

Genomic history of the seventh pandemic of cholera in Africa

François-Xavier Weill,^{1,2*} Daryl Domman,² Elisabeth Njamkepo,¹ Cheryl Tarr,³ Jean Rauzier,¹ Nizar Fawal,¹ Karen H. Keddy,^{4,5} Henrik Salje,^{6,7} Sandra Moore,⁸ Asish K. Mukhopadhyay,⁹ Raymond Bercion,^{10,11} Francisco J. Luquero,¹² Antoinette Ngandjio,¹³ Mireille Dosso,¹⁴ Elena Monakhova,¹⁵ Benoit Garin,^{11†} Christiane Bouchier,¹⁶ Carlo Pazzani,¹⁷ Ankur Mutreja,^{18,19} Roland Grunow,²⁰ Fati Sidikou,²¹ Laurence Bonte,^{22‡} Sébastien Breurec,^{10†} Maria Damian,²³ Berthe-Marie Njanpop-Lafourcade,²⁴ Guillaume Sapriel,^{25,26} Anne-Laure Page,¹² Monzer Hamze,²⁷ Myriam Henkens,^{28‡} Goutam Chowdhury,⁹ Martin Mengel,²⁴ Jean-Louis Koeck,^{29§} Jean-Michel Fournier,³⁰ Gordon Dougan,^{2,18} Patrick A. D. Grimont,³¹ Julian Parkhill,² Kathryn E. Holt,³² Renaud Piarroux,⁸ Thandavarayan Ramamurthy,¹⁹ Marie-Laure Quilici,^{1,30||} Nicholas R. Thomson^{2,33||}

The seventh cholera pandemic has heavily affected Africa, although the origin and continental spread of the disease remain undefined. We used genomic data from 1070 *Vibrio cholerae* O1 isolates, across 45 African countries and over a 49-year period, to show that past epidemics were attributable to a single expanded lineage. This lineage was introduced at least 11 times since 1970, into two main regions, West Africa and East/Southern Africa, causing epidemics that lasted up to 28 years. The last five introductions into Africa, all from Asia, involved multidrug-resistant sublineages that replaced antibiotic-susceptible sublineages after 2000. This phylogenetic framework describes the periodicity of lineage introduction and the stable routes of cholera spread, which should inform the rational design of control measures for cholera in Africa.

Cholera is an acute intestinal infection caused by the bacterium *Vibrio cholerae*, which produces cholera toxin (CTX), responsible for inducing a rapid and massive loss of body fluids. The seventh cholera pandemic (7P) began in 1961 in Indonesia, before spreading globally, in particular to South Asia (1963), Africa (1970), Latin America (1991), and the Caribbean (i.e., Haiti) (2010) (1). In 2008 to 2012, half a century after the onset of this pandemic, the global burden of cholera was estimated at 1.4 to 4.0 million cases of cholera annually, with 21,000 to 143,000 deaths (2). The agent responsible for the 7P belongs to the O1 serogroup (or more rarely, the O139 variant) and the El Tor biotype. Globally, 7P *V. cholerae* El Tor (7PET) isolates are genetically homogeneous and linked to a single source, the Bay of Bengal in South Asia (3). Phylogenetic analyses have identified at least three independent, but temporally overlapping, waves of global transmission during the 7P (3, 4).

Africa is the continent most affected by the current pandemic (5); however, little is known about the propagation routes of cholera in this region. Basic epidemiological data from individual countries have provided some insight into cholera dynamics (6, 7). Conventional typing methods such as serotyping (Ogawa and Inaba) theoretically provide additional data to discriminate lineages across outbreaks; however, they rapidly proved inadequate or even confusing, and the multiple molecular typing methods used in the pregenomic era (8–13) had low resolution or were generally unsuitable for longer-term phylogenetic inference. Furthermore, even in the genomic era, the use of representative African isolates has been very limited. Previous large-scale genomic studies on the 7P (3, 4) analyzed only 28 African isolates and provided a preliminary view of overall disease transmission on the continent. Other genomic studies dealing with epidemics in Africa have been published, but these studies focused on the disease exclusively at national or regional levels (14–17).

We reconstructed the spatiotemporal spread of cholera across the continent during the 7P by analyzing the genomes of 1070 *V. cholerae* El Tor isolates, including 714 new isolates sequenced in this study (table S1 and fig. S1). The newly sequenced African isolates ($n = 569$) were selected to represent the widest possible temporal and geographic distribution of cases reported to the World Health Organization (WHO) (figs. S2 and S3 and supplementary text note 1). In total, we analyzed the genomes of 651 7PET isolates collected in 45 of 54 African countries between 1966 and 2014 (table S2). Our sampling of cholera isolates in Africa over this time frame is representative of ~1.8 million (46.8%) of the ~3.8 million cases reported from Africa to the WHO between 1970 and 2014 (table S3 and figs. S2 and S3). We note that only a small number of isolates from East and Central Africa from the 1970s and 1980s were available for this study. Consequently, the duration and geographic spread of the imported sublineages (particularly for T3 and T4; see below) may have been underestimated, and some imported sublineages with a limited spatiotemporal spread may have remained undetected. Likewise, low-level and sporadic cases are also likely underrepresented. It is also noteworthy that this study, focusing on 7PET isolates, does not preclude the existence of non-7PET lineage isolates causing sporadic disease in Africa, as has been seen across Latin America [see companion paper by Domman et al. (18)] and in Mozambique (17).

We used two approaches to obtain a robust phylogenetic framework for the inference of propagation routes (19). Maximum likelihood phylogenetic analysis was performed on the 1070 genomes (Fig. 1A and fig. S4), with 9300 single-nucleotide variants (SNVs), evenly distributed over the genome (table S4). We detected a strong temporal signal

in the molecular data (fig. S5), allowing us to use a Bayesian phylogenetic approach to provide divergence times for a spatially and temporally representative subset of 228 isolates (Fig. 1B and fig. S6). These robust phylogenies enabled us to infer 11 introductions of 7PET into Africa. Each introduction resulted in mid- to long-term spread of cholera (introduction events T1, T3 to T12), and one export from Africa to Peru responsible for the Latin American epidemic in the 1990s [T2 (18)] (Fig. 2, table S5, and supplementary text note 2). All 651 African genomes were assigned to serogroup O1 on the basis of their *rfb* region (fig. S7).

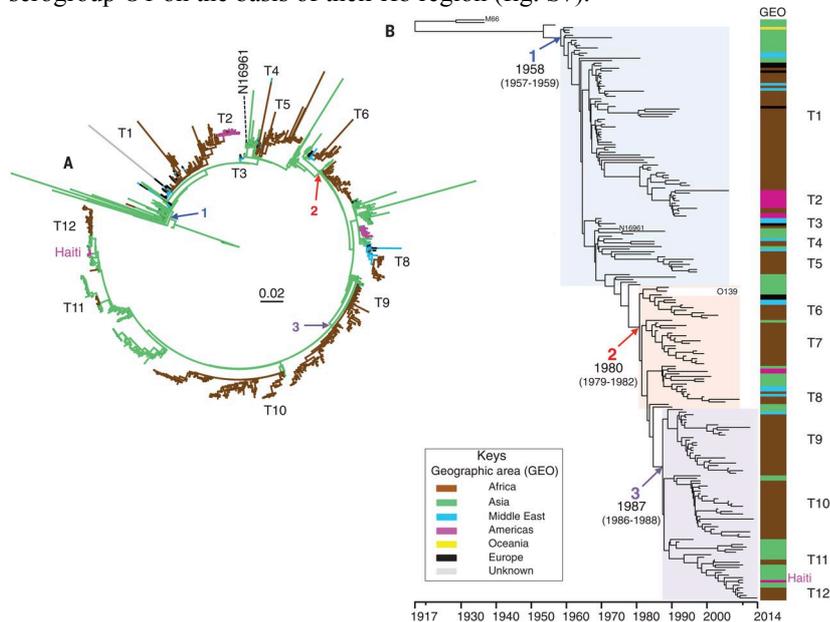


FIG. 1 Phylogeny of seventh pandemic *V. cholerae* El Tor isolates. (A) Maximum likelihood phylogeny of the 1070 genomes studied, including M66 as an outgroup. Branches are color-coded according to their geographic location, inferred by stochastic mapping of the geographic origin of each isolate onto the tree. The N16961 reference genome is indicated by a black dashed line. (B) Maximum clade credibility tree produced with BEAST for a subset of 228 representative isolates. The 12 introduction events involving Africa are indicated by the letter T. The three previously described waves (3) are indicated by colored arrows. The clades containing O139 isolates and that containing isolates from the 2010 outbreak in Haiti are shown.

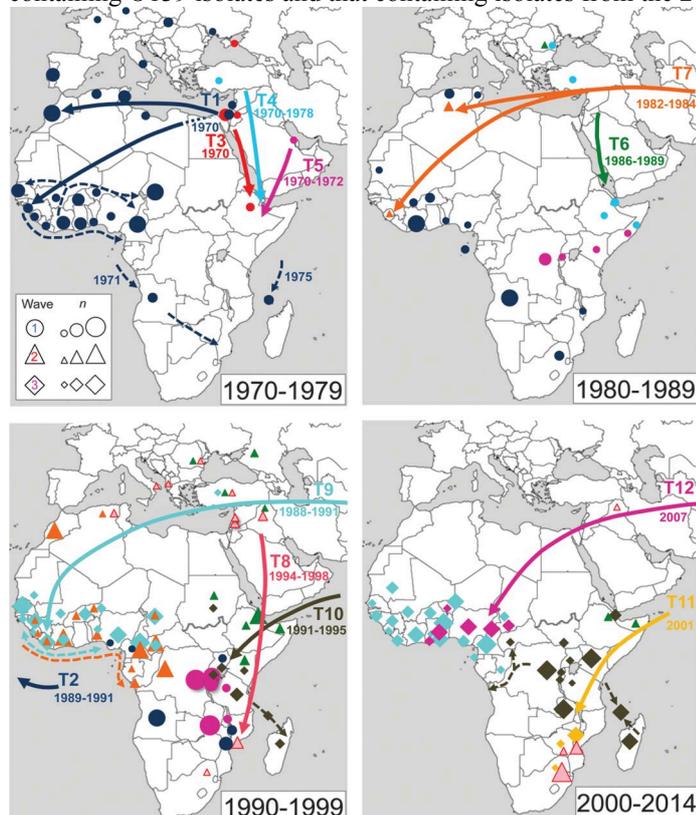


FIG. 2 Inferred propagation routes of seventh pandemic *V. cholerae* O1 El Tor populations to, from, and within Africa.

The 12 introduction events involving Africa are indicated by the letter T. The date ranges shown for introductions are the median values for the most recent common ancestor (MRCA) in years (taken from BEAST) with the first number indicating the median MRCA of the African isolates and their closest relative from the source location, and the second number indicating the median MRCA of the African isolates. Introductions and inferred secondary transmission chains are indicated by thick and dashed arrows, respectively. Secondary transmission chains for West Africa in 1970 to 1971 are based on published records. The geographic presence of the various lineages is indicated by a circle, triangle, and diamond for waves 1, 2, and 3, respectively, and colored according to the inferred transmission events. The size of the shapes is proportional to the number of genomes analyzed (see fig. S3).

We mapped the phylogenetic data onto historical records of African cholera outbreaks. The genomic analysis identified the two “El Tor strains” known to have invaded Africa in 1970: one in West Africa (serotype Ogawa) and the other in East Africa (serotype Inaba).

An “Ogawa strain” was first reported between July and September 1970, in cholera cases in the Black Sea region (Odessa and Kerch) of the former USSR, the Middle East (Lebanon, Israel, Turkey, Jordan), North Africa (Libya and Tunisia), and West Africa (Guinea, Sierra Leone, Liberia, Ghana) (1). This “Ogawa strain” corresponds to our isolates from the T1 introduction. These initial African T1 isolates were not directly related to Iraqi isolates from 1966 (the last isolates obtained during the western extension of the 7P before a lull that ended in 1970) but rather related to three Chinese isolates (1973–1981) and one Indian isolate (1980). This suggests that the resurgence of cholera in 1970 may not have been due to a resumption of westward progression from Iraq, but instead from a new introduction from South or East Asia into Russia and the Middle East. Cholera reached Angola in December 1971 (1). The origin of this outbreak has remained unclear, as (i) it occurred more than 1000 km from the closest country with a cholera outbreak at this time (Cameroon), and (ii) it was caused by a serotype Inaba strain. Our analysis suggests a West African origin for this outbreak, with T1 isolates displaying a non-synonymous SNV (G103A) in the *wbeT* gene. This mutation explains the serotype switch from Ogawa to Inaba (see supplementary text note 2, fig. S8, and tables S1 and S6). This sublineage continued to spread throughout Southern Africa during the 1970s, and it circulated in this region until the early 1990s. These findings are supported by the report of a cholera outbreak in the Mozambican seaport of Beira, imported by air from Angola during late 1973 (20). This particular pattern of long-distance transmission between Angola and Mozambique may be explained by a Portuguese decolonization war involving these two countries at the time. This T1 sublineage was also implicated in a large outbreak in Portugal in 1974, in which 2467 cases and 48 deaths were recorded (21). Colonial troops were stationed uphill from one of the springs supplying the city in which the outbreak began (21). These troops had been traveling back and forth between their base and Angola, Mozambique, and Portuguese Guinea (now Guinea Bissau) (21). All the other Western and Southern European isolates from the early 1970s were found to have originated from West or North Africa.

After the introduction of cholera into West Africa in 1970, we identified multiple subsequent introductions of cholera into this region, with further extensions to the Gulf of Guinea region and the Lake Chad Basin on at least three occasions over the next three decades: T7 (introduction dates: 1982 to 1984), T9 (1988 to 1991), and T12 (2007). The T9 introduction led to outbreaks with high attack rates, such as those that struck Guinea in 1994 (436 reported cholera cases per 100,000 population) and Liberia in 2003 (1241 per 100,000) (table S7).

The second introduction of cholera into Africa in 1970 occurred in East Africa, particularly in Ethiopia in November 1970 (1). The strain involved, another “Inaba strain,” corresponds to our T3 isolates with a premature stop codon (C157T) in the *wbeT* gene (fig. S8 and tables S1 and S6). The 1970 isolates from Ethiopia are most closely related to isolates from the same year from Jordan and Israel (22) (3- to 9-SNV separation), which is consistent with epidemiological data from the time that suggested an importation from the Middle East (1).

Over the course of the 7P, we identified six other introductions into East Africa (T4, T5, T6, T10) (mostly the Horn of Africa) and Southern Africa (T8, T11) (mostly Mozambique) (Fig. 2). Our data suggest that cholera around the African Great Lakes for a period of >10 years was due to the introduction of the T5 and T10 sublineages. The T5 sublineage was, in particular, associated with outbreaks in Rwandan refugee camps in 1994, whereas the T8 sublineage was associated with outbreaks in South Africa in 2001 to 2002 (125,000 cases) and in Zimbabwe in 2008 to 2009 (98,000 cases) (table S7). The T11 sublineage was also found associated with this latter outbreak.

The T1, T3, T4, T5, T6, T8, and, possibly, T9 introduction events involved 7PET sublineages originating from South or East Asia that were circulating in the Middle East before importation into West or East Africa (Figs. 1 and 2).

Moreover, epidemiological data for T1 isolates identified during an outbreak in the Comoros Islands (in the Indian Ocean, off the African coast) in 1975 linked them to pilgrims returning from the Hajj, a religious gathering involving millions of people (20), where cholera outbreaks were frequently described until 1989 (23). The most recent introductions (T10 to T12) appear to be imported directly from South Asia, as no Middle Eastern isolates were found to belong to the imported sublineages. This may be due to (i) an increase in the movement of people between Africa and South Asia or (ii) a lack of relevant samples, as the most recent Middle Eastern isolate analyzed was collected in 2003. We observed recurrent patterns of transmission for cholera epidemics in Africa. Separate 7PET sublineages from Asia were repeatedly introduced into two main regions: West Africa and East/Southern Africa. Epidemic waves then propagated regionally, in some instances spreading to Central Africa, over periods of a few years to 28 years (Fig. 3 and table S5). Only two notable instances of sublineage exchange between these two circulation hotspots were identified: (i) the spread of a sublineage between Angola and Mozambique during the Portuguese colonial war in the 1970s (20) and

(ii) the spread of a sublineage from the African Great Lakes Region to the western part of the Democratic Republic of the Congo (6) and the Central African Republic via the Congo River and its tributaries in 2011 to 2012 (Fig. 2).

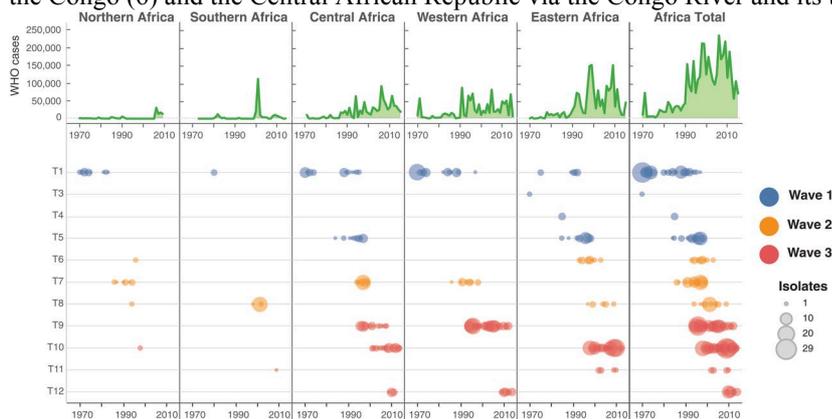


FIG. 3 Geographic and temporal distribution of seventh pandemic *V. cholerae* El Tor isolates from Africa according to their inferred introduction events (T1 to T12).

The annual number of cholera cases reported to the World Health Organization (WHO) is shown in the upper panels. The United Nations subregion scheme was used for the geographic breakdown. The size of the circle scales with the number of genomes analyzed per year.

Our study also provides a notable historical perspective on the evolution of antibiotic resistance in 7PET isolates.

Antibiotics have been used for decades in the treatment of cholera, as an adjunct to rehydration therapy, as they shorten the duration of diarrhea, thereby limiting bacterial spread [106 to 109 bacteria per milliliter of stool, with the loss of up to 1 liter of diarrheal fluid per hour (1)].

Our data suggest that African isolates became increasingly resistant to antibiotics over time (figs. S9 to S11). The first antibiotic-resistant African isolates in our collection were recovered in the early 1980s. No susceptible isolates have been collected since 2000 [96.4% susceptible isolates between 1970 and 1984 ($n = 80/83$), 36.9% between 1985 and 1999 ($n = 100/271$), and 0% between 2000 and 2014 ($n = 0/296$)]. We found that recent patterns of antibiotic resistance among 7PET sublineages in Africa have been largely driven by resistance determinants already present within the genomes of the imported sublineages, rather than by the independent local acquisition of resistance within this region (Fig. 4, A to C, and figs. S9 to S12).

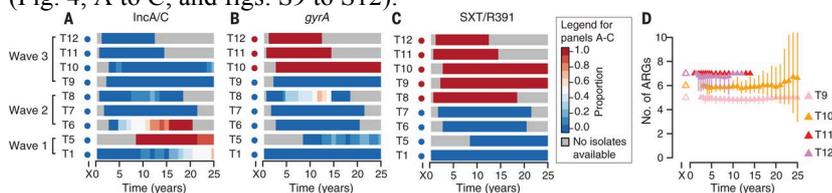


FIG. 4 Evolution of antibiotic resistance in seventh pandemic *V. cholerae* El Tor isolates from Africa.

For each introduction into Africa (T1, T5 to T12), the proportion of genomes at different time periods following introduction that contain (A) an IncA/C plasmid, (B) a *gyrA* mutation, and (C) a SXT/R391 genomic island. (D) For wave 3, the mean number of antibiotic resistance genes (ARGs) per isolate at different time periods following introduction. Each value is calculated over a 10-year window, with the point plotted at the midpoint of the period.

Sampling bias was prevented by randomly down-sampling the number of sequences to one sequence per country per year. In (A), (B), and (C), the color in the leftmost circle (at the value “X” on the x axis) represents the value for the closest genomic sequence obtained from outside Africa [for (A) to (C), a value of 1 implies that the sequence included the corresponding antibiotic resistance determinant, whereas 0 implies that the sequence lacked the determinant].

Where it did occur, the local acquisition of antibiotic resistance genes (ARGs) (table S5, fig. S9, supplementary text note 3) mostly involved multidrug-resistant (MDR) plasmids of the IncA/C group. IncA/C plasmids are the only stably maintained plasmids in the 7PET lineage (24). Almost all T5 isolates (1984 to 1998, wave 1) were found to contain an MDR IncA/C plasmid (Fig. 4A and fig. S12). This sublineage, which circulated through the African Great Lakes Region, may be derived from a strain described in Tanzania in 1977. At that time, the Tanzanian Ministry of Health used 1.79 metric tons of tetracycline to control an outbreak caused by a susceptible strain (25). Five months later, 76% of the isolates were resistant to tetracycline and other antibiotics owing to the acquisition of an IncA/C plasmid (25). We show that IncA/C plasmids were also acquired repeatedly between 1984 and 2014 in T1, T6, and T10 isolates collected in West or East Africa. These acquisition events may also be linked to mass chemoprophylaxis, as observed in Madagascar in 1999 (T10), when doxycycline was used for prophylaxis (26). The acquisition of resistance to the quinolone nalidixic acid, mediated by various point mutations in the DNA gyrase gene, *gyrA*, was identified three times in Central and East Africa (Fig. 4B, figs. S10 to S12, and table S1). The appearance of two independent mutations [resulting in the substitution of Ser83 with Arg (S83R) and Asp87 with Tyr (D87Y)] in T5 isolates may reflect the heavy use of nalidixic acid for the treatment of dysentery in Rwandan refugee camps, which experienced concomitant outbreaks of diarrhea caused by *Shigella dysenteriae* type 1 and *V. cholerae* O1 at this time (27).

Since 2000, almost all antibiotic-resistant 7PET sublineages circulating in Africa (T8 to T12) originated from South Asia and carry their ARGs on genomic islands SXT/R391 (28) or GI-15 (29) (Fig. 4C, figs. S6 and S11, table S5, and supplementary text note 3). These more recently imported sublineages also carry resistance mutations in *gyrA* [resulting in the replacement of Ser83 with Ile (S83I); all T10 to T12 isolates] and an additional mutation in *parC* [resulting in the substitution of Ser85 with Leu (S85L)] that decrease susceptibility to ciprofloxacin (all T11 and T12 isolates). The T9 to T12 isolates, which contain a similar SXT/R391 element (ICEVchInd5/ICEVchBan5), do not contain tetracycline resistance genes. The ARG content of these SXT/R391-containing sublineages has remained very stable for more than a decade since their introduction into Africa (Fig. 4D, fig. S12, and table S1). The apparent incompatibility between SXT/R391 and IncA/C plasmids in 7PET may be due to functional interference between these related elements (28). Our data demonstrate that the repeated introductions of 7PET sublineages into Africa describe the burden of disease (as measured by cases reported to the WHO) seen across this continent. These data are consistent with epidemiological studies, which have demonstrated that human-related factors play a much more important role in cholera dynamics in Africa than climatic and environmental factors (6, 7). Our data do not suggest that aquatic environmental reservoirs are the primary source of epidemic cholera in Africa, as has been suggested (30). Instead, these results highlight the role that humans play in the long-term spread and maintenance of the pathogen, whether by direct (human-to-human) or indirect (pollution of the environment with feces from cholera patients) transmission. Undoubtedly, the factors influencing the epidemiology and transmission of cholera are complex, but these data provide a detailed genetic context against which we can gauge the impact of interventions on future patterns of disease in this region.

Acknowledgments

This study was supported by the Institut Pasteur and the Institut Pasteur International Network, the Institut de Veille Sanitaire, the French government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-IBEID), the Fondation Le Roch-Les Mousquetaires, the Wellcome Trust through grant 098051 to the Sanger Institute, and the Indian Council of Medical Research, New Delhi, India. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. We thank C. Bréchet and H.-F. Thiéfaîne for their support; O. Vandenberg, Y. Germani, and M.-C. Fonkoua for providing isolates; D. Nedelec and E. E. Hansen for helpful discussions; S. Dupke and A. Smith for technical assistance; L. Ma, M. Santovenia, J. Winkjer, M. Aslett, A. J. Page, J. Keane, and the sequencing teams at the Institut Pasteur, the Wellcome Trust Sanger Institute, and the Centers for Disease Control and Prevention for sequencing the samples. F.J.L. and M.-L.Q. are members of the WHO Global Task Force on Cholera Control. F.J.L. is a past member of the WHO Strategic Advisory Group of Experts working group for cholera vaccines. J.P. is a member of the Institut Pasteur Scientific Council. J.P. has received consulting fees from Specific Technologies, Mountain View, California. Short-read sequences have been deposited in the European Nucleotide Archive (ENA) (www.ebi.ac.uk/ena) under study accession numbers PRJEB8764, PRJEB2215, and PRJEB19893. The whole-genome alignment for the 1070 genomes and other files have been deposited in FigShare: <https://figshare.com/s/d6c1c6f02eac0c9c871e>. Phylogeny and metadata can be viewed interactively at <https://microreact.org/project/globalcholera?tt=rc>.

Supplementary Material

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6364/785/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S12

Tables S1 to S9

References (31–77)

REFERENCES AND NOTES

1. D. Barua, W. B. Greenough, Eds., *Cholera* (Plenum, New York, 1992)
2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, *PLOS Negl. Trop. Dis.* 9, e0003832 (2015).
3. A. Mutreja et al., *Nature* 477, 462–465 (2011).
4. X. Didelot et al., *PLOS Genet.* 11, e1005072 (2015).
5. N. H. Gaffga, R. V. Tauxe, E. D. Mintz, *Am. J. Trop. Med. Hyg.* 77, 705–713 (2007).
6. S. Rebaudet, B. Sudre, B. Faucher, R. Piarroux, *J. Infect. Dis.* 208 (suppl. 1), S46–S54 (2013).
7. S. Rebaudet, B. Sudre, B. Faucher, R. Piarroux, *J. Infect. Dis.*

- 208 (suppl. 1), S98–S106 (2013).
8. S. Koblavi, F. Grimont, P. A. D. Grimont, *Res. Microbiol.* 141, 645–657 (1990).
 9. D. K. Karaolis, R. Lan, P. R. Reeves, *J. Bacteriol.* 176, 6199–6206 (1994).
 10. G. M. Evins et al., *J. Infect. Dis.* 172, 173–179 (1995).
 11. R. Lan, P. R. Reeves, *J. Clin. Microbiol.* 40, 172–181 (2002).
 12. C. Lam, S. Octavia, P. Reeves, L. Wang, R. Lan, *Emerg. Infect. Dis.* 16, 1130–1132 (2010).
 13. C. Lam, S. Octavia, P. R. Reeves, R. Lan, *BMC Microbiol.* 12, 82 (2012).
 14. J. Kiiru et al., *PLOS ONE* 8, e74829 (2013).
 15. S. Rebaudet et al., *PLOS Negl. Trop. Dis.* 8, e2898 (2014).
 16. R. S. Kaas et al., *PLOS ONE* 11, e0155691 (2016).
 17. M. Garrine et al., *PLOS Negl. Trop. Dis.* 11, e0005671 (2017).
 18. D. Domman et al., *Science* 358, XXX (2017).
 19. Materials and methods are available as supplementary materials.
 20. M. Echenberg, Ed., *Africa in the Time of Cholera: A History of Pandemics from 1817 to the Present* (Cambridge Univ. Press, Cambridge, UK, 2011).
 21. P. A. Blake et al., *Am. J. Epidemiol.* 105, 337–343 (1977).
 22. J. Cohen et al., *Lancet* 298, 86–89 (1971).
 23. Q. A. Ahmed, Y. M. Arabi, Z. A. Memish, *Lancet* 367, 1008–1015 (2006).
 24. K. Rahal, G. Gerbaud, D. H. Bouanchaud, *Ann. Microbiol. (Paris)* 129, 409–414 (1978).
 25. K. J. Towner, N. J. Pearson, F. S. Mhalu, F. O’Grady, *Bull. World Health Organ.* 58, 747–751 (1980).
 26. J. A. Dromigny et al., *Emerg. Infect. Dis.* 8, 336–338 (2002).
 27. E. Njamkepo et al., *Nat. Microbiol.* 1, 16027 (2016).
 28. R. A. Wozniak et al., *PLOS Genet.* 5, e1000786 (2009).
 29. C. J. Grim et al., *J. Bacteriol.* 192, 3524–3533 (2010).
 30. R. R. Colwell, *Science* 274, 2025–2031 (1996).