Genomic diversity of 2010 Haitian cholera outbreak strains

Nur A. Hasan[a,b,1], Seon Young Choi[a,b,1], Mark Eppinger[c,d,1], Philip W. Clark[a], Arlene Chen[a], Munirul Alam[e], Bradd J. Haley[a], Elisa Taviani[a], Erin Hine[c], Qi Su[c], Luke J. Tallon[c], Joseph B. Prosper[f], Keziah Furth[f], M. M. Hoq[g], Huai Li[b], Claire M. Fraser-Liggett[c], Alejandro Cravioto[e], Anwar Huq[a], Jacques Ravel[c,d], Thomas A. Cebula[b,h], and Rita R. Colwell[a,b,i,j,2]

[a]Maryland Pathogen Research Institute, University of Maryland, College Park, MD 20742; [b]CosmosID Inc., College Park, MD 20742; [c]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201; [f]Angel Missions Haiti, Salem, VA 24153; [e]Laboratory Sciences Division, International Centre for Diarrheal Disease Research, Bangladesh, Dhaka-1000, Bangladesh; [d]Department of Microbiology and Immunology, University of Maryland School of Medicine, College Park, MD 20742; [g]Department of Microbiology, University of Dhaka, Dhaka-1000, Bangladesh; [h]Department of Biology, Johns Hopkins University, Baltimore, MD 21218; [i]Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742; and [j]Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205

Contributed by Rita R. Colwell, May 17, 2012 (sent for review March 26, 2012)

The millions of deaths from cholera during the past 200 y, coupled with the morbidity and mortality of cholera in Haiti since October 2010, are grim reminders that *Vibrio cholerae*, the etiologic agent of cholera, remains a scourge. We report the isolation of both *V. cholerae* O1 and non-O1/O139 early in the Haiti cholera epidemic from samples collected from victims in 18 towns across eight Arrondissements of Haiti. The results showed two distinct populations of *V. cholerae* coexisted in Haiti early in the epidemic. As non-O1/O139 *V. cholerae* was the sole pathogen isolated from 21% of the clinical specimens, its role in this epidemic, either alone or in concert with *V. cholerae* O1, cannot be dismissed. A genomic approach was used to examine similarities and differences among the Haitian *V. cholerae* O1 and *V. cholerae* non-O1/O139 strains. A total of 47 *V. cholerae* O1 and 29 *V. cholerae* non-O1/O139 isolates from patients and the environment were sequenced. Comparative genome analyses of the 76 genomes and eight reference strains of *V. cholerae* isolated in concurrent epidemics outside Haiti and 27 *V. cholerae* genomes available in the public database demonstrated substantial diversity of *V. cholerae* and ongoing flux within its genome.

Cholera is a disease endemic in many developing countries, and the causative agent, *Vibrio cholerae*, is autochthonous to estuaries and river systems worldwide (1). The recent outbreak of cholera in Haiti is a stark reminder of how rapidly a cholera epidemic can emerge and spread through a population. On January 12, 2010, a devastating 7.0 M_w earthquake struck Haiti, and, in the ensuing 12 d, ~52 aftershocks with a magnitude of 4.5 M_w or greater were recorded. It is estimated that more than 250,000 people died, ~300,000 were injured, and 1 million were rendered homeless. A very hot summer season followed. Amid the clearing of rubble and rebuilding of infrastructure, the first case of cholera was diagnosed on October 21, 2010, and then the disease quickly erupted throughout Haiti. On November 5, 2010, the full force of Hurricane Tomas skirted the island, causing torrential rains and severe flooding. Cholera, not officially reported in Haiti since 1960 (2), erupted in epidemic proportions (3), and, in just 1 y after the initial cases, cholera has claimed more than 6,500 lives and sickened approximately 500,000 people (4). The series of events of 2010 clearly created a "perfect storm" for cholera, with assignment of attribution remaining controversial (4).

Identical pulsed field gel electrophoresis patterns for 13 bacterial isolates recovered from Haitian patients with cholera led to a conclusion that the Haitian isolates derived from a strain similar to *V. cholerae* isolated in South Asia (3). From a limited comparison of genomes, Chin et al. concluded the Haitian cholera epidemic was most likely a result of an introduction of a strain from a distant geographic location and derived from human activity (5). A United Nations Independent Panel of Experts, in a report on the 2010 cholera outbreak in Haiti, concluded that the Haiti cholera outbreak was a result of contamination of water upstream in the Artibonite River with a pathogenic strain of a South Asian type of *V. cholerae* (6). Another report (7) that used whole-genome sequence typing of Nepalese isolates and compared them with a few previously sequenced *V. cholerae* strains, interpreted the results as indicating that Nepal was the likely origin of the Haitian outbreak. A rebuttal, however, was published, which pointed out that South Asia, not Nepal, might be the origin of the Haitian cholera outbreak strains, stating correctly that proper attribution cannot be accomplished by an exclusionary approach (8). Notably, none of the Nepalese soldiers had been recorded to be ill during the outbreak in Haiti (4); therefore, involvement of Nepalese peacekeepers remains controversial (9). A further study that used whole-genome phylogeny and core genome SNPs (8) showed the Haiti outbreak strain to be genetically related to strains originating in India and Cameroon. Like earlier studies, however, the authors concluded that a definitive genetic origin for the outbreak in Haiti must remain speculative (8).

Although the molecular and genomic data in the previous reports cited here have been interpreted as being consistent with *V. cholerae* being introduced to Haiti by outside visitors (5, 7, 8), a definitive statement of source attribution cannot yet be made. To assess source attribution, a qualified reference database (including recent strains from Nepal, India, and Cameroon, and related *V. cholerae* strains from concurrent epidemics) representative of global phylogenetic diversity as well as well-documented population diversity within the heterogeneous clade and its near neighbor comparators, is critical. In the present study, the genomic diversity of Haitian isolates was examined, and these data are presented to provide a broader perspective of the complexity of the species, *V. cholerae*, particularly of those isolates from Haiti, in an effort to contribute to the understanding of the origin of the outbreak as indigenous or introduced.

Results and Discussion

Clinical samples from 81 patients who showed symptoms of cholera with varying severity were collected from 18 towns across

---

Author contributions: N.A.H., M.A., C.M.F.-L., A. Cravioto, A.H., J.R., T.A.C., and R.R.C. designed research; N.A.H., S.Y.C., M.E., P.W.C., A. Chen, B.J.H., E.T., E.H., Q.S., L.J.T., J.B.P., K.F., M.M.H., H.L., A.H., T.A.C., and R.R.C. performed research; N.A.H., S.Y.C., M.E., P.W.C., A. Chen, B.J.H., E.T., H.L., A.H., T.A.C., and R.R.C. analyzed data; and N.A.H., M.E., T.A.C., and R.R.C. wrote the paper.

The authors declare no conflict of interest.

[1] N.A.H., S.Y.C., and M.E. contributed equally to this work.

[2] To whom correspondence should be addressed. E-mail: rcolwell@umiacs.umd.edu.

See Author Summary on page 11480 (volume 109, number 29).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207359109/-/DCSupplemental.

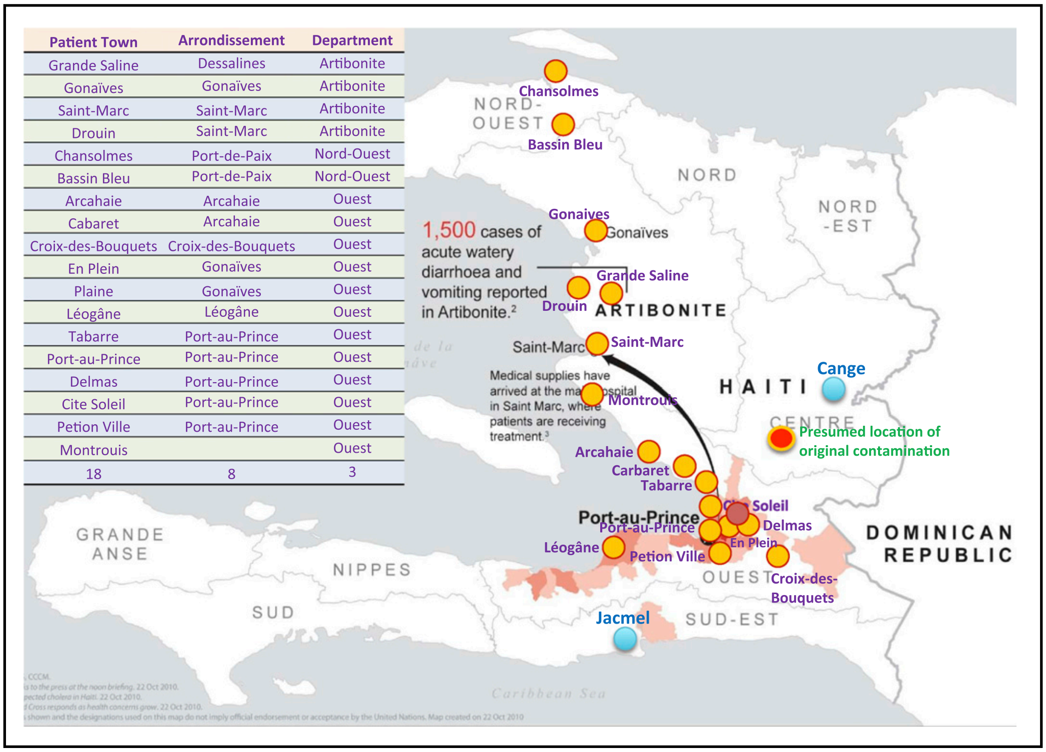
www.pnas.org/cgi/doi/10.1073/pnas.1207359109

eight Arrondissements of Haiti (Fig. 1 and Table S1), spanning 3 wk in November 2010, very early in the outbreak. Bacteriological analysis identified *V. cholerae* O1 and *V. cholerae* non-O1/O139 as pathogen in 48% ($n = 39$) and 21% ($n = 17$) of the samples, respectively. *V. cholerae* O1 and non-O1/O139 were cocultured from 7% of the O1 positive samples. Isolation of *V. cholerae* non-O1/O139, in addition to toxigenic *V. cholerae* O1, indicated that at least two disparate *V. cholerae* populations coexisted early in the outbreak. From the remaining 31% ($n = 25$) of the clinical samples, other enteric pathogens were also cultured, and subsequent analysis of a portion of the latter indicated the presence of opportunistic pathogens including *Aeromonas caviae*, *Aeromonas hydrophila*, and *Proteus mirabilis*. Environmental samples were also collected from two different towns in Haiti concurrent with the epidemics; bacteriological analysis of these samples identified *V. cholerae* non-O1/O139 from each of those samples.

Multilocus Variable-Number Tandem Repeat Analysis. Multilocus variable-number tandem repeat (VNTR) analysis (MLVA), a useful tool for tracking *V. cholerae* (10) and resolving distinct populations of clinical isolates from different geographic regions (11, 12), was used to examine the clinical and environmental *V. cholerae* O1 isolates from Haiti and other geographical locations (Table 1). Altered El Tor *V. cholerae* circulating in Southeast Asia has been reported to be a stable VNTR genotype over the years [Bangladesh, 2001–2005 (12), and India, 2004–2007 (11)], and our typing of the 9.3.6.n.n VNTR genotype supported the stability of that genotype. Haitian *V. cholerae* O1 (HCO1) strains, on the contrary, displayed a nearly identical VNTR genotype (Table 1), forming a tight clonal complex, as also reported previously (13). Investigation of *V. cholerae* strains from Zimbabwe, Zambia, Mexico, and Thailand confirmed, the Haitian *V. cholerae* strains contained a genetically distinct genotype (8.3.6.n.n) not observed previously, representing a unique VNTR genotype (Table 1). The MLVA data provide two valuable pieces of information. First, clonal transmission of VNTR alleles, or the limited VNTR diversity among the Haitian outbreak strains, indicates either a possible single source of contamination, in agreement with other findings (13); or a single clone of *V. cholerae* responsible for the epidemic. Secondly, Haitian strains exhibit a genetically distinct MLVA profile. Environmental *V. cholerae* non-O1/O139 isolates exhibited heterogeneous VNTR genotypes, with variability at all five loci (Table 1). Notable, however, was the observation that the clinical *V. cholerae* non-O1/O139 isolates revealed an identical VNTR genotype, akin to, but distinct from, the clinical *V. cholerae* O1 isolates.

Genomic Relatedness of Haitian Strains with Reference *V. cholerae* Strains. Whole-genome sequencing of 47 HCO1, 29 *V. cholerae* non-O1/O139, and eight reference strains of *V. cholerae* isolated outside Haiti in concurrent epidemics including in Bangladesh (CP1048, 2010) and Thailand (CP1042, 2010), as well as older epidemics (Table S1), was performed for this study. Genome comparisons of these 76 genomes and an additional 27 *V. cholerae* genomes available in the public database were conducted to ascertain the degree of genomic diversity within *V. cholerae* isolates from the same epidemics.

The Haitian genome sequences were found to be genetically similar to the recently emerged *V. cholerae* Altered El Tor strains CP1038 (Zimbabwe, 2009), CP1048 and CP1050 (Bangladesh, 2010), CP1041 (Zambia, 2004), and CP1042 (Thailand, 2010). These genomes belong to the seventh-pandemic clade, possessing a genomic backbone representative of *V. cholerae* O1 El Tor and *V. cholerae* O139, i.e., phylocore genome 1 clade (14, 15) with variations in their mobile genetic elements (MGEs) (Table S2). MGEs encode toxins, antibiotic resistance, virulence factors, and a wide range of proteins with as yet-unknown function. MGEs are resident in the natural genetic reservoir of *V. cholerae*



Patient Town	Arrondissement	Department
Grande Saline	Dessalines	Artibonite
Gonaïves	Gonaïves	Artibonite
Saint-Marc	Saint-Marc	Artibonite
Drouin	Saint-Marc	Artibonite
Chansolmes	Port-de-Paix	Nord-Ouest
Bassin Bleu	Port-de-Paix	Nord-Ouest
Arcahaie	Arcahaie	Ouest
Cabaret	Arcahaie	Ouest
Croix-des-Bouquets	Croix-des-Bouquets	Ouest
En Plein	Gonaïves	Ouest
Plaine	Gonaïves	Ouest
Léogâne	Léogâne	Ouest
Tabarre	Port-au-Prince	Ouest
Port-au-Prince	Port-au-Prince	Ouest
Delmas	Port-au-Prince	Ouest
Cite Soleil	Port-au-Prince	Ouest
Petion Ville	Port-au-Prince	Ouest
Montrouis		Ouest
18	8	3

1,500 cases of acute watery diarrhoea and vomiting reported in Artibonite.[2]

Medical supplies have arrived at the main hospital in Saint Marc, where patients are receiving treatment.[3]

Fig. 1. Source and distribution of isolates collected during the 2010 Haitian cholera outbreak. Yellow and blue circles indicate sites where clinical and environmental strains were collected, respectively. *Left*: Distribution of strains with respect to different locations of sample collection in Haiti.

Table 1. VNTR genotypes of a subset of *V. cholerae* strains isolated from Haiti and other geographic locations

ID	Origin, year	Source	Type	VC0147	VC0436-7*	VC1650	VCA0171	VCA0283
O395	India, 1965	C	O1 OGCL	7	4	3	24	14
N16961	Bangladesh, 1975	C	O1 INET	9	7	7	23	14
CIRS101	Bangladesh, 2002	C	O1 OGET	9	3	6	16	11
CP1032	Mexico, 1991	C	O1 OGET	8	7	8	13	17
CP1033	Mexico, 2000	C	O1 OGET	8	7	8	13	24
CP1038	Zimbabwe, 2009	C	O1 OGET	8	3	3	18	22
CP1041	Zambia, 2004	C	O1 OGET	8	7	6	9	10
CP1042	Thailand, 2010	C	O1 OGET	9	6	7	17	17
CP1048	Bangladesh, 2010	C	O1 OGET	9	3	6	22	17
AE-1786	Haiti, 2010	C	O1 OGET	8	3	6	14	20
AE-1792	Haiti, 2010	C	O1 OGET	8	3	6	14	ND
AE-1798	Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-7A1	Grande Saline, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-20A2	Saint-Marc, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-23A1	Bassin Bleu, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-38A1	Croix-des-Bouquets, Haiti, 2010	C	O1 OGET	8	3	6	12	16
HC-40A1	Arcahaie, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-43A1	Port-au-Prince, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-48A1	Port-au-Prince, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-49A2	Port-au-Prince, Haiti, 2010	C	O1 OGET	8	3	6	14	19
HC-61A1	Croix-des-Bouquets, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HC-61A3	Croix-des-Bouquets, Haiti, 2010	C	O1 OGET	8	3	6	14	ND
HC-70A1	Montrouis, Haiti, 2010	C	O1 OGET	8	3	6	14	12
HC-81A1	Cite Soleil, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HCUF01	Saint-Marc, Haiti, 2010	C	O1 OGET	8	3	6	14	20
HFU-02	Saint-Marc, Haiti, 2010	C	O1 OGET	8	3	6	14	19
HC-1A2	Saint-Marc, Haiti, 2010	C	Non-O1/O139	12	8	7	16	NA
HC-2A1	Saint-Marc, Haiti, 2010	C	Non-O1/O139	12	8	7	16	NA
HC-36A1	Tabarre, Haiti, 2010	C	Non-O1/O139	12	8	7	15	NA
HE-09	Cange, Haiti, 2010	E	Non-O1/O139	11	3	5	13	NA
HE39	Cange, Haiti, 2010	E	Non-O1/O139	8	3	2	13	NA
HE48	Cange, Haiti, 2010	E	Non-O1/O139	7	6	6	6	NA

Allele number indicates the number of repeat units of each locus. The repeat units of the MLVA loci are as follows: AACAGA at VC0147, GACCCTA at VC0436-7, GATAATCCA at VC1650, TGCTGT at VCA0171, and ACCAGA at VCA0283. C, clinical; E, environmental; INET, Inaba El Tor; NA, not applicable; ND, not determined; OGCL, Ogawa classical; OGET, Ogawa El Tor.

*Intergenic.

and related *Vibrio* spp., including *Vibrio parahaemolyticus* and *Vibrio mimicus* (14, 15). Thus, within a dynamic microbial gene pool, in which horizontal gene transfer is a significant driver of microbial evolution (16), MGEs play a pivotal role in the genesis of diversity and hypervariability of *Vibrio* genomes (14, 15). The MGE catalogue of the HCO1 strains demonstrates that they share genetically similar genomic and pathogenic islands (n = 18) with the recently isolated Altered El Tor *V. cholerae* strains from South Asia and Africa (Table S2). Like the concurrent epidemic strains, CTXΦ genes were located in the large chromosome (C-I) of the HCO1 isolates (Table S2). RS1 was detected in most of the HCO1 strains, but four strains appeared to lack RS1 (Table S2), as does strain (CP1041) from a 2004 epidemic in Zambia. In addition, all HCO1 isolates, except HC49A2, which also lacked RS1, harbor TLC (Toxin-linked cryptic plasmid, a genetic element that plays some role in the biology of CTXΦ, perhaps facilitating its acquisition or its replication). As anticipated, none of the Haitian *V. cholerae* non-O1/O139 strains had CTXΦ, RS1, or TLC.

Two genomic islands (GIs) in two HCO1 strains (GI-16 in AELH.1792 and GI-120 in HC-77A1) were located in the same insertion site (VCA0569-VCA0570) in the small chromosome (C-II) but absent in all concurrent epidemic strains examined in this study. GI-16 is a 6.6-kb island engendering *Vibrio* VSK prophage, a pre-CTX prophage, integration of which has previously been described in *V. cholerae* O139 MO10 (in C-I) (15) and non-O1/O139 AM-19226 and 1587 (in C-II) (16), but heretofore not reported in *V. cholerae* O1 to our knowledge. Further inspection identified GI-16, with a genetically similar set of coding sequences (CDS) on C-I (VC1451-VC1456), in six of the Haitian non-O1/O139 strains, including one strain (HC-2A1) containing GI-16 at both the C-I and C-II insertion sites (Table S2). GI-120, inserted at the same insertion site in HC-77A1, is a 31.5-kb island with 30 CDSs, encoding different transporters and stress response protein. Our search for this GI identified another Haitian strain, HC-78A1, a clinical non-O1/O139 isolate, harboring this GI at the same chromosomal location. Subsequent analysis indicated GI-120 might be the result of rearrangement of the C-I loci VC1416 to VC1446 into C-II. In addition, all HCO1 strains contained the 10 GIs (GI-01 to GI-10) previously reported in *V. cholerae* O1 (Table S2). In contrast, the 29 clinical and environmental *V. cholerae* non-O1/O139 strains revealed a number of GIs (n = 40) encoding genes for virulence, survival, stress response, and metabolism, and distributed across both chromosomes. The environmental strains possessed a substantially larger repertoire of GIs (Table S2).

Variations in CTXΦ-RS1. CTXΦ present in toxigenic *V. cholerae* consists of two functionally distinct gene clusters, core and RS2 region (17, 18). The core contains *ctx*AB, encoding cholera toxin (CT), together with five other genes—*psh*, *cep*, *pIIICTX*, *ace*, and *zot*—required for phage morphogenesis. RS2 encodes proteins with roles in replication (RstA), integration (RstB), and regulation (RstR) of site-specific recombination of CTXΦ (17, 19). In toxigenic *V. cholerae* El Tor strains, a satellite phage, RS1, invariably is present and located adjacent to CTXΦ. It is similar to RS2 except that it contains an additional gene, *rstC*, that encodes an antirepressor protein, promoting transcription of CTXΦ genes required for generating infectious particles (19). The CTXΦ and RS1 element array in the Haitian cholera isolates are similar, but not identical, to those found in the Altered El Tor *V. cholerae* strains, i.e., CIRS101.

We observed a single amino acid deletion (delta S), a GTA deletion at nucleotides 77 to 79 in *rst*B of the HCO1 strains (Table 2). This deletion, a unique feature for a classical biotype of *V. cholerae* (20), was not detected in any of the Altered El Tor, Hybrid, or El Tor variants of *V. cholerae*, except *V. cholerae* CP1048 (Bangladesh, 2010) and CP1032 (Mexico, 1991). Interestingly, *V. cholerae* INDRE 91/1 (Mexico, 1991) and RC27 (Indonesia, 1991) contained the GTA deletion, underscoring this deletion is not novel to Western Hemisphere isolates. Moreover, PCR amplification and sequence analysis of *rstB* gene identified unique polymorphisms (A to G at nucleotide 192 and G to A at nucleotide 108). The former was found only in the HCO1 strains, whereas the latter was also detected in the Mexico strains, *V. cholerae* INDRE91/1 and CP1032, suggesting the G-to-A polymorphism may be specific to Western Hemisphere strains, a conclusion that requires additional data.

Variation was also detected in the number of ToxR binding repeats (TTTTGAT) between *zot* and *ctx*A. *V. cholerae*. ToxR is a global transcriptional regulator of virulence gene expression, and this repeated sequence is required for ToxR binding and activation of the *ctx*AB promoter. The ToxR-binding site is located immediately upstream of *ctx*AB, and affinity of ToxR binding is influenced by the repeat sequences (21). The HCO1 strains contained five copies of this repeat (Table 2), compared with three or four copies in other Altered El Tor, Hybrid, or El Tor variants of *V. cholerae*, including recent epidemic strains (CP1048, CP1038, and CP1032), and seven copies were found in the classical biotype of *V. cholerae* O395. The presence of an increased number of ToxR binding repeats has been hypothesized to correlate with a severe form of cholera (22, 23), which may be the case in the Haitian epidemic (24).

A further feature of the Haitian strains was the presence of three mutations in *ctx*B compared with two in Altered El Tor Hybrid or El Tor variant strains of *V. cholerae* O1 (24). The same *ctxB* polymorphisms in *V. cholerae* have previously been reported for strains from Kolkata and Orissa, India, and most recently in *V. cholerae* isolates from Nigeria, Cameroon, Kenya, Sudan, Ethiopia, and Nepal (25, 26). Taken together, these data suggest that the ancestor to the strains responsible for the Haitian outbreak successfully used genetic changes in CTXΦ and RS1 element to affect disease severity, an attribute of the classical biotype, as well as better environmental fitness, an attribute of the El Tor biotype (5, 9–11).

***Vibrio* Pathogenicity Island-1.** *Vibrio* pathogenicity island-1 (VPI-1) codes for the toxin coregulated pilus involved in biofilm formation, attachment to chitinous organisms, and as the receptor for CTXφ (27). This island was found in all HCO1 isolates but not in clinical or environmental *V. cholerae* non-O1/O139 isolates. Interestingly, although devoid of VPI-1, three different GIs were identified at the VPI-1 site in 85% of non-O1/O139 isolates (Table S3). GI-41 was identified at the VPI-1 locus in one environmental and 15 clinical isolates, whereas other GIs, GI-102 and GI-81, were identified in other environmental isolates (Table S3). It is of interest that GI-81 represents a part of the *Vibrio* Seventh Pandemic (VSP) Island II (VSP-II) but is inserted at the VPI-1 site (Tables S2 and S3).

***Vibrio* Pathogenicity Island-2.** Most of the genomes have encoded sequences with at least some homology to the *Vibrio* pathogenicity island-2 (VPI-2) inserted at the canonical locus for this island (VC1758 and VC1810) (15). Whereas all *V. cholerae* O1 strains contained a complete VPI-2, many non-O1/O139 strains, both clinical and environmental, contained only the sialidase sequence of the sialic acid metabolism region of VPI-2. That is, the VPI-2 locus in non-O1/O139 strains lack most of the VPI-2 ORFs, including the type 1 restriction modification system and mu-like phage regions (Table S3). Interestingly, the sialidase has been shown to unmask the GM1 gangliosides of human intestinal epithelial cells, thus making them more available to CT (28). However, because many of these strains do not encode CT, the high level of conservation of sialidase among these strains suggests an additional role for sialidase in the life cycle of *Vibrio*. Notably, the clinical non-O1/O139 isolate HC1A2 contained GI-121 inserted at VPI-2 site and part of GI-09 encoding genes responsible for response regulator, autolysin sensor kinase, and ABC-type transport system. An environmental isolate, *V. cholerae* HE25, encoded a type III secretion system within the VPI-2 region. The type III secretion systems most frequently found in *V. cholerae* serogroups non-O1/non-O139 induce severe diarrhea in cholera infection models (29), and this secretion system has been reported in *V. cholerae* strains that encode and express the O1-antigen (15). Our results are consistent with previous studies that showed VPI-2 to be highly heterogeneous in clinical and environmental strains (15, 28).

VSP Islands. All HCO1, but none of the Haitian *V. cholerae* non-O1/O139 genomes examined in this study, carried VSP-I. Interestingly, an environmental *V. cholerae* non-O1/O139, HE16, carried a previously uncharacterized GI at the VSP-I insertion loci. Similarly, all HCO1 carried a variant VSP-II, first reported in *V. cholerae* O1 CIRS101, although none of the *V. cholerae*

Table 2. Sites of nucleotide variation in CTX prophages and RS1 elements

		CTX prophage/RS1 element																								
	Gene	*rstR*	*rstA*																*rstB*						*cep-zot*	*zot-ctxA**	*ctxB*
Strain	position	—	27	51	52	162	183	258	345	354	414	459	516	540	558	579	606	774	77–79	90	96	288	291	344	80 Sites	1,197 to −74[†]	—
N16961	CTX^{ET}	ET	C	T	T	C	C	G	G	C	T	C	G	A	C	T	T	C	GTA	A	T	A	C	G	E	4	E
O395	CTX^{CL}	CL	T	.	.	T	A	C	T	.	.	.	A	G	.	C	C	T	—	T	C	G	T	.	C	7	C
IB1346	CTX^{MOZ}	CL	T	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	E	3, 4	C
IB1617	CTX^{ENV}	Env	T	C	G	T	.	.	.	T	A	T	A	G	T	C	.	.	.	.	.	.	.	T	E	3, 4	C
CIRS101	CTX^{HYB}	ET	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	E	3	C
HCO1	CTX^{HYB}	ET	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	—	.	.	.	.	.	E	5	C^{6}

Dots (.) indicate sequence identical to that of *V. cholerae* N16961; (—) indicates deletion. E or ET, El Tor type; C or CL, classical type; C^{6}, cholera toxin type 6; Env, environmental type.

*ToxT binding site repeats.

[†] 74 nucleotide before the start of ctxA.

non-O1/O139 were found to carry this or any other variant at the VSP-II locus (Table S3). The VSP-II island in O1 strains was inserted at the usual insertion locus in the genome and, as observed in *V. cholerae* CIRS101, revealed many deletions compared with the VSP-II in other *V. cholerae* O1 seventh pandemic strains described previously (30).

Integrated Conjugated Elements. Integrated conjugated elements (ICEs) found in HCO1 strains were compared with ICE*Vch*Ban5 and ICE*Vch*Moz in *V. cholerae* O1 CIRS101 and B33, respectively. Analysis of their genetic organization revealed HCO1 ICE is >99% similar to ICE*Vch*Ban5, differing only by 5SNPS (31). In contrast, the majority of the Haitian *V. cholerae* non-O1/O139 strains carry an ICE sequence variant; only three strains lacked an insertion at this site (Table S3). All the ICE insertions encode an integrase gene of the P4 integrase family and integrate at *prfC* on C-I. Minor differences between these elements and ICE*Vch*Ban5 were observed at the junctions between sequence contigs. In contrast, the ICE found in Haitian clinical *V. cholerae* non-O1/O139 was distinct from any ICE reported to date, notably lacking the cluster of CDSs encoding antibiotic resistance. Although this might be expected in view of the higher variability observed in ICE present in *V. cholerae* non-O1/O139 from other geographical areas (32), it is interesting that the same element is present in each of the Haitian *V. cholerae* non-O1/O139 strains. Even more intriguing is the fact that these *V. cholerae* non-O1/O139 strains were of clinical origin and in 21% the only isolate cultured from the patient.

Mutations in *gyrA* and *parC*. The Haitian *V. cholerae* O1 isolates were resistant to sulfisoxazole, trimethoprim-sulfamethoxazole, furazolidone, streptomycin, and nalidixic acid, and displayed reduced susceptibility to ciprofloxacin (3, 5). The strains were sensitive to azithromycin and tetracycline. Analysis of the *gyrA* and *parC* (VC1258 and VC2430, respectively, in Table S3) genes revealed two point mutations in the Haitian strains, i.e., a ser83ile substitution in *gyrA* and a ser85leu substitution in *parC*. Both point mutations are associated with quinolone resistance in clinical *V. cholerae* and have been reported in India and most recently in Nigeria and Cameroon (33–37). We observed the same point mutations in *gyrA* and *parC* in recent isolates from Zimbabwe (CP1038), Thailand (CP1042), and Bangladesh (CP1048). In contrast, whereas *V. cholerae* O1 CIRS101 (Bangladesh, 2002) and CP1041 (Zambia, 2004) contained the same point mutation at *gyrA*, neither strain harbored the *parC* mutation, offering evidence of similarity with, but not identity to, *V. cholerae* in Asia and Africa.

Phylogenomics. The phylogeny of the Haitian *V. cholerae* isolates was inferred by constructing a genome-relatedness neighbor-

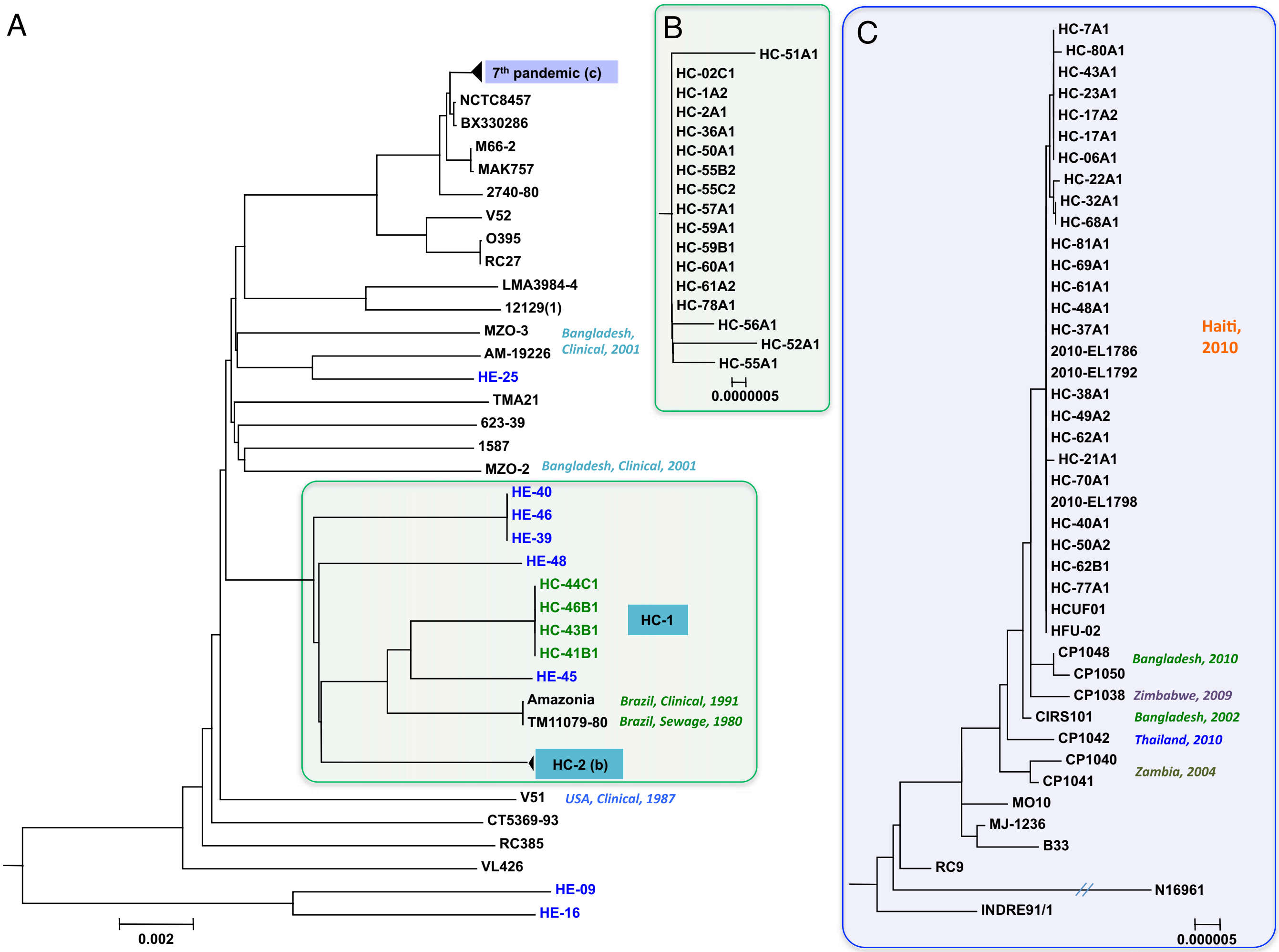


Fig. 2. Neighbor-joining trees showing phylogenetic relationships based on 992 orthologous protein-coding genes (~989,552 bp) of (*A*) 96 *V. cholerae* genomes representing diverse serogroups (strains with blue and green color represent Haitian environmental and clinical isolates, respectively), (*B*) *V. cholerae* non-O1/O139 strains representing clinical non-O1/O139 group 2, and (*C*) *V. cholerae* O1 and O139 strains representing the seventh pandemic cluster.

joining tree by using homologous alignment of 992 orthologous protein-coding genes (~989,552 bp) of 96 *V. cholerae* genomes, as a strict measure of the core *V. cholerae* genome. The evolutionary tree (Fig. 2*A*) corroborated the previously described 1.7-million-bp phylogeny of *V. cholerae* (15), showing serogroups of *V. cholerae* evolving into distinct evolutionary lineages. Significantly, 26 of 29 *V. cholerae* non-O1/O139 strains included in our phylogenomic analysis clustered with toxigenic *V. cholerae* O1 strains (i.e., Amazonia and TM11079-80; Fig. 2*A*) implicated in sporadic cases and severe epidemics originating in the Western Hemisphere. The phylogeny is concordant with the observation that non-O1/O139 strains isolated from the Haitian clinical samples harbor a battery of virulence/fitness factors organized on MGEs. That is, despite being non-O1/O139, these strains harbor the genomic backbone similar to some toxigenic *V. cholerae* O1, explaining the coexistence of non-O1/O139 with *V. cholerae* O1. Furthermore, the clinical non-O1/O139 isolates formed two closely related clusters (HC-1 and HC-2), with the majority of the non-O1/O139 isolates clustering together in HC-2 (Fig. 2*B*). Both clinical and environmental isolates clustered together, along with 1991 and 1980 *V. cholerae* O1 isolates from Brazil, indicating potential indigenous origin of these strains.

The HCO1 strains, on the contrary, clustered with other seventh pandemic *V. cholerae* strains in a single monophyletic clade (Fig. 2*A*). A tree constructed with only seventh pandemic strains (Fig. 2*C*) showed clustering of the Haitian strains together, as well as with clinical isolates from Zimbabwe (CP1038) and Bangladesh (CP1048 and CP1050), an observation in agreement with the recent report of shared ancestry of Haiti, India, and Cameroon isolates (8). HCO1 isolates subdivided into several closely related clusters within a tight monophyletic clade indicating stochastic biological transmission events resulting in genetic drift among the strains isolated during the first 3 wk of the outbreak. Genetic drift would account for radiation of ~50% of the isolates away from the basal Haitian isolates. To better demonstrate the heterogeneity of these clonal subpopulations, principal component analysis (PCA), a multivariate method useful for exploratory cluster visualization, was used. Based on the divergence of average nucleotide and tetra nucleotide frequency matrix using R, PCA analysis showed the Haitian strains formed a heterogeneous cloud of clonal subpopulations (Fig. S1), indicating that two closely related isolates from Bangladesh (CP1048) and Zimbabwe (CP1048) hard to exclude as maximum variations observed among the various Haitian strains must be included (Fig. S1).

Relatedness of Outbreak and Reference Isolates by Whole-Genome Mapping. Whole-genome maps were created for a set of Haitian *V. cholerae* isolates and isolates from concurrent epidemics elsewhere. Comparison of the whole-genome maps showed all HCO1 outbreak strains to be identical within the limits of the mapping technology resolution, suggesting the outbreak to be clonal. Detailed whole-genome map comparisons of outbreak isolates with closely related reference isolates CP1048 and CP1038 revealed substantial similarity, yet differences were observed (Fig. 3 and Fig. S2). Comparison of these regions with annotated sequences of a previously studied isolate, AE-1786, showed that, although most of the HCO1 and CP1048 (Bangladesh, 2010) contained TLC of ~18 kb, two strains AE-1786 (Haiti, 2010) and CP1038 (Zimbabwe, 2009), displayed a TLC of ~13 kb (Fig. 3*A*). Distinct polymorphism was also observed in the mannose-sensitive hemagglutinin (MSHA; Fig. 3*B*), a surface pilus contributing the ability of *V. cholerae* to adhere to plankton (38). Detailed whole-genome map comparisons demonstrated three Haitian strains—AE-1786, HC61A1, and HC46A1—harbored an ~18-kb tRNA cluster in MSHA, whereas other HCO1 and two closely related reference isolates—CP1048 and CP103—contained an approximately 24-kb cluster (Fig. 3*B*). A larger insert (~27 kb) was identified at the ICE locus in Haitian strain HC20A2, compared with the ~18-kb insert found in other HCO1 strains (Fig. 3*D*), as well as CP1048 and CP1038.

Genome similarity clustering (Fig. 4) of whole-genome maps for both chromosomes showed the outbreak isolates did not differ substantially from CP1038 and CP1048. However, other strains from Thailand (CP1042), Mexico (CP1032 and 1033), and Zambia (CP1041) differed by >2% and therefore clustered separately. All clinical and environmental non-O1/O139 strains clustered in a monophyletic group. The genome similarity clus-

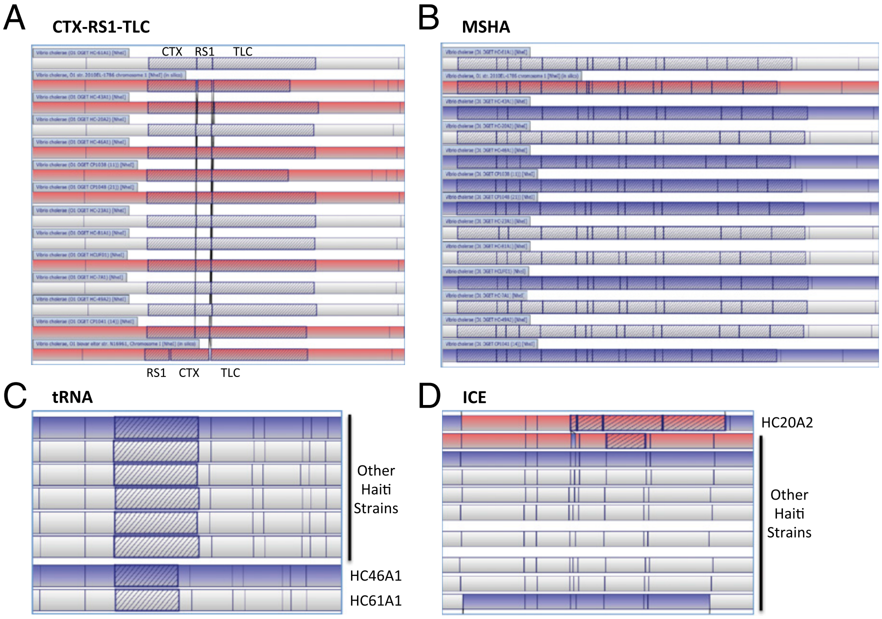


Fig. 3. Polymorphisms in Haitian outbreak and reference isolates as determined by whole-genome mapping. Polymorphism detected in (*A*) TLC element, (*B*) MSHA gene clusters, (*C*) tRNA region, and (*D*) ICE element. Highlighted by the blue colored crossed line box are regions where polymorphisms were detected.

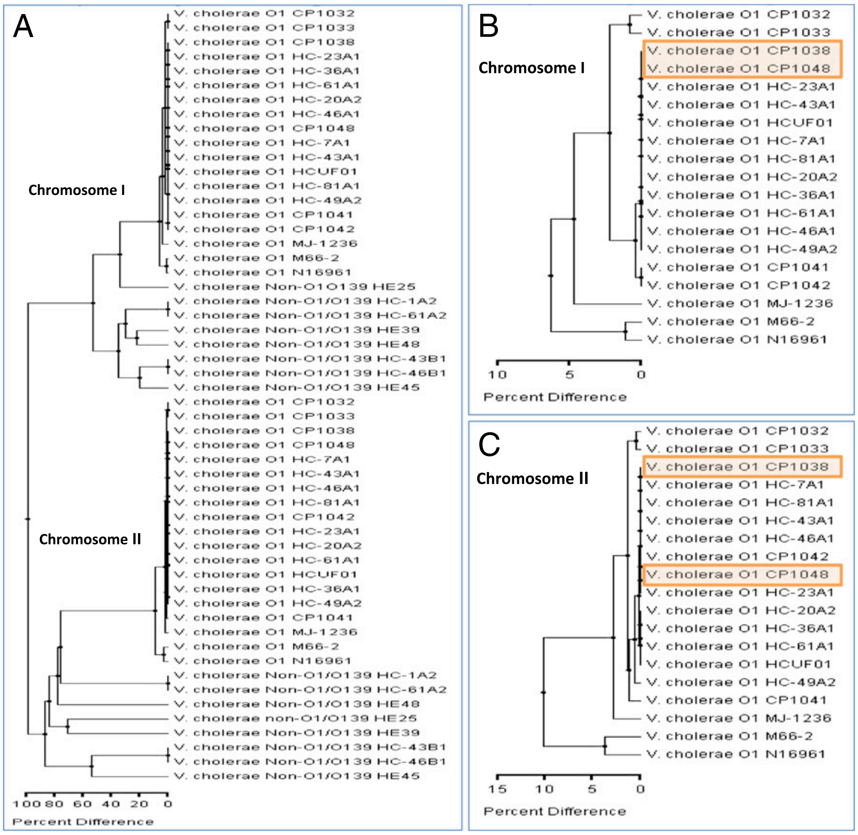


Fig. 4. Whole-genome map similarity cluster of Haitian outbreak isolates with reference strains by using unweighted pair group method with arithmetic mean: (*A*) Chromosome I and II of each isolates clustered together; (*B*) similarity cluster of *V. cholerae* O1 based on chromosome I; and (*C*) similarity cluster of *V. cholerae* O1 based on chromosome II.

tering confirmed heterogeneity of the *V. cholerae* non-O1/O139, as reported previously (38).

SNP-Derived Phylogeny. To establish a high-resolution phylogenomic framework, only 16 genomes were subjected to SNP discovery and SNP validation. Our developed bioinformatics pipeline takes into account genome coverage and corrects for paralogous genomic regions. This analysis yielded a panel of 277 high-quality SNPs (70 synonymous, 169 nonsynonymous, and 38 intergenic SNPs; Table S4) distributed stochastically within the chromosome I (213 SNPs) and II (64 SNPs), without indication of mutational hotspots (Figs. S3 and S4). The SNPs showed only two different alleles for every strain at each assayed position (Table S3). The high incidence of biallelism suggests limited diversity among the analyzed genomes. The high-resolution of the SNP-derived phylogenetic scheme allowed accurate placement of the Haitian *V. cholerae* strains, separated by 12 Haiti-specific SNPs from the rest of the strains included in the analysis (Fig. 5). Haitian *V. cholerae* strains formed a closely related cluster, with Haitian *V. cholerae* 1798 placing in the center of the radial expansion. The phylogeny shows genetic relationship of Haitian *V. cholerae* strains with *V. cholerae* CP1048, CP1038, CP1041, and CP1042 (Table 1). The Haitian *V. cholerae* strains feature as many as five isolate-specific SNPs indicative of microevolution within the timeframe of the Haitian cholera epidemic. Only a single shared SNP is carried in the Haitian *V. cholerae* from *V. cholerae* HC38A1 to HC48A1.

Conclusions

Genomic analysis of *V. cholerae* isolates obtained early in the Haitian cholera epidemic has provided evidence that two distinct *Vibrio* populations, *V. cholerae* O1 and *V. cholerae* non-O1/O139, contributed to the cholera epidemic in Haiti. Results of comprehensive genomic analysis showed that *V. cholerae* O1 populations were clonal, resembling epidemic isolates from South Asia and Africa. *V. cholerae* non-O1/O139 populations were not clonal, but most probably serve as a reservoir for genomic and pathogenicity islands. Results of our analysis demonstrated that *V. cholerae* non-O1/O139 populations in Haiti harbor a genomic backbone similar to that of toxigenic *V. cholerae* O1 circulating in the Western hemisphere. Core gene and SNP-derived phyloge-

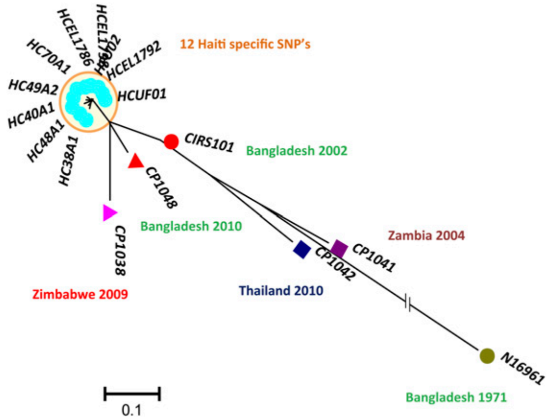


Fig. 5. SNP-based phylogenetic tree for *V. cholerae* O1, constructed by using 277 intra- and intergenic SNPs identified in the *V. cholerae* O1 population, revealing clustering and radial expansion of the Haitian *V. cholerae* strains.

nies suggest, and PCA findings reinforce, that, quite quickly, i.e., within a 3-wk period early in the cholera epidemic, substantial genomic diversity accumulated in the circulating population. Our study speaks to the critical need for an up-to-date, properly curated, and publicly available reference genomic database (including geographically diverse and circulating strains from cholera endemic regions) reflecting quality coverage of global phylogenetic diversity, as well as diversity within this heterogeneous clade and its near neighbors if investigations of future *V. cholerae* epidemics are to be effective and useful for public health applications. Moreover, it may also be necessary to generate the database at the time of a study, as there are evidence of genetic drift during the short time frame of routine laboratory passage of *V. cholerae*. Therefore, a need to resequence reference genome, as the apposite comparator, is also underscored if SNP data are to be used to assign attribution in outbreak investigations.

Materials and Methods

Traditional methods for isolating and identifying *V. cholerae* from environmental water and stool samples were used for bacteriological analysis of all samples. The identity of presumptive colonies was confirmed by using a battery of specific biochemical, serological, and molecular methods. Five loci for MLVA were analyzed by using methods described in previous studies (10–12). Genomic DNA of 76 confirmed *V. cholerae* strains along with eight reference strains were subjected to next generation whole-genome Illumina and hybrid Illumina/454 sequencing. Whole-genome maps were generated through OpGen,to facilitate assembly and a detailed study of polymorphisms and their respective genome localization in the studied population. Genome-to-genome comparisons, identification and characterization of MGEs, as well as core genome phylogenetics were performed by using methods described previously (14, 15). A NUCmer-based bioinformatics pipeline was used for discovery of fine SNPs and curated manually for any false-positive calls. Genomic regions with identities more than 98% on C-I and C-II were excluded, and regions that were not conserved among the strains were also excluded from further analysis. A further detailed description of the materials and methods is provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. The authors thank Jinna Choi, Elizabeth J. Sancomb, Jonathan Crabtree, Amy Egan, Naomi Sengamalay, and Lisa Sadzewicz for their assistance in sample processing, sequencing, and logistics; Stephanie Lansing (Department of Environmental Science and Technology, University of Maryland) and Michael McDonald (Center for Health Security and Infrastructure, University of Maryland) for providing environmental samples; and Judith A. Johnson, Afsar Ali, and J. Glenn Morris, Jr. (University of Florida) for two outbreak isolates of *V. cholerae*. This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Department of Health and Human Services, under Contract HHSN2722009000. This work was also supported in part by Department of Homeland Security Grant BAA 10-19, National Oceanic and Atmospheric Administration Grant SO660009, National Science Foundation/NIH Ecology of Infectious Diseases Program Grant EF-0813066, and NIH Grant 1R01A13912901.

1. Colwell RR (1996) Global climate and infectious disease: the cholera paradigm. *Science* 274:2025–2031.
2. Rottmann S (October 25, 2010) Haiti cholera outbreak causes not clear, experts say. *BBC News.* Available at http://www.bbc.co.uk/news/world-latin-america-11618352.
3. Centers for Disease Control and Prevention (2010) *Laboratory Test Results of Cholera Outbreak Strain in Haiti Announced* (CDC, Atlanta).
4. Farmer PE, Ivers LC (2012) Cholera in Haiti: The equity agenda and the future of tropical medicine. *Am J Trop Med Hyg* 86:7–8.
5. Chin CS, et al. (2011) The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364:33–42.
6. United Nations (2011) *Final Report of the Independent Panel of Experts on the Cholera Outbreak in Haiti, May 4, 2011.* Available at http://www.un.org/News/dh/infocus/haiti/UN-cholera-report-final.pdf. Accessed March, 2012.
7. Hendriksen RS, et al. (2011) Population genetics of Vibrio cholerae from Nepal in 2010: Evidence on the origin of the Haitian outbreak. *mBio* 2:e00157-11.
8. Reimer AR, et al.; V. cholerae Outbreak Genomics Task Force (2011) Comparative genomics of *Vibrio cholerae* from Haiti, Asia, and Africa. *Emerg Infect Dis* 17: 2113–2121.
9. Pun SB (2011) South Asia instead of Nepal may be the origin of the Haitian cholera outbreak strain. *mBio* 2:e00219-11.
10. Kendall EA, et al. (2010) Relatedness of *Vibrio cholerae* O1/O139 isolates from patients and their household contacts, determined by multilocus variable-number tandem-repeat analysis. *J Bacteriol* 192:4367–4376.
11. Ghosh R, et al. (2008) Epidemiological study of *Vibrio cholerae* using variable number of tandem repeats. *FEMS Microbiol Lett* 288:196–201.
12. Stine OC, et al. (2008) Seasonal cholera from multiple small outbreaks, rural Bangladesh. *Emerg Infect Dis* 14:831–833.
13. Ali A, et al. (2011) Recent clonal origin of cholera in Haiti. *Emerg Infect Dis* 17: 699–701.
14. Hasan NA, et al. (2010) Comparative genomics of clinical and environmental *Vibrio mimicus*. *Proc Natl Acad Sci USA* 107:21134–21139.
15. Chun J, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci USA* 106: 15442–15447.
16. Treangen TJ, Rocha EP (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* 7:e1001284.
17. Safa A, Nair GB, Kong RY (2010) Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol* 18:46–54.
18. Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272:1910–1914.
19. Waldor MK, Rubin EJ, Pearson GD, Kimsey H, Mekalanos JJ (1997) Regulation, replication, and integration functions of the *Vibrio cholerae* CTXphi are encoded by region RS2. *Mol Microbiol* 24:917–926.
20. Choi SY, et al. (2010) Classical RS1 and environmental RS1 elements in *Vibrio cholerae* O1 El Tor strains harbouring a tandem repeat of CTX prophage: Revisiting Mozambique in 2005. *J Med Microbiol* 59:302–308.
21. Pfau JD, Taylor RK (1996) Genetic footprint on the ToxR-binding site in the promoter for cholera toxin. *Mol Microbiol* 20:213–222.
22. Kaper JB, Morris JG, Jr., Levine MM (1995) Cholera. *Clin Microbiol Rev* 8:48–86.
23. Kay AB (1994) Isolation and identification of *Vibrio cholerae* O1 from fecal specimens. *In Vibrio cholerae and Cholera: Molecular to Global Perspectives* (ASM Press, Washington, DC), pp 3–24.
24. Nair GB, et al. (2006) Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. *J Clin Microbiol* 44:4211–4213.
25. M. Humphrys, et al. (2011) Cholera Toxin Genotyping Of Recent Outbreak Strains Of *Vibrio cholerae. Mismatch PCR 111th General Meeting American Society for Microbiology* (New Orleans, Louisiana), p 966.
26. Kumar P, et al. (2009) A large cholera outbreak due to a new cholera toxin variant of the *Vibrio cholerae* O1 El Tor biotype in Orissa, Eastern India. *J Med Microbiol* 58: 234–238.
27. Boyd EF, Moyer KE, Shi L, Waldor MK (2000) Infectious CTXPhi and the vibrio pathogenicity island prophage in *Vibrio mimicus*: Evidence for recent horizontal transfer between *V. mimicus* and *V. cholerae*. *Infect Immun* 68:1507–1513.
28. Almagro-Moreno S, Boyd EF (2009) Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine. *Infect Immun* 77: 3807–3816.
29. Shin OS, et al. (2011) Type III secretion is essential for the rapidly fatal diarrheal disease caused by non-O1, non-O139 *Vibrio cholerae*. *mBio* 2:e00106-11.
30. Taviani E, et al. (2010) Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis. *FEMS Microbiol Lett* 308:130–137.
31. Sjölund-Karlsson M, et al. (2011) Drug-resistance mechanisms in *Vibrio cholerae* O1 outbreak strain, Haiti, 2010. *Emerg Infect Dis* 17:2151–2154.
32. Taviani E, Grim CJ, Chun J, Huq A, Colwell RR (2009) Genomic analysis of a novel integrative conjugative element in *Vibrio cholerae*. *FEBS Lett* 583:3630–3636.
33. Baranwal S, Dey K, Ramamurthy T, Nair GB, Kundu M (2002) Role of active efflux in association with target gene mutations in fluoroquinolone resistance in clinical isolates of *Vibrio cholerae*. *Antimicrob Agents Chemother* 46:2676–2678.
34. Garg P, et al. (2000) Expanding multiple antibiotic resistance among clinical strains of *Vibrio cholerae* isolated from 1992-7 in Calcutta, India. *Epidemiol Infect* 124:393–399.
35. Krishna BV, Patil AB, Chandrasekhar MR (2006) Fluoroquinolone-resistant *Vibrio cholerae* isolated during a cholera outbreak in India. *Trans R Soc Trop Med Hyg* 100: 224–226.
36. Islam MS, Midzi SM, Charimari L, Cravioto A, Endtz HP (2009) Susceptibility to fluoroquinolones of *Vibrio cholerae* O1 isolated from diarrheal patients in Zimbabwe. *JAMA* 302:2321–2322.
37. Quilici ML, Massenet D, Gake B, Bwalki B, Olson DM (2010) *Vibrio cholerae* O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. *Emerg Infect Dis* 16: 1804–1805.
38. Chiavelli DA, Marsh JW, Taylor RK (2001) The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl Environ Microbiol* 67: 3220–3225.