Genomic Science in Understanding Cholera Outbreaks and Evolution of *Vibrio cholerae* as a Human Pathogen

William P. Robins[1] and **John J. Mekalanos**[1]

[1]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA

Abstract

Modern genomic and bioinformatic approaches have been applied to interrogate the *V. cholerae* genome, the role of genomic elements in cholera disease, and the origin, relatedness, and dissemination of epidemic strains. A universal attribute of choleragenic strains includes a repertoire of pathogenicity islands and virulence genes, namely the CTX–φ prophage and Toxin Co-regulated Pilus (TCP) in addition to other virulent genetic elements including those referred to as Seventh Pandemic Islands. During the last decade, the advent of Next Generation Sequencing (NGS) has provided highly resolved and often complete genomic sequences of epidemic isolates in addition to both clinical and environmental strains isolated from geographically unconnected regions. Genomic comparisons of these strains, as was completed during and following the Haitian outbreak in 2010, reveals that most epidemic strains appear closely related, regardless of region of origin. Non-O1 clinical or environmental strains may also possess some virulence islands, but phylogenic analysis of the core genome suggests they are more diverse and distantly related than those isolated during epidemics. Like Haiti, genomic studies that examine both the Vibrio core- and pan-genome in addition to Single Nucleotide Polymorphisms (SNPs) conclude that a number of epidemics are caused by strains that closely resemble those in Asia, and often appear to originate there and then spread globally. The accumulation of SNPs in the epidemic strains over time can then be applied to better understand the evolution of the *V. cholerae* genome as an etiological agent.

1 Introduction

Even with modern established treatments and preventative measures, *V. cholerae* continues to emerge as a dangerous pathogen. This is especially true in Southeast Asia where the yearly appearance of cholera cases follow predicted patterns or "seasons" (Russell 1925; Pascual et al. 2002). However, cholera epidemics have an unpredictable history of manifesting on a much larger scale and the accounts of global pandemics are well-defined and documented during the last two centuries. The seven recognized cholera pandemics can be traced from Asia and the Indo- Pacific region; the current 7th pandemic manifesting as two waves, one between 1961 and 1966 and the other spreading to much of the world after 1970 (Karaolis et al. 1994). Most pandemic strains possess the lipopolysaccharide antigen groups O1 and to a lesser extent O139. Apart from the antigen, O1/O139 strains are highly

Correspondence to: William P. Robins.

conserved at the nucleotide level and possess a number of common related genetic islands (GIs). The assemblage of GIs and subtle variance at the nucleotide level provides a useful scaffold for determining the genetic relatedness of both epidemic and environmental strains. A key theme in the study of cholera disease is how environmental factors and host-pathogen interactions influence genetic variability in *V. cholerae* genes and GIs and as well as their relatedness to virulence. The current repertoire of genomic tools has begun to answer these questions.

2 Conventional Genomics and Established Virulence Islands

The completion of the El Tor N16961 strain genome sequence in the year 2000 confirmed and identified the positions of the recognized 7th pandemic virulence islands (Heidelberg et al. 2000). Whereas established methods such as ribotyping (Wachsmuth et al. 1993), pulsed field gel electrophoresis (Weber et al. 1994), and multilocus enzyme analysis/sequence typing (Byun et al. 1999; Li et al. 2003; Lam et al. 2012) examined genomic complexity of isolated strains by probing the sequences of individual genes and regions or sizes of generated distinct fragments, the thoroughness of these examinations were limited by their design.

Though useful, this level of resolution was still inadequate; for example, it was only recently during the late 1990s when it was discovered that most if not all Vibrio species possessed two chromosomes (Trucksis et al. 1998; Yamaichi et al. 1999). In whole, the deciphered genomic sequence of this 7th pandemic strain has paved the way for tools such as microarray analysis (Dziejman et al. 2002) and parallel whole genome sequencing (Grim et al. 2010; Mutreja et al. 2011) to better understand genomic complexity of *V. cholerae* and gene expression at nucleotide resolution (Mandlik et al. 2011). Most recently, the de novo assembly of the closed and complete sequences of both chromosomes I and II by coupling different modern sequencing platforms such as Illumina, Sanger 454, and Pacific Biosciences exemplifies how technology has pushed Vibrio genomics past the limits of the basic molecular typing methods (Bashir et al. 2012).

The sequence and assemblage of both virulence-associated GIs and housekeeping genes now provides a highly resolved template that has proved useful for further interrogation. In the approximate 4 Mb of genomic DNA present in the large chromosome I (~3 MB) and small chromosome II (~1 MB) many of the core genes (~1500) are well-conserved in both O1 and non-O1 serotypes (Vesth et al. 2010). However, among these core genes are interspersed GIs and phage related sequences. Some of these are demonstrated to contribute to pathogenicity. Of the GIs found in various *V. cholerae* serotypes, we know of a handful that exhibit high conservation in more virulent isolates. Some GIs such as toxin co-regulated pilus (TCP) island and the integrated phage CTX–φ together regulate and encode expression of cholera toxin, the protein toxin that causes the most of the cholera diarrheal syndrome.

Using the genome reconstruction of the recent Haitian outbreak strain (H1, 2010) as a reference and comparator, and then mapping the nucleotide and protein identity from a diverse panel of Vibrio strains, the GIs specific to 7th pandemic strains are apparent when aligned and illustrated as a Blast Atlas (Fig. 1) (Vesth et al. 2013). Apart from these GIs, the

a bulk of the DNA sequence from 7th pandemic El Tor O1 isolates is nearly identical, thus making these islands useful for identification of virulent strains and genomic analysis. These genetic elements include Vibrio 7th Pandemic island I and II (VSPI and VSPII), TCP, Vibrio pathogenicity island II (VPI-II), CTX–φ and the Toxin linked cryptic (TLC) (Dziejman et al. 2002; Faruque and Mekalanos 2003; Chun et al. 2009). Another mobile element, SXT, was first identified in O139 clinical isolates during the 1990s. SXT is a self-transmissible integrating conjugative element that encodes antibiotic resistance and is now found to be present in almost all O1 clinical isolates post early 1990s (Waldor et al. 1996). In all Vibrio species, a large and hyper variable segment of chromosome II called the superintegron (SI) encodes more than 200 open reading frames that exist as individual mobile gene cassettes (MGC). The conservation of the collective GIs in sequenced clinical isolates has facilitated the use of genomic analysis to better elucidate their requisite role in virulence and pathogenesis.

Core and Pan-Genome

Collections of genome sequences are used to extrapolate the conserved minimal core and expanded pan genome/pan proteome within bacterial phylogenies. The core genome is demonstrated to be one the optimum datasets for determining phylogenic relationships (Rokas et al. 2003) and is appreciably more practical in determining phylogenetic reconstruction than using 16S/23S rRNA or a handful of essential housekeeping genes, especially when applied to examine closely related taxa. An initial comparative study of 23 diverse *V. cholerae* strains found 2,432 common core genes or orthologues and 6,953 total unique genes in the pan genome (Chun et al. 2009). This core genome assemblage from both O1 and non-O1 strains integrated essential and nonessential genes and excluded the major virulence islands. This report proposed 7th pandemic strains were one of 12 lineages in this group that share a very similar backbone and hence probably all originated from a common ancestor. One major difference between the predicted lineages in this study were GIs outside the core genome and it was proposed that a driving force in diversity of *V. cholerae* can be attributed to horizontal gene transfer of GIs. In similar work, Vesth et al. examined the core and extended pan genomes of 32 Vibrio isolates represented by 12 species (Vesth et al. 2010). Of the 18 represented *V. cholerae* representatives, core genes comprised nearly half the genome and the pan-genome spanned a total of 6,500 genes. The divergence within *V. cholerae* species was minor when compared to that for a broad representation of the Vibrio genus; the core genome decreased to only 1,000 families while the pan-genome expanded to ~20,000. When these aligned core genes were used to calculate phylogeny, *V. cholerae* strains were identified again as closely related, however, environmental or clinical *V. cholerae* strains diverged. Placement within the tree was generally correlated with virulence for the O1 strains; however the presence of these GIs did not absolutely dictate phylogeny. For example, M66-2 is a clinical isolate from 1937 Indonesia that clusters with toxigenic *V. cholerae*, but it that lacks the CTX–φ prophage, and thus may represent a direct ancestor to current pandemic/epidemic *V. cholerae* (Feng et al. 2008). This work also concluded that the clinical non-O1 strain 2740-80 appeared to be an intermediate between the clinical, toxigenic isolates, and those found generally in the environment. These features in the genomes of 2740-80 and M66-2 are relevant to the field because understanding the relationship between the core-genome variation and the influence of acquired GIs on

pathogenicity is an important aim of *V. cholerae* genomics. Because these bacteria can occupy multiple niches in both the estuarine biosphere and the human gut, the fitness for certain genes or GIs may vary in these environments and concomitantly be influenced by the genetic backbone. Thus, some strains may be better adapted for the only the aquatic environment and others also more fit for survival in the human gut. Surveys of environmental strains show that only a small percentage possess the phage that encodes cholera toxin and that a minority have the GI repertoire typically found in pandemic strains (Mukhopadhyay et al. 2001; Faruque et al. 2004; Rahman et al. 2008). It is also impossible to determine if these environmental strains simply represent contamination from cholera victims or free-living aquatic isolates (see below).

Distinct subgroups within the 7th pandemic El Tor clinical strains are identified. For example, during the last couple of decades, closely related O1 El Tor subgroups responsible for periodic epidemics appeared to originate in Southeast Asia and then to emanate as waves (Mutreja et al. 2011). Within the panel of clinical O1 genomes surveyed in this work nearly all genes are identical and the total number unique Single Nucleotide Polymorphisms (SNPs) between individual strains is enough to identify subtle sublineages. To illustrate phylogeny of 7th pandemic sublineages, using ~1.1 million nucleotides that encode a set of the unique nonredundant core genome (881 genes) in a panel of 56 strains from a number of independent studies and sources (Fig. 2b), it can be shown that the relatedness of 7th pandemic strains often correlate well with both date of isolation and geographical incidence and these construed phylogenies are similar to other independent and previous analyses. The strains isolated during the Peru epidemic of the early 1990s most closely resemble other strains isolated in Latin America during the 1990s. Separate from phylogeny based entirely on core genome analysis, this relationship is also evident in the genetic islands as unique deletions in VSP-II and by the presence of a prophage (Nusrin et al. 2009; de Sa Morais et al. 2012). Likewise, both MAK757 and M66-2 (1937, Indonesia) strains appear related and together diverge from current, more contemporary isolates. The entire group of O1 strains derived from Southern Asia (Pakistan, India, Nepal, and Thailand) in the period 2000–2011 clusters with the Haitian outbreak of 2010. Accordingly, the CDC and a number of independent investigators have concluded that the importation of *V. cholerae* into Haiti from Nepal or Southeast Asia is probable. Coincidently, a number of strains in this presented phylogenetic dataset (India 2005, Pakistan 2009, India 2010) include those isolated from stool samples in the USA from infected persons traveling abroad, thus directly demonstrating the opportunity for global dissemination of pathogenic strains. There are some exceptions; an Australian strain isolated in 1986 appears closely related to a 1910 Saudi Arabian sample. It is important to also keep in mind that single strain epidemiological oddities such as this example might find more parsimonious explanations by hypothesizing laboratory cross contamination rather than a complicated theory involving global transmission of a comparably "old" strain via either environmental or human sources. Clearly, the increased resolution of these predicted phylogenies will extend our understanding of the epidemiology. However, it is important to maintain perspective and note the similarity within pandemic O1 genomes as a group in contrast to others (Fig. 2a) when the phylogeny is extended to include a broad representation of non-O1 *V. cholerae* genomes. Thus, the occasional phylogenic mapping of a putative clinical or environmental

isolate of an O1 isolate in a otherwise diverged group of non-O1 strains, can be discounted as significant if the O1 isolate lacks most or all of the essential virulence genes present on typical genetic islands. Clearly, most O1 and O139 strains that cause clinical cholera and more specifically cholera epidemics are highly related from the genomic perspective and carry highly similar collection of virulence genes encoded by accessory genetic elements (Faruque and Mekalanos 2003).

TCP and CTX–ϕ Variation in O1 strains

In contrast to whole genome analysis, the targeted surveillance of mobile virulence islands can be applied to interrogate their role in pathogenesis. TCP (also called VPI) and CTX–φ are two virulent GIs most closely connected to cholera disease. Though acquired as separate entities, there exists a relationship between the two. The TCP gene cluster encodes a bundled type IV-like pilus that contributes to *V. cholerae* intestinal colonization during infection and also serves a dual role as the receptor for CTX–φ temperate phage acquisition (Kirn et al. 2000). The 6.9 kb genome of filamentous phage CTX–φ carries genes for both enterotoxin CTXA/B subunits in its core genome and thus can convert CTX–φ negative Vibrio into toxigenic bacteria (Waldor and Mekalanos 1996). Furthermore, the TCP-encoded ToxT/ToxS/TcpH gene products work with ToxR to regulate a number of virulence genes during infection including the CTX–φ genes for cholera toxin (Miller et al. 1989; Krukonis et al. 2000; Beck et al. 2004).

Some clinical O1 strains and non-O1 environmental *V. cholerae* isolates have been identified that possess either TCP or TCP/CTX together, but isolates that carry only CTX–φ without the TCP island are rare in nature (Li et al. 2003). It is widely accepted that TCP/CTX–φ positive O1 strains cause most modern cholera outbreaks while most non-O1/O139 cause typically less severe disease that is sporadic and seldom observed to cause epidemics although occasionally these do occur (Dalsgaard et al. 1995, 2001). The horizontal transfer of TCP and CTX–φ between O1 and non-O1/O139 strains is poorly documented and indeed this suggests that most non-O1/non-O139 serogroups do not gain benefit from these virulence elements. Given that the environment is dominated by non-O1/non-O139 strains, this observation suggests that these virulence factors probably do not promote fitness in the aquatic niche although there is some data that challenges this conclusion for at least chitin binding lectins (Kirn et al. 2005). Various typing methods and deep sequencing has enabled more sensitive surveillance of CTX–φ in strains isolated from the environment (Rivera et al. 2001; Singh et al. 2001; Pang et al. 2007; Awasthi et al. 2012; Hasan et al. 2012; Sealfon et al. 2012; Sellek et al. 2012). The prevalent assumption that the presence of toxigenic strains in water provides proof that toxigenic strain occupy a stable environmental niche needs to be readdressed with modern methods that might differentiate between contamination of the environment by nearby cholera victims as opposed to these strains maintaining themselves within the environment in the absence of infected humans.

The origin of TCP in Vibrio is unknown. Preliminary evidence suggested TCP may be a phage (Karaolis et al. 1999), but further investigation has failed to support this conclusion (Davis and Waldor 2003; Faruque et al. 2003). Given the observation that TCP assists biofilm differentiation on chitinaceous surfaces (Reguera and Kolter 2005) and chitin

induces natural competence (Meibom et al. 2005), the acquisition of the TCP island by different Vibrio strains could occur independent of a phage-mediated transduction events (Faruque et al. 2003).

Subtle variability in genetic organization of TCP has been noted in both epidemic and nonepidemic strains (Ghosh et al. 1997; Novais et al. 1999; Mukhopadhyay et al. 2001). The ~41 kb TCP island carries a putative integrase and transposase and is integrated at the tmRNA (*ssrA*) gene in chromosome I. The region is especially A/T rich when compared to that measured in Vibrio species and varies among *V. cholerae* strains. Most of the highly conserved region of the TCP island encodes genes required for the biosynthesis of a type IV pilus structure and a minority of the element regulates transcription of genes both inside and outside of the TCP element. A majority of divergent sequences in TCP are either noncoding or are identified as deletions and inserted elements/transposases (Hase and Mekalanos 1998; Yu and DiRita 1999; Mukhopadhyay et al. 2001).

The chromosomal integration structure of the temperate phage genome differs between classical and El Tor 7th pandemic strains and there are some identified amino acid changes at positions within key genes (Fig. 3). Integrated CTX–φ is found in the either of the chromosomes as a single copy or as a tandem array (Mekalanos 1983). In El Tor strains, CTX–φ, the adjacent RS1/RS2 regions, and TLC island are shown to be pivotal in the stable acquisition of CTX/ and maintaining the dif site, a sequence important for chromosomal dimer resolution in chromosome I (Rubin et al. 1998; Hassan et al. 2010). Classical strain O395 harbors the CTX–φ on both chromosomes I and II and lacks the adjacent RS1 cassette found in El Tor strain and instead has RS2 (Waldor et al. 1997; Davis et al. 2002). RS2 and RS1 are two very similar genetic regions that encode phage replication, regulation and integration genes for the CTX–φ prophage, however RS1 possesses an additional antirepressor gene *rstC*. Expression of *rstC* antirepressor may be important as it is shown to relieve RtsR repression of CTX genes. Presence of *rstC* has been shown to increase CTX–φ production as much as several thousandfold and also increases the production of rstA transcripts, some of which certainly extend through and include *ctxAB* (Davis et al. 2002). Furthermore, the variation between the RtsR repressor in El Tor and Classical strains is sufficient to confer biotype specificity (Kimsey and Waldor 1998). The other identified class of CTX–φ variation in lies within the ctxB gene. These are recognized nonsynonymous variants found in either classical and El tor (and variant) strains and have been utilized as benchmarks in genomic comparisons.

First noted in Calcutta 1990 strains and Bangladesh in 2001, a growing number of 7th pandemic El Tor strains harbor atypical CTX regions (Nair et al. 2002; Raychoudhuri et al. 2009). El Tor variants have become the predominant biotype isolated in Asia, Africa, and more recently Haiti (Nair et al. 2002, 2006; Ansaruzzaman et al. 2004; Safa et al. 2006). Instead of the consensus El-Tor CTX–φ RS1 sequence, these variants carry the *ctxB* gene that possesses one or several nonsynonymous variants in addition to a varying number of heptad consensus repeats in the consensus upstream ToxT binding site while often maintaining the El Tor RS1 region. The variants also possess specific amino acid residues in the *ctxB* gene that are also found in classical strains.

Preliminary functional and biochemical study of variants suggests they are located on exposed CtxB protein–protein interfaces within multimer CtxB pore and, therefore, could influence CtxB or CtxA binding dynamics (Shamini et al. 2011). The measured amount of cholera toxin produced by a panel of these is shown to be significantly increased when compared to classical and other El Tor strains (Ghosh-Banerjee et al. 2010; Son et al. 2011). However, the measured increase in toxin production and a concomitant hypervirulence by variants in the infant mouse model in this work were also attributed to increased levels of ToxT and TcpA.

Superintegron

Integrons are natural cloning systems that consist of open reading frames that use flanking sites and site-specific recombinases/integrases to entrap and acquire gene cassettes and promoters into a larger array of genes. The superintegron is a novel integron class first discovered in *V. cholerae* and is represented in current epidemic strains as a large assemblage of these integrated genes (more than 200) that spans more than 100KB of Chromosome II (Mazel et al. 1998). Many of the ORFs code genes that have unknown function and are largely nonessential for *V. cholerae* growth or pathogenesis. They are an interesting reservoir for horizontally exchanged, yet unrelated genes and their significance is not well-understood.

The evidence for essential or important genes in the SI in *V. cholerae* is generally lacking. Microarray and RNA-seq studies measure little to almost no expression across the entire region while growing in liquid media or during infection (Larocque et al. 2005; Mandlik et al. 2011). The only measurement for upregulation of these genes is during cell growth at high density or while undergoing stress (Yildiz et al. 2004). A handful of genes do encode putative virulence factors including the mannose-fucose-resistant hemagglutinin (*mrhA*), heat-stable toxin (*sto*), and a lipoprotein gene (Ogawa and Takeda 1993; Barker and Manning 1997; MacDonald et al. 2006). However, because of its dynamic nature and variation, the SI has been most useful for genomic analysis and a fingerprint for strain comparisons. Even closely related strains are sometimes found to possess minor variability. Gene cassettes can be duplicated and present as multiple copies in separate locations, probably produced by duplication and not horizontal transfer (Rowe-Magnus et al. 2003; Feng et al. 2008).

The presence of highly conserved flanking repeat sequences in SI has proved useful for amplification-based genomic analyses. The amplification of unique genes using primers that anneal to the flanked 59 bp repeat region has been used as a molecular fingerprint to identify and profile collections of *V. cholerae* strains (Labbate et al. 2007). The examined dynamic array of MGCs for 60 strains isolated between 1961 and 2008 found more diversity in those prior to 1980s and a predominant SI structure found in most current strains since the 1990s (Gao et al. 2011).

3 Application of Genomics

Regardless of serotype or strain, the treatment of cholera disease is the same, oral and intravenous rehydration with concurrent replacement of electrolytes. Antibiotic therapy is

recommended for severely ill patients but is not a first-line treatment. Though the most thorough current genomic analyses may fail to provide information relevant to the primary treatment of cholera, a case can be made for genomic surveillance in order to advance and adapt preventative measures for a number of reasons. First, the genomic content and SNPs can be compared to available sequence databases of environmental and global epidemic strains to better elucidate and determine the origin and dissemination. It is conceivable that the genome may someday be used in order to accurately predict magnitude of virulence. Second, the genomic analysis of emerging strains may be useful when evaluating the efficacy of vaccines and in their design. For example, O1 vaccine strains and strategies utilize expression and secretion of the nontoxic CtxB subunit to enhance efficacy (Thungapathra et al. 1999; Jani et al. 2002; Liang et al. 2003; Qadri et al. 2007; Yan et al. 2007). There is an emergence of the strains that encode classical-El Tor CTX islands with unique variants in the *ctxB* gene. The role of these amino acid residues during infection is unclear; incorporating this allele into vaccine strains may be useful.

New genomic tools and methods were applied to rapidly identify genomic features of the strain during the recent Haitian outbreak in 2010 (Chin et al. 2011). This represented the first cholera epidemic to be rapidly analyzed with a number of next generation sequencing platforms, even while some of these technologies were still in development. The emergence of this strain in Hispaniola, an island that had not experienced cholera in nearly 100 years posed several fundamental questions about the epidemiology and spread of *V. cholerae* that may in part be addressed with interrogation of the genomic content. It is obvious that the lack of sanitation and damaged infrastructure following the 2010 Haiti earthquake certainly exacerbated the spread of the cholera cases, but the understanding of the mode of emergence and transmission became an important aim.

One speculation was that the Haitian strain may have re-emerged from a reservoir in Latin America. Following the Peru pandemic that spread into Latin America in the 1990s, modern cholera epidemics in the Western Hemisphere have been generally sporadic but appear to be related. After the Peru pandemic, similar strains were found from Mexico to Brazil in years following and appeared to be monophyletic (Wachsmuth et al. 1993; Lam et al. 2010; Chin et al. 2011; Mutreja et al. 2011; de Sa Morais et al. 2012; Garza et al. 2012). The Latin American epidemic strains sequenced during this period possessed a unique 40 kb prophage inserted in within the alanine aminopeptidase gene (de Sa Morais et al. 2012). As a group these strains were found to be most closely resemble a strain from Angola in1989 and other independent phylogenetic analyses also hypothesized a common African origin which is possibly linked to immigration to Lima during the corresponding period before the 1991 epidemic (Mutreja et al. 2011).

The rapid sequencing and analysis of the Haitian strain in 2010 and all subsequent work provided almost irrefutable evidence for human transmission of a clonal strain of recent Asian origin (Ali et al. 2011; Chin et al. 2011; Hendriksen et al. 2011; Reimer et al. 2011; Sjolund-Karlsson et al. 2011; Talkington et al. 2011; Frerichs et al. 2012). The Haitian strain was distinctly atypical from Latin American strains isolated during and following the Peru 1991 epidemic and possessed a hybrid Classical/El Tor CTX and ICEVchInd5 type SXT, not yet detected in the Western hemisphere and more typical of those strains isolated in

Southeast Asia (Chin et al. 2011). Furthermore, an identified unique deletion in within VPS-II and an assemblage of genes found in the SI most closely matched strains previously characterized in Asia during the previous 8 years. Phylogenic clustering of a panel of El Tor isolates placed more distance between Haiti and both the Peru C6706 strain and the 1971 reference strain N16961 and highest similarity to two strains isolated in Bangladesh; a 2008 strain (M4), co-sequenced with the Haitian isolate, and CIRS101 isolated in 2002. Moreover, the first Haitian cases were noted in the isolated upper Artinobite river valley proximal to a United Nations camp occupied by troops from Nepal where cholera cases had been documented during the preceding months. This accidental importation and dissemination is supported by the published conclusion of several independent investigators and a UN special panel (UN 2011). Virtual definitive proof that the Haiti epidemic had an origin in Nepal came from the application of genomic analysis to isolates from obtained from Katmandu patients only weeks before the Haiti epidemic as described below.

In one of the most comprehensive and conclusive genomic analyses, the genomes of 24 Nepalese genomes were compared to ten previously sequenced genomes including three from the recent Haitian outbreak (Hendriksen et al. 2011). Using whole-genome sequence typing and phylogenetic analysis, a cluster of strains was found to be most closely related to the Haitian strains. Remarkably, two Nepal strains varied from the Hatian strains by only one or two base pair variations. Additional characterization of 77 Nepalese strains collected from ten different hospital laboratories from 2007 to 2010 clustered strains into four different groups using MLVA and *ctxB* gene typing (Shakya et al. 2012). Many of the MLVA patterns in this work matched clinical strains identified in adjacent Southeastern Asian countries and the 2008–2010 strains possessed the same CTX 3b-type toxin. A common theme in this work and other genetic studies is that regional outbreaks appear to be clonal or closely related and also resemble clinical strains isolated and identified in neighboring regions, suggesting regional dissemination. As demonstrated, it should be possible to develop phylogenies that encompass strains isolated from more regions that span multiple seasons or years to better follow evolution and dissemination of epidemic strains.

Current technology has enabled a more robust, deeper surveillance of variation in the pathogenic genome. The analytic value of these data includes understanding its rate of evolution. Recent work shows that we must be mindful in how we apply or interpret significance of measured genetic variation. Using sequence variants between an Indonesian prepandemic 1937 El Tor strain and also classical 6th and modern 7th pandemic strains Feng and colleagues attempted to calculate a "molecular clock" (Feng et al. 2008). The aim was to document changes that had occurred during the divergence of the three clones and estimate the rate of mutation. Though this dataset was limited to the Classical 6th pandemic strain O395 and two El Tor strains from 1937 to 1971, the results suggested mutation rates that were ~100 times higher than had been previously assumed. In other independent work, Mutreja et al. concluded that recent epidemic strains isolated during three waves of global transmission in the last 50 years have accumulated about 2.3–3.5 SNPs/year in the core genome. The selective factors that influence this molecular or evolutionary clock include but are not limited to interactions within the host and environmental factors such as bacteriophage. Both clinical data and mathematical studies support a model where lytic

phages may be important in ending outbreaks (Faruque et al. 2005a, b; Jensen et al. 2006). Recent work has evaluated genomic characterization of phages year-to-year in cholera prone areas using coupled-genomic approaches (Seed et al. 2011) and host-pathogen interactions using microarray and RNA-seq (Larocque et al. 2005; Mandlik et al. 2011).

4 Conclusions

Genomic science has greatly improved our understanding of pathogenic clones of *V. cholerae* and their relationship to relatively nonpathogenic environmental strains of this bacterial species. Clearly, gene content (largely driven by phage and GI acquisition, and superintegron and ICE element variation) defines a pathogenic O1 and O139 7th pandemic El Tor lineage of this organism that is separated distinctly from the most common ancestor of nontoxigenic non-O1/non-O139 strains present in environmental waters throughout the world (Boyd and Waldor 2002; Li et al. 2003; Pang et al. 2007; Rahman et al. 2008; Vesth et al. 2010). The suggestion that climate factors drive emergence of new pathogenic strains and genetic exchange between these distinctly different groups of *V. cholerae* (Hasan et al. 2012) is an interesting but largely speculative idea that simply has no strong support at the genome sequence level (Mekalanos et al. 2012; Katz et al. 2013) or epidemiological level (Gaudart et al. 2013a, b). Indeed, human activity (travel and poor sanitation) seems the most probable source of typical pathogenic clones globally over the last century and particularly over the three decades during which cholera has established itself as a threat to Africa, Latin America, and the Caribbean. These pathogenic clones certainly have and will undergo further evolution as they have in the case of the variant strains that now dominate cholera endemic and epidemic locales throughout the world. However, applying available genomic tools to other components in the aquatic environment (e.g. phage and microbiome) may someday define whether the key virulence genes of pathogenic *V. cholerae*, have an origin in a non-O1 *V. cholerae* (Haley et al. 2013) or even non-Vibrio bacterial species which may or not be pathogens of humans. Understanding the additional biological niches that *V. cholerae* virulence genes occupy will help define a better model for emergence of pathogenic clones of *V. cholerae.*

References

Ali A, Chen Y, Johnson JA, Redden E, Mayette Y, Rashid MH, et al. Recent clonal origin of cholera in Haiti. Emerg Infect Dis. 2011; 17:699–701. [PubMed: 21470464]

Ansaruzzaman M, Bhuiyan NA, Nair BG, Sack DA, Lucas M, Deen JL, et al. Cholera in Mozambique, variant of Vibrio cholerae. Emerg Infect Dis. 2004; 10:2057–2059. [PubMed: 16010751]

Awasthi SP, Asakura M, Chowdhury N, Neogi SB, Hinenoya A, Golbar HM, et al. Novel cholix toxin variants, an ADP-ribosylating toxin in Vibrio cholerae non-O1/non-O139 strains and their pathogenicity. Infect Immun. 2012; 81(2):531–541. [PubMed: 23230295]

Barker A, Manning PA. VlpA of Vibrio cholerae O1: the first bacterial member of the alpha 2-microglobulin lipocalin superfamily. Microbiology. 1997; 143(Pt 6):1805–1813. [PubMed: 9202455]

Bashir A, Klammer AA, Robins WP, Chin CS, Webster D, Paxinos E, et al. A hybrid approach for the automated finishing of bacterial genomes. Nat Biotechnol. 2012; 30(7):701–707. [PubMed: 22750883]

Beck NA, Krukonis ES, DiRita VJ. TcpH influences virulence gene expression in Vibrio cholerae by inhibiting degradation of the transcription activator TcpP. J Bacteriol. 2004; 186:8309–8316. [PubMed: 15576780]

Boyd EF, Waldor MK. Evolutionary and functional analyses of variants of the toxin- coregulated pilus protein TcpA from toxigenic Vibrio cholerae non-O1/non-O139 serogroup isolates. Microbiology. 2002; 148:1655–1666. [PubMed: 12055286]

Byun R, Elbourne LD, Lan R, Reeves PR. Evolutionary relationships of pathogenic clones of Vibrio cholerae by sequence analysis of four housekeeping genes. Infect Immun. 1999; 67:1116–1124. [PubMed: 10024551]

Chiang SL, Taylor RK, Koomey M, Mekalanos JJ. Single amino acid substitutions in the N-terminus of Vibrio cholerae TcpA affect colonization, autoagglutination, and serum resistance. Mol Microbiol. 1995; 17:1133–1142. [PubMed: 8594332]

Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. N Engl J Med. 2011; 364:33–42. [PubMed: 21142692]

Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. Proc Natl Acad Sci USA. 2009; 106:15442–15447. [PubMed: 19720995]

Craig L, Taylor RK, Pique ME, Adair BD, Arvai AS, Singh M, et al. Type IV pilin structure and assembly: X-ray and EM analyses of Vibrio cholerae toxin-coregulated pilus and Pseudomonas aeruginosa PAK pilin. Mol Cell. 2003; 11:1139–1150. [PubMed: 12769840]

Dalsgaard A, Albert MJ, Taylor DN, Shimada T, Meza R, Serichantalergs O, Echeverria P. Characterization of Vibrio cholerae non-O1 serogroups obtained from an outbreak of diarrhea in Lima. Peru J Clin Microbiol. 1995; 33:2715–2722.

Dalsgaard A, Serichantalergs O, Forslund A, Lin W, Mekalanos J, Mintz E, Shimada T, Wells JG. Clinical and environmental isolates of Vibrio cholerae serogroup O141 carry the CTX phage and the genes encoding the toxin-coregulated pili. J Clin Microbiol. 2001; 39:4086–4092. [PubMed: 11682534]

Davis BM, Kimsey HH, Kane AV, Waldor MK. A satellite phage-encoded antirepressor induces repressor aggregation and cholera toxin gene transfer. EMBO J. 2002; 21:4240–4249. [PubMed: 12169626]

Davis BM, Waldor MK. Filamentous phages linked to virulence of Vibrio cholerae. Curr Opin Microbiol. 2003; 6:35–42. [PubMed: 12615217]

de Sa Morais LL, Garza DR, Loureiro EC, Nunes KN, Vellasco RS, da Silva CP, et al. Complete genome sequence of a sucrose-nonfermenting epidemic strain of Vibrio cholerae O1 from Brazil. J Bacteriol. 2012; 194:2772. [PubMed: 22535947]

Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ. Comparative genomic analysis of Vibrio cholerae: genes that correlate with cholera endemic and pandemic disease. Proc Natl Acad Sci USA. 2002; 99:1556–1561. [PubMed: 11818571]

Faruque SM, Chowdhury N, Kamruzzaman M, Dziejman M, Rahman MH, Sack DA, et al. Genetic diversity and virulence potential of environmental Vibrio cholerae population in a cholera-endemic area. Proc Natl Acad Sci USA. 2004; 101:2123–2128. [PubMed: 14766976]

Faruque SM, Islam MJ, Ahmad QS, Faruque AS, Sack DA, Nair GB, et al. Self-limiting nature of seasonal cholera epidemics: role of host-mediated amplification of phage. Proc Natl Acad Sci USA. 2005a; 102:6119–6124. [PubMed: 15829587]

Faruque SM, Mekalanos JJ. Pathogenicity islands and phages in Vibrio cholerae evolution. Trends Microbiol. 2003; 11:505–510. [PubMed: 14607067]

Faruque SM, Zhu J, Asadulghani, Kamruzzaman M, Mekalanos JJ. Examination of diverse toxin-coregulated pilus-positive Vibrio cholerae strains fails to demonstrate evidence for Vibrio pathogenicity island phage. Infect Immun. 2003; 71:2993–2999. [PubMed: 12761075]

Faruque SM, Islam MJ, Ahmad QS, Faruque AS, Sack DA, Nair GB, Mekalanos JJ. Self- limiting nature of seasonal cholera epidemics: role of host-mediated amplification of phage. Proc Natl Acad Sci USA. 2005b; 102:6119–6124. [PubMed: 15829587]

Feng L, Reeves PR, Lan R, Ren Y, Gao C, Zhou Z, et al. A recalibrated molecular clock and independent origins for the cholera pandemic clones. PLoS ONE. 2008; 3:e4053. [PubMed: 19115014]

Frerichs RR, Keim PS, Barrais R, Piarroux R. Nepalese origin of cholera epidemic in Haiti. Clin Microbiol Infect. 2012; 18:E158–E163. [PubMed: 22510219]

Gao Y, Pang B, Wang HY, Zhou HJ, Cui ZG, Kan B. Structural variation of the superintegron in the toxigenic Vibrio cholerae O1 El Tor. Biomed Environ Sci. 2011; 24:579–592. [PubMed: 22365393]

Garza DR, Thompson CC, Loureiro EC, Dutilh BE, Inada DT, Junior EC, et al. Genome- wide study of the defective sucrose fermenter strain of Vibrio cholerae from the Latin American cholera epidemic. PLoS ONE. 2012; 7:e37283. [PubMed: 22662140]

Gaudart J, Moore S, Rebaudet S, Piarroux M, Barrais R, Boncy J, et al. Environmental factors influencing epidemic cholera. Am J Trop Med Hyg. 2013a; 89:1228–1230. [PubMed: 24306033]

Gaudart J, Rebaudet S, Barrais R, Boncy J, Faucher B, Piarroux M, et al. Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti. PLoS Negl Trop Dis. 2013b; 7:e2145. [PubMed: 23593516]

Ghosh-Banerjee J, Senoh M, Takahashi T, Hamabata T, Barman S, Koley H, et al. Cholera toxin production by the El Tor variant of Vibrio cholerae O1 compared to prototype El Tor and classical biotypes. J Clin Microbiol. 2010; 48:4283–4286. [PubMed: 20810767]

Ghosh C, Nandy RK, Dasgupta SK, Nair GB, Hall RH, Ghose AC. A search for cholera toxin (CT), toxin coregulated pilus (TCP), the regulatory element ToxR and other virulence factors in non-01/non-0139 Vibrio cholerae. Microb Pathog. 1997; 22:199–208. [PubMed: 9140915]

Grim CJ, Hasan NA, Taviani E, Haley B, Chun J, Brettin TS, et al. Genome sequence of hybrid Vibrio cholerae O1 MJ-1236, B-33, and CIRS101 and comparative genomics with *V. cholerae*. J Bacteriol. 2010; 192:3524–3533. [PubMed: 20348258]

Haley BJ, Choi SY, Hasan NA, Abdullah AS, Cebula TA, Huq A, et al. Genome sequences of clinical Vibrio cholerae isolates from an oyster-borne cholera outbreak in Florida. Genome Announc. 2013; 1(6):e00966–13. [PubMed: 24265497]

Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013; 30(5): 1229–1235. [PubMed: 23486614]

Hang L, John M, Asaduzzaman M, Bridges EA, Vanderspurt C, Kirn TJ, et al. Use of in vivo-induced antigen technology (IVIAT) to identify genes uniquely expressed during human infection with Vibrio cholerae. Proc Natl Acad Sci USA. 2003; 100:8508–8513. [PubMed: 12826608]

Hasan NA, Choi SY, Eppinger M, Clark PW, Chen A, Alam M, et al. Genomic diversity of 2010 Haitian cholera outbreak strains. Proc Natl Acad Sci USA. 2012; 109:e2010–e2017. [PubMed: 22711841]

Hase CC, Mekalanos JJ. TcpP protein is a positive regulator of virulence gene expression in Vibrio cholerae. Proc Natl Acad Sci USA. 1998; 95:730–734. [PubMed: 9435261]

Hassan F, Kamruzzaman M, Mekalanos JJ, Faruque SM. Satellite phage TLCphi enables toxigenic conversion by CTX phage through dif site alteration. Nature. 2010; 467:982–985. [PubMed: 20944629]

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature. 2000; 406:477–483. [PubMed: 10952301]

Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of Vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak. MBio. 2011; 2:e00157–11. [PubMed: 21862630]

Herrington DA, Hall RH, Losonsky G, Mekalanos JJ, Taylor RK, Levine MM. Toxin, toxin-coregulated pili, and the toxR regulon are essential for Vibrio cholerae pathogenesis in humans. J Exp Med. 1988; 168:1487–1492. [PubMed: 2902187]

Iredell JR, Manning PA. Biotype-specific tcpA genes in Vibrio cholerae. FEMS Microbiol Lett. 1994; 121:47–54. [PubMed: 7915998]

Jani D, Meena LS, Rizwan-ul-Haq QM, Singh Y, Sharma AK, Tyagi AK. Expression of cholera toxin B subunit in transgenic tomato plants. Transgenic Res. 2002; 11:447–454. [PubMed: 12437076]

Jensen MA, Faruque SM, Mekalanos JJ, Levin BR. Modeling the role of bacteriophage in the control of cholera outbreaks. Proc Natl Acad Sci USA. 2006; 103:4652–4657. [PubMed: 16537404]

Karaolis DK, Lan R, Reeves PR. Molecular evolution of the seventh-pandemic clone of Vibrio cholerae and its relationship to other pandemic and epidemic *V. cholerae* isolates. J Bacteriol. 1994; 176:6199–6206. [PubMed: 7928989]

Karaolis DK, Somara S, Maneval DR Jr, Johnson JA, Kaper JB. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature. 1999; 399:375–379. [PubMed: 10360577]

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–780. [PubMed: 23329690]

Katz LS, Petkau A, Beaulaurier J, Tyler S, Antonova ES, Turnsek MA, et al. Evolutionary dynamics of Vibrio cholerae O1 following a single-source introduction to Haiti. MBio. 2013; 4(4):e00398–13. [PubMed: 23820394]

Kimsey HH, Waldor MK. CTXphi immunity: application in the development of cholera vaccines. Proc Natl Acad Sci USA. 1998; 95:7035–7039. [PubMed: 9618534]

Kirn TJ, Jude BA, Taylor RK. A colonization factor links Vibrio cholerae environmental survival and human infection. Nature. 2005; 438:863–866. [PubMed: 16341015]

Kirn TJ, Lafferty MJ, Sandoe CM, Taylor RK. Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by Vibrio cholerae. Mol Microbiol. 2000; 35:896–910. [PubMed: 10692166]

Krukonis ES, Yu RR, Dirita VJ. The Vibrio cholerae ToxR/TcpP/ToxT virulence cascade: distinct roles for two membrane-localized transcriptional activators on a single promoter. Mol Microbiol. 2000; 38:67–84. [PubMed: 11029691]

Labbate M, Boucher Y, Joss MJ, Michael CA, Gillings MR, Stokes HW. Use of chromosomal integron arrays as a phylogenetic typing system for Vibrio cholerae pandemic strains. Microbiology. 2007; 153:1488–1498. [PubMed: 17464063]

Lam C, Octavia S, Reeves P, Wang L, Lan R. Evolution of seventh cholera pandemic and origin of 1991 epidemic, Latin America. Emerg Infect Dis. 2010; 16:1130–1132. [PubMed: 20587187]

Lam C, Octavia S, Reeves PR, Lan R. Multi-locus variable number tandem repeat analysis of 7th pandemic Vibrio cholerae. BMC Microbiol. 2012; 12:82. [PubMed: 22624829]

Larocque RC, Harris JB, Dziejman M, Li X, Khan AI, Faruque AS, et al. Transcriptional profiling of Vibrio cholerae recovered directly from patient specimens during early and late stages of human infection. Infect Immun. 2005; 73:4488–4493. [PubMed: 16040959]

Li J, Egelman EH, Craig L. Structure of the Vibrio cholerae Type IVb Pilus and stability comparison with the Neisseria gonorrhoeae type IVa pilus. J Mol Biol. 2012; 418:47–64. [PubMed: 22361030]

Li M, Kotetishvili M, Chen Y, Sozhamannan S. Comparative genomic analyses of the vibrio pathogenicity island and cholera toxin prophage regions in nonepidemic serogroup strains of Vibrio cholerae. Appl Environ Microbiol. 2003; 69:1728–1738. [PubMed: 12620865]

Liang W, Wang S, Yu F, Zhang L, Qi G, Liu Y, et al. Construction and evaluation of a safe, live, oral Vibrio cholerae vaccine candidate, IEM108. Infect Immun. 2003; 71:5498–5504. [PubMed: 14500467]

MacDonald D, Demarre G, Bouvier M, Mazel D, Gopaul DN. Structural basis for broad DNA-specificity in integron recombination. Nature. 2006; 440:1157–1162. [PubMed: 16641988]

Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, Waldor MK. RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. Cell Host Microbe. 2011; 10:165–174. [PubMed: 21843873]

Mazel D, Dychinco B, Webb VA, Davies J. A distinctive class of integron in the Vibrio cholerae genome. Science. 1998; 280:605–608. [PubMed: 9554855]

Meibom KL, Blokesch M, Dolganov NA, Wu CY, Schoolnik GK. Chitin induces natural competence in Vibrio cholerae. Science. 2005; 310:1824–1827. [PubMed: 16357262]

Mekalanos JJ. Duplication and amplification of toxin genes in Vibrio cholerae. Cell. 1983; 35(1):253–263. [PubMed: 6627396]

Mekalanos JJ, Robins W, Ussery DW, Davis BM, Schadt E, Waldor MK. Non-O1 Vibrio cholerae unlinked to cholera in Haiti. Proc Natl Acad Sci USA. 2012; 109:E3206. (author reply E7). [PubMed: 23035253]

Miller VL, DiRita VJ, Mekalanos JJ. Identification of toxS, a regulatory gene whose product enhances toxR-mediated activation of the cholera toxin promoter. J Bacteriol. 1989; 171:1288–1293. [PubMed: 2646275]

Mukhopadhyay AK, Chakraborty S, Takeda Y, Nair GB, Berg DE. Characterization of VPI pathogenicity island and CTXphi prophage in environmental strains of Vibrio cholerae. J Bacteriol. 2001; 183:4737–4746.

Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature. 2011; 477:462–465. [PubMed: 21866102]

Nair GB, Faruque SM, Bhuiyan NA, Kamruzzaman M, Siddique AK, Sack DA. New variants of Vibrio cholerae O1 biotype El Tor with attributes of the classical biotype from hospitalized patients with acute diarrhea in Bangladesh. J Clin Microbiol. 2002; 40:3296–3299. [PubMed: 12202569]

Nair GB, Qadri F, Holmgren J, Svennerholm AM, Safa A, Bhuiyan NA, et al. Cholera due to altered El Tor strains of Vibrio cholerae O1 in Bangladesh. J Clin Microbiol. 2006; 44:4211–4213. [PubMed: 16957040]

Novais RC, Coelho A, Salles CA, Vicente AC. Toxin-co-regulated pilus cluster in non-O1, non-toxigenic Vibrio cholerae: evidence of a third allele of pilin gene. FEMS Microbiol Lett. 1999; 171:49–55. [PubMed: 9987841]

Nusrin S, Gil AI, Bhuiyan NA, Safa A, Asakura M, Lanata CF, et al. Peruvian Vibrio cholerae O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. J Med Microbiol. 2009; 58:342–354. [PubMed: 19208885]

Ogawa A, Takeda T. The gene encoding the heat-stable enterotoxin of Vibrio cholerae is flanked by 123-base pair direct repeats. Microbiol Immunol. 1993; 37:607–616. [PubMed: 8246823]

Pang B, Yan M, Cui Z, Ye X, Diao B, Ren Y, et al. Genetic diversity of toxigenic and nontoxigenic Vibrio cholerae serogroups O1 and O139 revealed by array-based comparative genomic hybridization. J Bacteriol. 2007; 189:4837–4849. [PubMed: 17468246]

Pascual M, Bouma MJ, Dobson AP. Cholera and climate: revisiting the quantitative evidence. Microbes Infect. 2002; 4:237–245. [PubMed: 11880057]

Qadri F, Chowdhury MI, Faruque SM, Salam MA, Ahmed T, Begum YA, et al. Peru-15, a live attenuated oral cholera vaccine, is safe and immunogenic in Bangladeshi toddlers and infants. Vaccine. 2007; 25:231–238. [PubMed: 16996172]

Rahman MH, Biswas K, Hossain MA, Sack RB, Mekalanos JJ, Faruque SM. Distribution of genes for virulence and ecological fitness among diverse Vibrio cholerae population in a cholera endemic area: tracking the evolution of pathogenic strains. DNA Cell Biol. 2008; 27:347–355. [PubMed: 18462070]

Raychoudhuri A, Patra T, Ghosh K, Ramamurthy T, Nandy RK, Takeda Y, et al. Classical ctxB in Vibrio cholerae O1, Kolkata, India. Emerg Infect Dis. 2009; 15:131–132. [PubMed: 19116078]

Reguera G, Kolter R. Virulence and the environment: a novel role for Vibrio cholerae toxin-coregulated pili in biofilm formation on chitin. J Bacteriol. 2005; 187:3551–3555. [PubMed: 15866944]

Reimer AR, Van Domselaar G, Stroika S, Walker M, Kent H, Tarr C, et al. Comparative genomics of Vibrio cholerae from Haiti, Asia, and Africa. Emerg Infect Dis. 2011; 17:2113–2121. [PubMed: 22099115]

Rhine JA, Taylor RK. TcpA pilin sequences and colonization requirements for O1 and O139 Vibrio cholerae. Mol Microbiol. 1994; 13:1013–1020. [PubMed: 7854116]

Rivera IN, Chun J, Huq A, Sack RB, Colwell RR. Genotypes associated with virulence in environmental isolates of Vibrio cholerae. Appl Environ Microbiol. 2001; 67:2421–2429. [PubMed: 11375146]

Rokas A, Williams BL, King N, Carroll SB. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 2003; 425:798–804. [PubMed: 14574403]

Rowe-Magnus DA, Guerout AM, Biskri L, Bouige P, Mazel D. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. Genome Res. 2003; 13:428–442. [PubMed: 12618374]

Rubin EJ, Lin W, Mekalanos JJ, Waldor MK. Replication and integration of a Vibrio cholerae cryptic plasmid linked to the CTX prophage. Mol Microbiol. 1998; 28:1247–1254. [PubMed: 9680213]

Russell AJ. A statistical approach to the epidemiology of cholera in Madras presidency. Proc Natl Acad Sci USA. 1925; 11:653–657. [PubMed: 16587060]

Safa A, Bhuyian NA, Nusrin S, Ansaruzzaman M, Alam M, Hamabata T, et al. Genetic characteristics of Matlab variants of Vibrio cholerae O1 that are hybrids between classical and El Tor biotypes. J Med Microbiol. 2006; 55:1563–1569. [PubMed: 17030917]

Sealfon R, Gire S, Ellis C, Calderwood S, Qadri F, Hensley L, et al. High depth, whole- genome sequencing of cholera isolates from Haiti and the dominican republic. BMC Genom. 2012; 13:468.

Seed KD, Bodi KL, Kropinski AM, Ackermann HW, Calderwood SB, Qadri F, et al. Evidence of a dominant lineage of Vibrio cholerae-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. MBio. 2011; 2:e00334–10. [PubMed: 21304168]

Sellek RE, Niemcewicz M, Olsen JS, Bassy O, Lorenzo P, Marti L, et al. Phenotypic and genetic analyses of 111 clinical and environmental O1, O139, and non-O1/O139 Vibrio cholerae strains from different geographical areas. Epidemiol Infect. 2012; 140:1389–1399. [PubMed: 22074599]

Shakya G, Kim DW, Clemens JD, Malla S, Upadhyaya BP, Dumre SP, et al. Phenotypic and genetic characterization of Vibrio cholerae O1 clinical isolates collected through national antimicrobial resistance surveillance network in Nepal. World J Microbiol Biotechnol. 2012; 28:2671–2678. [PubMed: 22806193]

Shamini G, Ravichandran M, Sinnott JT, Somboonwit C, Sidhu HS, Shapshak P, et al. Structural inferences for cholera toxin mutations in Vibrio cholerae. Bioinformation. 2011; 6:1–9. [PubMed: 21464837]

Singh DV, Matte MH, Matte GR, Jiang S, Sabeena F, Shukla BN, et al. Molecular analysis of Vibrio cholerae O1, O139, non-O1, and non-O139 strains: clonal relationships between clinical and environmental isolates. Appl Environ Microbiol. 2001; 67:910–921. [PubMed: 11157262]

Sjolund-Karlsson M, Reimer A, Folster JP, Walker M, Dahourou GA, Batra DG, et al. Drug-resistance mechanisms in Vibrio cholerae O1 outbreak strain, Haiti, 2010. Emerg Infect Dis. 2011; 17:2151–2154. [PubMed: 22099122]

Son MS, Megli CJ, Kovacikova G, Qadri F, Taylor RK. Characterization of Vibrio cholerae O1 El Tor biotype variant clinical isolates from Bangladesh and Haiti, including a molecular genetic analysis of virulence genes. J Clin Microbiol. 2011; 49:3739–3749. [PubMed: 21880975]

Talkington D, Bopp C, Tarr C, Parsons MB, Dahourou G, Freeman M, et al. Characterization of toxigenic Vibrio cholerae from Haiti, 2010–2011. Emerg Infect Dis. 2011; 17:2122–2129. [PubMed: 22099116]

Thungapathra M, Sharma C, Gupta N, Ghosh RK, Mukhopadhyay A, Koley H, et al. Construction of a recombinant live oral vaccine from a non-toxigenic strain of Vibrio cholerae O1 serotype inaba biotype E1 Tor and assessment of its reactogenicity and immunogenicity in the rabbit model. Immunol Lett. 1999; 68:219–227. [PubMed: 10424424]

Trucksis M, Michalski J, Deng YK, Kaper JB. The Vibrio cholerae genome contains two unique circular chromosomes. Proc Natl Acad Sci USA. 1998; 95:14464–14469. [PubMed: 9826723]

UN. United Nation Final Report of the Independent Panel of Experts on the Cholera Outbreak in Haiti, May 4, 2011. United Nations: 2011. Available at http://www.un.org/News/dh/infocus/haiti/UN-cholera-report-final.pdf [Accessed March 2012]

Vesth T, Lagesen K, Acar O, Ussery D. CMG-biotools, a free workbench for basic comparative microbial genomics. PLoS ONE. 2013; 8:e60120. [PubMed: 23577086]

Vesth T, Wassenaar TM, Hallin PF, Snipen L, Lagesen K, Ussery DW. On the origins of a Vibrio species. Microb Ecol. 2010; 59:1–13. [PubMed: 19830476]

Wachsmuth IK, Evins GM, Fields PI, Olsvik O, Popovic T, Bopp CA, et al. The molecular epidemiology of cholera in Latin America. J Infect Dis. 1993; 167:621–626. [PubMed: 7680060]

Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science. 1996; 272:1910–1914. [PubMed: 8658163]

Waldor MK, Rubin EJ, Pearson GD, Kimsey H, Mekalanos JJ. Regulation, replication, and integration functions of the Vibrio cholerae CTXphi are encoded by region RS2. Mol Microbiol. 1997; 24:917–926. [PubMed: 9220000]

Waldor MK, Tschape H, Mekalanos JJ. A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in Vibrio cholerae O139. J Bacteriol. 1996; 178:4157–4165. [PubMed: 8763944]

Weber JT, Mintz ED, Canizares R, Semiglia A, Gomez I, Sempertegui R, et al. Epidemic cholera in Ecuador: multidrug-resistance and transmission by water and seafood. Epidemiol Infect. 1994; 112:1–11. [PubMed: 8119348]

Yamaichi Y, Iida T, Park KS, Yamamoto K, Honda T. Physical and genetic map of the genome of Vibrio parahaemolyticus: presence of two chromosomes in Vibrio species. Mol Microbiol. 1999; 31:1513–1521. [PubMed: 10200969]

Yan M, Liu G, Diao B, Qiu H, Zhang L, Liang W, et al. A Vibrio cholerae serogroup O1 vaccine candidate against CTX ET Phi infection. Vaccine. 2007; 25:4046–4055. [PubMed: 17428586]

Yildiz FH, Liu XS, Heydorn A, Schoolnik GK. Molecular analysis of rugosity in a Vibrio cholerae O1 El Tor phase variant. Mol Microbiol. 2004; 53:497–515. [PubMed: 15228530]

Yu RR, DiRita VJ. Analysis of an autoregulatory loop controlling ToxT, cholera toxin, and toxin-coregulated pilus production in Vibrio cholerae. J Bacteriol. 1999; 181:2584–2592. [PubMed: 10198025]
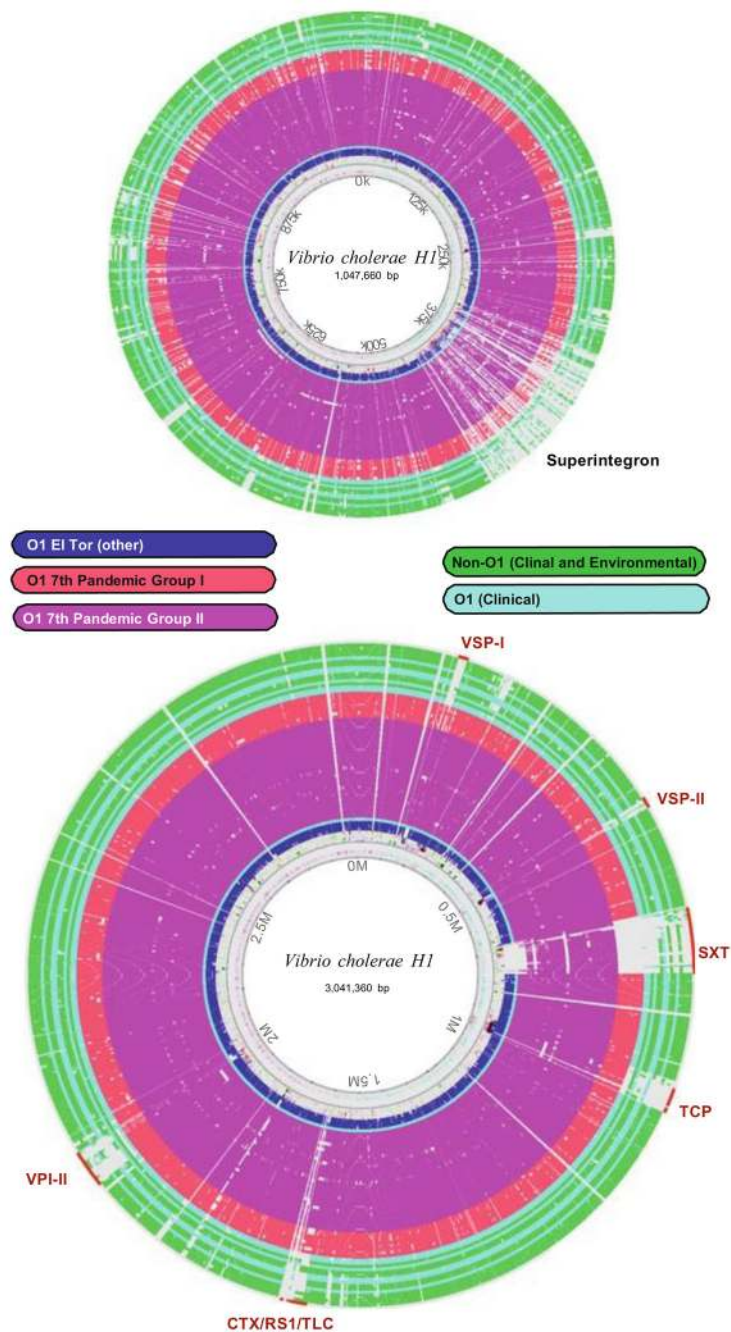
Fig. 1.
Genome blast atlas created with CMG Biotools (Vesth et al. 2013) showing the aligned
genomes of a panel of *V. cholerae* clinical and environmental isolates using the 2010 Haiti
epidemic strain as reference. Genomes are colored by both biotype and source and also
match genomes analyzed in the Fig. 2. Seventh pandemic islands and Superintegron are
labeled on both chromosomes
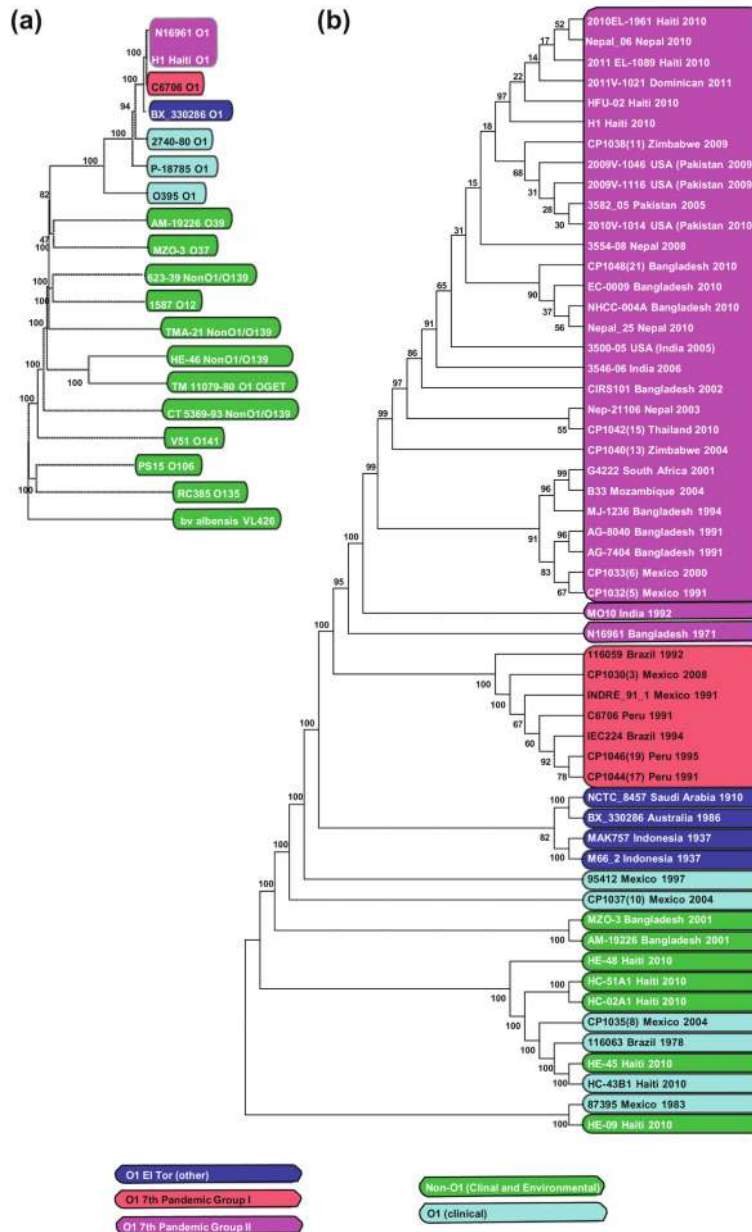
Fig. 2.
The bootstrap consensus tree (Neighbor-Joining) inferred from 500 replicates for a small diverse sample of O1 and non-O1 strains (a) and for a collection of clinical and environmental strains (b). Core genomes were extrapolated using CMG-biotools 2.2, aligned using MAAFT (vs 7.127), and trees were calculated and produced using MEGA (vs 5.10). Genomes are colored by biotype and source. CMG Biotools (vs 2.2, Vesth et al. 2013), MAAFT (vs 7.127, Katoh and Standely 2013) MEGA (vs 5.10, Hall 2013)
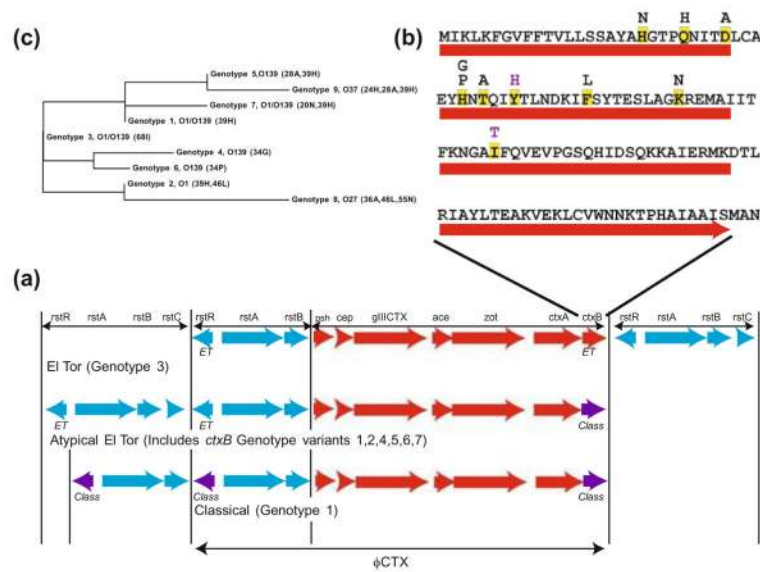
Fig. 3.
Different arrangement of CTX/RS1 region in El Tor, classical and variant strains (a). The amino acid variation in ctxB gene (b) is used to generate phylogenic tree based on the alignment of ctxB variants (c)