Journal of Biomolecular Structure and Dynamics
Vol. 30, No. 3, 2012, 338–346

Taylor & Francis
Taylor & Francis Group

Evolutionary perspective on the origin of Haitian cholera outbreak strain

Anirban Dasgupta[a], Rachana Banerjee[b], Santasabuj Das[c] and Surajit Basak[d]*

[a]*Department of Bioinformatics, West Bengal University of Technology, BF-142, Sector – I, Salt Lake, Kolkata 700064, India;* [b]*Department of Bio-Physics, Molecular Biology and Bioinformatics, University of Calcutta, 92, Acharya Prafulla Chandra Road, Kolkata 700009, India;* [c]*Biomedical Informatics Center, National Institute of Cholera and Enteric Diseases, P-33, C.I.T Road, Scheme-XM, Beliaghata, Kolkata 700010, India;* [d]*Department of Molecular Biology & Bioinformatics, Tripura University, Suryamaninagar, Tripura 799 022, India*

Communicated by Ramaswamy H. Sarma

(*Received 25 November 2011; final version received 27 February 2012*)

Cholera epidemic has not been reported in Haiti for at least 100 years, although cholera has been present in Latin America since 1991. Surprisingly, the recent cholera epidemic in Haiti (October 2010) recorded more than 250,000 cases and 4000 deaths in the first 6 months and became one of the most explosive and deadly cholera outbreak in recent history. In the present study, we conducted genomic analyses of pathogenicity islands of three Haitian *Vibrio cholerae* strains and compared them with nine different *V. cholerae* O1 El Tor genomes. Although CIRS101 is evolutionarily most similar to the Haitian strains, our study also provides some important differences in the genetic organization of pathogenicity islands of Haitian strains with CIRS101. Evolutionary analysis suggests that unusual functional constraints have been imposed on the Haitian strains and we hypothesize that amino acid substitution is more deleterious in Haitian strains than in nonHaitian strains.

Keywords: Cholera; Haitian outbreak; pathogenicity islands; comparative genomics; functional constraints

Introduction

Cholera, an acutely dehydrating diarrheal disease that can rapidly kill its victims, is caused by *Vibrio cholerae*, a gram-negative bacterium. Cholera is one of the most potent diarrheal diseases that continue to ravage many developing countries where endemic cholera remains a serious health threat and are particularly associated with poverty and poor sanitation (Lee, 2001; Sack, Sack, Nair, & Siddiquta, 2004). The recent cholera outbreak in Haiti placed this diarrheal infectious disease at the forefront of the global public health agenda. On 21 October 2010, the US centers for disease control and prevention (CDC) confirmed that cholera had returned to Haiti for the first time in more than 100 years. Within weeks, the disease had been identified in every province, and by the end of the year more than 150,000 cases and the deaths of 3500 people had been reported (Centers for Disease Control and Prevention [CDC], 2010). Since cholera outbreak had not been reported in Haiti for more than a century, the origin and evolution of the Haitian *V. cholerae* outbreak has been the subject of some controversy (Enserink, 2010).

Traditionally, *V. cholerae* strains are classified into serogroups on the basis of the structure of an outer-membrane O antigen and into biotypes on the basis of a variety of biochemical and microbiological tests (Chatterjee & Chaudhuri, 2003; Cho, Yi, Lee, Kim, & Chun, 2010; Kaper, Morris, & Levine, 1995). It is well documented that isolates of the sixth pandemic were almost exclusively of the *V. cholerae* O1 classical biotype whereas the current seventh pandemic has been caused by the O1 El Tor biotypes, first isolated on the Indonesian island of Sulawesi in 1961 (Chun et al., 2009). Since then, several novel genetic variants of *V. cholerae* O1 El Tor have emerged or re-emerged. Most notable was a new serogroup of the species, named O139 (Ramamurthy et al., 1993). This serogroup, first recognized in 1992 in India and Bangladesh, initially displaced the local existing O1 El Tor strains, though it is now restricted to Asia (Cholera Working Group, 1993;

*Corresponding author. Address: Biomedical Informatics Center, National Institute of Cholera and Enteric Diseases, P-33, C.I.T Road, Scheme-XM, Beliaghata, Kolkata 700010, India. Email: surajit@tripurauniv.in

ISSN 0739-1102 print/ISSN 1538-0254 online

http://dx.doi.org/10.1080/07391102.2012.680033
http://www.tandfonline.com

Faruque, Albert, & Mekalanos, 1998, 2003). The current pandemic in Asia and Africa is largely attributed to new variants of *V. cholerae* showing traits of both the classical and El Tor biotypes – and El Tor biotype strains producing the classical cholera toxin (CTX). Safa, Nair, and Kong (2010) coined the term, "atypical El Tor," for these clones, which include the Matlab types I, II, and III (Nair et al., 2002), altered El Tor (Nair et al., 2006), Mozambique El Tor (Ansaruzzaman et al., 2004; Lee et al., 2006), and hybrid El Tor strains (Safa, Sultana, Cam Dac, Mwansa, & Kong, 2008). Before the current epidemic, cases of cholera had not been reported in Haiti for many years (News BBC, 2010), and disease had not spread into Haiti during expansion of the El Tor pandemic into Latin America that began in Peru in 1991 (Ali et al., 2011). However, toxigenic *V. cholerae* O1 is present along the US Gulf Coast and in other coastal areas in the western hemisphere (Blake et al., 1980; Centers for Disease Control & Prevention, 2006). Analyses carried out by Haitian and US laboratories have indicated that the outbreak strain in Haiti is *V. cholerae* El Tor O1 (Ceccarelli, Spagnoletti, Cappuccinelli, Burrus, & Colombo, 2011) and thus is related to strains that are causing the ongoing seventh pandemic of cholera.

Comparisons of the relative rates of change at synonymous (silent) and nonsynonymous (replacement) sites have for many years been a central tenet of molecular evolution (Anbazhagan et al., 2010; Kimura, 1991; Wang, Liu, Wang, Huang, & Yu, 2011). The relative frequencies of these changes are determined by a complex blend of stochastic and selective forces acting at many different levels from a single site, codon, genome, or population (metagenome) Rocha et al., 2006. Estimates for dN (the number of nonsynonymous changes per nonsynonymous site) and dS (the number of synonymous changes per synonymous site) are typically interpreted in terms of the selective consequences of these changes (Banerjee, Basak, Gupta, & Ghosh, 2004; Basak & Ghosh, 2006; Banerjee, Roy, Ahmad, Das, & Basak 2012; Mukhopadhyay, Basak, & Ghosh, 2007; Sinha, Roy, Das, Das, & Basak, 2009).

Synonymous substitutions are usually regarded as neutral, or at least as having a much smaller effect on fitness than nonsynonymous substitutions. The dN/dS ratio resulting from the comparison between two orthologous genes therefore has both theoretical and practical implications as it can reveal the type of selection pressure acting on the genes. A low ratio ($dN/dS << 1$) indicates strong purifying (stabilizing) selection, whereas a high ratio ($dN/dS > 1$) indicates selection for diversification (positive selection). The calculation of dN/dS can therefore help to identify genes or domains under particular biochemical or ecological constraint, or conversely putative virulence factors or candidate vaccine targets subject to diversifying (frequency-dependent) selection from the host immune response (Hughes & Nei, 1988; Thomas et al., 2003; Vonderviszt, Kanto, Aizawa, & Namba, 1989; Wang, Zhang, Zhao, Wang, & Pan, 2010).

The present study is carried out to elucidate the evolutionary nature and mechanism of the clonal origin of Haitian strain among other genetically distinctive *V. cholerae* isolates using comparative genomics. Here, we will focus on genomic analysis of pathogenic islands of *V. cholerae* in light of genome wide comparison of available whole genome sequence data. Vibrio Pathogenicity Island is designated by VPI-1 and VPI-2. Vibrio Seventh Pandemic Island is designated by VSP-1 and VSP-2. Mannose-sensitive hemagglutinin (MSH) is an important host colonization factors, bacteriophage receptors, and mediators of DNA transfer. CTXphi is a filamentous phage that encodes CTX, the principal virulence factor of *V. cholerae*. There are two subunits of CTX genes, namely CTXA and CTXB. Toxin-linked cryptic (TLC) is a 4.7 kb DNA fragment which has some functional relationship to CTXphi for pathogenesis. Superintegron are very large integrons having gene-capture systems. SXT is a 100-kb integrative and conjugative element, encodes genes resistance to multiple antibiotics.

Materials and methods

Sequence retrieval and annotation

We considered the genome of three Haitian cholera outbreak isolates sequenced by the US CDC and deposited into the GenBank database (accession numbers, AELH 00000000.1, AELI00000000.1, and AELJ00000000.1). In the present manuscript, the three Haitian strains are represented by AELH, AELI, and AELJ. We compared the Haitian strain with nine *V. cholerae* O1 El Tor complete genomes (Table 1). The whole genome shotgun sequences for three Haitian strains were retrieved from GenBank WGS Projects. Draft contig sequences of Haitian strains available from WGS projects were fully contiguated into whole genome sequences using Perl program freely downloadable from ABACAS software (http://abacas.sourceforge.net) (Assefa, Keane, Otto, Newbold, & Berriman, 2009). The complete annotations of three Haitian genomes were performed using RAST server (http://RAST.nmpdr.org) (Aziz et al., 2008). The completely annotated nine *V. cholerae* genomes were downloaded directly from the pathosystems resource integration center database (https://patric.vbi.vt.edu) (Snyder et al., 2007).

Ortholog finding and comparison

To identify the orthologous genes for seven different pathogenicity islands (CTX phi, MSH, TLC, VPI1, VPI2, VSP1, and VSP2), SXT and superintegron regions we performed BLASTP search satisfying the conditions: identity > 90%, gap = 0, *e*-value < .0001, and overlap > 90%. All the genes for each pathogenicity island,

Table 1. Characteristics of *V. cholerae* strains analyzed in this study.

Strain	Serogroup	Biotype	Source of isolation	Year of isolation	Area of isolation
8457	O1	El tor	CLINICAL	1910	Saudi Arabia
M66	O1	El tor	CLINICAL	1937	Makassar
MAK757	O1	El tor	CLINICAL	1937	Celebes Island
N16961	O1	El tor	CLINICAL	1975	Bangladesh
RC9	O1	El tor	CLINICAL	1985	Kenya
INDRE	O1	El tor	CLINICAL	1991	Mexico
MJ1236	O1	El tor	CLINICAL	1994	Matlab
CIRS101	O1	El tor	CLINICAL	2002	Bangladesh
B33	O1	El tor	CLINICAL	2004	Mozambique
AELH	O1	El tor	CLINICAL	2010	Haiti
AELJ	O1	El tor	CLINICAL	2010	Haiti
AELI	O1	El tor	CLINICAL	2010	Haiti

SXT and superintegeron were aligned using Mauve genome alignment software (Rissman et al., 2009) available at http://gel.ahabs.wisc.edu/mauve.

Phylogenetic analyses

A set of orthologs of every pathogenicity island, SXT, and superintegron regions obtained from different *V. cholerae* strains were then aligned using the CLUSTALW2 program. The resultant multiple alignments were concatenated to generate genome scale alignments, which were subsequently used to reconstruct the neighbor-joining phylogenetic tree. The Kimura two parameter model (K2P) was used to generate the distance matrix to calculate K2P distance. Phylip was used for phylogenetic analysis (Felsenstein, 1997).

Evolutionary rate calculation

Pairwise synonymous (dS) and nonsynonymous (dN) distance between the orthologous genes was calculated by using the method of Nei and Gojobori (1986).

Results and discussion

Comparative analysis of gene arrangement of pathogenicity islands

A global alignment of all the pathogenicity islands was performed by Mauve after combining the genes present in individual pathogenicity island successively (Figure 1). During the alignment process, Mauve identifies conserved segments from different pathogenicity islands that appear to be internally free from genome rearrangements. Such regions are referred to as local collinear blocks and represented by different color-shaded schemes in Figure 1. It can be clearly visualized from Figure 1 about the presence/absence of an entire pathogenicity island in a particular strain, e.g. CTX phi is absent in M66 and SXT is absent in 8457, INDRE, MAK757, M66, and N16961. In addition, variation of size between pathogenicity islands among the strains indicates gene insertion/deletion processes e.g. SXT region is much larger while TLC region is much smaller for MJ1236 and B33 compared to other strains.

Analysis of single nucleotide variations and indels of V. cholerae **strains**

To specifically identify single nucleotide variations and indels, we again analyzed the genomic structures of every pathogenicity island separately and found that there exist some differences in CTX phi, TLC, and VPI2 regions which are discussed below.

In the CTX phi island, we noted that the gene encoding CTXB in Haiti isolates carries three nonsynonymous substitutions relative to N16961 as already reported by Chin et al. (2010). The three mutations enlisted are: His to Asp at position 20, Tyr to His at position 39, and Ile to Thr at position 68 (Figure 2). However, the mutation at position 20 is unique to Haitian strain as it is absent in all other *V. cholerae* O1 El Tor strains considered in this study.

Chin et al. (2010) discussed about the shared ancestry of the Haitian epidemic strain and recent South Asian strains of *V. cholerae*. In the present study, the sequence comparison and other literature sources reveal that the Haitian CTXB is not only similar to the South Asian and West African strains, but is exactly the same as cholera outbreak strain in Orissa during August–September 2007. CTXB of cholera outbreak strain in Orissa also contains the same mutation at position 20 as reported in the Haitian CTXB. It was demonstrated that the improved virulence of Orissa strain compared to the prototype El Tor strains is due to increase in toxin production (Ghosh-Banerjee et al., 2010). After the outbreak in Orissa in 2007, the same CTXB was found in the western Africa outbreak (Nigeria, Cameroon) in September–October 2009. Now the same CTXB has been the cause of the Haiti outbreak in 2010. This observation may provide valuable information about the geographical route of circulation of the Haitian strains originating from Indian subcontinent.

In the TLC region some differences has been noted in Haiti strain with respect to CIRS (Figure 3). VC1471

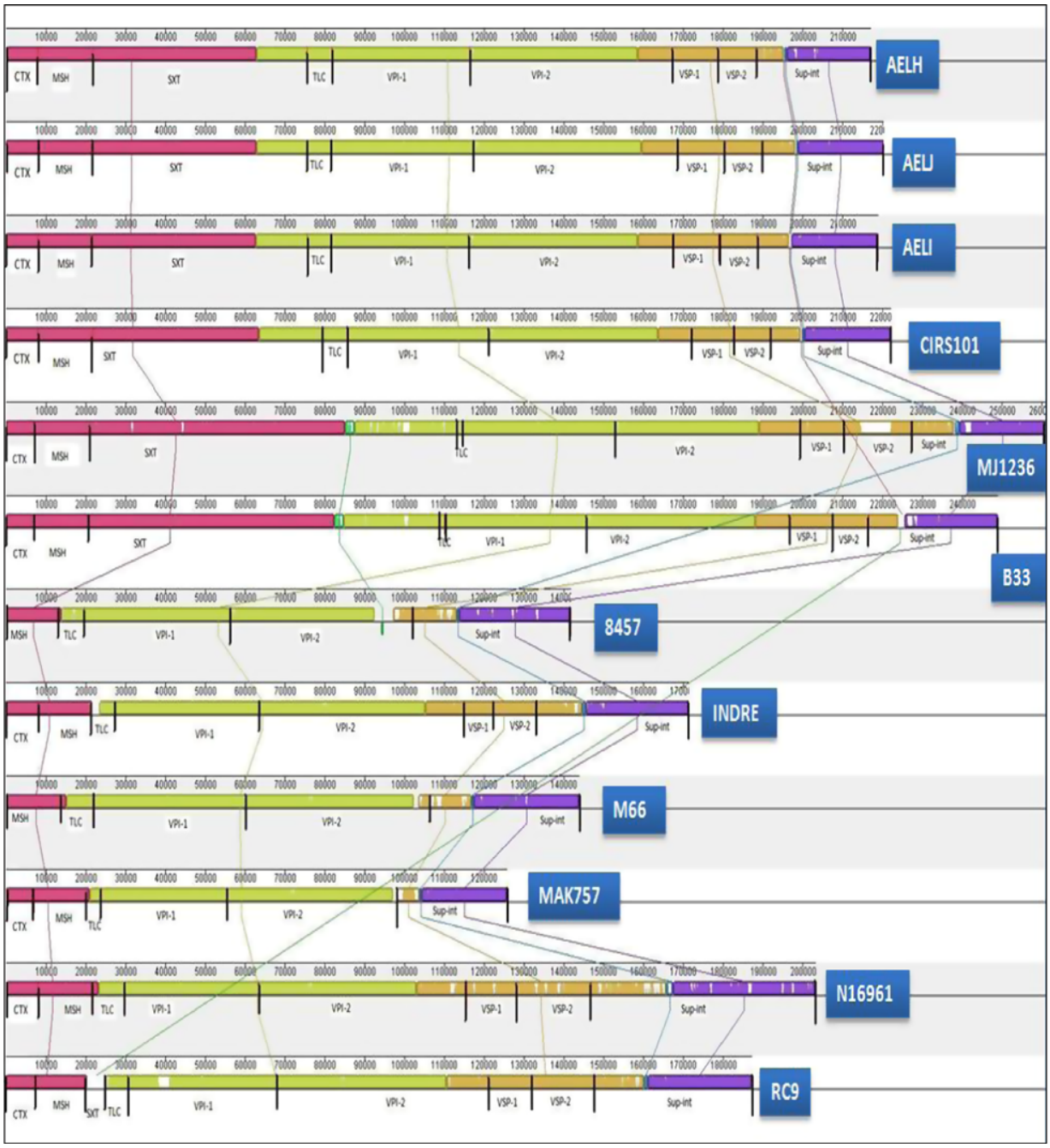

Figure 1. Arrangement of different pathogenicity islands of 12 *V. cholerae* strains.

is truncated in Haiti strain. Multiple copies of two CDSs (VC1477 and VC1478) are present in CIRS101. For five CDSs (VC1465, VC1467, VC1469, VC1473, and VC1475), nucleotide sequences are not completely identical between Haiti strain and CIRS101.

In the VPI-2 region, differences exist for four CDSs between Haiti and CIRS101 (Table 2). Deletion of 7 nucleotides and 12 nucleotides have been observed for Haiti strains compared to CIRS101 in VC1787 and VC1788, respectively. Multiple copies of two CDSs are present for VC1789 and VC1790 in Haiti compared to CIRS101.

SXT is a clinically important integrative and conjugative element that accounts for the dissemination of genes conferring resistance to several antibiotics in contempo-

```
AELH      MIKLKFGVFFTVLLSSAYANGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
AELI      MIKLKFGVFFTVLLSSAYANGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
AELJ      MIKLKFGVFFTVLLSSAYANGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
CIRS101   MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
B33       MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
MJ1236    MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
MAK757    MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIHTLNDKIFSYTESLAGKREMAI  60
N16961    MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAI  60
RC9       MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAI  60
INDRE     MIKLKFGVFFTVLLSSAYAHGTPQNIIDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAI  60
          ******************:******************:**********************

AELH      ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
AELI      ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
AELJ      ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
CIRS101   ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
B33       ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
MJ1236    ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
MAK757    ITFKNGATFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
N16961    ITFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
RC9       ITFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
INDRE     ITFKNGAIFQVEVPGSQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAI  120
          *******************************************************
```

Figure 2. Three nonsynonymous mutations are marked within box.

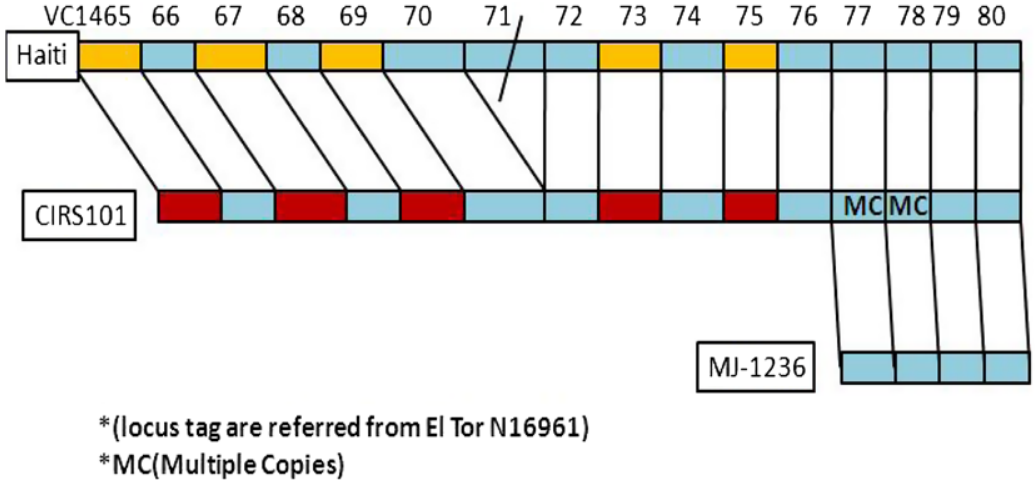

Figure 3. Schematic representation of genomic structure of TLC pathogenic island.

Table 2. Comparison of VPI2 island of Haitian strain with CIRS101 and Matlab variant.

Genes	AELH/AELJ/AELI	CIRS101	MJ1236
VC1787	7 nucleotides deletion	100% identical to El TOR NI16961	100% identical to El TOR NI16961
VC1788	12 nucleotides deletion	100% identical to El TOR NI16961	100% identical to El TOR NI16961
VC1789-VC1790	Single copy of gene	Multiple copies of gene	Single copy of gene
VC1795, VC1801–VC1803, VC1768	100% identical to El TOR NI16961	100% identical to El TOR NI16961	Absent
VC1775, VC1780–VC1784, VC1791, VC1798	100% identical to El TOR NI16961	100% identical to El TOR NI16961	Some portion deletion

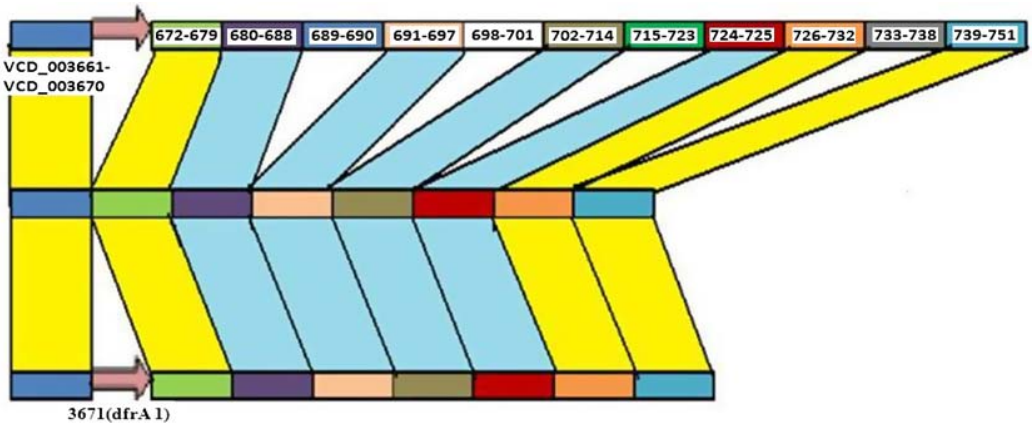

Figure 4. Schematic representation of genomic structure of SXT region. Bold arrow indicates presence of trimethoprim antibiotic resistance gene (dfrA1) in Haitian strain compared to CIRS101.

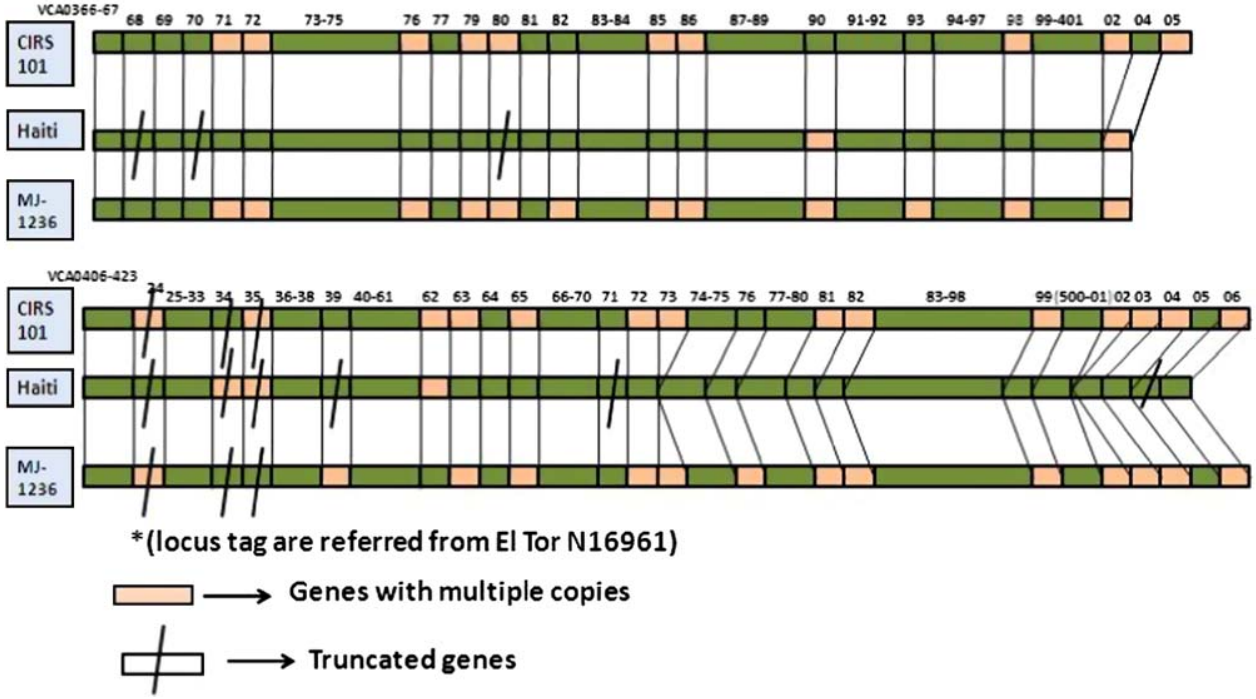

Figure 5. Schematic representation of genomic structure of superintegron region.

rary *V. cholerae* isolates. SXT region of Haiti strains are very similar with CIRS101 with one important deletion of VCD_003671 (dihydrofolate reductase, dfrA1) in CIRS101 (Figure 4). This gene mediates trimethoprim (antibiotic) in SXT and the presence of this gene in Haitian strain indicates greater antibiotic resistance of Haitian strains compared to CIRS101.

Superintegron region contains 137 genes. 102 genes of superintegron region of Haitian sequences are identical with both the CIRS101 as well as MJ1236. Some differences have been noted for 35 genes mainly due to the absence of multiple copies of genes and higher number of truncated genes in Haitian strains (Figure 5).

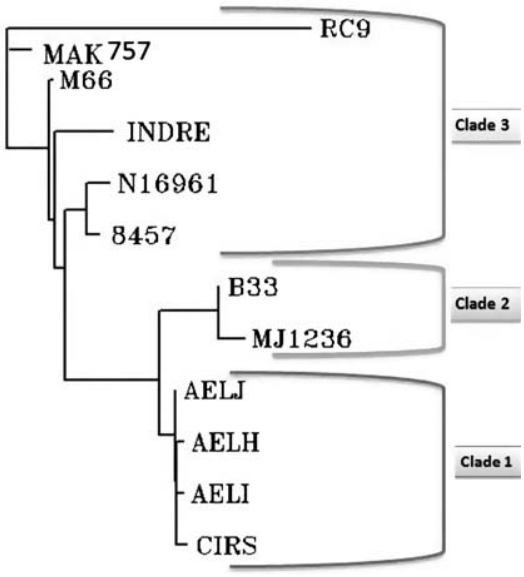

Figure 6. Phylogenetic tree constructed using ~225,000 bp region of pathogenicity islands of 12 *V. cholerae* O1 El Tor strains. The clade1 consists of Haitian strains (AELH, AELI, and AELJ) and CIRS101. All the altered El Tor variants including the Haitian strains form the clade 1 and clade 2.

Table 3. K2P distance between the three clades.

	Clade 1	Clade 2	Clade 3
Clade 1	0		
Clade 2	0.054	0	
Clade 3	0.126	0.156	0

Table 4. Evolutionary rate (dN/dS) and slopes between dN and dS of three clades.

Clade	dN/dS	Slope
1	.331	.097
2	.502	.560
3	.476	.346

Variation of evolutionary selection pressure among the V. cholerae **strains**

To understand the overall impact of different types of gene insertion, deletion, and mutation from evolutionary perspective, we constructed phylogenetic tree using neighbor-joining method based on the orthologous genes of pathogenic islands found in all available O1 El Tor *V. cholerae* strains (Figure 6). The phylogenetic tree was constructed using ~225,000 bp alignment. Only orthologous genes showing more than 95% protein sequence similarity to those of *V. cholerae* N16961 were selected. The phylogenetic tree unequivocally places three Haitian strains in the altered El Tor group. Among the altered El Tor variants, Haitian strains are similar to B33 (Mozambique strain, 2004) and MJ1236 (Matlab, Bangladesh, 1994). They are most similar to recent South Asian isolates CIRS101 from the 2002 outbreak in Bangladesh. Phylogenetic tree clearly reveals three distinct phyletic lineages on the basis of increasing order of evolutionary distance in terms of Kimura two-parameter model (K2P distance). The average difference of K2P distance suggests that the distance between clade1 and clade2 is almost half of the distance between clade1 and clade3 (Table 3). The K2P distance between the three clades thus suggests significant genomic diversity between the *V. cholerae* strains of three phyletic lineages.

We constructed core patho-genome for each of the three clades. The number of orthologous genes present in clade1, clade2, and clade3 are 216, 260, and 111, respectively. Table 4 shows that clade1 has lowest dN/dS value compared to clade2 and clade3. It suggests that the proteins of clade1 are subject to unusually strong purifying selection, leading to a reduced overall level of amino acid evolution per mutational event. Thus, the survival value conferred by adaptive protein features of pathogenic islands is expected to result in strong purifying natural selection on genes of clade1 with respect to clade2 and clade3. We then fitted linear regression lines between dN and dS through the origin on the assumption that dN and dS are both initially zero at the moment of lineage divergence. The slopes for the three lines of clade1, clade2, and clade3 are 0.097, 0.560, and 0.346, respectively. These results suggest that dN increases for a given increase in dS ~ 5.77 times faster in clade2 than in clade1. Similarly, dN increases for a given increase in dS ~ 3.56 times faster in clade2 than in clade1. These results suggest that proteins of clade1 are subject to unusually strong functional constraint with respect to clade2 and clade3, which is reflected in a reduced level of nonsynonymous nucleotide substitution for a given level of synonymous substitution. This, in turn, provides additional support for the hypothesis that unusual functional constraints have been imposed on the primary structure of Haitian protein sequences.

Conclusion

We hypothesize that average amino acid substitution is more deleterious in Haitian strains than in nonHaitian strains. On this hypothesis, the especially deleterious nature of mutation in Haitian sequences would have favored the evolutionary fixations of modifiers that specifically decrease the rates of base-substitution mutagenesis. Strong selection against amino acid selection in Haitian strains would impact their molecular evolution in two

ways. First, because most selectively neutral mutations that reached fixation are base substitutions and not indels, the overall rate of molecular evolution would be reduced. Second, within the generally reduced rate of molecular evolution, the fixation of nonsynonymous substitution would be reduced compared to that of synonymous substitutions. The present study provides support for the latter prediction.

Acknowledgment

This work is funded through an extramural project (IRIS ID: 2001-0496D) of ICMR, Government of India.

References

Ali, A., Chen, Y., Johnson, J.A., Redden, E., Mayette, Y., Rashid, M.H., … Morris, J.G. (2011). Recent clonal origin of cholera in Haiti. *Emerging Infectious Diseases, 17*(4), 699–701.

Anbazhagan, P., Purushottam, M., Kumar, H.B.K., Mukherjee, O., Jain, S., & Sowdhamini, R. (2010). Phylogenetic analysis and selection pressures of 5-HT receptors in human and non-human primates: Receptor of an ancient neurotransmitter. *Journal of Biomolecular Structure & Dynamics, 27*, 581–598.

Ansaruzzaman, M., Bhuiyan, N.A., Nair, G.B., Sack, D.A., Lucas, M., Deen, J.L., … Chaignat, C.-L. (2004). Cholera in Mozambique, variant of Vibrio cholerae. *Emerging Infectious Diseases, 10*, 2057–2059.

Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., & Berriman, M. (2009). ABACAS: Algorithm-based automatic contiguation of assembled sequences. *Bioinformatics, 25*, 1968–1969.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., … Zagnitko, O. (2008). The RAST server: Rapid annotations using subsystems technology. *BMC Genomics, 9*, 75.

Banerjee, T., Basak, S., Gupta, S.K., & Ghosh, T.C. (2004). Evolutionary forces in shaping the codon and amino acid usages in Blochmannia floridanus. *Journal of Biomolecular Structure & Dynamics, 22*, 13–23.

Banerjee, R., Roy, A., Ahmad, F., Das, S., & Basak, S. (2012). Evolutionary patterning of hemagglutinin gene sequence of 2009 H1N1 pandemic. *Journal of Biomolecular Structure & Dynamics, 29*(4), 733–742.

Basak, S., & Ghosh, T.C. (2006). Temperature adaptation of synonymous codon usage in different functional categories of genes: A comparative study between homologous genes of *Methanococcus jannaschii* and *Methanococcus maripaludis*. *FEBS Letters, 580*, 3895–3899.

Blake, P.A., Allegra, D.T., Snyder, J.D., Barrett, T.J., McFarland, L., Caraway, C.T, … Feldman, R.A. (1980). Cholera-a possible endemic focus in the United States. *New England Journal of Medicine, 302*, 305–309.

Ceccarelli, D., Spagnoletti, M., Cappuccinelli, P., Burrus, V., & Colombo, M.M. (2011). Origin of Vibrio cholerae in Haiti. *Lancet, 11*, 262.

Centers for Disease Control and Prevention (2006). Two cases of toxigenic *Vibrio cholerae* O1 infection after Hurricanes Katrina and Rita – Louisiana, October 2005. *MMWR Morbidity Mortality Weekly Report, 55*, 31–32.

Centers for Disease Control and Prevention (CDC) (2010). Update: Cholera outbreak – Haiti, 2010. *MMWR Morbidity Mortality Weekly Report, 59*, 1473–1479.

Chatterjee, S.N., & Chaudhuri, K. (2003). Lipopolysaccharides of *Vibrio cholerae* I. Physical and chemical characterization. *Biochimica et Biophysica Acta, 1639*, 65–79.

Chin, C.S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C, Jean-Charles, R.R., … Waldor, M.K. (2010). The origin of the Haitian cholera outbreak strain. *New England Journal of Medicine, 364*, 33–42.

Cho, Y., Yi, H., Lee, J.H., Kim, D.W., & Chun, J. (2010). Genomic evolution of *Vibrio cholerae*. *Current Opinion in Microbiology, 13*, 1–6.

Cholera Working Group (1993). Large epidemic of cholera-like disease in Bangladesh caused by Vibrio cholerae O139 synonym Bengal. Cholera working group, International center for diarrhoeal disease research, Bangladesh. *Lancet, 342*, 387–390.

Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J, … Colwell, R.R. (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proceedings of the National Academy Sciences of the United States of America, 106*, 15442–15447.

Enserink, M. (2010). Infectious diseases. Haiti's outbreak is latest in cholera's new global assault. *Science, 330*, 738–739.

Faruque, S.M., Albert, M.J., & Mekalanos, J.J. (1998). Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiology and Molecular Biology Reviews, 62*, 1301–1314.

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 Sciences of the United States of America, 100*, 1304–1309.

Felsenstein, J. (1997). An alternating least squares approach to inferring phylogenies from pairwise distances. *Systematic Biology, 46*, 101–111.

Ghosh-Banerjee, J., Senoh, M., Takahashi, T., Hamabata, T., Barman, S., Koley, H., … Takeda, Y. (2010). Cholera toxin production by the El Tor variant of Vibrio cholerae O1 compared to prototype El Tor and classical biotypes. *Journal of Clinical Microbiology, 48*, 4283–4286.

Hughes, A.L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature, 335*, 167–170.

Kaper, J.B.Jr, Morris, J.G., & Levine, M.M. (1995). Cholera. *Clinical Microbiology Reviews, 8*, 48–86.

Kimura, M. (1991). Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proceedings of the National Academy Sciences of the United States of America, 88*, 5969–5973.

Lee, K. (2001). The global dimensions of cholera. *Global Change & Human Health, 2*, 6–17.

Lee, J.H., Han, K.H., Choi, S.Y., Lucas, M.E., Mondlane, C., & Ansaruzzaman, M. (2006). Multilocus sequence typing (MLST) analysis of Vibrio cholerae O1 El Tor isolates from Mozambique that harbour the classical CTX prophage. *Journal of Medical Microbiology, 55*, 165–170.

Mukhopadhyay, P., Basak, S., & Ghosh, T.C. (2007). Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes. *Gene, 400*, 71–81.

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

Nair, G.B., Safa, A., Bhuiyan, N.A., Nusrin, S., Murphy, D., Nicol, C., … Vally, H. (2006). Isolation of Vibrio cholerae O1 strains similar to pre-seventh pandemic El Tor strains during an outbreak of gastrointestinal disease in an island resort in Fiji. *Journal of Medical Microbiology, 55*, 1559–1562.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution, 3*, 418–426.

News BBC. *Haiti cholera outbreak causes not clear, experts say.* Retrieved October 28, 2010 [cited 2010 Dec 3] from http://www.bbc.co.uk/news/world-latin-america-11618352.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S.K., Nair, G.B., & Shimada, T. (1993). Emergence of novel strain of Vibrio cholerae with epidemic potential in southern and eastern India. *Lancet, 341*, 703–704.

Rissman, A.I., Mau, B., Biehl, B.S., Darling, A.E., Glasner, J.D., & Perna, N.T. (2009). Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics, 25*, 2071–2073.

Rocha, E.P., Smith, J.M., Hurst, L.D., Holden, M.T., Copper, J.E., Smith, N.H., … Feil, E.J. (2006). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of Theoretical Biology, 239*, 226–235.

Sack, D.A., Sack, R.B., Nair, G.B., & Siddiquta, A.K. (2004). Cholera. *Lancet, 363*, 223–233.

Safa, A., Nair, G.B., & Kong, R.Y. (2010). Evolution of new variants of Vibrio cholerae O1. *Trends in Microbiology, 18*, 46–54.

Safa, A., Sultana, J., Cam Dac, P., Mwansa, J.C., & Kong, R.Y. (2008). Vibrio cholerae O1 hybrid El Tor strains, Asia and Africa. *Emerging Infectious Diseases, 14*, 987–988.

Sinha, N.K., Roy, A., Das, B., Das, S., & Basak, S. (2009). Evolutionary complexities of swine flu H1N1 gene sequences of 2009. *Biochemical and Biophysical Research Communications, 390*, 349–351.

Snyder, E.E., Kampanya, N., Lu, J., Nordberg, E.K., Karur, H.R, & Shukla, M. (2007). PATRIC: The VBI PathoSystems Resource Integration Center. *Nucleic Acids Research, 5*, D401–D406.

Thomas, M.A., Weston, B., Joseph, M., Wu, W., Nekrutenko, A., & Tonellato, P.J. (2003). Evolutionary dynamics of oncogenes and tumor suppressor genes: Higher intensities of purifying selection than other genes. *Molecular Biology and Evolution, 20*, 964–968.

Vonderviszt, F., Kanto, S., Aizawa, S., & Namba, K. (1989). Terminal regions of flagellin are disordered in solution. *Journal of Molecular Biology, 209*, 127–133.

Wang, D., Liu, F., Wang, L., Huang, S., & Yu, J. (2011). Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biology Direct, 6*, 13.

Wang, M., Zhang, X., Zhao, H., Wang, Q., & Pan, Y. (2010). Comparative analysis of vertebrate PEPT1 and PEPT2 genes. *Genetica, 138*, 587–599.