Genomic diversity of 2010 Haitian cholera outbreak strains

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

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

AUTHOR SUMMARY

Cholera is a disease endemic in many developing countries, and its causative agent, \textit{Vibrio cholerae}, is indigenous to estuaries and river systems worldwide (1). The devastating 2010 cholera epidemic in Haiti generated worldwide attention. Assignment of potential source was explored in several published works, but a definitive assignment was not established. In this study, genomes of \textit{V. cholerae} isolates from Haitian patients with cholera were examined to explore diversity within this species. \textit{V. cholerae} O1 and non-O1/O139 were isolated repeatedly from patients in early weeks of the outbreak, underscoring the difficulty in verifying the ultimate source and origin of outbreak strain(s). Our analysis indicates the importance of maintaining up-to-date public databases of pathogen genome sequences to combat and even prevent future outbreaks of cholera.

In the present study, we report the analysis of clinical samples from 81 patients showing symptoms of cholera with varying severity from 18 towns across eight Arrondissements of Haiti, spanning 3 wk in November 2010, very early in the outbreak. Bacteriological analysis identified \textit{V. cholerae} O1 and \textit{V. cholerae} non-O1/O139 as pathogen in 48% and 21% of the samples, respectively. \textit{V. cholerae} O1 and non-O1/O139 were cocultured from 7% of the O1 positive samples. We sequenced whole genomes of 47 \textit{V. cholerae} O1 and 29 \textit{V. cholerae} non-O1/O139 isolates from patients and the environment in Haiti as well as eight reference strains of \textit{V. cholerae} O1 isolated in concurrent epidemics outside Haiti. Our analysis of these Haitian isolates, along with 27 \textit{V. cholerae} genomes (2, 3) available in public databases, demonstrates the utility of combining multiple approaches [multilocus variable-number tandem repeat (VNTR) analysis, SNP, mobile genetic elements, core genome phylogeny, and whole-genome mapping, formerly known as optical mapping] in defining genetic diversity for a given population. The results revealed that two distinct \textit{Vibrio} populations, \textit{V. cholerae} O1 and \textit{V. cholerae} non-O1/O139, contributed to the cholera epidemic in Haiti. Because non-O1/O139 \textit{V. cholerae} was the sole pathogen isolated from 21% of the clinical specimens, its role in this epidemic, either alone or in concert with \textit{V. cholerae} O1, cannot be dismissed.

Results of comprehensive genomic analysis showed that \textit{V. cholerae} O1 populations were clonal, resembling epidemic isolates from South Asia and Africa, whereas \textit{V. cholerae} non-O1/O139 populations were not clonal and harbor a genomic backbone similar to that of toxigenic \textit{V. cholerae} O1 circulating in the western hemisphere (Fig. P1). Of interest, \textit{V. cholerae} O1 strains from Zimbabwe, Zambia, Mexico, Thailand, Bangladesh, and Haiti contained a distinct genotype (8.3.6.n,n) not observed previously. Environmental \textit{V. cholerae} non-O1/O139 isolates exhibited diverse VNTR genotypes with variability at all five loci. However, the clinical \textit{V. cholerae} non-O1/O139 isolates possessed an identical VNTR genotype, akin to, but distinct from, the clinical \textit{V. cholerae} O1 isolates. We observed a single amino acid residue deletion in the rstB gene of the Haitian \textit{V. cholerae} O1 strains. This deletion, a unique feature for a classical biotype of \textit{V. cholerae}, was not detected in any of the Altered El Tor, Hybrid, or El Tor variants of \textit{V. cholerae}, except \textit{V. cholerae} CP1048 (Bangladesh, 2010) and CP1032 (Mexico, 1991), proving that this deletion is not

\textcopyright 2012 National Academy of Sciences. All rights reserved.

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


Cite this Author Summary as: PNAS 10.1073/pnas.1207359109.
singly a characteristic of western hemisphere isolates. Moreover, PCR amplification and sequence analysis of \textit{rstB} gene identified two unique polymorphisms. Whole-genome mapping of the selected isolates also demonstrated polymorphisms within the Haitian outbreak strains and between Haitian and the other reference strains. Genome similarity clustering of both chromosomes reinforced the conclusion that the Haitian outbreak isolates did not differ substantially from isolates from Zimbabwe (CP1038) and Bangladesh (CP1048). The core genome phylogeny (Fig. P1) of the Haitian \textit{V. cholerae} isolates demonstrates that the majority of \textit{V. cholerae} non-O1/O139 strains cluster with toxigenic \textit{V. cholerae} O1 strains (e.g., Amazonia and TM11079-80) (2, 4) implicated in sporadic cases and epidemics originating in the western hemisphere (Brazil, 1991 and 1980), indicating potential indigenous origin of these strains. The Haitian \textit{V. cholerae} O1 strains, on the contrary, tightly clustered together, along with the clinical isolates from Zimbabwe (CP1038) and Bangladesh (CP1048 and CP1050). This observation agrees with a recent report of shared ancestry of Haiti, India, and Cameroon isolates. Phylogenetic analysis also reveals that within the short span of 3 wk, early in the cholera epidemic, substantial genomic diversity accumulated in the circulating population (Fig. P1).

The present 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 regions where cholera is endemic) reflecting quality coverage of global phylogenetic diversity as well as diversity within this heterogeneous clade and its phylogenetic near-neighbors. The need for such a qualified database is critically important if investigations of future \textit{V. cholerae} epidemics are to be effective and timely for public health applications. Moreover, it may also be necessary to generate the database during an epidemic, as the apposite comparator if SNP data are to be used to assign attribution in outbreak investigations.