Genomic Microevolution of *Vibrio cholerae* O1, Lake Tanganyika Basin, Africa

Yaovi M.G. Hounmanou, Elisabeth Njamkepo, Jean Rauzier, Karin Gallandat, Aurélie Jeandron, Guyguy Kamwiziku, Klaudia Porten, Francisco Luquero, Aaron Aruna Abedi, Baron Bashige Rumedeka, Berthe Miwanda, Martin Michael, Placide Welo Okitayemba, Jaime Mufitini Saidi, Renaud Piarroux, François-Xavier Weill, Anders Dalsgaard, Marie-Laure Quilici

Africa's Lake Tanganyika basin is a cholera hotspot. During 2001–2020, *Vibrio cholerae* O1 isolates obtained from the Democratic Republic of the Congo side of the lake belonged to 2 of the 5 clades of the AFR10 sublineage. One clade became predominant after acquiring a *parC* mutation that decreased susceptibility to ciprofloxacin.

Cholera is an acute life-threatening diarrheal disease responsible for ≈4.3 million cases and 142,000 deaths annually worldwide (*1*). Excluding epidemic peaks in Haiti and Yemen (*2*,*3*), most cases of cholera originate from sub-Saharan Africa, predominantly the African Great Lakes Region (AGLR); specifically, the countries of the Lake Tanganyika basin (*4*). Many recurrent cholera outbreaks in the Democratic Republic of the Congo (DRC), Tanzania, Burundi, and Zambia have been linked to a common hotspot area around the Lake Tanganyika basin (*5*–*8*).

Author affiliations: University of Copenhagen, Frederiksberg, Denmark (Y.M.G. Hounmanou, A. Dalsgaard); Institut Pasteur, Université Paris Cité, Paris, France (E. Njamkepo, J. Rauzier, F.-X. Weill, M.-L. Quilici); London School of Hygiene and Tropical Medicine, London, UK (K. Gallandat, A. Jeandron); University of Kinshasa, Kinshasa, Democratic Republic of the Congo (G. Kamwiziku); Epicentre, Paris (K. Porten, F. Luquero); Ministry of Public Health, Kinshasa (A. Aruna Abedi, P. Welo Okitayemba); Ministry of Public Health, Uvira, Democratic Republic of the Congo (B. Bashige Rumedeka, J. Mufitini Saidi); Institut National de Recherche Biomédicale, Kinshasa (B. Miwanda); Sokoine University of Agriculture College of Veterinary Medicine and Biomedical Sciences, Morogoro, Tanzania (M. Michael); Sorbonne Université, Inserm UMR 1136, Assistance Publique-Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, Paris (R. Piarroux)

DOI: https://doi.org/10.3201/eid2901.220641

By the end of 2018, the World Health Organization had noted a steady decline in cholera cases throughout the world, including the AGLR (*9*). Continuous genomic surveillance of circulating *Vibrio cholerae* bacteria strains is required to understand the transmission dynamics and genetic evolution of *V. cholerae* and potentially to guide prevention and response interventions to continue the trend toward decreasing case numbers, in line with the global cholera roadmap to 2030 (*10*). One lineage, seventh pandemic *V. cholerae* O1 El Tor (7PET), is responsible for the current pandemic, which began in 1961 (*11*); Africa was hit by 7PET in 1970 (*11*). During 1970–2014, ≥11 different 7PET sublineages were introduced from South Asia into Africa, and sublineage AFR10 (previously T10) replaced AFR5 (previously T5) in the AGLR in the late 1990s (*11*). Sublineage AFR13 (previously T13) was identified in East Africa (Tanzania, Uganda, Kenya) and Zimbabwe (*12*). We tracked the 7PET populations circulating in the Lake Tanganyika basin by studying recent *V. cholerae* O1 isolates collected in the region by conventional bacteriology and genomics and placing these genomes in a broader phylogenetic context to elucidate their evolutionary history.

The Study

We analyzed 96 *V. cholerae* O1 isolates collected during 2015–2020 in DRC (86 clinical isolates, including 39 collected in 2018–2020) and Tanzania (10 environmental isolates from fish and lake water) (Appendix 1, https://wwwnc.cdc.gov/EID/article/29/1/22-0641-App1.pdf; Appendix 2 Table 1, https://wwwnc.cdc.gov/EID/article/29/1/22-0641-App2.xlsx). We subjected the isolates to antimicrobial susceptibility testing, whole-genome sequencing, genomic characterization, and phylogenetic analyses, as previously

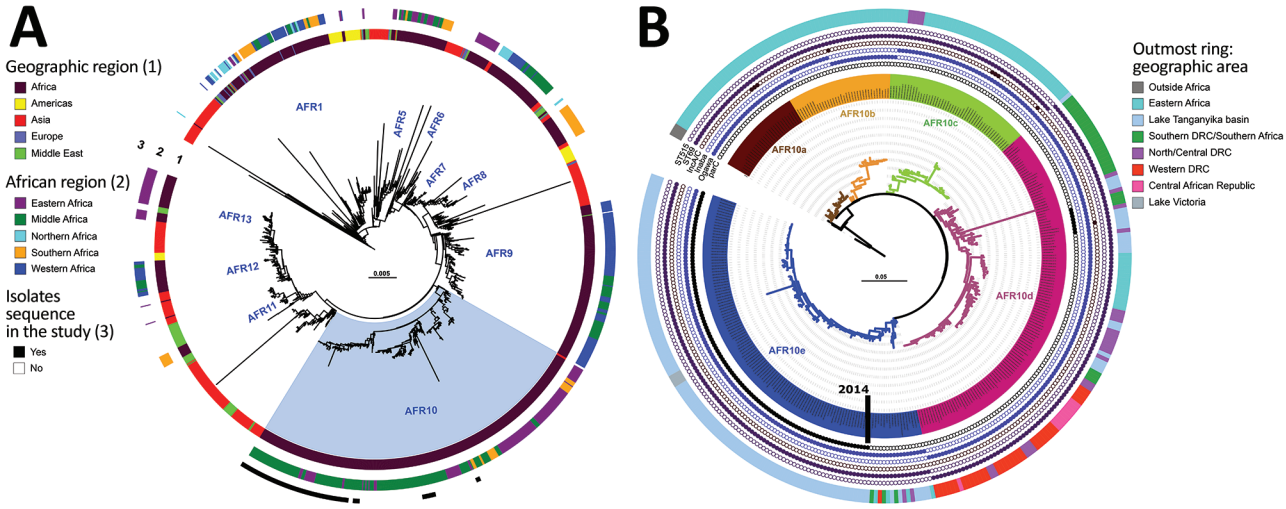


Figure 1. Phylogenomics of clinical and environmental *Vibrio cholerae* O1 El Tor isolates from the Lake Tanganyika basin, Africa. A) Maximum-likelihood phylogeny of 1,366 seventh pandemic *V. cholerae* O1 El Tor (7PET) genomes with strain A6 as the outgroup. The different sublineages introduced into Africa are indicated. Light blue indicates AFR10 sublineage. Rings 1 and 2 show geographic origin of isolates; ring 3 shows isolates sequenced in this study. B) Maximum-likelihood tree for 357 AFR10 isolates, with strain N16961 as an outgroup. The 5 clades are color-coded: AFR10a, brown; AFR10b, yellow; AFR10c, green; AFR10d, pink; and AFR10e, blue. The outermost ring indicates the geographic locations of the different isolates in the tree. Filled circles indicate the presence of ST69 or ST515, Ogawa and Inaba serotypes, IncA/C plasmid, and the S85L mutation in *par*C; open circles indicate their absence. MLST, multilocus sequence typing; ST, sequence type.

described (*11*,*12*) (Appendix 1). We performed a phylogenetic analysis of these genomes within a global collection of 1,366 7PET *V. cholerae* O1 genomes (Appendix 2 Table 2), including another 130 genomes from DRC collected during 1984–2017. We based the final maximum-likelihood phylogenetic tree on 10,352 single-nucleotide variants distributed over the nonrepetitive, nonrecombinant core genome (Figure 1, panel A).

Phylogenetic analysis of the 96 genomes of *V. cholerae* O1 isolates showed that all belonged to 7PET sublineage AFR10 (Figure 1, panel A). Within the limits of our sampling, sublineage AFR5, which circulated actively in the AGLR during the 1980s–1990s (*11*), appears to be extinct in the region, whereas sublineage AFR13, reported in 2015 in Uganda and Tanzania (*12*), has not yet spread to DRC. Since 1998, the endemicity of the AFR10 sublineage in the Tanganyika basin and surrounding countries has led to microevolution; an analysis of 357 AFR10 genomes from our global dataset revealed the presence of 5 clades, AFR10a–AFR10e (Figure 1, panel B; Appendix Figure 1). Clades AFR10a, AFR10b, and AFR10c were mostly associated with the eastern AGLR countries. The isolates of these clades were of sequence type (ST) 69 (Figure 1, panel B). Clades AFR10d and AFR10e predominated in DRC and the Lake Tanganyika basin. Clade AFR10d is of ST515, essentially Inaba serotype, and was widespread in DRC and neighboring countries (Figure 2, panel A), as previously reported (*13*). It was the only clade found in the Lake Kivu and Lake Edward basins. AFR10e strains are of ST69, Ogawa serotype, and were essentially restricted to the Lake Tanganyika basin, confirming previous findings (*13*). A further pangenome analysis of the AFR10 isolates revealed no clade-specific gain or loss of genes (Appendix Figure 2). AFR10e strains have gradually replaced AFR10d strains in the region since 2014; all *V. cholerae* O1 strains obtained from the Tanganyika basin by 2017, as well as those obtained from the lake itself in 2018 and 2019, were AFR10e strains (Figure 2 panel B). Epidemiologic studies identified cholera hotspots in the AGLR as a source of major countrywide outbreaks reaching the capital, Kinshasa, and the Atlantic coast, via the Congo River, in 2011, 2012, and 2016 (*5*). These outbreaks were caused primarily by clade AFR10d, which has a wider geographic distribution than clade AFR10e (Figure 2) (*14*).

One striking characteristic of AFR10e isolates was the presence of a mutation in the quinolone resistance–determining region of the topoisomerase IV subunit A gene, *parC* (S85L), that led to higher MIC values (0.25–0.38 mg/L) for ciprofloxacin. These isolates, which have a lower susceptibility

to ciprofloxacin than wild-type populations, would be classified as either resistant (MIC 0.38 mg/L) or susceptible (MIC 0.25 mg/L) in accordance with clinical breakpoints published by the European Committee on Antimicrobial Susceptibility Testing (https://www.eucast.org/clinical_breakpoints) in 2022, as inferred from the study of *Vibrio* strains and not Enterobacteriaceae strains. This *parC* mutation, sporadically reported in other AFR10 clades (Table; Figure 1, panel B; Appendix 2 Table 1), has been a distinctive characteristic of AFR10e isolates since 2014 (Figure 2, panel B). It was the second mutation affecting susceptibility to quinolones and fluroquinolones to be found in this AFR10e clade; the first was a mutation in the DNA gyrase subunit A gene, *gyrA* (S83I), present in all AFR10 isolates. This additional mutation does not seem to be associated with the specific use of fluoroquinolones for treating cholera outbreaks, because the antimicrobial drugs used for first-line cholera control in DRC are tetracyclines and macrolides, to which AFR10e isolates remain susceptible. Instead, the mutation may result from widespread self-medication with antimicrobials, a common practice in many sub-Saharan Africa countries including DRC (*15*).

All 96 isolates analyzed had known mutations of the *VC_0715* and *VC_A0637* genes conferring nitrofuran resistance (Table), consistent with previous findings (*2*). The isolates also carried the SXT/R391 genomic element ICE*Vch*Ind5, encoding resistance to streptomycin (*strAB*), sulfonamides (*sul2*), chloramphenicol (*flo*R), trimethoprim and the O/129 vibriostatic agent (*dfrA1*), and trimethoprim–sulfamethoxazole (*sul2* and *dfrA1*), with concordance between the phenotypic and genotypic data (Table; Appendix 2 Table 1).

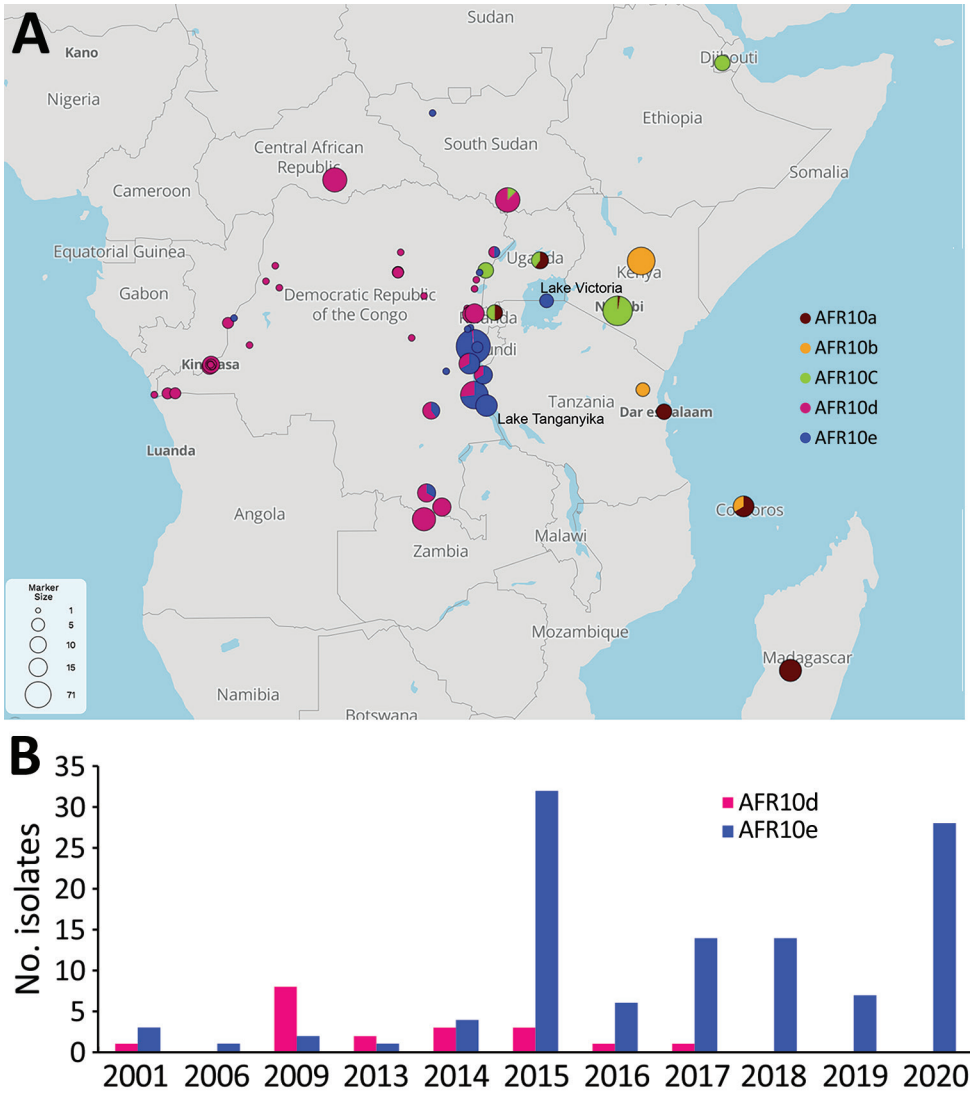


Figure 2. A) Spatiotemporal dynamics of the AFR10 clades of *Vibrio cholerae* O1 in the African Great Lakes Region, Africa, 1998–2020. Circle size indicates the number of isolates at the location concerned. The 5 AFR10 clades are color-coded: AFR10a, brown; AFR10b, yellow; AFR10c, green; AFR10d, pink; and AFR10e, blue. B) *V. cholerae* O1 isolates from the Lake Tanganyika basin. All AFR10 isolates from Bujumbura (Burundi), Kigoma (Tanzania) and the South-Kivu province (DRC) were considered to be Lake Tanganyika basin isolates.

Table. Characteristics of *Vibrio cholerae* O1 isolates from humans and the environment around Lake Tanganyika, Africa, 2015–2020*

Characteristic	2015	2016	2017	2018	2019	2020
No. isolates	26	7	14	14	7	28
AFR10 clade	AFR10d (1) AFR10e (25)	AFR10d (5) AFR10e (2)	AFR10e	AFR10e	AFR10e	AFR10e
Serotype	Inaba (1) Ogawa (25)	Inaba (5) Ogawa (2)	Ogawa	Ogawa	Ogawa	Ogawa
Source	Human	Human	Human	Human (10) Fish (2) Water (2)	Human (1) Fish (4) Water (2)	Human
MLST	ST515 (1) ST69 (25)	ST515 (5) ST69 (2)	ST69	ST69	ST69	ST69
Antimicrobial resistance determinants						
SXT/R391 element	ICE*Vch*Ind5	ICE*Vch*Ind5	ICE*Vch*Ind5†	ICE*Vch*Ind5	ICE*Vch*Ind5	ICE*Vch*Ind5
gyrA	S83I	S83I	S83I	S83I	S83I	S83I
parC	S85L	S85L (2), WT (5)	S85L	S85L	S85L	S85L
VC_0715	R169C	R169C	R169C	R169C	R169C	R169C
VC_A0637	Q5Stop	Q5Stop	Q5Stop	Q5Stop	Q5Stop	Q5Stop
AMR	AMR1‡ (26)	AMR1‡ (2) AMR2§ (5)	AMR1‡ (13) AMR3¶ (1)	AMR1 (14)‡	AMR1 (7)‡	AMR1 (28)‡

*Numbers in parentheses indicate number of isolates for each designation. AMR, antimicrobial resistance patterns; CIP, ciprofloxacin; DS, decreased susceptibility; FUR, nitrofurantoin; MLST, multilocus sequence typing; NAL, nalidixic acid; O129, cross-resistance to the vibriostatic agent O/129 and trimethoprim; PMB, polymyxin B; SSS, sulfonamides; ST, sequence type; STR, streptomycin; SXT, trimethoprim/sulfamethoxazole
†Deletion of the *strA*, *strB*, *floR*, and *sul2* genes in isolate CNRVC170308 (all isolates with intact ICE*Vch*Ind5 harbor the *strA*, *strB*, *dfrA1*, *floR*, and *sul2* resistance genes)
‡Resistance to STR, SSS, O129, SXT, FUR, PMB, NAL, and CIP^{DS}.
§Resistance to STR, SSS, O129, SXT, FUR, PMB, and NAL.
¶Resistance to O129, FUR, PMB, NAL, CIP^{DS}.

Conclusions

We found that the cholera outbreaks in the eastern part of DRC during 2001–2020 were caused by *V. cholerae* O1 sublineage AFR10, which was introduced into East Africa from South Asia in the late 1990s. The AFR13 sublineage was already reported in 2015 in Tanzania, including the city of Kigoma, located on the shore of Lake Tanganyika, but had not been detected in DRC as of 2022. The AFR10 isolates of this region belong principally to 2 clades, AFR10d (Inaba, ST515) and AFR10e (Ogawa, ST69). AFR10d was responsible for outbreaks reported in the western part of DRC in 2011–2017 and neighboring countries; AFR10e (Ogawa, ST69) was restricted to the Lake Tanganyika basin, in which reduced susceptibility to ciprofloxacin has been seen since 2014. Lake Tanganyika seems to serve as a transmission channel, favoring the establishment of AFR10e in local human populations. Further investigation, including studies of population movement, should reveal why AFR10e clade has remained within the Lake Tanganyika basin. The replacement of other clades by this antimicrobial-resistant clade in this area highlights the need for more systematic documentation of antimicrobial drug use and the implementation of adapted stewardship programs, particularly in outbreak responses. Overall, these findings highlight the need for continuous genomic surveillance and for coordinated communication between countries for effective interventions.

The International Foundation for Sciences funded the fieldwork for the isolation of environmental isolates (grant no. 1-2-A-6100-1). A.D.'s laboratory work is supported by core funding from the University of Copenhagen. The London School of Hygiene and Tropical Medicine study was cofunded by the French Agency for Development and the Veolia Foundation. The laboratory of F.-X.W. is part of the Integrative Biology of Emerging Infectious Diseases Laboratory of Excellence funded by the Government of France's "Investissement d'Avenir" programme (grant no. ANR-10-LABX-62-IBEID). The CNRVC is cofunded by Santé Publique France and the Institut Pasteur.

About the Author

Dr. Hounmanou is a postdoctoral fellow specializing in One Health at the University of Copenhagen. His primary research interests are microbial genomics, antimicrobial resistance, and routes of transmission between animals, humans, and bodies of water.

References

1. Weil AA, Ryan ET. Cholera: recent updates. Curr Opin Infect Dis. 2018;31:455–61. https://doi.org/10.1097/QCO.0000000000000474
2. Weill FX, Domman D, Njamkepo E, Almesbahi AA, Naji M, Nasher SS, et al. Genomic insights into the 2016–2017 cholera epidemic in Yemen. Nature. 2019;565:230–3. https://doi.org/10.1038/s41586-018-0818-3
3. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian

outbreak. MBio. 2011;2:e00157-11. https://doi.org/10.1128/mBio.00157-11
4. Lessler J, Moore SM, Luquero FJ, McKay HS, Grais R, Henkens M, et al. Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales. Lancet. 2018;391:1908–15 https://doi.org/10.1016/S0140-6736(17)33050-7
5. Ingelbeen B, Hendrickx D, Miwanda B, van der Sande MAB, Mossoko M, Vochten H, et al. Recurrent cholera outbreaks, Democratic Republic of the Congo, 2008–2017. Emerg Infect Dis. 2019;25:856–64. https://doi.org/10.3201/eid2505.181141
6. Hounmanou YMG, Mølbak K, Kähler J, Mdegela RH, Olsen JE, Dalsgaard A. Cholera hotspots and surveillance constraints contributing to recurrent epidemics in Tanzania. BMC Res Notes. 2019;12:664. https://doi.org/10.1186/s13104-019-4731-0
7. Debes AK, Shaffer AM, Ndikumana T, Liesse I, Ribaira E, Djumo C, et al. Cholera hot-spots and contextual factors in Burundi, planning for elimination. Trop Med Infect Dis. 2021;6:76. https://doi.org/10.3390/tropicalmed6020076
8. Mwaba J, Debes AK, Shea P, Mukonka V, Chewe O, Chisenga C, et al. Identification of cholera hotspots in Zambia: a spatiotemporal analysis of cholera data from 2008 to 2017. PLoS Negl Trop Dis. 2020;14:e0008227. https://doi.org/10.1371/journal.pntd.0008227
9. World Health Organization. Drop in cholera cases worldwide, as key endemic countries report gains in cholera control. 2019 [cited 2020 Apr 22]. https://www.who.int/news-room/detail/19-12-2019-drop-in-cholera-cases-worldwide-as-key-endemic-countries-report-gains-in-cholera-control
10. World Health Organization. Ending cholera. a global roadmap to 2030. 2017 [cited 2020 Apr 22]. https://www.gtfcc.org/wp-content/uploads/2019/10/gtfcc-ending-cholera-a-global-roadmap-to-2030.pdf
11. Weill FX, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, et al. Genomic history of the seventh pandemic of cholera in Africa. Science. 2017;358:785–9. https://doi.org/10.1126/science.aad5901
12. Hounmanou YMG, Leekitcharoenphon P, Kudirkiene E, Mdegela RH, Hendriksen RS, Olsen JE, et al. Genomic insights into *Vibrio cholerae* O1 responsible for cholera epidemics in Tanzania between 1993 and 2017. PLoS Negl Trop Dis. 2019;13:e0007934. https://doi.org/10.1371/journal.pntd.0007934
13. Irenge LM, Ambroise J, Mitangala PN, Bearzatto B, Kabangwa RKS, Durant JF, et al. Genomic analysis of pathogenic isolates of *Vibrio cholerae* from eastern Democratic Republic of the Congo (2014–2017). PLoS Negl Trop Dis. 2020;14:e0007642. https://doi.org/10.1371/journal.pntd.0007642
14. Breurec S, Franck T, Njamkepo E, Mbecko JR, Rauzier J, Sanke-Waïgana H, et al. Seventh Pandemic *Vibrio cholerae* O1 sublineages, Central African Republic. Emerg Infect Dis. 2021;27:262–6. https://doi.org/10.3201/eid2701.200375
15. Belachew SA, Hall L, Selvey LA. Non-prescription dispensing of antibiotic agents among community drug retail outlets in sub-Saharan African countries: a systematic review and meta-analysis. Antimicrob Resist Infect Control. 2021;10:13. https://doi.org/10.1186/s13756-020-00880-w

Address for correspondence: Marie-Laure Quilici, Institut Pasteur, Unité des Bactéries Pathogènes Entériques, 28 rue du Dr. Roux, 75724 Paris CEDEX 15, France; email: quilici@pasteur.fr

EID Podcast

Tracking *Bordetella pertussis*, Austria, 2018–20

Whooping cough is a reemerging, potentially deadly disease spread by a bacterium known as *Bordetella pertussis*. Fortunately, this respiratory infection is largely preventable with vaccination.

However, nature doesn't stay still, new antigenic variants of this bacterium are evolving and spreading.

In this EID podcast, Dr. Adriana Cabal Rosel, a public health microbiologist at the Austrian Agency for Health and Food Safety, describes a new surveillance system to track down these emerging variants in Austria.

Visit our website to listen: https://go.usa.gov/xshSV

EMERGING INFECTIOUS DISEASES®

Article DOI: https://doi.org/10.3201/eid2901.220641

Genomic Microevolution of *Vibrio cholerae* O1, Lake Tanganyika Basin, Africa

Appendix 1

Detailed Materials and Methods

Environmental Isolates from Lake Tanganyika

Fieldwork was conducted during September 2018–March 2019, a period including a dry and a rainy season, at 2 sites on the Tanzania side of Lake Tanganyika: Kibirizi, 2.2 km north of Kigoma, and Katonga, 4–5 km south of Kigoma Bay. Samples were collected after the major cholera outbreak that occurred in 2015–2018 in all 26 regions of the United Republic of Tanzania (https://www.who.int/emergencies/disease-outbreak-news/item/12-january-2018-cholera-tanzania-en). Since April 2018, the number of reported cases has fallen considerably in Kigoma, but cholera cases were still being reported right up until the end of 2018 (https://reliefweb.int/report/zimbabwe/bulletin-cholera-and-awd-outbreaks-eastern-and-southern-africa-regional-update-5) and early 2019 (https://reliefweb.int/report/mozambique/bulletin-cholera-and-awd-outbreaks-eastern-and-southern-africa-regional-update).

In total, 140 individual fish (*Oreochromis tanganicae*), each weighing ≈300 g and captured from deep offshore waters, were purchased directly from fishing boats, as soon as they landed at the 2 landing sites at Kibirizi and Katonga. During sample processing, each fish was subdivided into 2 subsamples corresponding to the gills and intestines, yielding a total of 280 subsamples for analysis.

We also collected 60 water samples (2 L each) over the sampling period. Each sample was collected ≥500 m from the shore, at the Kibirizi and Katonga landing sites, at a depth of ≈1 m from the surface. The samples were analyzed according to the protocols described in our previous articles, including an initial enrichment step in alkaline peptone water (*1*,*2*). The identity of the bacterial isolates was confirmed by PCR targeting the *V. cholerae*–specific outer membrane protein gene *ompW* and the cholera enterotoxin *ctxA* gene. The samples were then

serotyped with polyvalent O1 antiserum. Ten of the *V. cholerae* isolates recovered were confirmed to be *ctxA*-positive *V. cholerae* O1: 6 from fish (gills) and 4 from water. These 10 isolates were shipped to the University of Copenhagen, where they were tested for antimicrobial resistance patterns as described below. DNA was extracted with a Maxwell automatic purification machine and kit (Promega, https://www.promega.com) according to the manufacturer's protocol and were fully sequenced with the Illumina MiSeq platform (*2*,*3*), with 40× to 75× coverage. The paired-end reads obtained were submitted to the European Nucleotide Archive under accession number PRJEB37131.

Clinical *V. cholerae* O1 Isolates from the Democratic Republic of Congo (DRC)

Isolates collected in 2015 and 2016 (n = 17) from Uvira, Kinshasa, Lubumbashi, and Bunia were provided by Kinshasa University. Stool samples were collected by Médecins Sans Frontières France at the Kalemie Cholera Treatment Center (CTC) during October 2015–February 2016, as part of the cholera surveillance program, following the use of a bivalent oral cholera vaccine in the city of Kalemie. Samples from Uvira were collected during 2017–2020 in the context of a study conducted by the London School of Hygiene and Tropical Medicine (LSHTM) with the aim of evaluating the impact of urban water supply improvements on cholera and other diarrheal diseases in Uvira, DRC (www.clinicaltrials.gov identifier: NCT02928341, ethics approval from the LSHTM 8913, 10603 and from the Ethics Committee from the School of Public Health, University of Kinshasa, DRC ESP/CE/088/2015). Rectal swabs from patients with severe acute diarrhea admitted to the CTC at Uvira General Hospital were subjected to a six-hour enrichment procedure in alkaline peptone water (APW) at room temperature and were tested with a rapid diagnostic test (Crystal VC, Span Diagnostics, https://span-diagnostics.com). Rectal swab samples subjected to enrichment in APW that tested positive with cholera rapid diagnostic tests were preserved on a 6 mm strip of filter paper in a capped tube containing a few drops of saline ("wet filter sample"), and were transferred, in batches, from Uvira to Institut Pasteur at room temperature, in accordance with IATA transport restrictions for category B infectious substances.

Bacterial isolates, stool samples and enriched samples were sent to the Centre National de Référence des Vibrions et du Cholera (CNRVC), Institut Pasteur, Paris, France, for the confirmation of cholera diagnosis by culture according to standard protocols (*4*) and/or PCR (*5*)

and for further characterization to support epidemiologic surveillance; 16 isolates from stools (Kalemie) and 53 from enriched samples (Uvira) were recovered and added to the study.

Total DNA was extracted from the clinical isolates with the Maxwell 16-cell DNA purification kit (Promega, https://www.promega.com) in accordance with the manufacturer's instructions. Whole-genome sequencing was carried out on the Biomics and PIBnet sequencing platforms of the Institut Pasteur and at the *Institut du Cerveau* platform (Paris), on Illumina NextSeq or Novaseq platforms (Illumina, https://www.illumina.com) generating 100–250 bp paired-end reads, yielding 110x to 728x coverage. All the publicly available genomes added to the analysis are referenced in Appendix 2 Table 2. Short reads from the clinical isolates were submitted to ENA under accession number PRJEB47296.

Antimicrobial Susceptibility Testing

All the *V. cholerae* isolates analyzed in this study were subjected to antimicrobial susceptibility testing at the University of Copenhagen (environmental isolates) and the CNRVC, Institut Pasteur, Paris, France (clinical isolates). Antimicrobial susceptibility testing was performed by the disk diffusion method, on Mueller-Hinton agar (Bio-Rad, https://www.bio-rad.com), in accordance with EUCAST guidelines (https://www.eucast.org). The antimicrobial drugs tested were: ampicillin, cefalotin, cefotaxime, streptomycin, chloramphenicol, azithromycin, sulfonamides, trimethoprim-sulfamethoxazole, vibriostatic agent O/129, tetracycline, nalidixic acid, ciprofloxacin, nitrofurantoin, and polymyxin B. *Escherichia coli* CIP76.24 (ATCC25922) was used as a control. The MICs of nalidixic acid and ciprofloxacin were determined with Etests (bioMérieux, https://www.biomerieux.com). Interpretation of the MIC results for ciprofloxacin was based on the EUCAST 2022 clinical breakpoints for *Vibrio* species (https://www.eucast.org/clinical_breakpoints).

Genome Analysis

Trimmed reads were assembled with SPAdes version 3.15.2 (*6*). *In silico* multilocus sequence typing (MLST) for *V. cholerae* was performed with PubMLST (https://pubmlst.org/bigsdb?db = pubmlst_vcholerae_seqdef). Specific genetic markers for *V. cholerae* were analyzed with BLAST version 2.2.26. against reference sequences of the *rfb* O1gene, *ctxB*, *wbeT*, and the whole VSP-II locus, as previously described (*7*). Acquired antimicrobial resistance genes were predicted with ResFinder version 4.0.1

(https://cge.food.dtu.dk/services/ResFinder), BLAST analysis against GI-15, Tn*7*, and SXT/R391 integrative and conjugative elements, and PlasmidFinder version 2.1.1. (https://cge.food.dtu.dk/services/PlasmidFinder). The presence of mutations in genes associated with resistance to quinolones (*gyrA*, *gyrB, parC, parE*) and nitrofurans (*VC_0715* and *VC_A0637*) was investigated with in-house tools, using BLAST version 2.2.26, as previously described (*8*).

Phylogenetic Analysis

Single-nucleotide variants (SNVs) were called with Snippy version 4.6.0 /Freebayes version 1.3.2 (https://github.com/tseemann/snippy) under the following constraints: mapping quality of 60, a minimum base quality of 13, a minimum read coverage of 4, and 75% read concordance at a locus for a variant to be reported relative to the *V. cholerae* O1 El Tor N16961 reference genome. An alignment of core genome SNVs was produced in Snippy version 4.1.0 for phylogeny inference. Repetitive (insertion sequences and the TLC-RS1-CTX region) and recombinogenic (VSP-II) regions in the alignment were masked (*9*). Putative recombinogenic regions were detected and masked with Gubbins version 2.4.1 (*10*). A maximum likelihood (ML) phylogenetic tree was built, with RAxML version 8.2.12, under the GTR model, with 200 bootstraps (*11*). The final tree was rooted on the A6 genome and visualized with iTOL v5 (*12*) or FigTree version 1.4.2. (http://tree.bio.ed.ac.uk/software/figtree).

Pangenome Analysis

We annotated 355 of the 357 genomes of the AFR10 sublineage (the assemblies for the Tanz_100 and Tanz_20 strains being excluded due to light contamination with non-*Vibrio* species sequences) with Prokka version 1.13.4 (*13*). The annotated GFF3 files were then used as an input for the Roary (version.3.7.0) (*14*) pangenome analysis tool in a Linux interface. We used the binary presence/absence data of the accessory genome produced in Roary, together with the tree of AFR10 strains, for visualization in Phandango (*15*). The results are shown in Appendix 1 Figure 2.

References

1. Hounmanou YMG, Mdegela RH, Dougnon TV, Mhongole OJ, Mayila ES, Malakalinga J, et al. Toxigenic *Vibrio cholerae* O1 in vegetables and fish raised in wastewater irrigated fields and stabilization ponds during a non-cholera outbreak period in Morogoro, Tanzania: an environmental health study. BMC Res Notes. 2016;9:466. PubMed https://doi.org/10.1186/s13104-016-2283-0

2. Hounmanou YMG, Leekitcharoenphon P, Hendriksen RS, Dougnon TV, Mdegela RH, Olsen JE, et al. Surveillance and genomics of toxigenic *Vibrio cholerae* O1 from fish, phytoplankton and water in Lake Victoria, Tanzania. Front Microbiol. 2019;10: 901. https://www.frontiersin.org/articles/10.3389/fmicb.2019.00901/full

3. Hounmanou YMG, Leekitcharoenphon P, Kudirkiene E, Mdegela RH, Hendriksen RS, Olsen JE, et al. Genomic insights into *Vibrio cholerae* O1 responsible for cholera epidemics in Tanzania between 1993 and 2017. PLoS Negl Trop Dis. 2019;13:e0007934. PubMed https://doi.org/10.1371/journal.pntd.0007934

4. Dodin A, Fournier JM. Diagnosis of the cholera vibrio. Laboratory methods for the diagnosis of cholera vibrio and other vibrios. Institut Pasteur Paris (France); 1992. p. 59–82.

5. Fields PI, Popovic T, Wachsmuth K, Olsvik O. Use of polymerase chain reaction for detection of toxigenic *Vibrio cholerae* O1 strains from the Latin American cholera epidemic. J Clin Microbiol. 1992;30:2118–21. PubMed https://doi.org/10.1128/jcm.30.8.2118-2121.1992

6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77. PubMed https://doi.org/10.1089/cmb.2012.0021

7. Breurec S, Franck T, Njamkepo E, Mbecko JR, Rauzier J, Sanke-Waïgana H, et al. Seventh pandemic *Vibrio cholerae* O1 sublineages, Central African Republic. Emerg Infect Dis. 2021;27:262–6. PubMed https://doi.org/10.3201/eid2701.200375

8. Weill FX, Domman D, Njamkepo E, Almesbahi AA, Naji M, Nasher SS, et al. Genomic insights into the 2016–2017 cholera epidemic in Yemen. Nature. 2019;565:230–3. PubMed https://doi.org/10.1038/s41586-018-0818-3

9. Weill FX, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, et al. Genomic history of the seventh pandemic of cholera in Africa. Science. 2017;358:785–9. PubMed https://doi.org/10.1126/science.aad5901

10. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15–15. PubMed https://doi.org/10.1093/nar/gku1196

11. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3. PubMed https://doi.org/10.1093/bioinformatics/btu033

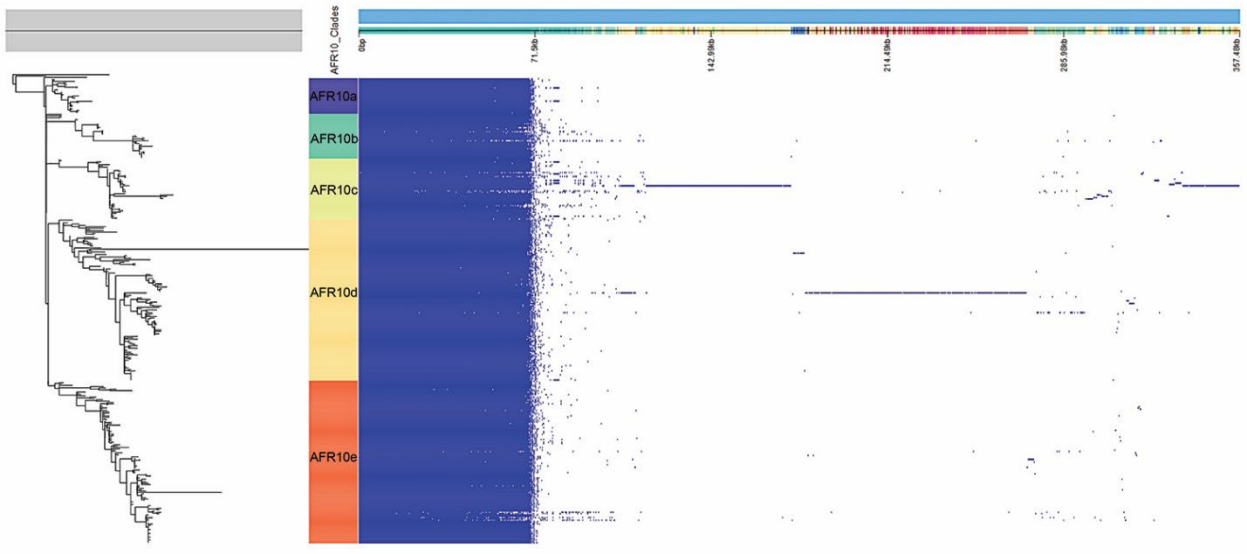
12. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–245. https://doi.org/10.1093/nar/gkw290

13. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. PubMed https://doi.org/10.1093/bioinformatics/btu153

14. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3. PubMed https://doi.org/10.1093/bioinformatics/btv421

15. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics. 2018;34:292–3. PubMed https://doi.org/10.1093/bioinformatics/btx610

Appendix Figure 1. Detailed phylogenetic tree for the AFR10 sublineage isolates with bootstrap values at nodes. Branches are color-coded: AFR10a, brown; AFR10b, yellow; AFR10c, green; AFR10d, pink; and AFR10e, blue.



Appendix Figure 2. Pangenome analysis of 355 AFR10 isolates, using Roary version 3.7.0. On the left, the maximum-likelihood tree as in Appendix Figure 1, except for the 2 contaminated genomes, which were excluded. The 5 clades, AFR10a, AFR10b, AFR10c, AFR10d, and AFR10e, are color coded. On the right, pangenomic matrix, sorted from core genes on the left to accessory genes on the right, shows the absence (white) and presence (blue) of genes. The multicolor line (top) indicates the size of contigs in kilobases (kb).