ORIGINAL ARTICLE

The Origin of the Haitian Cholera Outbreak Strain

Chen-Shan Chin, Ph.D., Jon Sorenson, Ph.D., Jason B. Harris, M.D., William P. Robins, Ph.D., Richelle C. Charles, M.D., Roger R. Jean-Charles, M.D., James Bullard, Ph.D., Dale R. Webster, Ph.D., Andrew Kasarskis, Ph.D., Paul Peluso, Ph.D., Ellen E. Paxinos, Ph.D., Yoshiharu Yamaichi, Ph.D., Stephen B. Calderwood, M.D., John J. Mekalanos, Ph.D., Eric E. Schadt, Ph.D., and Matthew K. Waldor, M.D., Ph.D.

ABSTRACT

BACKGROUND

Although cholera has been present in Latin America since 1991, it had not been epidemic in Haiti for at least 100 years. Recently, however, there has been a severe outbreak of cholera in Haiti.

METHODS

We used third-generation single-molecule real-time DNA sequencing to determine the genome sequences of 2 clinical *Vibrio cholerae* isolates from the current outbreak in Haiti, 1 strain that caused cholera in Latin America in 1991, and 2 strains isolated in South Asia in 2002 and 2008. Using primary sequence data, we compared the genomes of these 5 strains and a set of previously obtained partial genomic sequences of 23 diverse strains of *V. cholerae* to assess the likely origin of the cholera outbreak in Haiti.

RESULTS

Both single-nucleotide variations and the presence and structure of hypervariable chromosomal elements indicate that there is a close relationship between the Haitian isolates and variant *V. cholerae* El Tor O1 strains isolated in Bangladesh in 2002 and 2008. In contrast, analysis of genomic variation of the Haitian isolates reveals a more distant relationship with circulating South American isolates.

CONCLUSIONS

The Haitian epidemic is probably the result of the introduction, through human activity, of a *V. cholerae* strain from a distant geographic source. (Funded by the National Institute of Allergy and Infectious Diseases and the Howard Hughes Medical Institute.)

From Pacific Biosciences, Menlo Park, CA (C.-S.C., J.S., J.B., D.R.W., A.K., P.P., E.E.P., E.E.S.); the Division of Infectious Diseases, Massachusetts General Hospital (J.B.H., R.C.C., S.B.C.), Channing Laboratory, Brigham and Women's Hospital (Y.Y., M.K.W.), the Departments of Pediatrics (J.B.H.), Medicine (R.C.C., Y.Y., S.B.C., M.K.W.), Microbiology (W.P.R., S.B.C., J.J.M., M.K.W.), and Molecular Genetics (W.P.R., S.B.C., J.J.M., M.K.W.), Harvard Medical School, and the Howard Hughes Medical Institute (M.K.W.) — all in Boston; and Fondation pour le Développement des Universités et de la Recherche en Haïti, Port-au-Prince, Haiti (R.R.J.-C.). Address reprint requests to Dr. Schadt at eschadt@pacificbiosciences.com.

Drs. Chin, Sorenson, Harris, and Robins contributed equally to this article.

This article (10.1056/NEJMoa1012928) was published on December 9, 2010, at NEJM.org.

N Engl J Med 2010.

Copyright © 2010 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

The outbreak of cholera that began in Haiti in late October 2010 illustrates the continued public health threat of this ancient scourge.[1] Cholera, an acutely dehydrating diarrheal disease that can rapidly kill its victims, is caused by *Vibrio cholerae*, a gram-negative bacterium.[2] This disease, which is usually transmitted through contaminated water, can and has spread in an explosive fashion. In the weeks since cases were first confirmed in the Artibonite province of Haiti on October 19, 2010, the disease has reached all 10 provinces in Haiti and has spread to the neighboring Dominican Republic on the island of Hispaniola. Of the more than 93,000 persons who have been sickened from the outbreak, more than 2100 have died, according to the Haitian Ministry of Public Health and Population (www.mspp.gouv.ht/site/index.php), and it is thought that the epidemic has not yet peaked.[3] Cholera epidemics had not been reported in Haiti for more than a century, and the origin of the Haitian *V. cholerae* outbreak has been the subject of some controversy.[4]

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 microbiologic tests. The ongoing seventh pandemic of cholera is caused by the *V. cholerae* El Tor biotype of serogroup O1 (El Tor O1),[5] which has replaced the previous "classical" biotype and has spread globally since its appearance in Indonesia in 1961. It reached the Americas in 1991, beginning in Peru and then spreading throughout much of South America and Central America, where it has since become endemic[6]; however, the strains of *V. cholerae* El Tor O1 that are now endemic in South America and Central America had not previously been reported to have caused cholera on Hispaniola. Analyses carried out by Haitian and U.S. laboratories have indicated that the current outbreak strain in Haiti is also *V. cholerae* El Tor O1 and thus is related to strains that are causing the ongoing seventh pandemic of cholera.

Both genetic and phenotypic diversity have arisen among circulating strains of *V. cholerae* El Tor O1, reflecting the acquisition, loss, or alteration of mobile genetic elements (for this and other key terms, see the Glossary), including CTX phage, which bears the genes encoding cholera toxin[7]; genomic islands[8]; and SXT-family integrative and conjugative elements, which often encode resistance to several antibiotics.[9] Single-nucleotide variations (SNVs) and insertions and deletions have also been detected in the core *V. cholerae* genome.[10,11] Such heterogeneity has been used to group strains and to model and understand their transmission around the globe[10,11] and is most comprehensively captured by sequencing genomic DNA. Second-generation DNA-sequencing technologies, although greatly productive, require a week or more to generate DNA sequence at high coverage and produce reads that are much shorter than those produced with first-generation sequencing technologies — making it difficult to characterize DNA variation in repeat regions.[12] Third-generation single-molecule real-time sequencing involves direct observation of the DNA polymerase while it synthesizes a strand of DNA; thus, it is much faster than previously developed methods and provides a comparatively long read length.[13,14] We therefore used a third-generation, single-molecule, real-time DNA sequencing method[13,14] to determine the genome sequences of two Haitian *V. cholerae* isolates and three additional *V. cholerae* clinical isolates from other regions of the world, allowing us to determine the probable origin of the cholera outbreak strain in Haiti.

METHODS

PATIENTS AND SAMPLES

Samples of spontaneously passed stool from two patients who had received a clinical diagnosis of cholera were cultured. Both patients received standard medical treatment for cholera, as appropriate to their clinical conditions. Bacterial isolates (H1 and H2) were shipped to the United States, with the use of an import license for this purpose (2010-10-108) that was provided through the Centers for Disease Control and Prevention (CDC). Isolates were identified as *V. cholerae* and were determined to be susceptible to tetracycline and erythromycin but resistant to trimethoprim–sulfamethoxazole and nalidixic acid. The use of bacterial isolates derived from discarded stool samples and that do not have individual patient identifiers is exempt from regulations regarding research on human subjects. Existing clinical isolates from the 1991 outbreak in Peru, strain C6706 (C6); the 2008 outbreak in Bangladesh, strain MDC126 (M4); and the 1971 outbreak in Bangla-

10.1056/NEJMoa1012928 NEJM.ORG

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

Glossary

Allele: A particular form of the DNA sequence of a gene.
Coverage: The average number of times a nucleotide in a sequenced genome is covered by reads generated by the sequencing instrument.
CTX phage: A filamentous bacteriophage that encodes cholera toxin, the principal virulence factor of *Vibrio cholerae*.
Deletion (mutation): A mutation involving the loss of genetic material that can be small, involving a single missing DNA base pair, or large, involving a piece of a chromosome.
DNA library: The collection of templates generated from a single DNA sample — in this case, from purified genomic DNA sheared to a target size of 2 kb. Each template is a double-stranded DNA template capped by hairpin loops at both ends.
DNA polymerase: An enzyme that catalyzes the polymerization of deoxyribonucleotides into a DNA strand, best known for its role in DNA replication, in which the polymerase "reads" an intact DNA strand as a template and uses it to synthesize the new strand.
First-generation sequencing technologies: A DNA sequencing approach developed by Frederick Sanger in 1975 in which different-sized fragments of DNA are generated, each starting from the same location and ending with a particular base, and labeled with an indicator corresponding to that base. All fragments are distributed in the order of their length by means of capillary electrophoresis. The DNA sequence is then revealed by the relative position, and the color, of each fragment.
Genomic island: A region of the chromosome that is thought to have been acquired by horizontal gene transfer but is no longer mobile.
Hyper-recombinant chromosomal elements: Regions in a chromosome in which recombination occurs significantly more frequently than the average rate of recombination over the entire genome.
Insertion (mutation): A type of mutation involving the addition of genetic material that can be small, involving a single extra DNA base pair, or large, involving a piece of a chromosome.
Integrative and conjugative element: A self-transmissible mobile genetic element that has plasmidlike and phagelike features; it is transferred through conjugation, and once it is transmitted, it integrates into the chromosome of the new host. Integrative and conjugative elements are increasingly recognized as contributing to lateral gene flow in prokaryotes.
Mobile elements: DNA elements, including plasmids, phages, and integrative and conjugative elements, that are able to move from cell to cell.
Nonsynonymous substitution: A substitution of one nucleotide base for another in an open reading frame, resulting in a modified amino acid sequence.
Open reading frame: A polynucleotide sequence that begins with an initiation (methionine ATG) codon and ends with a nonsense codon. All open reading frames have the potential to encode a protein or polypeptide, although many may not actually do so.
Read: A polynucleotide sequence, typically on the order of 30 to 3000 nucleotides in length, that is generated as output from the primary analysis of data from a sequencing run.
Read length: The total number of bases produced from a single molecule read.
Relative fitness: The average number of progeny from one strain that survive after one generation, as compared with the average number of progeny that survive from competing strains.
Repeated DNA sequences: Stretches of DNA that repeat themselves throughout a genome, either in tandem or interspersed along the genome; they can comprise up to 50% or more of the DNA of an organism. Repeated DNA regions can code for an end product, can have a structural function (such as telomeres), or can comprise sequences with no known function.
Second-generation DNA sequencing technologies: Currently available technologies for DNA sequencing that can simultaneously sequence multiple areas of the genome at massively high throughput and at low cost.
Single-nucleotide variation (SNV): A single-nucleotide variation in a genetic sequence; a common form of variation in the human genome.
SMRTbell construct: A basic DNA template construct in which hairpin loops are ligated to both ends of double-stranded DNA fragments of a particular size, creating a linear DNA structure that is topologically circular.
Structural variation: Operationally defined as genomic alterations that involve segments of DNA that are larger than 1 kb and can be microscopic or submicroscopic. Examples include copy-number variants, segmental duplication, low-copy repeats, inversion, translocation, and segmental uniparental disomy.
Superintegron: Integrons are gene-capture systems; all *V. cholerae* strains have very large integrons, referred to as superintegrons, on their second chromosome.
SXT: A 100-kb integrative and conjugative element, which was first isolated from a 1992 *Vibrio cholerae* O139 clinical isolate and which encodes resistance to multiple antibiotics.

desh, strain N16961 (N5) were cultured as described in the Supplementary Appendix, available with the full text of this article at NEJM.org.

DNA PREPARATION AND SEQUENCING

We isolated genomic DNA from each of the C6, N5, M4, H1, and H2 strains and sequenced it using previously described methods.[15] More specifically, we constructed DNA libraries comprising SMRTbell constructs, each of which was bound to a DNA polymerase and sequenced in a manner similar to that described previously,[16] using the PacBio *RS* sequencing system (Pacific Biosciences). For additional details regarding DNA

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

sequencing, resequencing analysis, and detection of DNA variations, see the Supplementary Appendix. Methods for the reconstruction of phylogenetic trees and the characterization of VSP-2 (a genomic island), SXT, and the superintegron are provided in the Supplementary Appendix.

RESULTS

STRUCTURES OF FIVE *V. CHOLERAE* GENOMES

The H1 and H2 isolates were sequenced in less than 24 hours, with enough DNA sequencing reads generated in this time to cover the genomes 60 and 32 times, respectively. C6, M4, and N5 were similarly rapidly sequenced at coverages of 28, 37, and 36, respectively. (Table 1 in the Supplementary Appendix). We used previously obtained genome sequences of N16961,[17] CIRS101,[11] and MJ-1236[11] as reference genomes to facilitate genomewide characterization of the five sequenced isolates. When we mapped raw sequencing reads to the canonical N16961 reference, we identified copy-number variation — typically in hyper-recombinant genomic regions — affecting ribosomal RNAs, the *V. cholerae* superintegron, the SXT-integrative and conjugative element, and the seventh-pandemic genomic islands (VSP-1 and VSP-2).[18] The five isolates showed a high degree of similarity, as well as notable structural variation (Fig. 1). The structures of the H1 and H2 genomes were identical (Fig. 1). The sequence from sample N5 matched the canonical reference strain from which it was purportedly cultured.

SNVS AND THE RELATEDNESS OF THE *V. CHOLERAE* STRAINS

A comparison of the SNVs of each strain also indicated that H1 and H2 were essentially identical and were more similar to the M4 strain from Asia than to the C6 strain from Peru or the canonical N16961 reference (Table 2 in the Supplementary Appendix). Although we used data from 20-times coverage to determine the SNVs present in each genome (GenBank accession number, SRP004712) for the comparative analyses, the key SNVs highlighted in Figure 1 were apparent after achieving 12-times coverage of the genomes of these isolates; we obtained 12-times coverage of the five genomes within 3 hours of sequencing.

In our initial assessment of the relatedness of the five sequenced isolates, we analyzed a set of 1588 conserved orthologous genes (encompassing approximately 1.8 Mb of DNA) that were previously reported to resolve the relatedness of different *V. cholerae* strains[10] of diverse origin. We aligned the consensus sequences of those 1.8 Mb from C6, N5, M4, H1, and H2 with those of 23 previously sequenced *V. cholerae* strains[10] and constructed a phylogenetic tree that unequivocally places H1 and H2 in the seventh-pandemic group. Although the Haitian strains are similar to isolates from Latin America (C6 from the 1991 outbreak in Peru) and Africa (B33 from the 2004 outbreak in Mozambique), they are most closely related to recent South Asian isolates (M4 from the 2008 outbreak in Bangladesh and CIRS101 from the 2002 outbreak in Bangladesh) (Fig. 2A). H1 and H2 are only distantly related to the U.S. Gulf Coast isolates, such as strain 2740-80; the latter does not even cluster with seventh-pandemic strains (Fig. 2A).

Thirty SNVs have previously been shown to differentiate six groups within the seventh-pandemic strains.[19] We compared the alleles of these SNVs from each of the five isolates with those from 78 cholera strains from the seventh pandemic and 3 cholera strains isolated before the seventh pandemic[19] and constructed a phylogenetic tree (Fig. 2B). Six groups from the seventh pandemic are readily identified in this tree, with H1 and H2 falling into group V, which also includes variant strains from Bangladesh (CIRS101

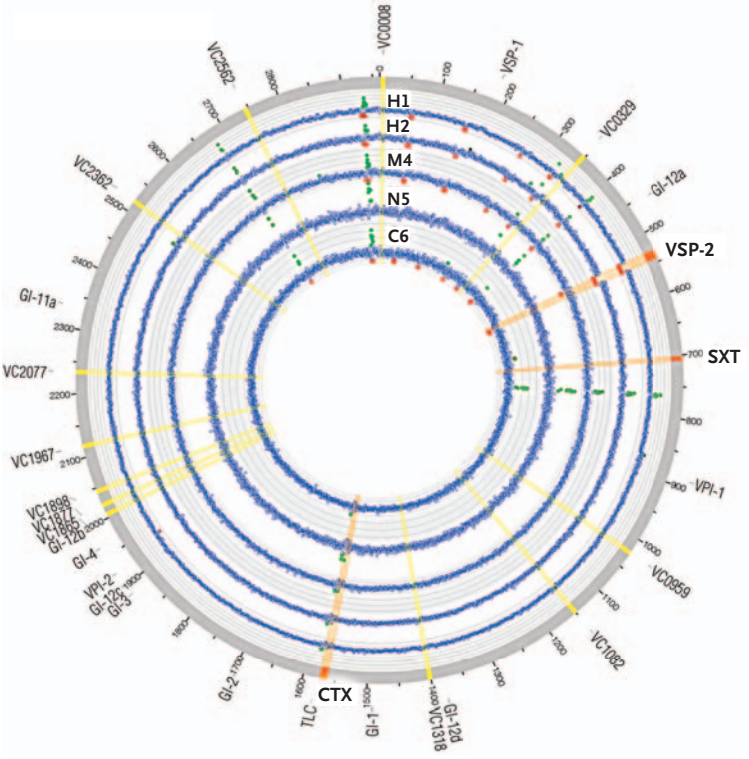
Figure 1 (facing page). Sequence Depth of Coverage for the Five *Vibrio cholerae* Isolates.

The figure shows the observed sequencing depth of coverage for the five isolates we sequenced (H1, H2, M4, N5, and C6), relative to the published sequence from the two chromosomes of *V. cholerae* N16961. Areas in the genome in which the read coverage was more than 4 SD higher than the background coverage are plotted with green points (repetitive regions). Areas in the genome in which the read coverage was more than 4 SD lower than the background coverage are plotted with red points (missing segments). Regions in the outer ring show known strain markers that allow the typing of these five isolates with respect to each other and to published strains. The locations of discriminating markers of single-nucleotide variations (SNVs) are shown as yellow bands, and the locations of discriminating mobile elements are shown as orange bands. The identifiers outside the outermost circle correspond to the positions of known mobile elements[10] and strain-specific SNV markers.[19]

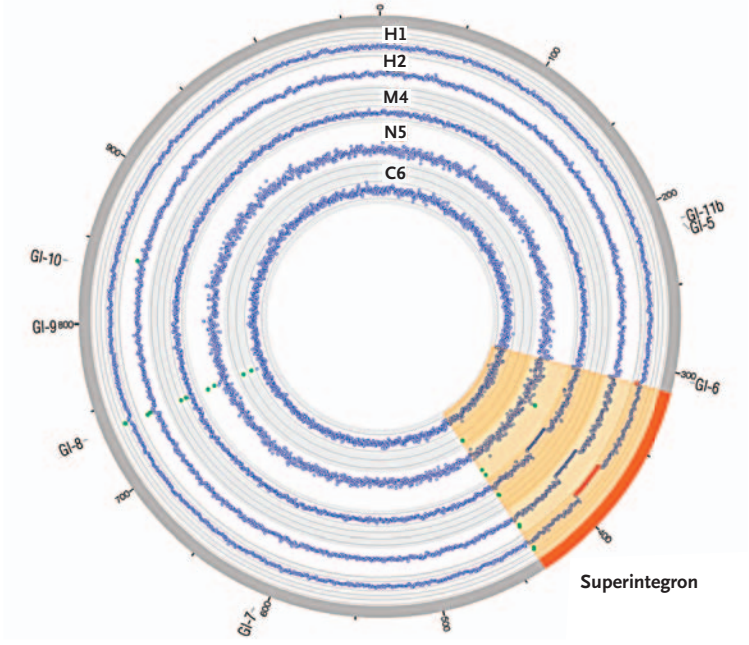
10.1056/NEJMoa1012928 NEJM.ORG

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

A Large Chromosome



B Small Chromosome



The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

and M4). The phylogeny highlights the distance between strains of group V and those of group II, the latter of which consists mainly of Latin American strains (including C6, the strain we sequenced) and African strains isolated between 1970 and 1998. It supports the conclusion that Haitian *V. cholerae* is more closely related to contemporary South Asian strains of *V. cholerae* than to Latin American strains. The placement of C6 in group II is consistent with a previously proposed hypothesis that Latin American strains of *V. cholerae* may have been introduced from Africa.[19]

Figure 2 (facing page). Reconstructing Phylogenetic Relationships among *V. cholerae* Strains.

Panel A shows the phylogenetic relationships among pandemic *V. cholerae* strains on the basis of single-nucleotide variations identified among all strains for which a set of 1588 orthologous genes has been completely sequenced.[10] The magnified inset represents strains in the seventh pandemic, including H1, H2, M4, C6, and N5. Panel B shows the phylogenetic relationships among a broad set of seventh-pandemic *V. cholerae* strains.[19] The phylogenetic tree is rooted with three pre–seventh-pandemic strains.

STRUCTURAL VARIATIONS IN THE SUPERINTEGRON, VSP-2, AND SXT

Analyses of insertions and deletions in hyper-recombinant chromosomal elements, which are often mobile elements, can be used to complement the analysis of SNV markers in the establishment of the lineage of a given strain.[10] We therefore assessed the sequences of 20 previously described hyper-recombinant chromosomal elements[10] in the genomes of C6, N5, M4, H1, and H2 (see Fig. 1 for the locations of these elements). The long read lengths that we obtained (the average read length of unfiltered H1 and H2 sequences was 954 bp, with 5% of the reads exceeding 2800 bases) are ideal for identifying structural variation, especially in the context of repeated DNA sequences. Of the 20 regions we examined, most were structurally conserved in the five strains we sequenced — consistent with the coverage results in Figure 1. However, we did observe structural variation in 3 of the 20 regions: superintegron, VSP-2, and SXT.

A map of the superintegron region from strains C6, N5, M4, and H1 is shown in Figure 3A. The superintegrons of C6 and N5 are structurally identical to that of the canonical reference strain N16961 (Table 1). In contrast, the superintegron structures of M4 and H1 are distinct from those of C6 and N5 (i.e., N16961); both M4 and H1 lack a segment that contains 41 open reading frames (Table 3 in the Supplementary Appendix). M4 is also missing a single open reading frame that is present in the H1 superintegron; otherwise, their genomic structures in this region are identical. Because the SNV data suggested that H1 (and H2) are more closely related to CIRS101 than to M4 (Fig. 2A), we also compared superintegron regions of the H1 and CIRS101 strains and found them to be structurally identical.

H1, M4, and C6 lack different overlapping segments of the VSP-2 region relative to N16961 (Table 3 and Fig. 1 in the Supplementary Appendix). The pattern of deletion in the VSP-2 sequence of CIRS101 is identical to that of H1, but not to that of M4, providing additional evidence that H1 is more closely related to CIRS101 than to M4 (Table 1).

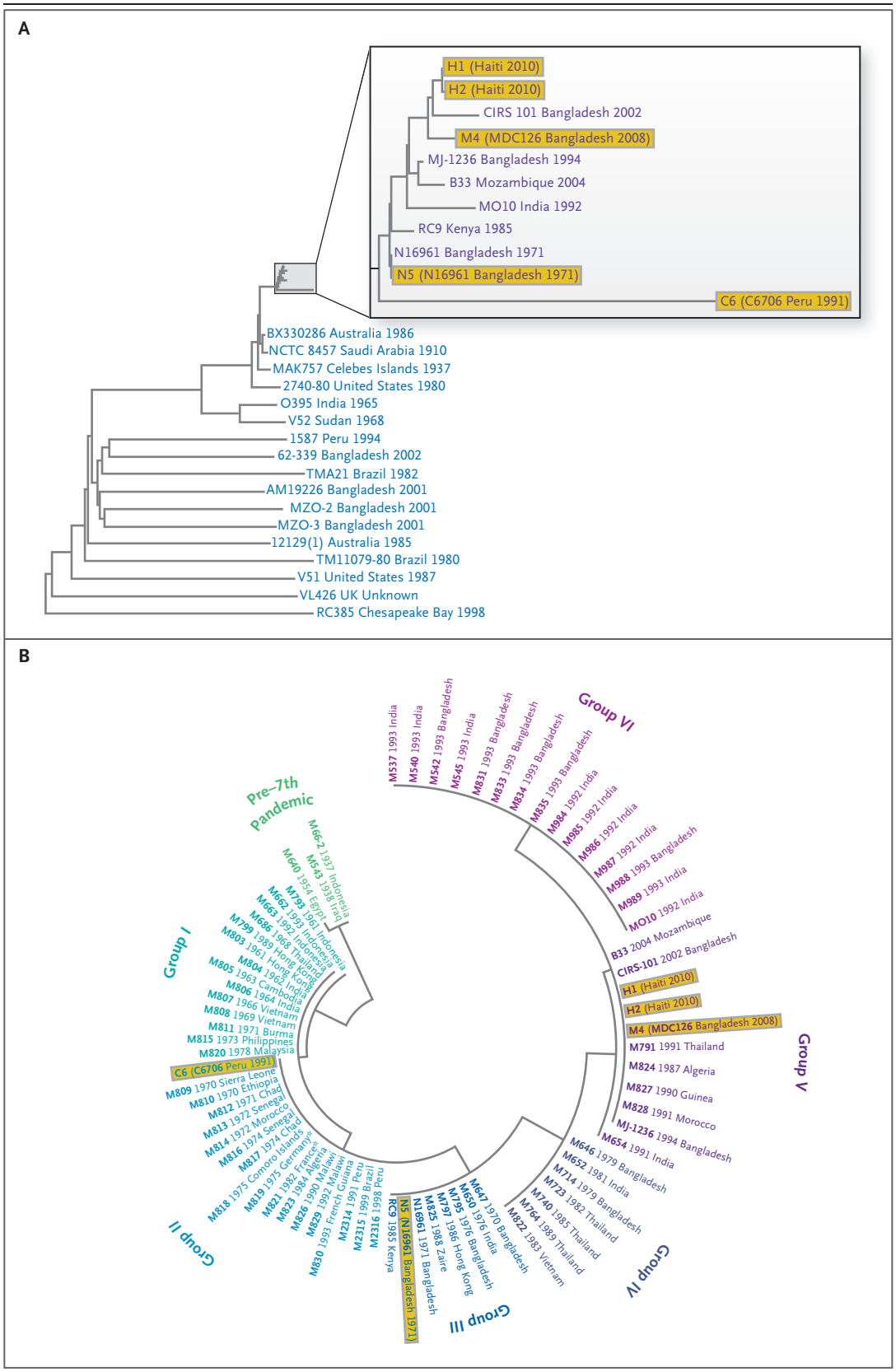
SXT is a clinically important integrative and conjugative element that accounts for the dissemination of genes conferring resistance to several antibiotics in contemporary *V. cholerae* isolates.[20] N16961 and Latin American epidemic strains (including C6706) are known to lack SXT and remain susceptible to antibiotics; not surprisingly, no reads from N5 or C6 mapped to a reference SXT sequence derived from the MJ-1236 strain (Table 1). However, structural analyses revealed that M4 and H1 contained very similar SXT elements and that both lack a closely related subset of the SXT genes that is present in MJ-1236 (Fig. 3B, and Table 3 in the Supplementary Appendix).

SNVs with biologic and epidemiologic significance have accumulated in the CTX prophage region. The gene encoding cholera toxin B subunit (*ctxB*) in isolate H1 (and H2) carries three nonsynonymous substitutions relative to N16961 (Fig. 3C). Two of these changes are characteristic of *ctxB* in classical strains of the sixth pandemic, and they have been detected in recent El Tor O1 strains (including CIRS101) from South Asia.[11] M4, H1, and H2 carry these two *ctxB* mutations. The third mutant allele, predicting the substitution of histidine with asparagine at position 20, (last line, Fig. 3C) has previously been observed only in El Tor variant strains from South Asia[21] and in very recent isolates from West Africa.[22]

RELATEDNESS OF H1 AND H2 TO OTHER ISOLATES OF HAITIAN *V. CHOLERAE*

We compared the sequences of H1 and H2 to the unassembled genome sequence data of three in-

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.



The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

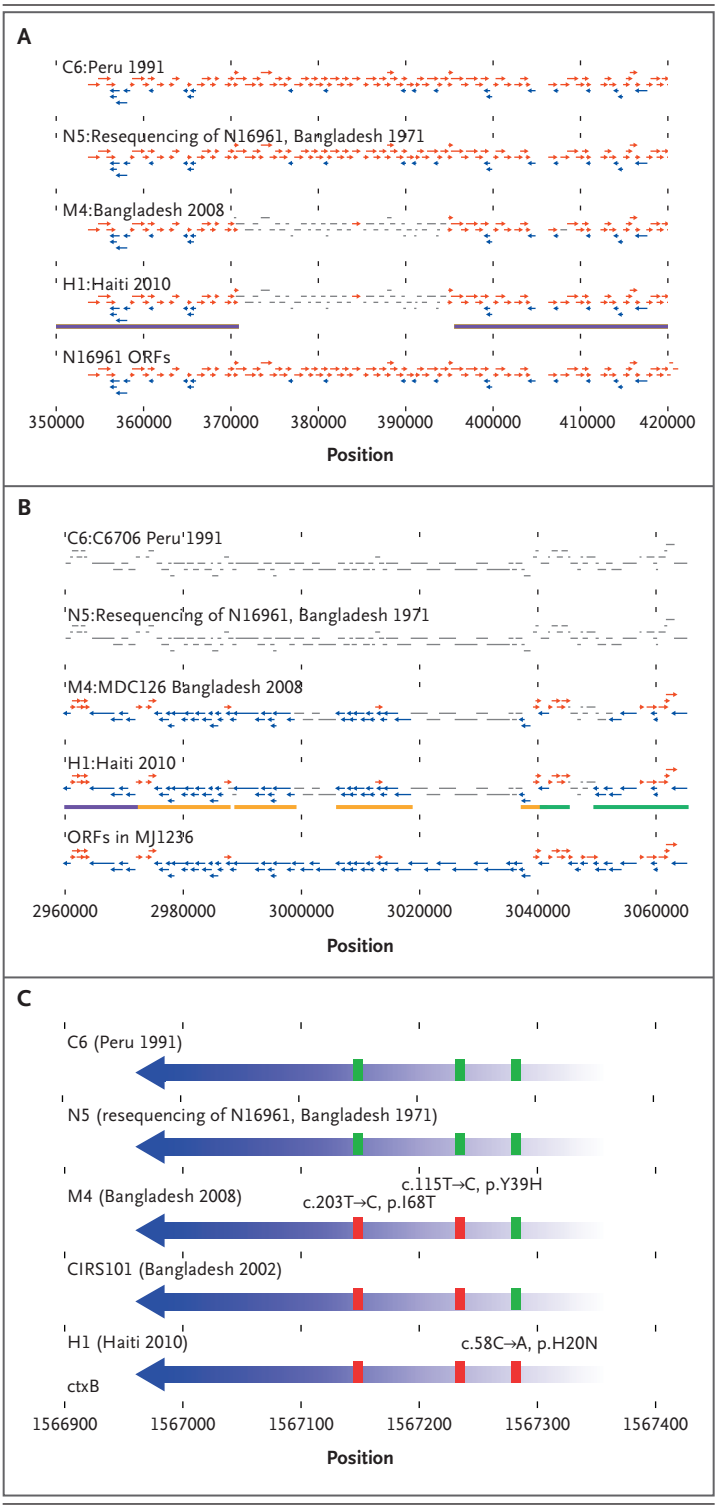


Figure 3. Gene Maps of the Superintegron, SXT, and *ctxB* Regions in the C6, N5, M4, and H1 *V. cholerae* Strains.

Representations of the coverage with respect to open reading frames (ORFs) in each of these regions are shown. Open reading frames that are shown in red are mapped to the positive strand, and those shown in blue are mapped to the negative strand; gray indicates that the open reading frame was supported by no reads or by a very low number of reads for the indicated sample. Coverages for H1 and H2 were identical over all regions; therefore, only H1 is shown in the figure. Colored bars below the H1 plots indicate the CIRS101 genomic region to which the reference sequence for the indicated element maps. Gaps indicate segments in the reference that are missing from CIRS101. For the SI region (Panel A), no coverage gaps were observed between the N16961 and the C6 and N5 strains, whereas a single coverage gap was observed in M4 and H1 from base positions 370682 to 395783 covering 41 open reading frames. For SXT (Panel B), none of the reads from N5 or C6 mapped to the MJ-1236 SXT reference sequence; in the case of H1, four coverage gaps were observed, covering 27 open reading frames. M4 had three gaps overlapping with three of the H1 gaps, but the M4 gaps covered 4 open reading frames in addition to 25 of the 27 open reading frames covered by the H1 gaps. Also shown (Panel C) is the location of all variant calls found in the cholera enterotoxin subunit B (VC_1456) open reading frame. The boxes indicate three sites in which the alleles represented in H1, CIRS101, and M4 (red) differ from the alleles in the N16961 and C6 sequences (green) and lead to nonsynonymous changes. In Panel C, c denotes the nucleotide position, and p the amino acid position.

dependently isolated Haitian strains that have been deposited by the CDC into the GenBank database (accession numbers, AELH00000000.1, AELI00000000.1, and AELJ00000000.1). H1, H2, and the three isolates obtained by the CDC are virtually identical in all the regions previously shown to harbor structural variation.[10] The three coding mutations found in the ctxB gene of H1 and H2 are also present in each of the three CDC strains.

DISCUSSION

The *V. cholerae* strain responsible for the expanding cholera epidemic in Haiti is nearly identical to so-called variant seventh-pandemic El Tor O1 strains that are predominant in South Asia, including Bangladesh.[23,24] The shared ancestry of the Haitian epidemic strain and recent South Asian strains of *V. cholerae* is distinct from that of circulating Latin American and East African strains of *V. cholerae*. Patterns of DNA from Haitian strains and *V. cholerae* strains in a large collection held by the CDC, as determined by means of pulsed-field gel electrophoresis, also suggested that the Haitian strains of *V. cholerae* are most similar to re-

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

Table 1. Gene Content of Hypervariable Elements in H1, H2, M4, and C6.*

Isolate	Has CTX	Missing ORFs in VSP-II region	Missing ORFs in Superintegron	Missing ORFs in SXT_MJ-1236	Missing ORFs in SXT_CIRS101
Sample					
H1 and H2	Yes	Yes†	Yes	Yes†	No
M4	Yes	Yes†	Yes	Yes†	Yes
N5	Yes	No	No	No hit‡	No hit‡
C6	Yes	Yes†	No	No hit‡	No hit‡
Reference genomes					
CIRS101	Yes	Yes	Yes	ND	No
N16961	Yes	No	No	No SXT	No SXT
B33	Yes	No	Yes	No	No
MJ-1236	Yes	No	Yes	No§	No§

* ND denotes not determined, and ORF open reading frame.
† The missing regions are different in each of these.
‡ No hit indicates that no SXT ORF was observed.
§ Reference sequence comprises subsequences derived from MJ-1236.

cent South Asian *V. cholerae* strains.[3] Our comparative analysis of the H1 and H2 strains and three CDC isolates indicate that the Haitian cholera epidemic is clonal. Collectively, our data strongly suggest that the Haitian epidemic began with introduction of a *V. cholerae* strain into Haiti by human activity from a distant geographic source.

Our data distinguish the Haitian strains from those circulating in Latin America and the U.S. Gulf Coast and thus do not support the hypothesis that the Haitian strain arose from the local aquatic environment.[25,26] It is therefore unlikely that climatic events led to the Haitian epidemic, as has been suggested in the case of other cholera epidemics.[27,28] Understanding exactly how this South Asian variant strain of *V. cholerae* was introduced to Haiti will require further epidemiologic investigation.

The Haitian outbreak strains can be distinguished from earlier seventh-pandemic strains by several genetic polymorphisms, including those in *ctxB*. Alterations in the *ctxB* sequence in the context of other structural variations (e.g., within SXT and VSP-2) are hallmarks of the variant strains that have emerged in South Asia. Because these variant strains replaced previously dominant strains of the seventh pandemic in South Asia, it has been hypothesized that their unique genetic composition increases their relative fitness, perhaps as a consequence of increased pathogenicity.[21,23] Specifically, by causing more severe dehydrating disease, variant strains increase their own dissemination through the increased production of infectious stools by their human hosts.[24]

Our findings have policy implications for public health officials who are considering the deployment of vaccines or other measures for controlling cholera.[29,30] The apparent introduction of cholera into Haiti through human activity emphasizes the concept that predicting outbreaks of infectious diseases requires a global rather than a local assessment of risk factors.

The accidental introduction of South Asian variant *V. cholerae* El Tor into Haiti may have consequences beyond Haiti. The apparently higher relative fitness[23,24] and increased antibiotic resistance of the South Asian strains and the ability of those strains to cause severe cholera[23] suggest that the South Asian variant *V. cholerae* El Tor that is now in Haiti could displace the resident El Tor O1 seventh-pandemic strains in Latin America. It is likely that the Caribbean ecosystem may now be host to a set of genes, including classical biotype-like cholera toxin genes and the STX integrative and conjugative element, that were previously absent from this region. Clearly, the provision of adequate sanitation and clean water is essential for preventing the further spread of the Haitian cholera epidemic.[3] Vaccination would also help to prevent the spread of disease, although cholera vaccines are in short supply. Our findings suggest that public health measures to counter the spread of cholera[30-32] in Hispaniola

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.

could minimize the dissemination of the new South Asian strain and the virulence genes that it carries beyond the shores of this Caribbean island.

Work at Harvard Medical School and Brigham and Women's Hospital was supported by grants from the National Institute of Allergy and Infectious Diseases (AI-018045, to Dr. Mekalanos; and AI R37-042347) and by a grant from the Howard Hughes Medical Institute to Dr. Waldor. Work at Massachusetts General Hospital was supported by a grant from the National Institute of Allergy and Infectious Diseases (AI058935, to Dr. Calderwood).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the organizations and persons who continue to provide outstanding patient care in this outbreak and the Massachusetts General Hospital Office of Disaster Response/Center for Global Health and Project Hope, which allowed some of our team to assist as volunteers in this cholera outbreak; Steve Turner (Pacific Biosciences) for discussion and advice and Kristin Robertshaw for assistance in rendering Figure 2; Ali Bashir, Simon Chang, Janice Cheng, Pei-Lin Hsiung, Armuta Joshi, Dimitris Iliopolous, Aaron Klammer, Deborah Kwo, Brianna LaMay, Steven Lin, Aseneth Lopez, Khai Luong, John Major, Patrick Marks, Phillip McClurg, Emilia Mollova, Huy Nguyen, Andy Pham, Ruben Pingue, Homero Rey, Robert Sebra, Marie Valdovino, Susana Wang, and Jackie Yen at Pacific Biosciences for their assistance in rapidly preparing and sequencing the cholera samples and aiding in the analyses; and Brigid Davis, Wen Zheng, Dan Portnoy, and Steve Lory for discussion of our results and the manuscript.

REFERENCES

1. Cholera vaccines: WHO position paper. Wkly Epidemiol Rec 2010;85:117-28.
2. Sack DA, Sack RB, Nair GB, Siddique AK. Cholera. Lancet 2004;363:223-33.
3. Centers for Disease Control and Prevention (CDC), Update: cholera outbreak — Haiti, 2010. MMWR Morb Mortal Wkly Rep 2010;59:1473-9.
4. Enserink M. Infectious diseases: Haiti's outbreak is latest in cholera's new global assault. Science 2010;330:738-9.
5. Faruque SM, Albert MJ, Mekalanos JJ. Epidemiology, genetics, and ecology of toxigenic Vibrio cholerae. Microbiol Mol Biol Rev 1998;62:1301-14.
6. Wachsmuth IK, Evins GM, Fields PI, et al. The molecular epidemiology of cholera in Latin America. J Infect Dis 1993; 167:621-6.
7. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 1996;272:1910-4.
8. Taviani E, Grim CJ, Choi J, et al. Discovery of novel Vibrio cholerae VSP-II genomic islands using comparative genomic analysis. FEMS Microbiol Lett 2010; 308:130-7.
9. Wozniak RA, Fouts DE, Spagnoletti M, et al. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet 2009;5(12): e1000786.
10. Chun J, Grim CJ, Hasan NA, et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. Proc Natl Acad Sci U S A 2009;106:15442-7.
11. Grim CJ, Hasan NA, Taviani E, et al. Genome sequence of hybrid Vibrio cholerae O1 MJ-1236, B-33, and CIRS101 and comparative genomics with V. cholerae. J Bacteriol 2010;192:3524-33.
12. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet 2010;19:R227-R240.
13. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009;323:133-8.
14. Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP. Computational solutions to large-scale data management and analysis. Nat Rev Genet 2010;11:647-57.
15. Travers KJ, Chin CS, Rank DR, Eid JS, Turner SW. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 2010;38(15):e159.
16. Korlach J, Bjornson KP, Chaudhuri BP, et al. Real-time DNA sequencing from single polymerase molecules. Methods Enzymol 2010;472:431-55.
17. Heidelberg JF, Eisen JA, Nelson WC, et al. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature 2000;406:477-83.
18. Nusrin S, Gil AI, Bhuiyan NA, 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;58:342-54.
19. 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;16:1130-2.
20. Burrus V, Marrero J, Waldor MK. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. Plasmid 2006;55:173-83.
21. Goel AK, Jain M, Kumar P, Bhadauria S, Kmboj DV, Singh L. A new variant of Vibrio cholerae O1 El Tor causing cholera in India. J Infect 2008;57:280-1.
22. Quilici ML, Massenet D, Gake B, Bwalki B, Olson DM. Vibrio cholerae O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. Emerg Infect Dis 2010;16:1804-5.
23. 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;40:3296-9.
24. Siddique AK, Nair GB, Alam M, et al. El Tor cholera with severe disease: a new threat to Asia and beyond. Epidemiol Infect 2010;138:347-52.
25. Cholera in Haiti — the climate connection. Circle of Blue Water News. November 11, 2010. (http://www.circleofblue.org/waternews/2010/world/hold-cholera-in-haiti-the-climate-connection/.)
26. Reed C. Cholera in Haiti likely to infect 400,000. Discovery News. November 24, 2010. (http://news.discovery.com/earth/cholera-in-haiti-likely-to-infect-400000.html.)
27. Constantin de Magny G, Colwell RR. Cholera and climate: a demonstrated relationship. Trans Am Clin Climatol Assoc 2009;120:119-28.
28. Constantin de Magny G, Murtugudde R, Sapiano MR, et al. Environmental signatures associated with cholera epidemics. Proc Natl Acad Sci U S A 2008;105: 17676-81.
29. Glass RI, Claeson M, Blake PA, Waldman RJ, Pierce NF. Cholera in Africa: lessons on transmission and control for Latin America. Lancet 1991;338:791-5.
30. Enserink M. Public health: no vaccines in the time of cholera. Science 2010; 329:1462-3.
31. Ryan ET, Calderwood SB, Qadri F. Live attenuated oral cholera vaccines. Expert Rev Vaccines 2006;5:483-94.
32. Sur D, Nair GB, Lopez AL, Clemens JD, Katoch VM, Ganguly NK. Oral cholera vaccines — a call for action. Indian J Med Res 2010;131:1-3.

Copyright © 2011 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from www.nejm.org on December 11, 2010. For personal use only. No other uses without permission.
From the NEJM Archive Copyright © 2010 Massachusetts Medical Society.