Genomic history of the seventh pandemic of cholera in Africa


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) (7). 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 generally, 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 underestimated. 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.

2Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Paris, 75015, France. 2Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. 3Centers for Disease Control and Prevention, Escherichia and Shigella Reference Unit, Atlanta, GA 30333, USA. 4Centre for Enteric Diseases, National Institute for Communicable Diseases, Johannesburg 2193, South Africa. 5Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa. 6Institut Pasteur, Mathematical Modelling of Infectious Diseases, Paris, 75015, France. 7Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. 8Laboratoire de Parasitologie-Mycologie, CHU Timone, Université de la Méditerranée, Marseille, 13385, France. 9National Institute of Cholera and Enteric Diseases (NICED), Kolkata, West Bengal 700010, India. 10Institut Pasteur de Bangui, BP 923, Bangui, Central African Republic. 11Institut Pasteur de Dakar, BP 220, Dakar, Senegal. 12Epicentre, Paris, 75011, France. 13Centre Pasteur du Cambouran, BP 1274, Yaoundé, Cameroon. 14Bacteriological and Virology Department, Institut Pasteur, Abidjan, Côte d’Ivoire. 15Rostov-on-Don Research Institute for Plague Control, Rostov-on-Don, 344022, Russia. 16Institut Pasteur, Plate-forme Génomique (PFG), Paris, 75015, France. 17Université de Bari “A. Moro”, Department of Biology, Bari, 70125, Italy. 18Department of Medicine, University of Cambridge, Addenbrooke’s Hospital, Cambridge CB2 0SP, UK. 19Translational Health Science and Technology Institute (THSTI), Faridabad, Haryana 121001, India. 20Robert Koch Institute, 13353 Berlin, Germany. 21Centre de Recherche Medicale et Sanitaire (CEREMS), BP 10887, Niamey, Niger. 22Médecins Sans Frontières (MSF), Paris, 75011, France. 23Cantacuzino National Institute of Research-Development for Microbiology and Immunology, Bucharest, Romania. 24Agence de Médecine Prévontive (AMP), Paris, 75015, France. 25Université de Versailles Saint-Quentin-en-Yvelines, UFR des sciences de la santé Simone Veil, Montigny-le-Bretonneux, 78130, France. 26Atelier de Bioinformatique, ISYEB, UMR 7205, Paris, 75015, France. 27Laboratoire Microbiologie Sante et Environnement (LMSEE), EDST-FSP, Université Libanais, Tripoli, Lebanon. 28Médecins Sans Frontières, Brussels, B 1050, Belgium. 29Centre Hospitalier des Armées Bouffard, Djibouti, Republic of Djibouti. 30Institut Pasteur, Unité du Choléra et des Vibrions, Paris, 75015, France. 31Institut Pasteur, Unité Biodiversité des Bactéries Pathogènes Emergentes, Paris, 75015, France. 32Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia. 33London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

*Corresponding author. Email: francois-xavier.weill@pasteur.fr (F.-X.W.) Present address: University Hospital of Pointe-à-Pitre/Abymes, University of Antilles, Pointe-à-Pitre, Guadeloupe, France. †These authors contributed equally to this work.
Each introduction resulted in mid- to long-term spread of cholera (introduction events T1, T3 to T7), 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).

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 (CI57T).

**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). Weill et al., Science 358, 785–789 (2017)
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 diarrhoea 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.

REFERENCES AND NOTES

2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, 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.

REFERENCES AND NOTES

2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, 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.

REFERENCES AND NOTES

2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, 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.

REFERENCES AND NOTES

2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, 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.

REFERENCES AND NOTES

2. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, 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.
Genomic history of the seventh pandemic of cholera in Africa


Science 358 (6364), 785-789.
DOI: 10.1126/science.aad5901

Wave upon wave of disease

The cholera pathogen, *Vibrio cholerae*, is considered to be ubiquitous in water systems, making the design of eradication measures apparently fruitless. Nevertheless, local and global *Vibrio* populations remain distinct. Now, Weill *et al.* and Domman *et al.* show that a surprising diversity between continents has been established. Latin America and Africa bear different variants of cholera toxin with different transmission dynamics and ecological niches. The data are not consistent with the establishment of long-term reservoirs of pandemic cholera or with a relationship to climate events. *Science*, this issue p. 785, p. 789