxis
MFS transporter	CEQ48_05115-CEQ48_05190	Missing	+	-	-	+	Ion transport
Vibriobactin siderophore	Missing	VC0771-VC0780	-	+	+	-	Iron acquisition
Phosphoenol pyruvate synthase (PEP)	Missing	VCA03609-VCA03612	-	+	+	-	Pyruvate metabolism
Na(+)/H(+) antiporter subunit	Missing	VCA02875-VCA02882	-	+	+	-	Sodium transport

Description of the novel species:

Vibrio tarriae [tarr'i.ae. N.L. gen. fem. n. tarriae of tarr., named after Cheryl Tarr, in recognition of her contribution to the study of *Vibrio* diversity and exploration of atypical vibrios]

Cells are Gram-negative, curved, motile rods, 0.64–0.78×1.48–1.68 μm in size, which produce convex, smooth, circular, entire, cream colonies on TSB agar and yellow sucrose fermenting *V. cholerae* like colonies on TCBS agar. Growth is observed in TSB at 30 °C with salt concentrations in the range of 0–6.5 % NaCl; no growth occurs in the presence of 10 % NaCl. Growth is also observed in TSB with 1.5 % total NaCl concentration at a temperature range of 30–40 °C, and no growth occurs at 4 °C and 45 °C. The ability to utilize D-lactose, *N*-acetyl-D-glucosamine and pectin as well as the lack of ability to utilize D-mannose, D-serine, citric acid, D-ribose, propionic acid, D-alanin, mono-methyl succinate and caproic acid as the sole carbon and energy source differentiates *V. tarriae* from *V. cholerae.* On the other hand, *V. tarriae* can be differentiated from *V. paracholerae* by their ability to utilize N-acetyl-D-glucosamine and inability to utilize α-cyclodextrin as the sole carbon and energy source.

V. tarriae can produce indole and β-glucosidase as well as reduce nitrate to nitrite, positive for glucose fermentation, lysine decarboxylase and ornithine decarboxylase. Produces acid from the fermentation of mannitol but not arabinose. Arginine dihydrolase- and urease-negative. Forms V. cholerae-like yellow circular colonies on TCBS agar and flat, smooth, circular colonies of

creamy-white colour on TSB agar. Positive for carbon utilization from a-D-glucose, β-galactose, citrate, D-fructose, D-glucoronic acid, D-maltose, D-trehalose, dextrin, gelatin, N-acetyl-D-galactosamine, sucrose, D-lactose and pectin. Diversity between strains exists for utilization of L-asparagine, L-glutamine, Glycyl-L-aspartic acid, N-Acetyl-D-Glucosamine, Glycerol, D,L-malic acid, D-Glucose-1-Phosphate, L-Serine, L-Threonine, Succinic Acid, D-Glucuronic Acid, Tween 40, D-Psicose, D-Galactose, D-Gluconic Acid, α-Keto-Glutaric Acid, α-Hydroxy Glutaric Acid-γ-Lactone, Bromo Succinic Acid, L-Alanyl-Glycine, L-Aspartic Acid, D,L-α -Glycerol-Phosphate, L-Proline, β-Methyl-D-Glucoside, N-Acetyl- β -D-Mannosamine, L-Galactonic Acid- γ -Lactone, Methyl Pyruvate, D-Mannitol, L-Glutamic Acid, thymidine, Gentiobiose, raffinose, Capric acid, salicin, D-glucosamine and laminarin.

The type strain 2521-89^T (=DSM112461^T =CCUG75318^T) was isolated from a lake of New Mexico in USA. The G+C content of the type strain, based on whole-genome sequencing, is 47.1 %.

5.5. References

Alam, M., Hasan, N. A., Sultana, M., Nair, G. B., Sadique, A., Faruque, A. S., et al. (2010). Diagnostic limitations to accurate diagnosis of cholera. *J Clin Microbiol* 48, 3918–3922. doi:10.1128/JCM.00616-10.

Angiuoli, S. V., and Salzberg, S. L. (2011). Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics*. doi:10.1093/bioinformatics/btq665.

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*.

doi:10.1186/1471-2164-9-75.

Baker-Austin, C., Oliver, J. D., Alam, M., Ali, A., Waldor, M. K., Qadri, F., et al. (2018). Vibrio spp. infections. *Nat. Rev. Dis. Prim.* doi:10.1038/s41572-018-0005-8.

Boyd, E. F., Carpenter, M. R., Chowdhury, N., Cohen, A. L., Haines-Menges, B. L., Kalburge, S. S., et al. (2015). Post-Genomic Analysis of Members of the Family Vibrionaceae. *Microbiol. Spectr.* 3, 10.1128/microbiolspec.VE-0009–2014. doi:10.1128/microbiolspec.VE-0009-2014.

Bueno, E., Sit, B., Waldor, M. K., and Cava, F. (2018). Anaerobic nitrate reduction divergently governs population expansion of the enteropathogen *Vibrio cholerae*. *Nat. Microbiol.* 3, 1346–1353. doi:10.1038/s41564-018-0253-0.

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* doi:10.1038/srep24373.

Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., et al. (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. doi:10.1073/pnas.0907787106.

Colwell, R. R. (2004). Infectious disease and environment: cholera as a paradigm for waterborne disease. *Int Microbiol* 7, 285–289.

Farmer III, J. J., Michael Janda, J., Brenner, F. W., Cameron, D. N., and Birkhead, K. M. (2015). "Vibrio," in *Bergey's Manual of Systematics of Archaea and Bacteria* (American Cancer Society), 1–79. doi:https://doi.org/10.1002/9781118960608.gbm01078.

Haley, B. J., Grim, C. J., Hasan, N. A., Choi, S. Y., Chun, J., Brettin, T. S., et al. (2010). Comparative genomic analysis reveals evidence of two novel Vibrio species closely related to V. cholerae. *BMC Microbiol.* doi:10.1186/1471-2180-10-154.

Hasan, N. A., Grim, C. J., Haley, B. J., Chun, J., Alam, M., Taviani, E., et al. (2010). Comparative genomics of clinical and environmental *Vibrio mimicus*. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1013825107.

Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae* . *Nature* 406, 477–483. doi:10.1038/35020000.

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.* doi:10.1128/aem.45.1.275-283.1983.

Islam, M. T., Alam, M., and Boucher, Y. (2017a). Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *Int. Microbiol.* doi:10.2436/20.1501.01.291.

Islam, M. T., Alam, M., and Boucher, Y. (2017b). Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *Int Microbiol* 20, 106–115. doi:10.2436/20.1501.01.291.

Islam, M. T., Liang, K., Im, M. S., Winkjer, J., Busby, S., Tarr, C. L., et al. (2018). Draft Genome Sequences of Nine Vibrio sp. Isolates from across the United States Closely Related to *Vibrio cholerae*. *Microbiol Resour Announc* 7. doi:10.1128/MRA.00965-18.

Islam, M. T., Nasreen, T., Kirchberger, P., Liang, K. Y. H., Orata, F. D., Johura, F.-T., et al.

Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp. nov.) with a history of association with humans. *Appl. Environ. Microbiol.* 0, AEM.00422-21. doi:10.1128/AEM.00422-21.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. doi:10.1093/bioinformatics/bts199.

Kirchberger, P. C., Orata, F. D., Nasreen, T., Kauffman, K. M., Tarr, C. L., Case, R. J., et al. (2020). Culture-independent tracking of *Vibrio cholerae* lineages reveals complex spatiotemporal dynamics in a natural population. *Environ. Microbiol.* doi:10.1111/1462-2920.14921.

Kirchberger, P. C., Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., et al. (2014). Vibrio metoecus sp. nov., a close relative of *Vibrio cholerae* isolated from coastalbrackish ponds and clinical specimens. *Int. J. Syst. Evol. Microbiol.* 64, 3208–3214. doi:10.1099/ijs.0.060145-0.

Liang, K., Islam, M. T., Hussain, N., Winkjer, N. S., Im, M. S., Rowe, L. A., et al. (2019). Draft Genome Sequences of Eight *Vibrio* sp. Clinical Isolates from across the United States That Form a Basal Sister Clade to *Vibrio cholerae*. *Microbiol. Resour. Announc.* 8, e01473-18. doi:10.1128/MRA.01473-18.

Liang, K., Orata, F. D., Winkjer, N. S., Rowe, L. A., Tarr, C. L., and Boucher, Y. (2017). Complete Genome Sequence of Vibrio sp. Strain 2521-89, a Close Relative of *Vibrio cholerae* Isolated from Lake Water in New Mexico, USA. *Genome Announc.* 5.

doi:10.1128/genomeA.00905-17.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. doi:10.1186/1471-2105-14-60.

Oliver, J. D., Warner, R. A., and Cleland, D. R. (1983). Distribution of *Vibrio vulnificus* and Other Lactose-Fermenting Vibrios in the Marine Environment. *Appl. Environ. Microbiol.* 45, 985–998. doi:10.1128/aem.45.3.985-998.1983.

Orata, F. D., Kirchberger, P. C., Méheust, R., Jed Barlow, E., Tarr, C. L., and Boucher, Y. (2015). The dynamics of genetic interactions between Vibrio metoecus and *Vibrio cholerae*, two close relatives co-occurring in the environment. *Genome Biol. Evol.* doi:10.1093/gbe/evv193.

Orata, F. D., Xu, Y., Gladney, L. M., Rishishwar, L., Case, R. J., Boucher, Y., et al. (2016). Characterization of clinical and environmental isolates of Vibrio cidicii sp. nov., a close relative of Vibrio navarrensis. *Int. J. Syst. Evol. Microbiol.* 66, 4148–4155. doi:10.1099/ijsem.0.001327.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0906412106.

Stamatakis, A. (2014). RAxML. *Manual/tutorial*.

Thode, S. K., Rojek, E., Kozlowski, M., Ahmad, R., and Haugen, P. (2018). Distribution of siderophore gene systems on a Vibrionaceae phylogeny: Database searches, phylogenetic

analyses and evolutionary perspectives. *PLoS One* 13, 1–20. doi:10.1371/journal.pone.0191860.

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., et al. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl. Environ. Microbiol.* 71, 5107–5115. doi:10.1128/AEM.71.9.5107-5115.2005.

Wyckoff, E. E., Mey, A. R., and Payne, S. M. (2007). Iron acquisition in *Vibrio cholerae*. in *BioMetals* doi:10.1007/s10534-006-9073-4.

CHAPTER 6: Concluding remarks

CHAPTER 6

6.1. Brief summary

The research in this thesis aims to explore the natural diversity and population compositions of *Vibrio cholerae* and the clade containing species closely related to *V. cholerae* in the context of association to humans. Chapter 2 reassesses the theories and hypotheses regarding the emergence, dispersal and evolution of the pandemic generating (PG) lineage of *V.cholerae* in the light of recently available genome sequence data. In chapter 3, intra-species diversity of *V. cholerae* in a cholera endemic location is described and compared with that of a cholera free location. A combination of culture-based and culture independent techniques revealed a novel sister species co-existing with *V. cholerae* in the environment of cholera endemic Dhaka: *V. paracholerae*. The dynamics of genetic interactions between *V. cholerae* and *V. paracholerae* are also analyzed, and unique phenotypic and genetic properties that distinguish the two sister species are described. The potential of *V. paracholerae* as a human pathogen and an agent impacting the evolution of *V. cholerae* is also discussed. Chapter 4 investigates the impact of environmental factors on the population composition of *V. cholerae* and *V. paracholerae* in cholera endemic Dhaka. It is revealed that salinity is the most important factor to govern the population composition in that environmental setting and human population have a potential role to play in impacting the intra species diversity and competitions. Finally, Chapter 5 discovers and describes a novel sister species to *V. cholerae* isolated from both clinical environmental samples from a non-cholera endemic region (USA): *V. tarriae* sp. nov..

6.2. Potential impact of the findings:

6.2.1. Key points of the time and events for the emergence of PG *V. cholerae*

The case of PG *V. cholerae* can be considered as a prime example of human pathogen originating from an environmental population of bacteria and massively impacting humanity. All the known cholera pandemics have been caused by this single phylogenetic lineage. How and when the divergence of this lineage happened and what events led to its evolution, persistence and dispersal have been illusive so far. Recently, as a result of some extra-ordinary sampling efforts in cholera endemic and non-endemic regions, some intermediate lineages to the pandemic generating and environmental *V. cholerae* have been found. Hence, estimation of the time and pathway for the emergence of the pandemic generating *V. cholerae* lineage from the environmental ancestors have become more accurate. In chapter 2 of this thesis, the timelines and orders of the events leading to the emergence of pandemic generating lineage was estimated considering the newly available information and a consolidated pathway was reconstructed to explain the emergence, evolution and biogeographic dispersal of this very important lineage of *V. cholerae*. This knowledge will help deciphering the key events of the history of cholera pandemics and translate the observations into hypotheses to explain the emergence of pandemic causing agents in future.

6.2.2. Population dynamics of *V. cholerae* in cholera endemic region:

My research describes for the first time the intra-species diversity of *V. cholerae* in a cholera endemic region using a culture- independent method. In chapter 3, I have applied a recently developed technique of amplicon sequencing in conjunction with qPCR to describe the population composition and dynamics of *V. cholerae* in a natural setting. This will pave the way for the use of these techniques to analyze environmental populations of *V. cholerae* or similar bacteria. Chapter 3 compares the population compositions of *V. cholerae* in cholera-endemic and

cholera-free regions and finds out potential differences which would serve as baselines for future environmental surveillance of this group of bacteria.

6.2.3. *V. paracholerae*: a species potentially pathogenic to human and contributing to the evolution of *V. cholerae*:

Chapter 3 of this thesis connects the novel species *V. paracholerae* to a historical case during the second world war, where a 'cholera-like' bacteria was isolated from a choleraic diarrheal patient. It also identifies this globally existing bacteria as associated to humans. The history, biology and genomic information highlights the importance of this novel species as potential human pathogen. Being the closest known sister species to *V. cholerae* to date, *V. paracholerae* will be an important model for the study of evolution of the former. The study presents preliminary evidence of genetic interactions and potential impacts on the emergence of novel pathogenic groups of *V. cholerae* (i.e. O139), which underscores the importance of this novel species as a future study model.

6.2.4. The impact of global climate change on microbial populations

Vibrios are heavily influenced by the environment, and incidence of *vibrio* related human diseases have spiked in recent years, supposedly due to the rise of global temperature and related geo-climatic events. This study (chapter 4) looks at the potential environmental factors impacting the natural population of *V. cholerae* in cholera endemic Dhaka. In such a region, salinity could be the most important factor driving intraspecies and intra-genera dynamics. There has been growing consensus on the effect of salinity on the evolution of aquatic bacterial population. This

study sheds light on how climate change can impact the natural bacterial population even at a subspecies level. This knowledge will help assessing risk factors and strategies to minimize the effect of global climate change on lives on earth.

6.2.5. The phylogeny of the "Cholerae clade"

The study of *V. cholerae* has been very important in understanding human pathogens originating from the environment. The impact of cholera on our current knowledge of aquatic bacteriology has been enormous. Consequently, the phylogenetic clade *V. cholerae* fall in has gained special attention. However, there was ambiguity on the phylogenetic structure of the clade. With the two new sister clades: *V. paracholerae* (chapter 3) and *V. tarriae* sp. nov. (chapter 5), added to the phylogeny, the evolutionary context of the 'Cholerae clade' has become much clearer. Members of this clade are all capable of surviving in fresh water or low salt water, which is in contrast with most other species of the *Vibrio* genus. Evolution of this clade and its divergence from other clades of the *Vibrio* genus can be a fascinating field of study. This research provides insights on the commonalities and differences of these closely related sister species and a consolidated view of the clade which would propel future research on this exciting field.

REFERENCES

REFERENCES

Abd, H., Weintraub, A., and Sandström, G. (2004). Interaction between *Vibrio cholerae* and Acanthamoeba castellanii. *Microb. Ecol. Health Dis.* doi:10.1080/08910600410029190.

Akanda, A. S., Jutla, A. S., Gute, D. M., Sack, R. B., Alam, M., Huq, A., et al. Population Vulnerability to Biannual Cholera Outbreaks and Associated Macro-Scale Drivers in the Bengal Delta. *Am. Soc. Trop. Med. Hyg.* 89, 950–959. doi:10.4269/ajtmh.12-0492.

Alam, M., Hasan, N. A., Sadique, A., Bhuiyan, N. A., Ahmed, K. U., Nusrin, S., et al. (2006). Seasonal cholera caused by *Vibrio cholerae* serogroups O1 and O139 in the coastal aquatic environment of Bangladesh. *Appl Env. Microbiol* 72, 4096–4104. doi:10.1128/AEM.00066-06.

Alam, M., Hasan, N. A., Sultana, M., Nair, G. B., Sadique, A., Faruque, A. S., et al. (2010a). Diagnostic limitations to accurate diagnosis of cholera. *J Clin Microbiol* 48, 3918–3922. doi:10.1128/JCM.00616-10.

Alam, M., Islam, A., Bhuiyan, N., Rahim, N., Hossain, A., Khan, G. Y., et al. (2011). Clonal transmission, dual peak, and off-season cholera in Bangladesh. *Infect. Ecol. Epidemiol.* doi:10.3402/iee.v1i0.7273.

Alam, M., Islam, M. T., Rashed, S. M., Johura, F., Bhuiyan, N. A., Delgado, G., et al. (2012). *Vibrio cholerae* classical biotype strains reveal distinct signatures in Mexico. *J. Clin. Microbiol.* 50, 2212–2216. doi:10.1128/JCM.00189-12.

Alam, M., Nusrin, S., Islam, A., Bhuiyan, N. A., Rahim, N., Delgado, G., et al. (2010b). 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. doi:10.1128/JCM.00866-

10.

Alam, M., Sultana, M., Balakrish Nair, G., Siddique, A. K., Hasan, N. A., Bradley Sack, R., et al. (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.* doi:10.1073/pnas.0705599104.

Aldunate, M., De la Iglesia, R., Bertagnolli, A. D., and Ulloa, O. (2018). Oxygen modulates bacterial community composition in the coastal upwelling waters off central Chile. *Deep. Res. Part II Top. Stud. Oceanogr.* doi:10.1016/j.dsr2.2018.02.001.

Ali, M., Lopez, A. L., Ae You, Y., Eun Kim, Y., Sah, B., Maskery, B., et al. (2012). The global burden of cholera. *Bull. World Health Organ.* doi:10.2471/blt.11.093427.

Almagro-Moreno, S., and Taylor, R. K. (2013). Cholera: Environmental Reservoirs and Impact on Disease Transmission. *Microbiol. Spectr.* 1. doi:10.1128/microbiolspec.OH-0003-2012.

Angiuoli, S. V., and Salzberg, S. L. (2011). Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics*. doi:10.1093/bioinformatics/btq665.

Arnosti, C. (2014). Patterns of Microbially Driven Carbon Cycling in the Ocean: Links between Extracellular Enzymes and Microbial Communities. *Adv. Oceanogr.* doi:10.1155/2014/706082.

Auch, A. F., Klenk, H. P., and Göker, M. (2010a). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genomic Sci.* doi:10.4056/sigs.541628.

Auch, A. F., von Jan, M., Klenk, H. P., and Goker, M. (2010b). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2, 117–134. doi:10.4056/sigs.531120.

Azarian, T., Ali, A., Johnson, J. A., Jubair, M., Cella, E., Ciccozzi, M., et al. (2016). Non-toxigenic environmental *Vibrio cholerae* O1 strain from Haiti provides evidence of pre-pandemic cholera in Hispaniola. *Sci. Rep.* 6, 36115. doi:10.1038/srep36115.

Azarian, T., Ali, A., Johnson, J. A., Mohr, D., Prosperi, M., Veras, N. M., et al. (2014). Phylodynamic Analysis of Clinical and Environmental *Vibrio cholerae* Isolates from Haiti Reveals Diversification Driven by Positive Selection. *MBio* 5, e01824-14. doi:10.1128/mBio.01824-14.

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*. doi:10.1186/1471-2164-9-75.

Bachmann, V., Kostiuk, B., Unterweger, D., Diaz-Satizabal, L., Ogg, S., and Pukatzki, S. (2015). Bile Salts Modulate the Mucin-Activated Type VI Secretion System of Pandemic *Vibrio cholerae*. *PLoS Negl. Trop. Dis.* 9, e0004031–e0004031. doi:10.1371/journal.pntd.0004031.

Baharoglu, Z., Bikard, D., and Mazel, D. (2010). Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet.* 6, e1001165–e1001165. doi:10.1371/journal.pgen.1001165.

Baker-Austin, C., Oliver, J. D., Alam, M., Ali, A., Waldor, M. K., Qadri, F., et al. (2018). Vibrio spp. infections. *Nat. Rev. Dis. Prim.* doi:10.1038/s41572-018-0005-8.

Baker-Austin, C., Trinanes, J., Gonzalez-Escalona, N., and Martinez-Urtaza, J. (2017). Non-Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends Microbiol.* 25, 76–84. doi:10.1016/j.tim.2016.09.008.

Barcat, J. A. (2014). [Filippo Pacini and cholera, 1854]. *Medicina (B. Aires).* 74, 77–79.

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

428.

Bina, X. R., Provenzano, D., Nguyen, N., and Bina, J. E. (2008). *Vibrio cholerae* RND family efflux systems are required for antimicrobial resistance, optimal virulence factor production, and colonization of the infant mouse small intestine. *Infect. Immun.* doi:10.1128/IAI.01620-07.

Bishop-Lilly, K. A., Johnson, S. L., Verratti, K., Luu, T., Khiani, A., Awosika, J., et al. (2014). Genome sequencing of 15 clinical Vibrio isolates, including 13 non-o1/non-o139 serogroup strains. *Genome Announc* 2. doi:10.1128/genomeA.00893-14.

Blackwell, K. D., and Oliver, J. D. (2008). The ecology of Vibrio vulnificus, *Vibrio cholerae*, and Vibrio parahaemolyticus in North Carolina Estuaries. *J. Microbiol.* doi:10.1007/s12275-007-0216-2.

Bochner, B. R., Gadzinski, P., and Panomitros, E. (2001). Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* doi:10.1101/gr.186501.

Boucher, Y. (2016). Sustained Local Diversity of *Vibrio cholerae* O1 Biotypes in a Previously Cholera-Free Country. *MBio* 7, e00570-16. doi:10.1128/mBio.00570-16.

Boucher, Y., Orata, F. D., and Alam, M. (2015a). The out-of-the-delta hypothesis: dense human populations in low-lying river deltas served as agents for the evolution of a deadly pathogen. *Front Microbiol* 6, 1120. doi:10.3389/fmicb.2015.01120.

Boucher, Y., Orata, F. D., and Alam, M. (2015b). The out-of-the-delta hypothesis: Dense human populations in low-lying river deltas served as agents for the evolution of a deadly pathogen. *Front. Microbiol.* doi:10.3389/fmicb.2015.01120.

Boyd, E. F., Carpenter, M. R., Chowdhury, N., Cohen, A. L., Haines-Menges, B. L., Kalburge, S.

S., et al. (2015). Post-Genomic Analysis of Members of the Family Vibrionaceae. *Microbiol. Spectr.* 3, 10.1128/microbiolspec.VE-0009–2014. doi:10.1128/microbiolspec.VE-0009-2014.

Brayton, P. R., Bode, R. B., Colwell, R. R., MacDonell, M. T., Hall, H. L., Grimes, D. J., et al. (1986). Vibrio cincinnatiensis sp. nov., a new human pathogen. *J. Clin. Microbiol.* doi:10.1128/jcm.23.1.104-108.1986.

Bueno, E., Pinedo, V., and Cava, F. (2020). Adaptation of *Vibrio cholerae* to Hypoxic Environments. *Front. Microbiol.* doi:10.3389/fmicb.2020.00739.

Bueno, E., Sit, B., Waldor, M. K., and Cava, F. (2018). Anaerobic nitrate reduction divergently governs population expansion of the enteropathogen *Vibrio cholerae*. *Nat. Microbiol.* 3, 1346–1353. doi:10.1038/s41564-018-0253-0.

Burrus, V., Quezada-Calvillo, R., Marrero, J., and Waldor, M. K. (2006). SXT-Related Integrating Conjugative Element in New World *Vibrio cholerae*. *Appl. Environ. Microbiol.* 72, 3054–3057. doi:10.1128/AEM.72.4.3054-3057.2006.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. doi:10.1038/nmeth.3869.

Chatterjee, S. N., and Chaudhuri, K. (2003). Lipopolysaccharides of *Vibrio cholerae*: I. Physical and chemical characterization. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1639, 65–79. doi:10.1016/j.bbadis.2003.08.004.

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* doi:10.1038/srep24373.

Chiang, S.-R., and Chuang, Y.-C. (2003). Vibrio vulnificus infection: clinical manifestations,

pathogenesis, and antimicrobial therapy. *J. Microbiol. Immunol. Infect.* 36, 81–88.

Choi, S. Y., Rashed, S. M., Hasan, N. A., Alam, M., Islam, T., Sadique, A., et al. (2016). Phylogenetic diversity of *Vibrio cholerae* associated with endemic cholera in Mexico from 1991 to 2008. *MBio*. doi:10.1128/mBio.02160-15.

Chowdhury, M. A. R., Xu, B., Montilla, R., Hasan, J. A. K., Huq, A., and Colwell, R. R. (1995). A simplified immunofluorescence technique for detection of viable cells of *Vibrio cholerae* O1 and O139. *J. Microbiol. Methods*. doi:10.1016/0167-7012(95)00066-6.

Chun, J., Grim, C. J., Hasan, N. A., Je, H. L., Seon, Y. C., Haley, B. J., et al. (2009a). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0907787106.

Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., et al. (2009b). 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. doi:10.1073/pnas.0907787106.

Chun, J., Huq, A., and Colwell, R. R. (1999). Analysis of 16S-23S rRNA intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. *Appl. Environ. Microbiol.* doi:10.1128/aem.65.5.2202-2208.1999.

Citil, B. E., Derin, S., Sankur, F., Sahan, M., and Citil, M. U. (2015). *Vibrio alginolyticus* Associated Chronic Myringitis Acquired in Mediterranean Waters of Turkey. *Case Rep. Infect. Dis.* 2015, 187212. doi:10.1155/2015/187212.

Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., et al. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol* 68, 2386–2392. doi:10.1099/ijsem.0.002809.

Colwell, R. R. (1970). Polyphasic taxonomy of the genus vibrio: numerical taxonomy of *Vibrio cholerae*, Vibrio parahaemolyticus, and related Vibrio species. *J. Bacteriol.* doi:10.1128/jb.104.1.410-433.1970.

Colwell, R. R. (2000). Viable but nonculturable bacteria: A survival strategy. *J. Infect. Chemother.* doi:10.1007/PL00012151.

Colwell, R. R. (2004a). Infectious disease and environment: cholera as a paradigm for waterborne disease. *Int Microbiol* 7, 285–289.

Colwell, R. R. (2004b). Infectious disease and environment: Cholera as a paradigm for waterborne disease. *Int. Microbiol.* doi:10.2436/im.v7i4.9485.

Conner, J. G., Teschler, J. K., Jones, C. J., and Yildiz, F. H. (2016). Staying Alive: *Vibrio cholerae*'s Cycle of Environmental Survival, Transmission, and Dissemination. *Microbiol. Spectr.* 4. doi:10.1128/microbiolspec.VMBF-0015-2015.

Constantin de Magny, G., Murtugudde, R., Sapiano, M. R. P., Nizam, A., Brown, C. W., Busalacchi, A. J., et al. (2008). Environmental signatures associated with cholera epidemics. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17676–17681. doi:10.1073/pnas.0809654105.

Coyne, J. A. (1992). Genetics and speciation. *Nature* 355, 511–515. doi:10.1038/355511a0.

Dalia, A. B., Seed, K. D., Calderwood, S. B., and Camilli, A. (2015). A globally distributed mobile genetic element inhibits natural transformation of *Vibrio cholerae*. *Proc. Natl. Acad. Sci.* 112,10485–10490. doi:10.1073/pnas.1509097112.

Davis, B. R., Fanning, G. R., Madden, J. M., Steigerwalt, A. G., Bradford, H. B., Smith, H. L., et al. (1981). Characterization of biochemically atypical *Vibrio cholerae* strains and designation of a new pathogenic species, *Vibrio mimicus*. *J. Clin. Microbiol.* doi:10.1128/jcm.14.6.631-639.1981.

DAVIS, G. H., and PARK, R. W. (1962). A taxonomic study of certain bacteria currently classified as Vibrio species. *J. Gen. Microbiol.* doi:10.1099/00221287-27-1-101.

Deen, J., Mengel, M. A., and Clemens, J. D. (2020). Epidemiology of cholera. *Vaccine*. doi:10.1016/j.vaccine.2019.07.078.

Dorman, M. J., Kane, L., Domman, D., Turnbull, J. D., Cormie, C., Fazal, M. A., et al. (2019). The history, genome and biology of NCTC 30: A non-pandemic *Vibrio cholerae* isolate from World War One. *Proc. R. Soc. B Biol. Sci.* doi:10.1098/rspb.2018.2025.

Dziejman, M., Serruto, D., Tam, V. C., Sturtevant, D., Diraphat, P., Faruque, S. M., et al. (2005). Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0409918102.

Edgar, C. R. C. (2013). USEARCH. *J. Natl. Cancer Inst.*

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. doi:10.1186/1476-072X-7-31.

Escobar, L. E., Ryan, S. J., Stewart-Ibarra, A. M., Finkelstein, J. L., King, C. A., Qiao, H., et al. (2015). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Trop.* 149, 202–211. doi:10.1016/j.actatropica.2015.05.028.

Fakruddin, M., Mannan, K. S. Bin, and Andrews, S. (2013). Viable but Nonculturable Bacteria: Food Safety and Public Health Perspective. *ISRN Microbiol.* 2013, 703813. doi:10.1155/2013/703813.

Farmer III, J. J., Michael Janda, J., Brenner, F. W., Cameron, D. N., and Birkhead, K. M. (2015). "Vibrio," in *Bergey's Manual of Systematics of Archaea and Bacteria* (American Cancer Society), 1–79. doi:https://doi.org/10.1002/9781118960608.gbm01078.

Farmer, J. J. (2006). "The Family Vibrionaceae," in *The Prokaryotes: A Handbook on the Biology*

of Bacteria Volume 6: Proteobacteria: Gamma Subclass, eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (New York, NY: Springer New York), 495–507. doi:10.1007/0-387-30746-X_17.

Faruque, S. M., Albert, M. J., and Mekalanos, J. J. (1998). Epidemiology, Genetics, and Ecology of Toxigenic*Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* doi:10.1128/mmbr.62.4.1301-1314.1998.

Faruque, S. M., Sack, D. A., Sack, R. B., Colwell, R. R., Takeda, Y., and Nair, G. B. (2003). Emergence and evolution of *Vibrio cholerae* O139. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0337468100.

Fiedler, G., Pajatsch, M., and Böck, A. (1996). Genetics of a novel starch utilisation pathway present in Klebsiella oxytoca. *J. Mol. Biol.* doi:10.1006/jmbi.1996.0085.

Franco, A., Rückert, C., Blom, J., Busche, T., Reichert, J., Schubert, P., et al. (2020). High diversity of Vibrio spp. associated with different ecological niches in a marine aquaria system and description of Vibrio aquimaris sp. nov. *Syst. Appl. Microbiol.* doi:10.1016/j.syapm.2020.126123.

Froelich, B. A., and Daines, D. A. (2020). In hot water: effects of climate change on Vibrio–human interactions. *Environ. Microbiol.* 22, 4101–4111. doi:https://doi.org/10.1111/1462-2920.14967.

Fu, Y., Waldor, M. K., and Mekalanos, J. J. (2013). Tn-Seq analysis of *Vibrio cholerae* intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe* 14, 652–663. doi:10.1016/j.chom.2013.11.001.

Galanis, E., Otterstatter, M., and Taylor, M. (2020). Measuring the impact of sea surface temperature on the human incidence of Vibrio sp. infection in British Columbia, Canada,

1992–2017. *Environ. Heal.* 19, 58. doi:10.1186/s12940-020-00605-x.

Gardner, A. D., and Venkatraman, K. V. (1935). The antigens of the cholera group of vibrios. *J. Hyg. (Lond).* doi:10.1017/S0022172400032265.

Garrine, M., Mandomando, I., Vubil, D., Nhampossa, T., Acacio, S., Li, S., et al. (2017). Minimal genetic change in *Vibrio cholerae* in Mozambique over time: Multilocus variable number tandem repeat analysis and whole genome sequencing. *PLoS Negl Trop Dis* 11, e0005671. doi:10.1371/journal.pntd.0005671.

Georgiades, K., and Raoult, D. (2011). Defining pathogenic bacterial species in the genomic era. *Front. Microbiol.* 1, 151. doi:10.3389/fmicb.2010.00151.

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* doi:10.1038/nrg.2016.49.

Gregory, G. J., and Boyd, E. F. (2021). Stressed out: Bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from Vibrionaceae. *Comput. Struct. Biotechnol. J.* 19, 1014–1027. doi:https://doi.org/10.1016/j.csbj.2021.01.030.

Hafiza Khatun, Nishat Falgunee, M. J. R. K. (2017). Analyzing urban population density gradient of Dhaka Metropolitan. *Geogr. OnlineTM Malaysian J. Soc. Sp.* 11, 1–13.

Haley, B. J., Grim, C. J., Hasan, N. A., Choi, S. Y., Chun, J., Brettin, T. S., et al. (2010). Comparative genomic analysis reveals evidence of two novel Vibrio species closely related to V. cholerae. *BMC Microbiol.* doi:10.1186/1471-2180-10-154.

Hasan, M. K., Shahriar, A., and Jim, K. U. (2019). Water pollution in Bangladesh and its impact on public health. *Heliyon*. doi:10.1016/j.heliyon.2019.e02145.

Hasan, N. A., Grim, C. J., Haley, B. J., Chun, J., Alam, M., Taviani, E., et al. (2010). Comparative genomics of clinical and environmental *Vibrio mimicus*. *Proc. Natl. Acad. Sci. U. S. A.*

doi:10.1073/pnas.1013825107.

Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae* . *Nature* 406, 477–483. doi:10.1038/35020000.

Hickman, F. W., Farmer, J. J., Hollis, D. G., Fanning, G. R., Steigerwalt, A. G., Weaver, R. E., et al. (1982). Identification of Vibrio hollisae sp. nov. from patients with diarrhea. *J. Clin. Microbiol.* 15, 395–401. doi:10.1128/jcm.15.3.395-401.1982.

Hu, D., Liu, B., Feng, L., Ding, P., Guo, X., Wang, M., et al. (2016). Origins of the current seventh cholera pandemic. *Proc. Natl. Acad. Sci. U. S. A.* 113, E7730–E7739. doi:10.1073/pnas.1608732113.

Huang, Z., Yu, K., Fang, Y., Dai, H., Cai, H., Li, Z., et al. (2020). Comparative Genomics and Transcriptomics Analyses Reveal a Unique Environmental Adaptability of Vibrio fujianensis. *Microorganisms* 8, 555. doi:10.3390/microorganisms8040555.

Hunt, D. E., Gevers, D., Vahora, N. M., and Polz, M. F. (2008). Conservation of the chitin utilization pathway in the Vibrionaceae. *Appl. Environ. Microbiol.* 74, 44–51. doi:10.1128/AEM.01412-07.

Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., et al. (2005). Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* doi:10.1128/AEM.71.8.4645-4654.2005.

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.* doi:10.1128/aem.45.1.275-283.1983.

Huq, A., Yunus, M., Sohel, S. S., Bhuiya, A., Emch, M., Luby, S. P., et al. (2010). Simple sari

cloth filtration of water is sustainable and continues to protect villagers from cholera in Matlab, Bangladesh. *MBio* 1. doi:10.1128/mBio.00034-10.

Hussain, N. A. S., Kirchberger, P. C., Case, R. J., and Boucher, Y. F. (2021). Modular Molecular Weaponry Plays a Key Role in Competition Within an Environmental *Vibrio cholerae* Population. *Front. Microbiol.* 12, 671092. doi:10.3389/fmicb.2021.671092.

Islam, M. S. (1990). Effect of various biophysicochemical conditions on toxigenicity of *Vibrio cholerae* 01 during survival with a green alga, Rhizoclonium fontanum, in an artificial aquaticenvironment. *Can. J. Microbiol.* 36, 464–468. doi:10.1139/m90-081.

Islam, M. T., Alam, M., and Boucher, Y. (2017a). Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *Int Microbiol* 20, 106–115. doi:10.2436/20.1501.01.291.

Islam, M. T., Alam, M., and Boucher, Y. (2017b). Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *Int. Microbiol.* doi:10.2436/20.1501.01.291.

Islam, M. T., Liang, K., Im, M. S., Winkjer, J., Busby, S., Tarr, C. L., et al. (2018). Draft Genome Sequences of Nine Vibrio sp. Isolates from across the United States Closely Related to *Vibrio cholerae*. *Microbiol Resour Announc* 7. doi:10.1128/MRA.00965-18.

Islam, M. T., Nasreen, T., Kirchberger, P., Liang, K. Y. H., Orata, F. D., Johura, F.-T., et al. Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp. nov.) with a history of association with humans. *Appl. Environ. Microbiol.* 0, AEM.00422-21. doi:10.1128/AEM.00422-21.

Jäckel, C., Hammerl, J. A., Arslan, H.-H.-T., Göllner, C., Vom Ort, N., Taureck, K., et al. (2020). Phenotypic and Genotypic Characterization of Veterinary Vibrio cincinnatiensis Isolates. *Microorganisms* 8, 739. doi:10.3390/microorganisms8050739.

Jiang, S. C., and Fu, W. (2001). Seasonal abundance and distribution of *Vibrio cholerae* in coastal waters quantified by a 16S-23S intergenic spacer probe. *Microb. Ecol.* doi:10.1007/s00248-001-0029-5.

Jin, Y., Zhou, J., Zhou, J., Hu, M., Zhang, Q., Kong, N., et al. (2020). Genome-based classification of Burkholderia cepacia complex provides new insight into its taxonomic status. *Biol. Direct* 15, 6. doi:10.1186/s13062-020-0258-5.

Jolley, K. A., and Maiden, M. C. J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. doi:10.1186/1471-2105-11-595.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993. doi:10.1038/nature06536.

Jutla, A., Whitcombe, E., Hasan, N., Haley, B., Akanda, A., Huq, A., et al. (2013). Environmental factors influencing epidemic cholera. *Am. J. Trop. Med. Hyg.* doi:10.4269/ajtmh.12-0721.

Kamp, H. D., Patimalla-Dipali, B., Lazinski, D. W., Wallace-Gadsden, F., and Camilli, A. (2013). Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *PLoS Pathog.* 9, e1003800. doi:10.1371/journal.ppat.1003800.

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

Katz, L. S., Petkau, A., Beaulaurier, J., Tyler, S., Antonova, E. S., Turnsek, M. A., et al. (2013). Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. *MBio* 4, e00398-13. doi:10.1128/mBio.00398-13.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. doi:10.1093/bioinformatics/bts199.

Kimes, N. E., Grim, C. J., Johnson, W. R., Hasan, N. A., Tall, B. D., Kothary, M. H., et al. (2012). Temperature regulation of virulence factors in the pathogen Vibrio coralliilyticus. *ISME J.* 6, 835–846. doi:10.1038/ismej.2011.154.

King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature* 454, 877–880. doi:10.1038/nature07084.

Kirchberger, P. C., Orata, F. D., Barlow, E. J., Kauffman, K. M., Case, R. J., Polz, M. F., et al. (2016a). A small number of phylogenetically distinct clonal complexes dominate a coastal *Vibrio cholerae* population. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01177-16.

Kirchberger, P. C., Orata, F. D., Barlow, E. J., Kauffman, K. M., Case, R. J., Polz, M. F., et al. (2016b). A Small Number of Phylogenetically Distinct Clonal Complexes Dominate a Coastal *Vibrio cholerae* Population. *Appl Env. Microbiol* 82, 5576–5586. doi:10.1128/AEM.01177-16.

Kirchberger, P. C., Orata, F. D., Nasreen, T., Kauffman, K. M., Tarr, C. L., Case, R. J., et al. (2020a). Culture-independent tracking of *Vibrio cholerae* lineages reveals complex spatiotemporal dynamics in a natural population. *Env. Microbiol*. doi:10.1111/1462-2920.14921.

Kirchberger, P. C., Orata, F. D., Nasreen, T., Kauffman, K. M., Tarr, C. L., Case, R. J., et al. (2020b). Culture-independent tracking of *Vibrio cholerae* lineages reveals complex spatiotemporal dynamics in a natural population. *Environ. Microbiol.* doi:10.1111/1462-2920.14921.

Kirchberger, P. C., Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., et al. (2014a). Vibrio metoecus sp. nov., a close relative of *Vibrio cholerae* isolated from coastal brackish ponds and clinical specimens. *Int. J. Syst. Evol. Microbiol.* 64, 3208–3214.

doi:10.1099/ijs.0.060145-0.

Kirchberger, P. C., Turnsek, M., Hunt, D. E., Haley, B. J., Colwell, R. R., Polz, M. F., et al. (2014b). Vibrio metoecus sp. nov., a close relative of *Vibrio cholerae* isolated from coastal brackish ponds and clinical specimens. *Int J Syst Evol Microbiol* 64, 3208–3214. doi:10.1099/ijs.0.060145-0.

Kirchberger, P. C., Unterweger, D., Provenzano, D., Pukatzki, S., and Boucher, Y. (2017). Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. *Sci. Rep.* doi:10.1038/srep45133.

Kirschner, A. K. T., Schlesinger, J., Farnleitner, A. H., Hornek, R., Süß, B., Golda, B., et al. (2008). Rapid Growth of Planktonic *Vibrio cholerae* Non-O1/Non-O139 Strains in a Large Alkaline Lake in Austria: Dependence on Temperature and Dissolved Organic Carbon Quality. *Appl. Environ. Microbiol.* 74, 2004–2015. doi:10.1128/AEM.01739-07.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01043-13.

Labbate, M., Orata, F. D., Petty, N. K., Jayatilleke, N. D., King, W. L., Kirchberger, P. C., et al. (2016). A genomic island in *Vibrio cholerae* with VPI-1 site-specific recombination characteristics contains CRISPR-Cas and type VI secretion modules. *Sci. Rep.* 6, 36891. doi:10.1038/srep36891.

Lawson, R. G., and Jurs, P. C. (1990). New Index for Clustering Tendency and Its Application to Chemical Problems. *J. Chem. Inf. Comput. Sci.* doi:10.1021/ci00065a010.

Le Roux, F., and Blokesch, M. (2018). Eco-evolutionary Dynamics Linked to Horizontal Gene

Transfer in Vibrios. *Annu. Rev. Microbiol.* 72, 89–110. doi:10.1146/annurev-micro-090817-062148.

Lee, K. M., Park, Y., Bari, W., Yoon, M. Y., Go, J., Kim, S. C., et al. (2012). Activation of cholera toxin production by anaerobic respiration of trimethylamine N-oxide in *Vibrio cholerae*. *J. Biol. Chem.* doi:10.1074/jbc.M112.394932.

LEE, J. V, SHREAD, P., FURNISS, A. L., and BRYANT, T. N. (1981). Taxonomy and Description of Vibrio fluvialis sp. nov. (Synonym Group F Vibrios, Group EF6). *J. Appl. Bacteriol.* 50, 73–94. doi:https://doi.org/10.1111/j.1365-2672.1981.tb00873.x.

León Robles, A., Acedo Félix, E., Gomez-Gil, B., Quiñones Ramírez, E. I., Nevárez-Martínez, M., and Noriega-Orozco, L. (2013). Relationship of aquatic environmental factors with the abundance of *Vibrio cholerae*, Vibrio parahaemolyticus, *Vibrio mimicus* and Vibrio vulnificus in the coastal area of Guaymas, Sonora, Mexico. *J. Water Health*. doi:10.2166/wh.2013.160.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*. doi:10.1093/bioinformatics/btl529.

Levade, I., Terrat, Y., Leducq, J.-B., Weil, A. A., Mayo-Smith, L. M., Chowdhury, F., et al. (2017). *Vibrio cholerae* genomic diversity within and between patients. *Microb. genomics* 3. doi:10.1099/mgen.0.000142.

Liang, K., Islam, M. T., Hussain, N., Winkjer, N. S., Im, M. S., Rowe, L. A., et al. (2019). Draft Genome Sequences of Eight *Vibrio* sp. Clinical Isolates from across the United States That Form a Basal Sister Clade to *Vibrio cholerae*. *Microbiol. Resour. Announc.* 8, e01473-18. doi:10.1128/MRA.01473-18.

Liang, K., Orata, F. D., Winkjer, N. S., Rowe, L. A., Tarr, C. L., and Boucher, Y. (2017). Complete

Genome Sequence of Vibrio sp. Strain 2521-89, a Close Relative of *Vibrio cholerae* Isolated from Lake Water in New Mexico, USA. *Genome Announc.* 5. doi:10.1128/genomeA.00905-17.

Liang, S. Y., and Messenger, N. (2018). Infectious Diseases After Hydrologic Disasters. *Emerg. Med. Clin. North Am.* 36, 835–851. doi:10.1016/j.emc.2018.07.002.

Lin, W., Fullner, K. J., Clayton, R., Sexton, J. A., Rogers, M. B., Calia, K. E., et al. (1999). Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the choleratoxin prophage. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.96.3.1071.

Lipp, E. K., Huq, A., and Colwell, R. R. (2002). Effects of global climate on infectious disease: The cholera model. *Clin. Microbiol. Rev.* doi:10.1128/CMR.15.4.757-770.2002.

Logar-Henderson, C., Ling, R., Tuite, A. R., and Fisman, D. N. (2019). Effects of large-scale oceanic phenomena on non-cholera vibriosis incidence in the United States: implications for climate change. *Epidemiol. Infect.* 147, e243–e243. doi:10.1017/S0950268819001316.

Longini, I. M., 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.* doi:10.1086/341206.

Lugo, M. R., and Merrill, A. R. (2015). The father, son and cholix toxin: The third member of the DT group mono-ADP-ribosyltransferase toxin family. *Toxins (Basel).* doi:10.3390/toxins7082757.

Lutz, C., Erken, M., Noorian, P., Sun, S., and McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. *Front. Microbiol.* doi:10.3389/fmicb.2013.00375.

MacIntyre, D. L., Miyata, S. T., Kitaoka, M., and Pukatzki, S. (2010). The *Vibrio cholerae* type

VI secretion system displays antimicrobial properties. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1012931107.

Mahmuduzzaman, M., Ahmed, Z. U., Nuruzzaman, A. K. M., and Ahmed, F. R. S. (2014). Causes of Salinity Intrusion in Coastal Belt of Bangladesh. *Int. J. Plant Res.*

Mansour, G., and Esseku, H. (2017). Situation analysis of the urban sanitation sector in Ghana. *WSUP - Water Sanit. Urban Poor*.

Matz, C., and Kjelleberg, S. (2005). Off the hook - How bacteria survive protozoan grazing. *Trends Microbiol.* doi:10.1016/j.tim.2005.05.009.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. doi:10.1186/1471-2105-14-60.

Meier-Kolthoff, J. P., Klenk, H.-P., and Göker, M. (2014). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol.* 64, 352–356. doi:https://doi.org/10.1099/ijs.0.056994-0.

Merrell, D. S., Butler, S. M., Qadri, F., Dolganov, N. A., Alam, A., Cohen, M. B., et al. (2002). Host-induced epidemic spread of the cholera bacterium. *Nature* 417, 642–645. doi:10.1038/nature00778.

Miyata, S. T., Kitaoka, M., Wieteska, L., Frech, C., Chen, N., and Pukatzki, S. (2010). The *Vibrio cholerae* Type VI Secretion System: Evaluating its Role in the Human Disease Cholera. *Front. Microbiol.* 1, 117. doi:10.3389/fmicb.2010.00117.

Mohammad Tarequl Islam, Tania Nasreen, Kevin Y. H. Liang, Fatema-Tuz Johura, Paul C. Kirchberger, Marzia Sultana, Rebecca J. Case, M. A. and Y. F. B. Dynamic Subspecies Population Structure of *Vibrio cholerae* in Dhaka, Bangladesh. *Prepr.*

https//www.researchsquare.com/article/rs-364004/v1.

Moore, W. E. C., Stackebrandt, E., Kandler, O., Colwell, R. R., Krichevsky, M. I., Truper, H. G., et al. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/00207713-37-4-463.

Mutreja, A., Kim, D. W., Thomson, N., Connor, T. R., Hee, J., Lebens, M., et al. (2013). Evidence for multiple waves of global transmission within the seventh cholera pandemic. *Nature*. doi:10.1038/nature10392.Evidence.

Nasreen, T., Hussain, N. A. S., Islam, M. T., Orata, F. D., Kirchberger, P. C., Case, R. J., et al. (2020). Simultaneous quantification of vibrio metoecus and *Vibrio cholerae* with its o1 serogroup and toxigenic subpopulations in environmental reservoirs. *Pathogens*. doi:10.3390/pathogens9121053.

Nelson, E. J., Harris, J. B., Morris Jr, J. G., Calderwood, S. B., and Camilli, A. (2009). Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat. Rev. Microbiol.* 7, 693–702. doi:10.1038/nrmicro2204.

Nelson, K. B., Karver, J., Kullman, C., and Graham, J. P. (2014). User perceptions of shared sanitation among rural households in Indonesia and Bangladesh. *PLoS One*. doi:10.1371/journal.pone.0103886.

Neogi, S. B., Chowdhury, N., Awasthi, S. P., Asakura, M., Okuno, K., Mahmud, Z. H., et al. (2019). Novel cholera toxin variant and ToxT regulon in environmental *Vibrio mimicus* isolates: Potential resources for the evolution of *Vibrio cholerae* hybrid strains. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01977-18.

Neogi, S. B., Lara, R., Alam, M., Harder, J., Yamasaki, S., and Colwell, R. R. (2018). Environmental and hydroclimatic factors influencing Vibrio populations in the estuarine

zone of the Bengal delta. *Environ. Monit. Assess.* 190, 565. doi:10.1007/s10661-018-6925-7.

Nora Hussein, Paul Kirchberger, Y. F. B. Type six secretion mediated competetion in *Vibrio cholerae*. *Front. Microbiol.*

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P. M., et al. (2008). The vegan package. *Community Ecol. Packag.*

Oliver, J. D., Warner, R. A., and Cleland, D. R. (1983). Distribution of *Vibrio vulnificus* and Other Lactose-Fermenting Vibrios in the Marine Environment. *Appl. Environ. Microbiol.* 45, 985–998. doi:10.1128/aem.45.3.985-998.1983.

Orata, F. D., Keim, P. S., and Boucher, Y. (2014). The 2010 Cholera Outbreak in Haiti: How Science Solved a Controversy. *PLOS Pathog.* 10, 1–5. doi:10.1371/journal.ppat.1003967.

Orata, F. D., Kirchberger, P. C., Méheust, R., Jed Barlow, E., Tarr, C. L., and Boucher, Y. (2015). The dynamics of genetic interactions between Vibrio metoecus and *Vibrio cholerae*, two close relatives co-occurring in the environment. *Genome Biol. Evol.* doi:10.1093/gbe/evv193.

Orata, F. D., Xu, Y., Gladney, L. M., Rishishwar, L., Case, R. J., Boucher, Y., et al. (2016). Characterization of clinical and environmental isolates of Vibrio cidicii sp. nov., a close relative of Vibrio navarrensis. *Int. J. Syst. Evol. Microbiol.* 66, 4148–4155. doi:10.1099/ijsem.0.001327.

Owen, R., and Pitcher, D. (1985). Current methods for estimating DNA base composition and levels of DNA-DNA hybridization. *Soc. Appl. Bacteriol. Tech. Ser.*, 67–93.

Payne, S. M., Mey, A. R., and Wyckoff, E. E. (2016). Vibrio Iron Transport: Evolutionary Adaptation to Life in Multiple Environments. *Microbiol. Mol. Biol. Rev.* doi:10.1128/mmbr.00046-15.

Polz, M. F., Alm, E. J., and Hanage, W. P. (2013). Horizontal gene transfer and the evolution of

bacterial and archaeal population structure. *Trends Genet.* doi:10.1016/j.tig.2012.12.006.

Pruzzo, C., Vezzulli, L., and Colwell, R. R. (2008). Global impact of *Vibrio cholerae* interactions with chitin. *Environ. Microbiol.* 10, 1400–1410. doi:https://doi.org/10.1111/j.1462- 2920.2007.01559.x.

Pukatzki, S., and Provenzano, D. (2013). *Vibrio cholerae* as a predator: lessons from evolutionary principles. *Front. Microbiol.* 4, 384. doi:10.3389/fmicb.2013.00384.

R Development Core Team, R. (2011). *R: A Language and Environment for Statistical Computing*. doi:10.1007/978-3-540-74686-7.

Rafique, R., Rashid, M. U., Monira, S., Rahman, Z., Mahmud, M. T., Mustafiz, M., et al. (2016). Transmission of Infectious *Vibrio cholerae* through Drinking Water among the Household Contacts of Cholera Patients (CHoBI7 Trial). *Front Microbiol* 7, 1635. doi:10.3389/fmicb.2016.01635.

Ramamurthy, T., Chowdhury, G., Pazhani, G., and Shinoda, S. (2014). Vibrio fluvialis: an emerging human pathogen. *Front. Microbiol.* 5, 91. doi:10.3389/fmicb.2014.00091.

Rashid, R. Bin, Ferdous, J., Tulsiani, S., Mackie Jensen, P. K., and Begum, A. (2017). Development and validation of a novel real-time assay for the detection and quantification of *Vibrio cholerae*. *Front. Public Heal.* doi:10.3389/FPUBH.2017.00109.

Reguera, G., and Kolter, R. (2005). Virulence and the Environment: a Novel Role for *Vibrio cholerae* Toxin-Coregulated Pili in Biofilm Formation on Chitin. *J. Bacteriol.* 187, 3551–3555. doi:10.1128/JB.187.10.3551-3555.2005.

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106, 19126–19131. doi:10.1073/pnas.0906412106.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.0906412106.

Riley, M. A., and Lizotte-Waniewski, M. (2009). Population genomics and the bacterial species concept. *Methods Mol. Biol.* 532, 367–377. doi:10.1007/978-1-60327-853-9_21.

Ripley, B. D. (2001). The R Project in Statistical Computing. *MSOR Connect.* doi:10.11120/msor.2001.01010023.

Rocha, E. P. C. (2018). Neutral Theory, Microbial Practice: Challenges in Bacterial Population Genetics. *Mol. Biol. Evol.* 35, 1338–1347. doi:10.1093/molbev/msy078.

Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67. doi:10.1111/j.1574-6976.2001.tb00571.x.

Routh, M. D., Zalucki, Y., Su, C. C., Long, F., Zhang, Q., Shafer, W. M., et al. (2010). Efflux pumps of the Resistance-nodulation-division family: A perspective of their structure, function, and regulation in gram-negative bacteria. *Adv. Enzymol. Relat. Areas Mol. Biol.* doi:10.1002/9780470920541.ch3.

Roux, F. Le, Wegner, K. M., Baker-Austin, C., Vezzulli, L., Osorio, C. R., Amaro, C., et al. (2015). The emergence of Vibrio pathogens in Europe: ecology, evolution, and pathogenesis (Paris, 11–12th March 2015). *Front. Microbiol.* 6, 830. doi:10.3389/fmicb.2015.00830.

Rusydi, A. F. (2018). Correlation between conductivity and total dissolved solid in various type of water: A review. in *IOP Conference Series: Earth and Environmental Science (2018)* doi:10.1088/1755-1315/118/1/012019.

Safa, A., Nair, G. B., and Kong, R. Y. C. (2010). Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol.* 18, 46–54. doi:10.1016/j.tim.2009.10.003.

Sant'anna, F. H., Reiter, K. C., Almeida, P. de F., and Passaglia, L. M. P. (2020). Systematic

review of descriptions of novel bacterial species: Evaluation of the twenty-first century taxonomy through text mining. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/ijsem.0.004070.

Sawabe, T., Kita-Tsukamoto, K., and Thompson, F. L. (2007). Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. *J. Bacteriol.* doi:10.1128/JB.00693-07.

Schirmeister, F., Wieczorek, A., Dieckmann, R., Taureck, K., and Strauch, E. (2014). Evaluation of molecular methods to discriminate the closely related species Vibrio fluvialis and Vibrio furnissii. *Int. J. Med. Microbiol.* 304, 851–857. doi:https://doi.org/10.1016/j.ijmm.2014.09.001.

Schliep, K., Lopez, P., Lapointe, F. J., and Bapteste, É. (2011). Harvesting evolutionary signals in a forest of prokaryotic gene trees. *Mol. Biol. Evol.* doi:10.1093/molbev/msq323.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* doi:10.1128/AEM.01541-09.

Shapiro, B. J., Levade, I., Kovacikova, G., Taylor, R. K., and Almagro-Moreno, S. (2016). Origins of pandemic *Vibrio cholerae* from environmental gene pools. *Nat Microbiol* 2, 16240. doi:10.1038/nmicrobiol.2016.240.

Shin, J. H., Shin, M. G., Suh, S. P., Ryang, D. W., Rew, J. S., and Nolte, F. S. (1996). Primary Vibrio damsela Septicemia. *Clin. Infect. Dis.* 22, 856–857. doi:10.1093/clinids/22.5.856.

Spietz, R. L., Williams, C. M., Rocap, G., and Horner-Devine, M. C. (2015). A dissolved oxygen threshold for shifts in bacterial community structure in a seasonally hypoxic estuary. *PLoS One*. doi:10.1371/journal.pone.0135731.

Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA

reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* doi:10.1099/00207713-44-4-846.

Stamatakis, A. (2014a). RAxML. *Manual/tutorial*.

Stamatakis, A. (2014b). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. doi:10.1093/bioinformatics/btu033.

Stine, O. C., Alam, M., Tang, L., Nair, G. B., Siddique, A. K., Faruque, S. M., et al. (2008). Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerg. Infect. Dis.* doi:10.3201/eid1405.071116.

Sultana, M., Nusrin, S., Hasan, N. A., Sadique, A., Ahmed, K. U., Islam, A., et al. (2018). Biofilms Comprise a Component of the Annual Cycle of *Vibrio cholerae* in the Bay of Bengal Estuary. *MBio* 9. doi:10.1128/mBio.00483-18.

Sun, T., and Altenbuchner, J. (2010). Characterization of a mannose utilization system in bacillus subtilis. *J. Bacteriol.* doi:10.1128/JB.01673-09.

Takemura, A. F., Chien, D. M., and Polz, M. F. (2014). Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* doi:10.3389/fmicb.2014.00038.

Tamayo, R., Patimalla, B., and Camilli, A. (2010). Growth in a Biofilm Induces a Hyperinfectious Phenotype in *Vibrio cholerae*. *Infect. Immun.* 78, 3560–3569. doi:10.1128/IAI.00048-10.

Thode, S. K., Rojek, E., Kozlowski, M., Ahmad, R., and Haugen, P. (2018). Distribution of siderophore gene systems on a Vibrionaceae phylogeny: Database searches, phylogenetic analyses and evolutionary perspectives. *PLoS One* 13, 1–20. doi:10.1371/journal.pone.0191860.

Thompson, C. C., Vicente, A. C. P., Souza, R. C., Vasconcelos, A. T. R., Vesth, T., Alves, N., et

al. (2009). Genomic taxonomy of vibrios. *BMC Evol. Biol.* doi:10.1186/1471-2148-9-258.

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., et al. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl. Environ. Microbiol.* 71, 5107–5115. doi:10.1128/AEM.71.9.5107-5115.2005.

Thompson, F. L., Iida, T., and Swings, J. (2004). Biodiversity of Vibrios. *Microbiol. Mol. Biol. Rev.* doi:10.1128/mmbr.68.3.403-431.2004.

Torondel, B., Ensink, J. H. J., Gundogdu, O., Ijaz, U. Z., Parkhill, J., Abdelahi, F., et al. (2016). Assessment of the influence of intrinsic environmental and geographical factors on the bacterial ecology of pit latrines. *Microb. Biotechnol.* doi:10.1111/1751-7915.12334.

Trastoy, R., Manso, T., Fernández-García, L., Blasco, L., Ambroa, A., Pérez Del Molino, M. L., et al. (2018). Mechanisms of bacterial tolerance and persistence in the gastrointestinal and respiratory environments. *Clin. Microbiol. Rev.* doi:10.1128/CMR.00023-18.

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.* doi:10.1186/s13059-014-0524-x.

Turner, J. W., Good, B., Cole, D., and Lipp, E. K. (2009). Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME J.* 3, 1082–1092. doi:10.1038/ismej.2009.50.

Unterweger, D., Miyata, S. T., Bachmann, V., Brooks, T. M., Mullins, T., Kostiuk, B., et al. (2014a). The *Vibrio cholerae* type VI secretion system employs diverse effector modules forintraspecific competition. *Nat. Commun.* doi:10.1038/ncomms4549.

Unterweger, D., Miyata, S. T., Bachmann, V., Brooks, T. M., Mullins, T., Kostiuk, B., et al. (2014b). The *Vibrio cholerae* type VI secretion system employs diverse effector modules for intraspecific competition. *Nat Commun* 5, 3549. doi:10.1038/ncomms4549.

VanInsberghe, D., Arevalo, P., Chien, D., and Polz, M. F. (2020). How can microbial population genomics inform community ecology? *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* 375, 20190253. doi:10.1098/rstb.2019.0253.

Vezzulli, L., Brettar, I., Pezzati, E., Reid, P. C., Colwell, R. R., Höfle, M. G., et al. (2012). Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *ISME J.* 6, 21–30. doi:10.1038/ismej.2011.89.

Vezzulli, L., Grande, C., Reid, P. C., Hélaouët, P., Edwards, M., Höfle, M. G., et al. (2016). Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.1609157113.

Vezzulli, L., Pruzzo, C., Huq, A., and Colwell, R. R. (2010). Environmental reservoirs of *Vibrio cholerae* and their role in cholera. *Environ. Microbiol. Rep.* doi:10.1111/j.1758-2229.2009.00128.x.

Waldor, M. K., and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1914. doi:10.1126/science.272.5270.1910.

Wallet, F., Tachon, M., Nseir, S., Courcol, R. J., and Roussel-Delvallez, M. (2005). Vibrio metschnikovii pneumonia. *Emerg. Infect. Dis.* 11, 1641–1642. doi:10.3201/eid1110.050177.

Weber, G. G., Kortmann, J., Narberhaus, F., and Klose, K. E. (2014). RNA thermometer controls temperature-dependent virulence factor expression in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14241–14246. doi:10.1073/pnas.1411570111.

Westra, E. R., van Houte, S., Gandon, S., and Whitaker, R. (2019). The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* 374, 20190101. doi:10.1098/rstb.2019.0101.

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The

primary kingdoms. *Proc. Natl. Acad. Sci.* 74, 5088–5090. doi:10.1073/pnas.74.11.5088.

Wright, J. J., Lee, S., Zaikova, E., Walsh, D. A., and Hallam, S. J. (2009). DNA extraction from 0.22 μM Sterivex filters and cesium chloride density gradient centrifugation. *J. Vis. Exp.* doi:10.3791/1352.

Wyckoff, E. E., Mey, A. R., and Payne, S. M. (2007). Iron acquisition in *Vibrio cholerae*. in *BioMetals* doi:10.1007/s10534-006-9073-4.

Yamasaki, S., Shimizu, T., Hoshino, K., Ho, S.-T., Shimada, T., Nair, G. B., et al. (1999). The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are closely related to those of *Vibrio cholerae* O22. *Gene* 237, 321–332. doi:https://doi.org/10.1016/S0378-1119(99)00344-3.

Yildiz, F. H., and Visick, K. L. (2009). Vibrio biofilms: so much the same yet so different. *Trends Microbiol.* 17, 109–118. doi:10.1016/j.tim.2008.12.004.

Zheng, B., Jiang, X., Cheng, H., Guo, L., Zhang, J., Xu, H., et al. (2017). Genome characterization of two bile-isolated Vibrio fluvialis strains: An insight into pathogenicity and bile salt adaption. *Sci. Rep.* doi:10.1038/s41598-017-12304-8.

APPENDICES

Apendix A: List of publications

This is the complete list of authored and co-authored publications in the course of my doctorate program (2015–2021). The list includes manuscripts that have been published, in press, or in preparation (i.e., full draft of manuscript available) as of the writing of this thesis.

A.1. Published

1. Islam MT, Nasreen T, Kirchberger P, Liang KYH, Orata FD, Johura FT, Hussain NAS, Im MS, Tarr CL, Alam M, Boucher YF. Population analysis of *Vibrio cholerae* in aquatic reservoirs reveals a novel sister species *(Vibrio paracholerae* sp. nov.) with a history of association with humans. *Applied Environmental Microbiology* 2021 Jun 16: AEM0042221. doi: 10.1128/AEM.00422-21.

2. Tania Nasreen, Nora AS Hussain, **Mohammad Tarequl Islam**, Fabini D Orata, Paul C Kirchberger, Rebecca J Case, Munirul Alam, Stephanie K Yanow, Yann F Boucher. Simultaneous quantification of Vibrio metoecus and *Vibrio cholerae* with its O1 serogroup and toxigenic subpopulations in environmental reservoirs. *Pathogens* 9 (12), 1053.

3. KYH Liang, FD Orata, **MT Islam**, T Nasreen, M Alam, CL Tarr, YF Boucher. A Vibrio cholerae core genome multilocus sequence typing scheme to facilitate the epidemiological study of cholera. *Journal of bacteriology* 202 (24), e00086-20.

4. KYH Liang, FD Orata, **MT Islam**, T Nasreen, M Alam, CL Tarr, YF Boucher. Draft Genome Sequences of Eight *Vibrio* sp. Clinical Isolates from across the United States That Form a Basal Sister Clade to *Vibrio cholerae*. *Microbiology resource announcements* 8 (3),e01473-18.

5. **MT Islam**, K Liang, MS Im, J Winkjer, S Busby, CL Tarr, Y Boucher. Draft Genome Sequences of Nine *Vibrio* sp. Isolates from across the United States Closely Related to *Vibrio cholerae*. *Microbiology resource announcements* 7 (21), e00965-18.

6. **MT Islam**, M Alam, Y Boucher. Emergence, ecology and dispersal of the pandemic generating *Vibrio cholerae* lineage. *International Microbiology* 20 (3), 106-15.

7. Seon Young Choi, Shah M Rashed, Nur A Hasan, Munirul Alam, **Tarequl Islam**, Abdus Sadique, Fatema-Tuz Johura, Mark Eppinger, Jacques Ravel, Anwar Huq, Alejandro Cravioto, Rita R Colwell. Phylogenetic diversity of *Vibrio cholerae* associatedwith endemic cholera in Mexico from 1991 to 2008. MBio 7 (2), e02160-15

8. Mahamud-ur Rashid, Shah M Rashed, **Tarequl Islam**, Fatema-Tuz Johura, Haruo Watanabe, Makoto Ohnishi, Munirul Alam. ctxB1 outcompetes ctxB7 in *Vibrio cholerae*O1, Bangladesh. Journal of medical microbiology 65 (1), 101-103

A.2. Accepted

1. Tania Nasreen*, **Mohammad Islam***, Kevin YH Liang, Fatema-Tuz Johura, Paul C Kirchberger, Marzia Sultana, Rebecca J Case, Munirul Alam, Yann F Boucher.

Dynamic Subspecies Population Structure of *Vibrio cholerae* in Dhaka, Bangladesh. *Microbial Ecology*. Accepted in July 2021.

A.3. In preparation

1. Mohammad Tarequl Islam, Kevin Liang, Monica S. Im, Jonathan Winkjer , Shelby Busby , Cheryl L. Tarr, and Yann F. Boucher. *Vibrio tarriae*: A lactose-fermenting relative of *Vibrio cholerae* associated with opportunistic infections in humans.

2. Mohammad Tarequl Islam, Moussa djouda, Cheryl Tarr and Yann F. Boucher. *Vibrio paracholerae* co-exists with *V. cholerae* in the natural environment of Northern Cameroon.

3. Mohammad Tarequl Islam, Munirul Alam and Yann F. Boucher. "The cholerae clade": How many sisters are there?

4. Mohammad Tarequl Islam, Munirul Alam, Cheryll Tarr and Yann F. Boucher. *Vibrio paracholerae* contributed to the evolution of *Vibrio cholerae* O139.

Apendix B: Supplementary data for chapter 3

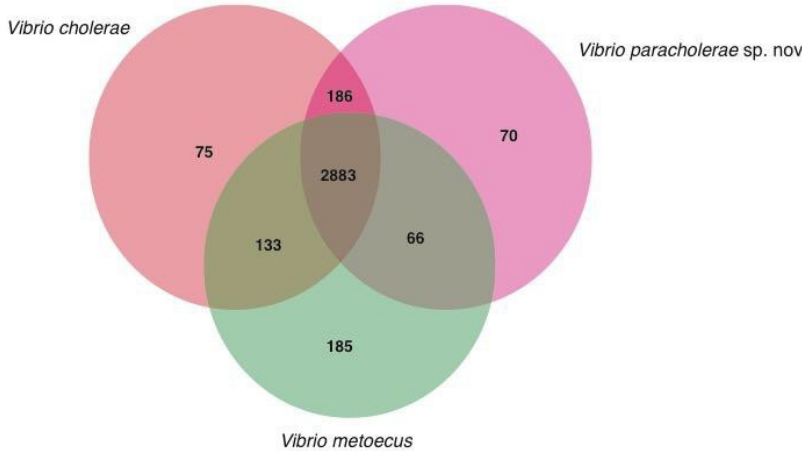


Fig. B.1. Venn diagram of the orthologous protein coding gene family comparison of *Vibrio paracholerae* sp. nov. with *V. cholerae* and *V. metoecus*. Shared and unique gene clusters were calculated using OrthoVenn2 (https://orthovenn2.bioinfotoolkits.net/home).

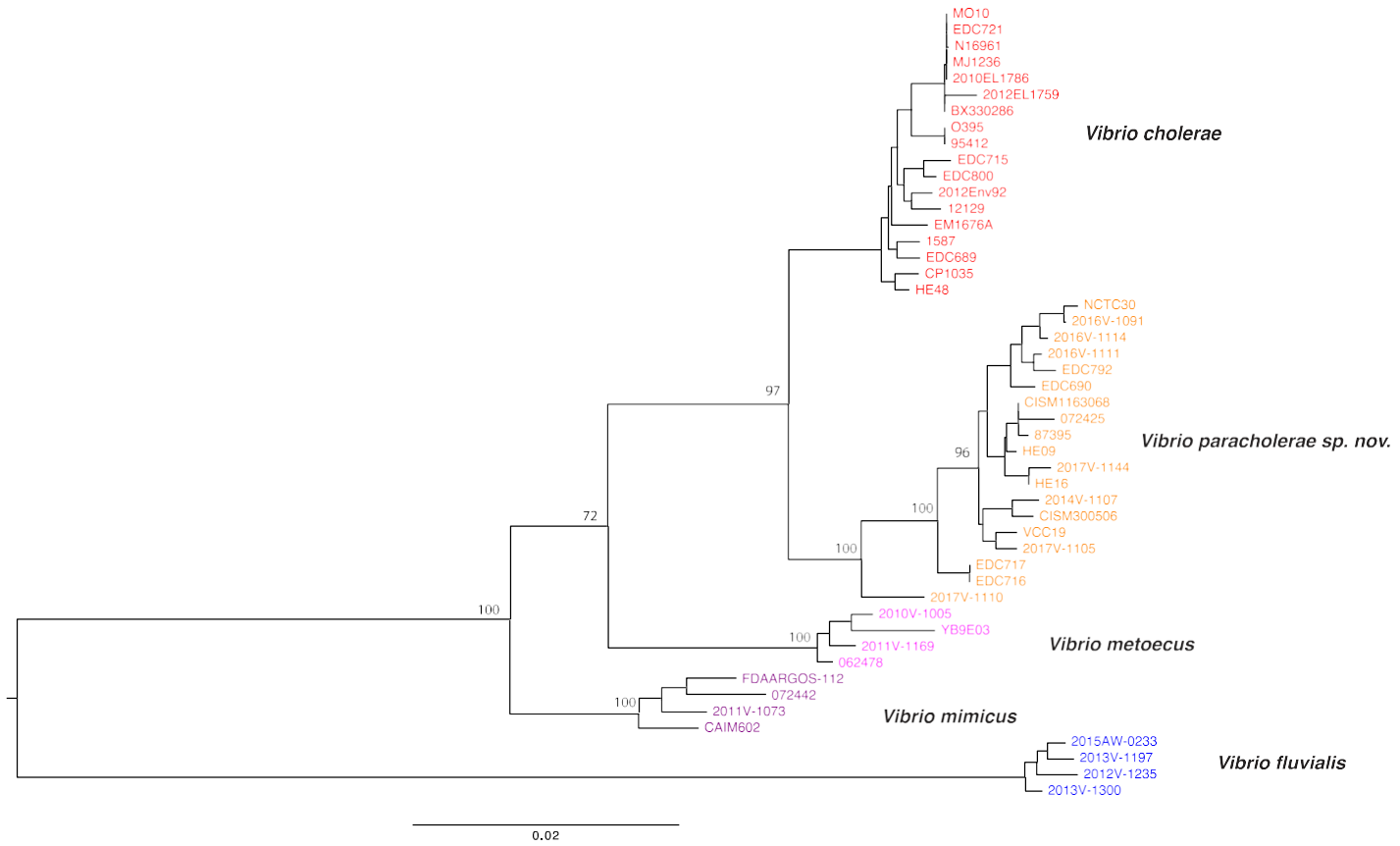
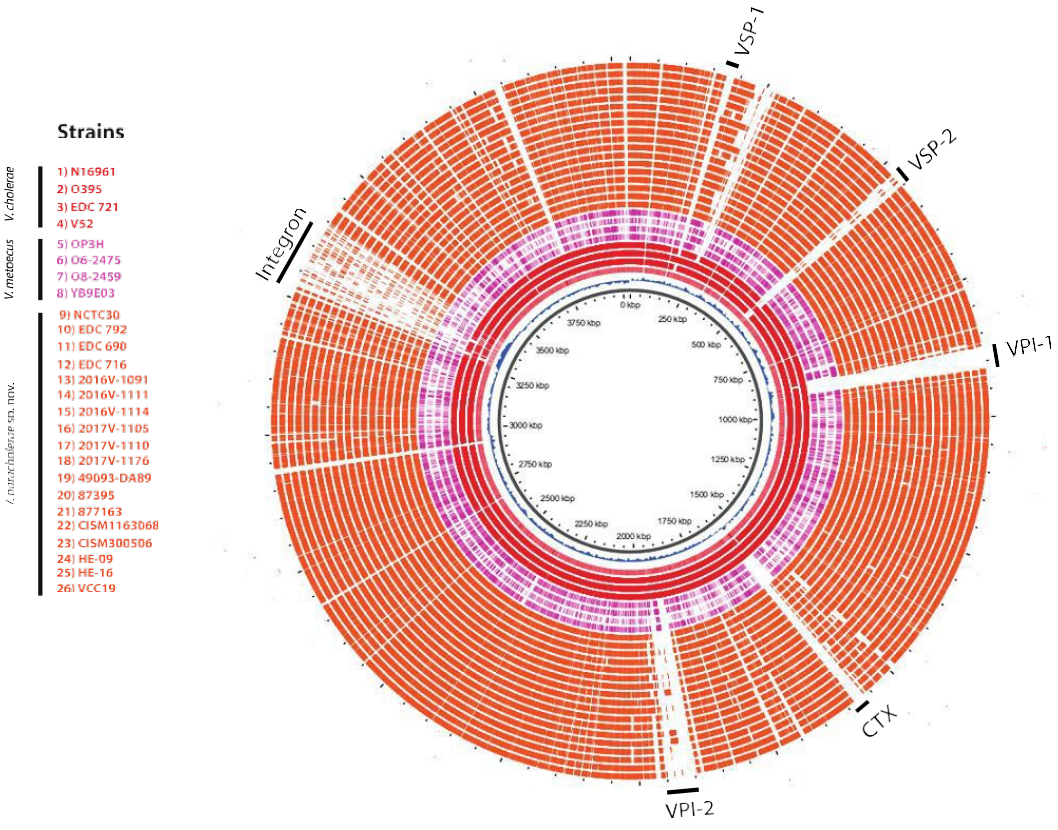


Fig. B.2. Multilocus sequence analysis (MLSA) of *Vibrio paracholerae* sp. nov. and its sister species. The maximum-likelihood (PhyML) phylogenetic tree was constructed from the alignment (7392 bp) of four protein coding housekeeping genes (*rpoA*, *rpoB*, *recA*, *pyrH*). Bootstrap support values over 70% are indicated on the corresponding nodes.



Fig, B.3. BLAST atlas of *V. paracholerae* sp. nov., *V. cholerae* and *V. metoecus* strains. *V. cholerae* strain N16961 was used as the reference sequence for BLASTN comparisons. Each colored ring represents a genome starting from N16961. Outermost black bars indicate the major genomic islands of V. cholerae. VSP, Vibrio seventh pandemic island, VPI, Vibrio pathogenicity island, CTX, cholera toxin prophage.

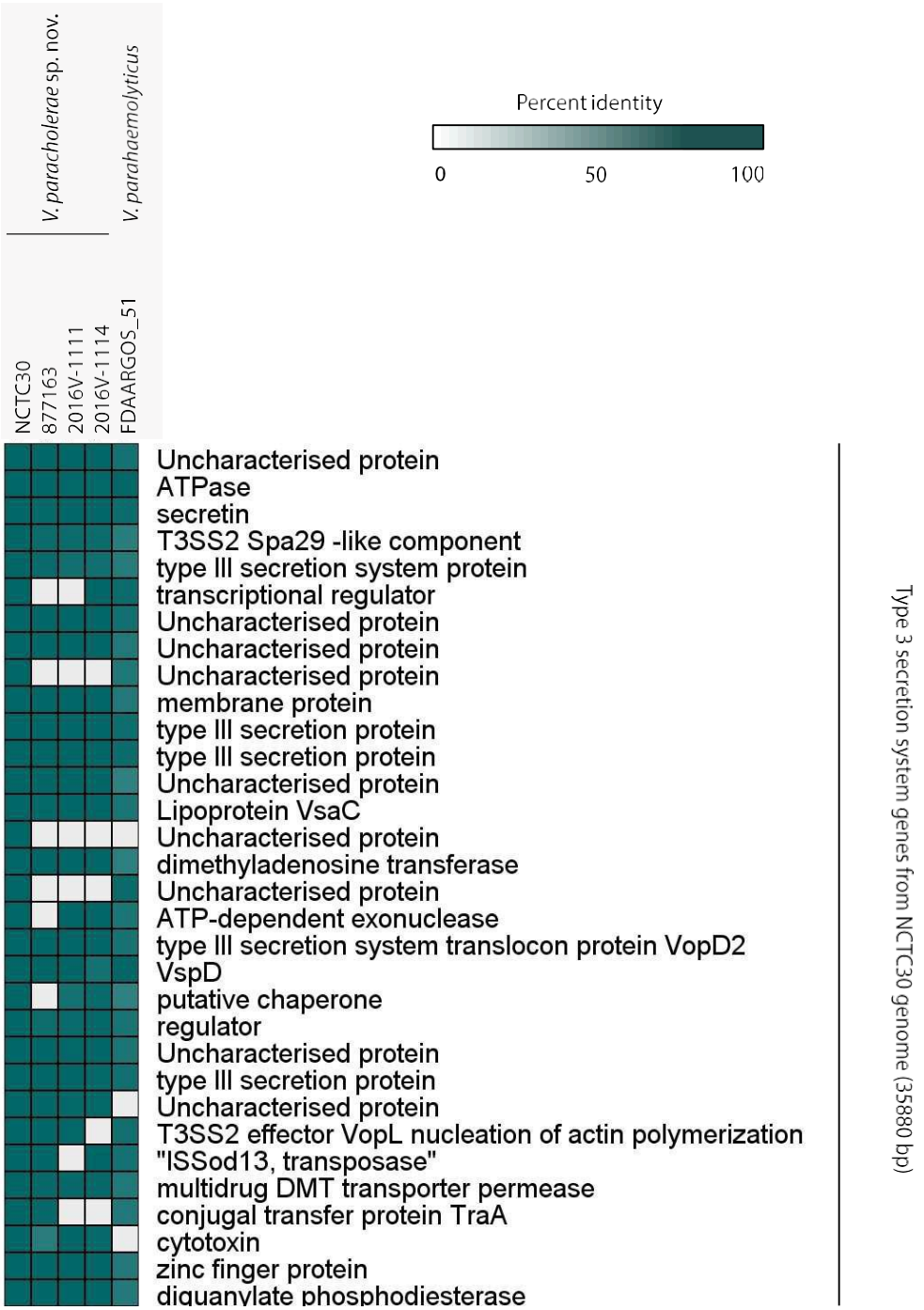


Fig. B.4. Comparison of T3SS island found in *Vibrio paracholerae* sp. nov. and *Vibrio parahaemolyticus*. Genes found in T3SS island of NCTC30 genome were screened in the genomes of three T3SS positive *V. paracholerae* sp. nov. strains and *V. parahaemolyticus* strain FDAARGOS_51 using BLASTN. Green box indicates presence and white box indicates absence of the genes.

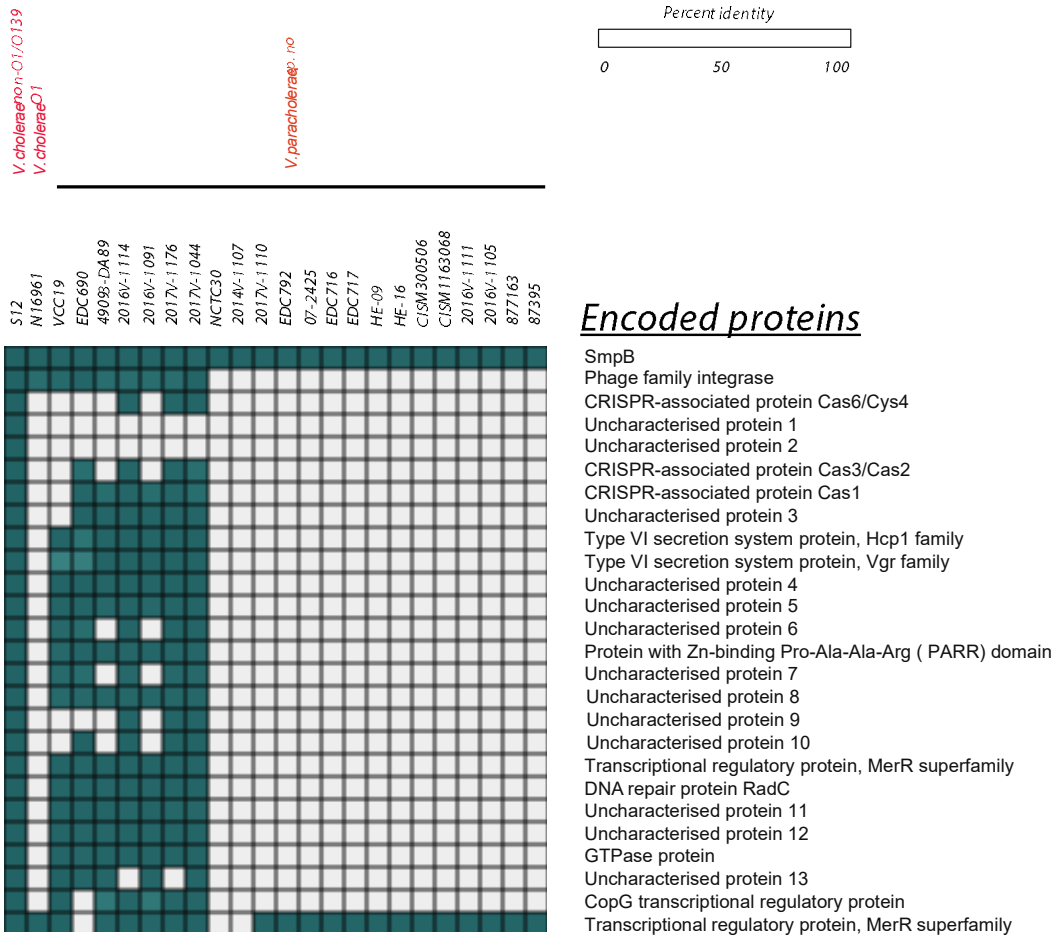


Fig. B.5. Presence of GIVchS12 island within VPI-1 insertion site in *V. paracholerae* sp. nov. isolates. Regions of nucleotide identity to VPI-1 and surrounding regions in V. cholerae N16961 are shown. Presence of Protein coding genes in GIVchS12 island described by Maurizio *et al* were looked for using BLASTN in *V. paracholerae* sp. nov. genomes (Labbate et al., 2016). Green box indicates presence and white box indicates absence of genes.

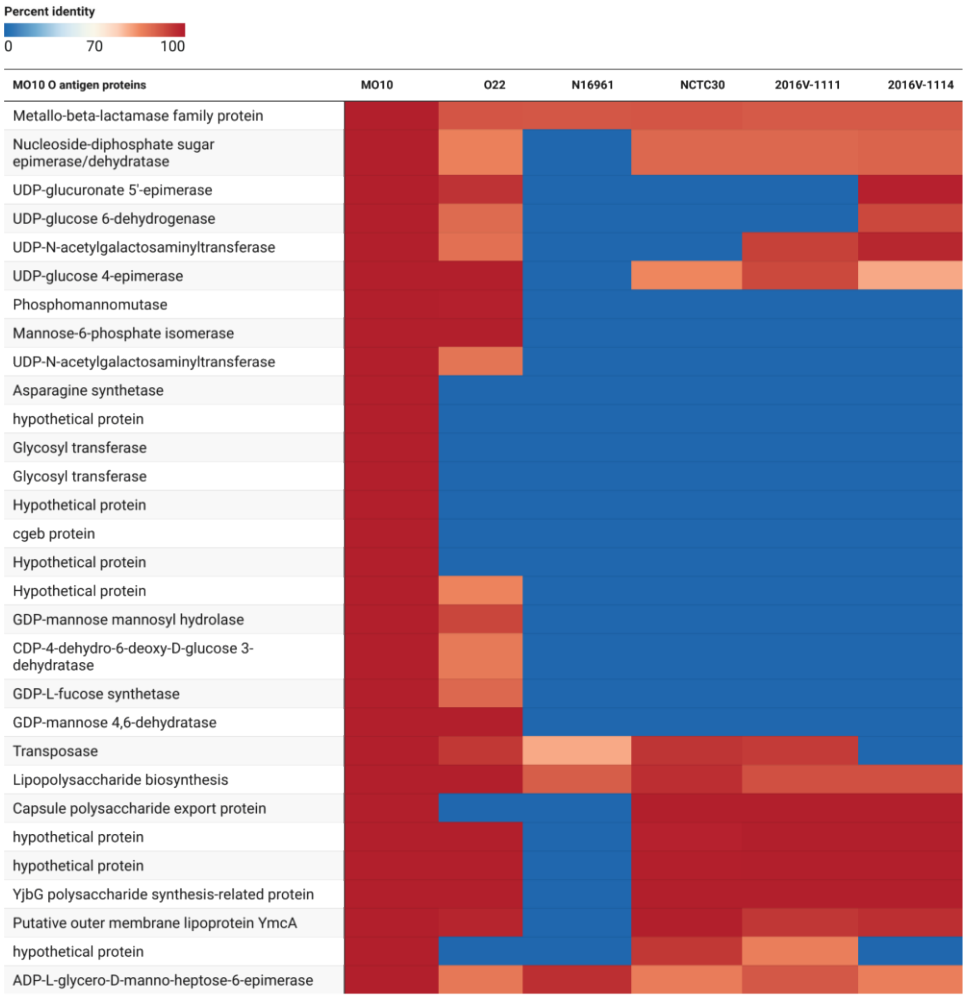


Figure B. 6: Comparison of O-antigen biosynthesis region of *V. cholerae* O139 with *V. cholerae* O22, *V. cholerae* O1 and *V. paracholerae* sp. nov. MO10: reference strain for *V. cholerae* O139; O22: O-antigen region for *V. cholerae* serogroup O22; N16961: reference strain for V. cholerae O1, NCTC30: *V.paracholerae* sp. nov., 2016V-1111: *V.paracholerae* sp. nov; 2016V-1114: *V.paracholerae* sp. nov. O-antigen regions were extracted and compared using BLASTN.

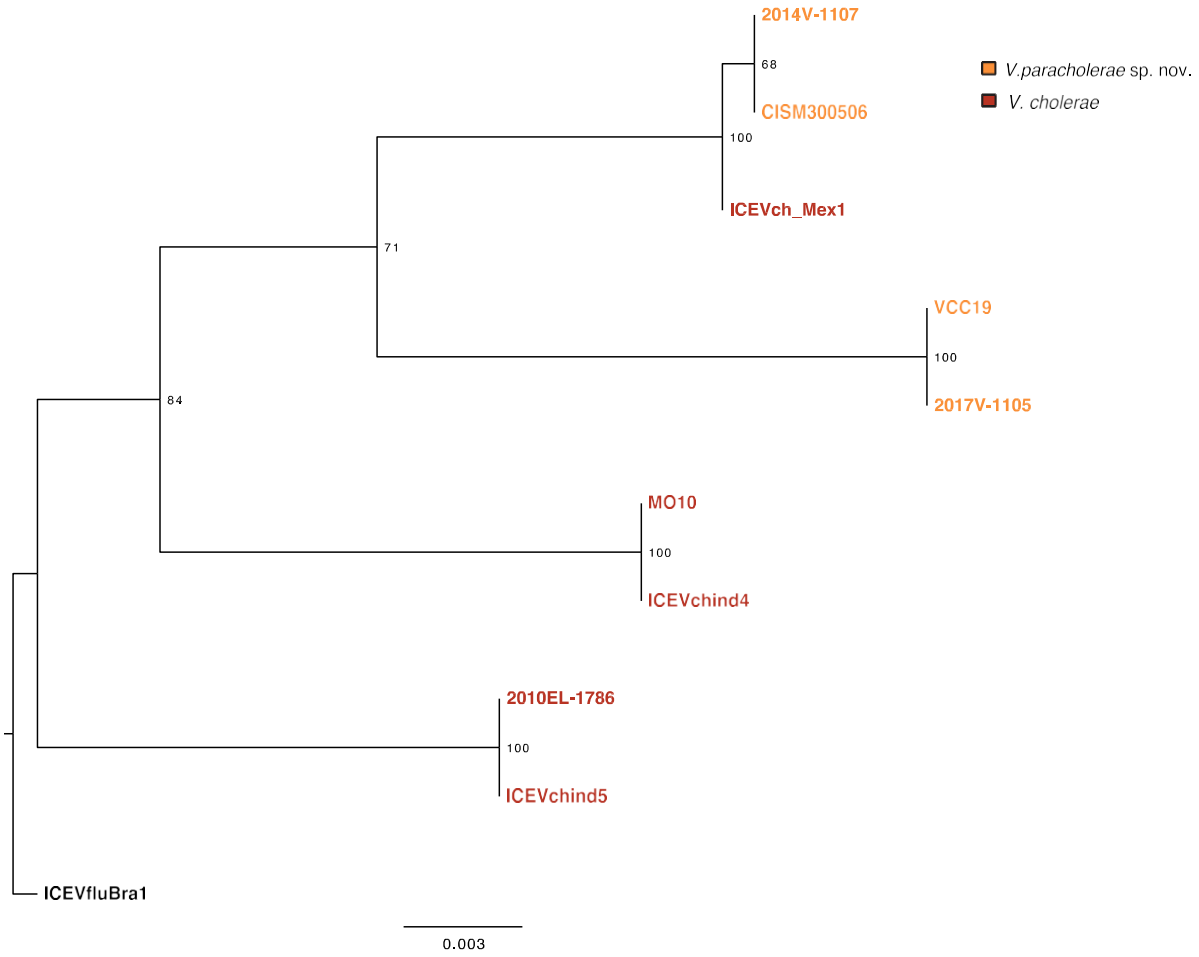


Figure B.7. Maximum likelihood tree of the 1243 bp integrase gene (int) found in integretive conjugative element (ICE) of *V. paracholerae* sp. nov. and *V. cholerae* strains. Sequence of integrase gene found in *V. fluvialis* (ICEVfluBra1) was used as outgroup to root the tree. Bootstrap values are shown on the corresponding nodes.

Table B.1. Physiochemical data ranges of the two sampling locations used in the study.

Location	Number of samples	Salinity range (ppt)	DO range (mg/L)	pH range	Temperature range (°C)
Dhaka	60	0-0.8	0.19-5.05	6.4-7.7	27.1-33.6
Oyster Pond	40	0-6	5.4-8.6	5.5-7.5	10.5-28

Table B.2. Responses towards chemical/heavy metals of *Vibrio paracholerae* sp. nov., differentiating it from its closest relatives *Vibrio cholerae.* R: resistant; S: sensitive; ND: not done.

Chemical/ heavy metal	***Vibrio paracholerae* sp. nov.**				***Vibrio cholerae***			
	EDC 690	**EDC 716**	**EDC 792**	**2016V-1091**	**N16961**	**V52**	**YB3B05**	**YB8E08**
Cadmium chloride	R	R	R	ND	S	S	ND	ND
Sodium selenite	R	R	R	ND	S	S	ND	ND
Dichlofuanid	R	R	R	ND	S	S	ND	ND

Table B.3. Presence of potential virulence traits in *V. paracholerae* sp. nov isolates in comparison to *V. cholerae* reference strains

Strain	Phylogenetic group	CTX-VPI1-VPI2	RTX toxin	SXT element	T3SS	RND efflux gene cluster	Cholix toxin (chxA)	β-lactamase
N16961	*V. cholerae* O1 El Tor (PG)	+	+	-	-	-	-	-
O395	*V. cholerae* O1 classical (PG)	+	+	-	-	-	-	-
MO10	*V. cholerae* O139 (PG)	+	+	+	-	-	-	-
12129	*V. cholerae* non-O1/O139 (non-PG)	-	+	-	+	-	-	-
VCC19	*V. paracholerae* sp. nov.	-	+	+	-	+	-	+
877163		-	+	-	+	+	+	-
EDC-792		-	+	-	-	+	+	-
EDC-690		-	+	-	-	-	+	-
EDC-716		-	+	-	-	+	-	-
EDC-717		-	+	-	-	+	-	-
HE09		-	+	-	-	+	+	-
HE16		-	+	-	-	+	-	-
CISM300506		-	+	+	-	+	+	+
CISM1163068		-	+	+	-	+	+	+
49093-DA89		-	+	-	-	+	-	+

SIO		-	+	-	-	-	+	-
2017V-1144		-	+	-	-	+	-	+
2014V-1107		-	+	+	-	+	+	-
2017V-1105		-	+	+	-	+	+	+
2016V-1114		-	+	-	+	+	-	-
2016V-1111		-	+	-	+	+	-	-
2017V-1176		-	+	-	-	+	-	+
2016V-1091		-	+	-	-	+	+	-
2017V-1110		-	+	-	-	-	-	-
87395		-	+	-	-	+	+	+
07-2425		-	-	-	-	+	-	-
NCTC 30		-	+	-	+	+	-	+

Table B.4. Information of genomes falling into the *V. paracholerae* sp. nov clade disclosed in the NCBI database after 2019. dDDH: digital DNA-DNA hybridization value; Clin: clinical; Env: environmental.

Strain	Origin	Source	dDDH with *V. paracholerae* sp. nov (EDC792)	NCBI genome accession number
N2784	China	Clin	85.2	VSHN01000030.1
N2770	China	Clin	85.7	VSHB01000078.1
N2768	China	Clin	89.4	VSGZ01000037.1
EL2338	China	Clin	89.1	VMPB01000147.1
2204	Brazil	Env	85.8	VHOF01000002.1
2290	Brazil	Env	79.1	VHOE01000001.1
N2748	China	Clin	83.7	VSGI01000001.1
EL2403	China	Clin	83.2	VMOL01000089.1
A110912Z3	Austria	Env	82.4	VIQC01000020.1
N2807	China	Clin	83	VSID01000044.1
N2795	China	Clin	83.2	VSHY01000044.1
N2794	China	Clin	83.4	VSHX01000029.1
N2791	China	Clin	83.3	VSHU01000058.1
A3_296	Brazil	Clin	80.6	QBJE01000015.1
FORC_076	South korea	Clin	86	NZ_CP026531

Table B.5. Type VI secretion system effector and immunity gene combinations in *V. paracholerae* sp. nov. strains. Genes in three loci are identified and named using the scheme described in Kirchberger *et al* (Kirchberger et al., 2017)

Strain	T6SS effector immunity locus		
	aux1	aux2	main
07-2425	Aacc	Dddd	LliE*ecga
2014V-1107	Cc	Dd	Aa
2016V-1091	Ccccc	Aa	FfbE*ec
2016V-1111	Cccc	Dd	Hh
2016V-1114	Cccc	Dd	Hh
2017V-1105	Cc	Dd	Aa
2017V-1110	Cc	Aa	Gg
2017V-1144	AAcc	Dd	Ll
2017V-1176	Cccc	Aa	Aa
NCTC30	Aacc	Dd	Ii
CISM300506	Cc	Dd	Aa
CISM1163068	Aacc	Dddd	LliE*ecga
EDC 690	Aaaaac	Dd	Aa
EDC 792	Ccccc	Dd	Hh
EDC 717	Cccc	Aa	Ii
EDC 716	Cccc	Aa	Ii
49093-DA89	Aaccc	Dddd	LliE*ecga
HE-09	Aacc	Dddd	Aa
87395*	Aacc	Dddd	LliE*ecga
HE-16	Aacc	Dd	Lli
VCC19	Cc	Dd	Aa
877163	Aac	Dd	Ii

Upper case: Effector

Lower case: Immunity

* Denotes truncated E-type effector

Apendix C: Supplementary figure for chapter 4

Table C.1. Genomes representative to *viuB* alleles found in Dhaka, Bangladesh used to construct the phylogenetic tree and DDH analysis

Name of the isolate	*viuB* type	Phylogeny	Country of Origin	Accession number
EDC_690	6	LB	Bangladesh	WUWI00000000
Vc_229135	6	LB	Spain	SRR7062498
EDC_716	7	LB	Bangladesh	WYBY00000000
EDC_717	7	LB	Bangladesh	WYBZ00000000
Vc_2016V_1091	5	LB	Unknown	GCA_003312065.1
Vc_PNUSAV000170	5	LB	USA	SRR6456907
Vc_236140	10	Non-PG	India	SRR7062513
Vc_229143	10	Non-PG	India	SRR7062619
Vc_YB2H11	39	Non-PG	USA	GCA_003349365.1
Vc_YB1A01	39	Non-PG	USA	GCA_001402185.1
Vc_EDC_800	45	Non-PG	Bangladesh	WYCA00000000
Vc_EDC_688	45	Non-PG	Bangladesh	WYCS00000000
Vc_BD21	32	Non-PG	Bangladesh	GCA_003348245.1
Vc_BD23	32	Non-PG	Bangladesh	GCA_003348215.1
Vc_L11	51	Non-PG	Sweden	GCA_001718105.1
Vc_NHCC-008D	65	Non-PG	Bangladesh	GCA_000348425.2
Vc_493492	65	Non-PG	Thailand	SRR7062586
Vc_EDC_754	73	El Tor	Bangladesh	WYCF00000000
Vc_EDC_755	73	El Tor	Bangladesh	WYCE00000000
Vc_BD04	25	Non-PG	Bangladesh	GCA_003348485.1
Vc_BD34	25	Non-PG	Bangladesh	GCA_003348055.1

LB denotes long branch

Table C.2. Month wise isolation data of *V. cholerae* O1 and *V. cholerae* non-O1/O139 in Dhaka. Identity of the isolates were confirmed by conventional PCR for *V. cholerae* and serogroup O1

Month	Site	*V. cholerae* O1	*V. cholerae* non-O1/O139
July, 2015	S1	-	+
	S2	-	+
	S3	-	-
	S4	-	-
	S5	+	+
	S6	+	+
	S7	-	+
August, 2015	S1	-	+
	S2	-	+
	S3	-	+
	S4	-	-
	S5	-	+
	S6	-	+
	S7	-	+
September, 2015	S1	-	+
	S2	-	+
	S3	-	+
	S4	-	+
	S5	-	+
	S6	-	-
	S7	-	+
October, 2015	S1	+	+
	S2	-	+
	S3	-	+
	S4	-	+

	S5	+	+
	S6	+	+
	S7	-	+
	S1	-	+
	S2	-	+
	S3	-	+
	S4	-	+
	S5	-	+
	S6	-	+
November, 2015	S7	+	+
	S1	-	+
	S2	-	+
	S3	-	+
	S4	-	+
	S5	-	+
	S6	-	+
December, 2015	S7	-	+
	S1	-	+
	S2	-	+
	S3	-	-
	S4	-	+
	S5	-	-
	S6	-	+
January, 2016	S7	-	+
	S1	-	+
	S2	-	+
	S3	-	+
	S4	-	+
February, 2016	S5	-	+

	S6	-	+
	S7	-	+
	S1	-	+
	S2	-	+
	S3	-	+
	S4	-	+
	S5	-	+
	S6	+	+
March, 2016	S7	-	+

Table C.3. Average values of environmental parameters in sampling locations in Dhaka

Site	pH	DO (mg/L)	Con (µs/cm)	TDS (mg/L)	Salinity(ppt)	WT ^{0}c	AT ^{0}c
Site 1	6.86	0.95	749.94	379.77	0.34	30.18	30.86
Site 2	7.03	2.70	639.63	289.57	0.23	30.47	31.11
Site 3	7.03	2.79	598.56	299.12	0.23	30.50	31.14
Site 4	7.03	2.78	587.28	293.6	0.23	30.56	31.20
Site 5	7.0	2.80	559.98	279.87	0.21	30.70	31.32
Site 6	7.0	3.04	541.77	270.83	0.21	30.73	31.4
Site 7	6.91	0.90	848.11	426.25	0.35	30.79	31.41

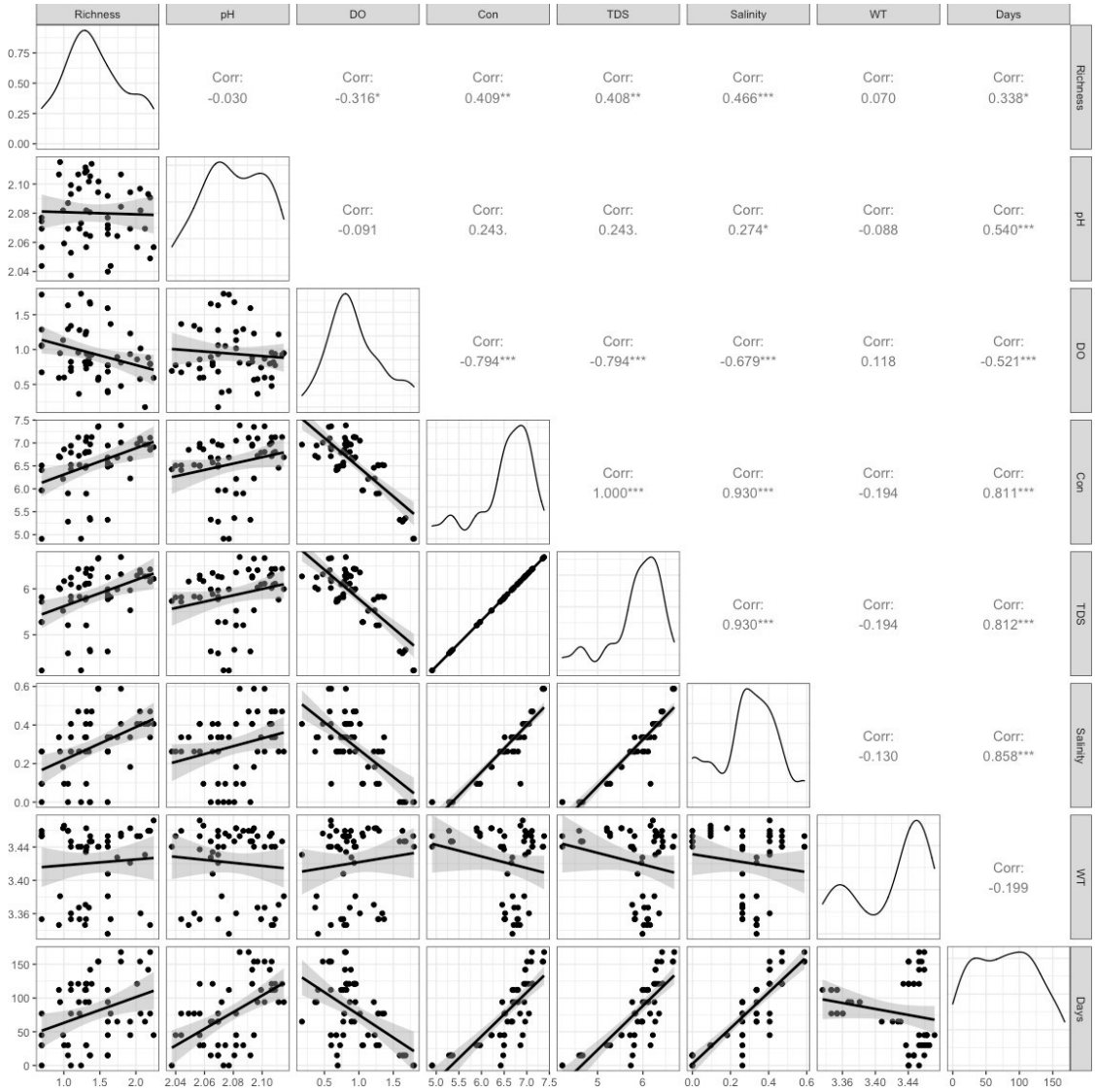


Fig. C.1. Pearson correlation analysis describing correlations among different environmental variables and richness of *V. cholerae* population defined by *viuB* alelles. DO, dissolved oxygen; Con, conductivity; TDS, total dissolved solids; WT, water temperatures; Corr, correlation co-efficient. Significant values are marked by asterisks (*).

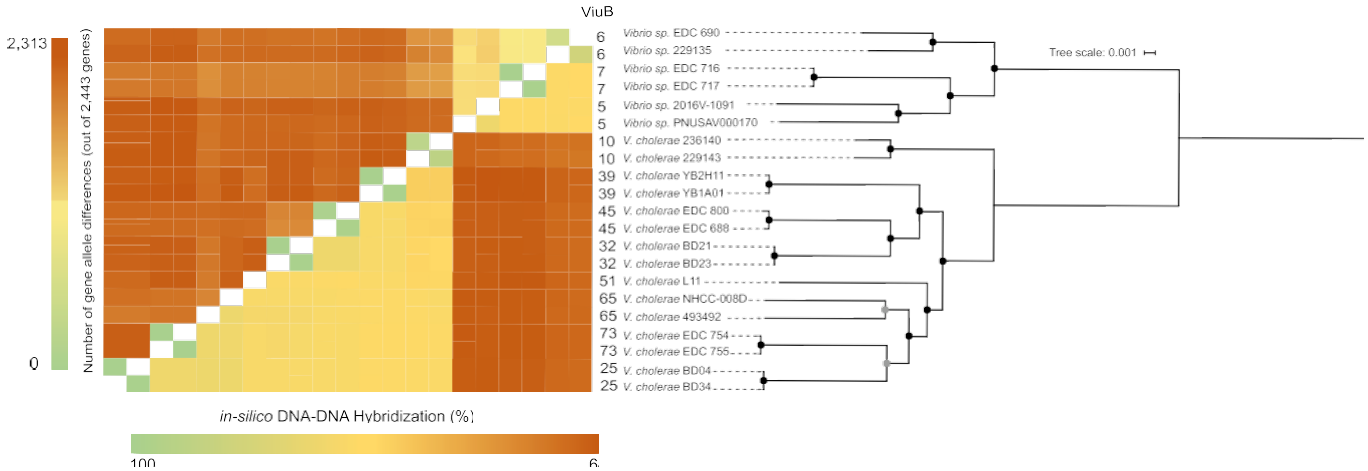


Fig. C.2. Genomic and phylogenetic distinction between strains carrying the *viuB*-05, *viuB*-06 *and* *viuB*-07 *alleles* and those carrying other alleles. At most two representative strains for each *viuB* allele were chosen. All genomes were aligned using Mugsy v1r2.3 with default parameters. A maximum likelihood phylogenetic tree was reconstructed using RAxML v8.2.11 based on this alignment under the GTR GAMMA model. Branch support was evaluated with 100 bootstrap replicates and is indicated as black (≥95%) and grey (≥50%) circles. Allelic differences based on the *V. cholerae* cgMLST scheme shown on the left of the matrix and *in-silico* DNA-DNA Hybridization values on the right. Self-comparisons are ignored. *viuB*-73 is exclusively associated with the pandemic generating lineage, which includes the El Tor biotype responsible for the current cholera pandemic. Strains carrying the *viuB*-05, *viuB*-06 and *viuB*-07 alleles were clearly distinct from the rest of the *V. cholerae* isolates, as they form a clear monophyletic clade by themselves. cgMLST allelic differences and *in-silico* DDH data also showed a clear boundary between isolates carrying these *viuB* alleles and isolates carrying other alleles. *viuB*-79 is not included in this analysis as no genomes with this allele were available.

Apendix D: Supplementary data for chapter 5

Table D.1. Average patristic distance among the strains of *Vibrio* species.

	V. cholerae	*V. paracholerae*	*V. tarriae*	*V. metoecus*	*V. mimicus*	*V. fluvialis*	*V. furnissii*
V. cholerae	0.03	0.07	0.06	0.235	0.278	0.93	0.94
V. paracholerae	0.08	0.03	0.11	0.27	0.32	0.97	0.98
V. tarriae	0.066	0.094	0.03	0.22	0.271	0.92	0.929
V. metoecus	0.232	0.265	0.231	0.038	0.245	0.895	0.904
V. mimicus	0.285	0.309	0.264	0.239	0.03	0.86	0.87
V. fluvialis	0.93	0.95	0.92	0.91	0.861	0.035	0.292
V. furnissi	0.94	0.984	0.929	0.904	0.864	0.292	0.02

References

Kirchberger, P. C., Unterweger, D., Provenzano, D., Pukatzki, S., & Boucher, Y. (2017). Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. *Scientific Reports*. https://doi.org/10.1038/srep45133

Labbate, M., Orata, F. D., Petty, N. K., Jayatilleke, N. D., King, W. L., Kirchberger, P. C., Allen, C., Mann, G., Mutreja, A., Thomson, N. R., Boucher, Y., & Charles, I. G. (2016). A genomic island in *Vibrio cholerae* with VPI-1 site-specific recombination characteristics contains CRISPR-Cas and type VI secretion modules. *Scientific Reports*, *6*(1), 36891. https://doi.org/10.1038/srep36891