Emergence and Evolutionary Response of *Vibrio cholerae* to Novel Bacteriophage, Democratic Republic of the Congo[1]

Meer T. Alam,[2] Carla Mavian,[2] Taylor K. Paisie, Massimiliano S. Tagliamonte, Melanie N. Cash, Angus Angermeyer, Kimberley D. Seed, Andrew Camilli, Felicien Masanga Maisha, R. Kabangwa Kakongo Senga, Marco Salemi, J. Glenn Morris, Jr., Afsar Ali

Cholera causes substantial illness and death in Africa. We analyzed 24 toxigenic *Vibrio cholerae* O1 strains isolated in 2015–2017 from patients in the Great Lakes region of the Democratic Republic of the Congo. Strains originating in southern Asia appeared to be part of the T10 introduction event in eastern Africa. We identified 2 main strain lineages, most recently a lineage corresponding to sequence type 515, a *V. cholerae* cluster previously reported in the Lake Kivu region. In 41% of fecal samples from cholera patients, we also identified a novel ICP1 (Bangladesh cholera phage 1) bacteriophage, genetically distinct from ICP1 isolates previously detected in Asia. Bacteriophage resistance occurred in distinct clades along both internal and external branches of the cholera phylogeny. This bacteriophage appears to have served as a major driver for cholera evolution and spread, and its appearance highlights the complex evolutionary dynamic that occurs between predatory phage and bacterial host.

Cholera remains an ongoing public health threat on the continent of Africa, especially in the Democratic Republic of the Congo (DRC), which in 2020 reported the largest number of cases (19,789) of any country in the world with the exception of Yemen (*1*). Biotype El Tor strains of the seventh cholera pandemic (P7ET) were first reported in Africa in the early 1970s (*2*–*6*). Major outbreaks occurred in DRC in 2008, 2009, 2011–2012, 2013, and 2015–2017; in 2017 alone, an estimated 53,000 cholera cases with 1,145 deaths were reported from 20 of 26 provinces in DRC (*2*). Outbreaks have been most persistent in the eastern part of DRC in the Great Lakes region, along the Albertine Rift (*2*–*4*).

As reported elsewhere (*3*), strains appear to have been initially introduced into this area as part of what has been characterized as the T5 introduction (1970–1972) under the first wave of P7ET. In 1992, as part of the third wave of P7ET, the disease was reintroduced by a strain from southern Asia, in what has been designated as the T10 introduction event (*3*). Subsequent studies have documented persistence of T10 strains in this region, and ongoing cholera outbreaks in the Great Lakes region and spread of strains from this area suggest establishment of a regional focus of endemic disease derived from the T10 introduction (*6*).

Bacteriophages (phages) and their host bacteria follow predator–prey dynamics that drive co-evolution, resulting in the long-term persistence of both within ecosystems (*7*). Phage predation has been linked with seasonal patterns of cholera emergence and with clinical response to infection in humans (*8*–*11*). Phages are generally highly specific to their host; thus, both phage and susceptible host must maintain a dynamic equilibrium to coexist. Recently, it has been shown that mobile genetic elements associated with sulfamethoxazole/trimethoprim (SXT) antimicrobial resistance, designated as SXT integrative conjugative

Author affiliations: University of Florida Emerging Pathogens Institute, Gainesville, Florida, USA (M.T. Alam, C. Mavian, T.K. Paisie, M.S. Tagliamonte, M.N. Cash, F.M. Maisha, M. Salemi, J.G. Morris, Jr., A. Ali); University of Florida College of Medicine, Gainesville (C. Mavian, T.K. Paisie, M.S. Tagliamonte, M.N. Cash, M. Salemi, J.G. Morris, Jr.); University of California, Berkeley, California, USA (A. Angermeyer, K.D. Seed); Chan Zuckerberg Biohub, San Francisco, California, USA (K.D. Seed); Tufts University School of Medicine, Boston, Massachusetts, USA (A. Camilli); Appui Medical Integre aux Activites de Laboratoire (AMI-LABO), Goma, Democratic Republic of the Congo (R.K.K. Senga); University of Goma, Goma (R.K.K. Senga); University of Florida College of Public Health and Health Professions, Gainesville (M.T. Alam, A. Ali)

DOI: https://doi.org/10.3201/eid2812.220572

[1]Previously presented at Epidemics—8th International Conference on Infectious Diseases Dynamics [online], November 30–December 3, 2021.

[2]These authors contributed equally to this article.

elements (ICEs), can determine phage resistance in *V. cholerae* (*12*). Furthermore, another study has demonstrated that susceptibility to phage killing of marine *V. lentus* was mediated by as many as 6–12 mobile genetic elements (*13*). Taken together, these recent studies support the concept that phage/host in situ interplay has a major role in adaptation and evolution.

Using microbiologic, phylogenomic, and molecular clock analyses, we investigated endemic cholera in the DRC Great Lakes regional hotspot. We also explored the genetic resistance of these *V. cholerae* strains to a novel ICP1 (Bangladesh cholera phage 1) *V. cholerae* phage isolated in cholera patients in the region and genetically distinct from previous ICP1 phages detected in Asia (*14*,*15*).

Methodsphage/

Isolation and Characterization of Toxigenic *V. cholerae* O1 and Virulent Phages

In an initial study involving the isolation and characterization of toxigenic *V. cholerae* O1 strains, we collected fecal samples from suspected cholera patients admitted to cholera treatment centers around Goma, DRC, during 2015–2017 (Table). After collection, we brought the samples to the Laboratoire Provincial de Sante Publique du Nord-Kivu in Goma for microbiological and serologic analysis. We isolated bacteria and confirmed species using methods described elsewhere (*16*), then stored strains in soft Luria-Bertani (LB) Miller agar (0.7% agar) and sent them to the Emerging Pathogens Institute at the University of Florida (Gainesville, FL, USA) for sequencing.

In a second study, we tried to isolate phages preying on *V. cholerae* O1 strains from fecal samples obtained in 2016–2017 from 41 additional cholera patients. We centrifuged cholera rice-water fecal samples at 5,000 × *g* for 10 minutes and filtered resultant supernatant through a 0.22-μm syringe filter, stored them at 4°C in a sterile microfuge tube, and sent them to the Emerging Pathogens Institute for analysis. To identify virulent phages, we tested each filtered fecal sample using standard plaque assay against *V. cholerae* O1 AGC-15, a strain we randomly selected

Table. Characteristics of toxigenic *Vibrio cholerae* O1 strains isolated from the Democratic Republic of the Congo, 2015–2017*

Strain	Isolation date	Province/location	Serotype		Susceptibility of *V. cholerae* to ICP1_2017_A_DRC†	Mutation in O1 antigen and other genes‡	SRA ID
			Ogawa	Inaba			
AGC-1	2015 Apr 30	North Kivu/Kirotshe	–	+	S	–	SRR15192533
AGC-2	2015 May 18	Goma/Buhimba	–	+	S	–	SRR15192532
AGC-3	2015 May 20	Mutwanga	–	+	R	*rfbD*	SRR15192521
AGC-4	2015 Mar 07	Goma/Buhimba	–	+	R	*rfbN*	SRR15192516
AGC-5	2015 Mar 20	Goma/Buhimba	–	+	S	–	SRR15192515
AGC-6	2015 Jul 26	Goma/Buhimba	–	+	R	*rfbV*, VC0559 (hypothetical), *rplE*, *phrA*, *fliD*, VC0672 (hypothetical)	SRR15192514
AGC-7	2015 Jun 06	Goma/Buhimba	–	+	S	–	SRR15192513
AGC-8	2015 Aug 06	Goma/Buhimba	–	+	S	–	SRR15192512
AGC-9	2016 Jun 20	Maniema/Kabambare	+	–	S	–	SRR15192511
AGC-10	2016 Aug 09	Karisimbi/Hop Millitaire	–	+	R	*rfbD*	SRR15192510
AGC-11	2016 May 28	Alimbongo	–	+	R	*rfbD*	SRR15192531
AGC-12	2016 Jul 27	South Kivu/Fizi	+	–	S	–	SRR15192530
AGC-13	2016 Aug 08	Maniema/Kimbilulenge	+	–	S	–	SRR15192529
AGC-14	2017 May 18	Kirotshe/Rubaya	–	+	S	–	SRR15192528
AGC-15	2017 May 31	Rutshuru/Hgr	–	+	S	–	SRR15192527
AGC-16	2017 Jun 10	Rutshuru/Hgr	–	+	S	–	SRR15192526
AGC-17	2017 Jul 01	Nyiragongo/Turunga	–	+	S	–	SRR15192525
AGC-18	2017 Jul 03	Goma/Hop.Provincial	–	+	S§	*manA*	SRR15192524
AGC-19	2017 Jul 03	Goma/Hop.Provincial	–	+	S	–	SRR15192523
AGC-20	2019 Jul 03	Goma/Hop.Provincial	–	+	S	–	SRR15192522
AGC-21	2017 Jul 06	Karisimbi/Prison centrale	–	+	S	–	SRR15192520
AGC-22	2017 Jul 14	Karisimbi/Majengo	–	+	S§	*manA*	SRR15192519
AGC-23	2017 Jul 19	Karisimbi/Majengo	–	+	R	*rfbB*	SRR15192518
AGC-24	2017 Jul 15	Karisimbi/Majengo	–	+	S	*rfbU*	SRR15192517

*R, resistant; S, susceptible; +, positive; –, negative
†Susceptibility to a virulent ICP1 phage (ICP1_2017_A_DRC) determined by strains yielding either complete resistance or forming turbid plaques in response to phage infection in plaque assay. The penultimate column indicates which strains had mutations in the O1-antigen biosynthetic complex and in other genes in the chromosome, with the mutated gene designated. AGC-18, AGC-22, and AGC-24 sustained 1, 1, and 18 bp deletion mutations in the indicated gene(s), resulting in a frame shift mutation in that gene, but all other ICP1 phage-resistant isolates sustained ≥1 missense mutation in the O-antigen biosynthetic gene cluster.
‡As detected by analysis using single-nucleotide polymorphism, insertion/deletion, or both.
§Plaques were turbid as described elsewhere (*29*).

from the DRC isolates from the first part of the study (Table). AGC-15 has the wild-type *ompU* sequence, which encodes the receptor for ICP2; it also has the wild-type O1-antigen biosynthetic genetic region that serves as the receptor for ICP1 and ICP3 (*14*) and lacks any PLE elements mediating immunity to ICP1 (*17*). For phage purification, we picked a single clear plaque using a Pasteur pipette into 1 mL of LB broth and incubated it overnight at 4°C to enable the phage to diffuse out of the soft agar. We made high-titer stocks of purified phage by infecting AGC-15 with phage in LB broth culture.

Whole-Genome Mapping and High-Quality Single-Nucleotide Polymorphism Calling

We performed whole-genome sequencing on the 24 *V. cholerae* O1 isolates from the first part of the study with the Illumina MiSeq for 500 cycles (Qui); we further conducted high-quality single-nucleotide polymorphism (hqSNP) calling (Appendix, https://wwwnc.cdc.gov/EID/article/28/12/220572-App1.pdf). The final genomewide hqSNP alignment included 120 T10 sublineage *V. cholerae* genome sequences: 24 strains collected as part of our study (Table); 71 from publicly available genomes from outbreaks in eastern DRC during 2014–2016 (*6*); 6 archival and publicly available DRC genomes collected during 2001–2013; 17 genomes collected across Africa during 1998–2014; and 2 publicly available genomes from India, ancestors of T10 sublineage (*3*) (Appendix Table 1). We performed multilocus sequence typing analysis using the online tool PubMLST (K. Jolley, unpub data, https://doi.org/10.12688/wellcomeopenres.14826.1) (Appendix Table 2).

Phylogeography

All datasets used in this study passed phylogenetic quality checks (Appendix Figure 1). To explore the origins of strains in the eastern portion of DRC and neighboring countries we used the Bayesian phylogeographic coalescent-based method implemented in BEAST version 1.10.4 software (*18*–*20*). The reconstruction of *V. cholerae* O1 spatiotemporal spread from different locations through Bayesian phylogeography requires calibration of a molecular clock. We estimated evolutionary rates implementing a Hasegawa-Kishino-Yano nucleotide substitution model (*21*) with empirical base frequencies, gamma distribution of site-specific rate heterogeneity, and ascertainment bias correction (*22*), testing a constant demographic prior against nonparametric demographic models, Gaussian Markov random field Skyride (*23*) and Bayesian Skyline plot (*24*), to rule out spurious changes in effective population size inferred by a nonparametric model, which would, in turn, effect timing of divergence events (*25*). We obtained the weighted average of synonymous (*dS*) and nonsynonymous substitution rates (*dN*) in the protein-coding regions of the *V. cholerae* O1 genome for all internal and external branches from a subset of 200 Bayesian maximum credibility clade (MCC) trees randomly obtained from the posterior distribution of trees, as described elsewhere (*26*,*27*).

Whole genome sequencing, genome assembly and annotation of DRC phages

We sequenced 8 plaque-purified phages isolated from 8 independent patient fecal samples with Illumina MiSeq for 50 cycles. We obtained >200-fold coverage that helped with de novo assembly of each phage genome into 1 complete contig using CLC Genomics Workbench (QIAGEN, https://www.qiagen.com). We manually confirmed and corrected low-coverage or problem areas as needed to ensure authentic genome assembly. We annotated phage genomes as described elsewhere (*15*) and deposited sequences into the National Center for Biotechnology Information Sequence Read Archive (BioProject identification no. PRJNA748018; Appendix Table 3).

Results

Of the 24 toxigenic *V. cholerae* O1 strains isolated from fecal samples from cholera patients attending cholera treatment centers during 2015–2017 in the Goma region, 21 (87.5%) were serotype Inaba and 3 (12.5%) serotype Ogawa (Table). All strains in wave 3 were *ctxB* genotype-I and within the T10 introductory clade (*3*,*6*). Consistent with findings published elsewhere (*3*), our MCC tree (Figure 1; Appendix Figure 2) indicated a mean time for the most recent common ancestor (tMRCA) of the T10 *V. cholerae* sublineage introduced to Africa of March 1994 (95% highest posterior density [HPD] September 1991–February 1996). In addition, our analysis showed that subsequent independent introductions (spillover events denoted by asterisks in Figure 1) in the DRC Great Lakes region likely occurred from Rwanda. The first spillover, in May 2001 (95% HPD September 1999–June 2001), is represented by a DRC isolate, ERR1878097_CD_2003, that branches out of a lineage circulating in Rwanda (Figure 1). The other event resulted in 2 major monophyletic clades that match multilocus sequence types reported elsewhere (*6*).

The tMRCA of the first major lineage (denoted as I in Figure 1), February 2009 (95% HPD November 2005–September 2011), corresponds to the sequence

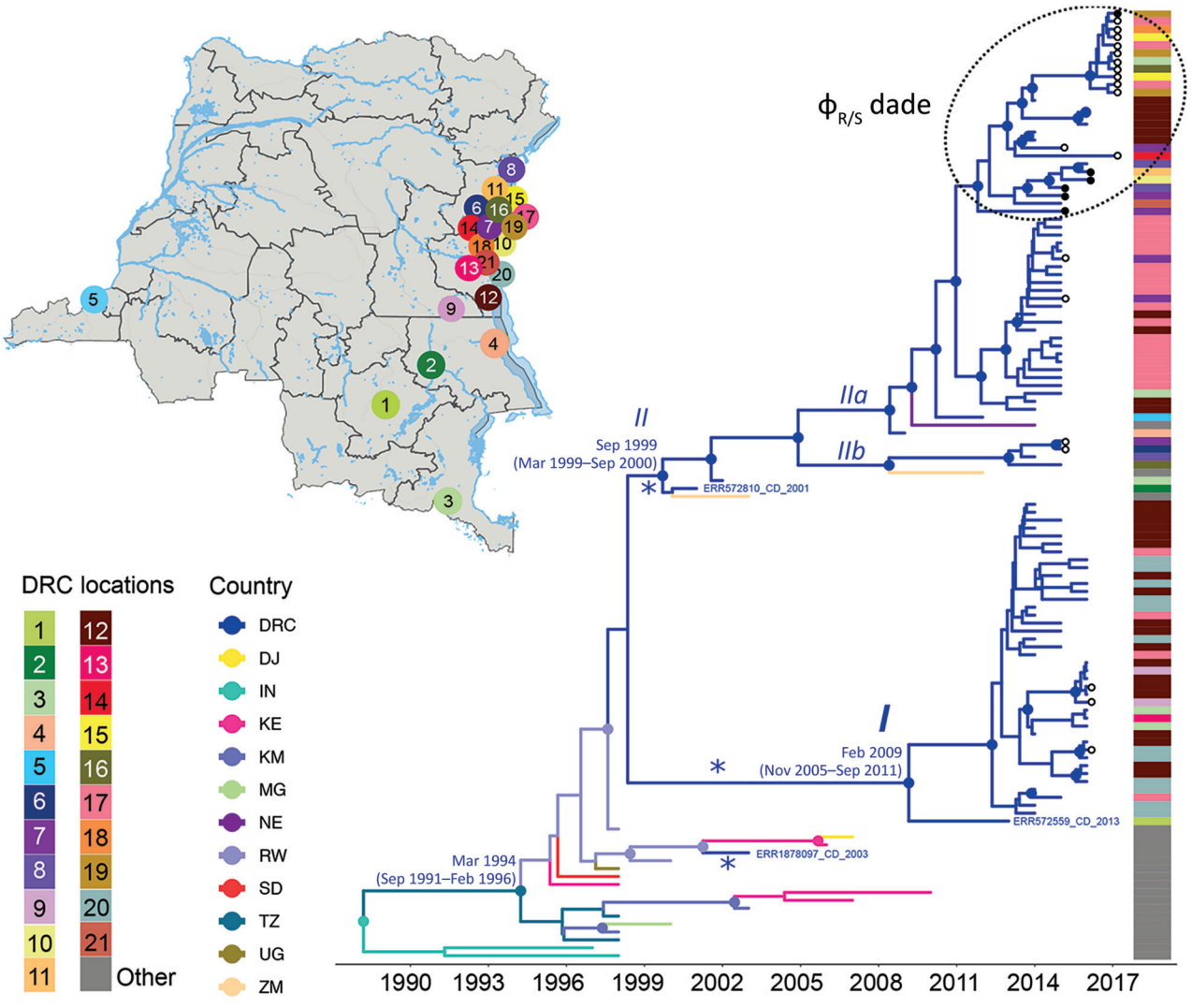


Figure 1. Spatiotemporal evolution and dissemination of *Vibrio cholerae* epidemic in the Democratic Republic of the Congo, 2015–2017. Sampling locations of *V. cholerae* strains sequenced in this study are indicated on the map. Each sampling location is coded by color and number, defined in the key; location colors are indicated for each tip in the maximum clade credibility tree as a heatmap (exact locations in Appendix Table 2, https://wwwnc.cdc.gov/EID/article/28/12/22-0572-App1.pdf). The tree was inferred from full genome *V. cholerae* isolates from DRC, neighboring countries, and Asia. Branches are scaled in time and colored by country of origin. Circles in internal nodes indicate posterior probability support >0.9, and the colors indicate ancestral countries inferred by Bayesian phylogeographic reconstruction. Circles at tips indicates the strains collected and sequenced in this study, with black circles designating phage-resistant strains. Notations *I*, *II*, *IIa*, and *IIb* indicate well-supported lineages and sublineages circulating in DRC during outbreaks. Asterisks (*) indicate potential spillover events within the Great Lakes region originated from neighboring countries. The tree with full tip labels is provided in Appendix Figure 2.

type 69 cluster (*6*) identified during the first reported outbreaks of cholera in DRC during 2008 and 2009. This cluster included the only Ogawa serotype isolates present in our collection. However, the long branch separating the isolate from the Tanganyika province (ERR572559_CD_2013) strain at the base of the monophyletic clade raises the possibility of unsampled *V. cholerae* strains (either from Rwanda or other neighboring countries) that could constitute missing linkage between this DRC lineage and its actual ancestor. The second major lineage (denoted as II in Figure 1) contains all of the Inaba serotype strains collected. According to the molecular clock calibration, lineage II tMRCA dates to September 1999 (95% HPD March 1999–September 2000). The monophyletic clade also includes 3 strains identified in Zambia and Niger, which were likely the result of spillover events from DRC. This clade further divides into 2

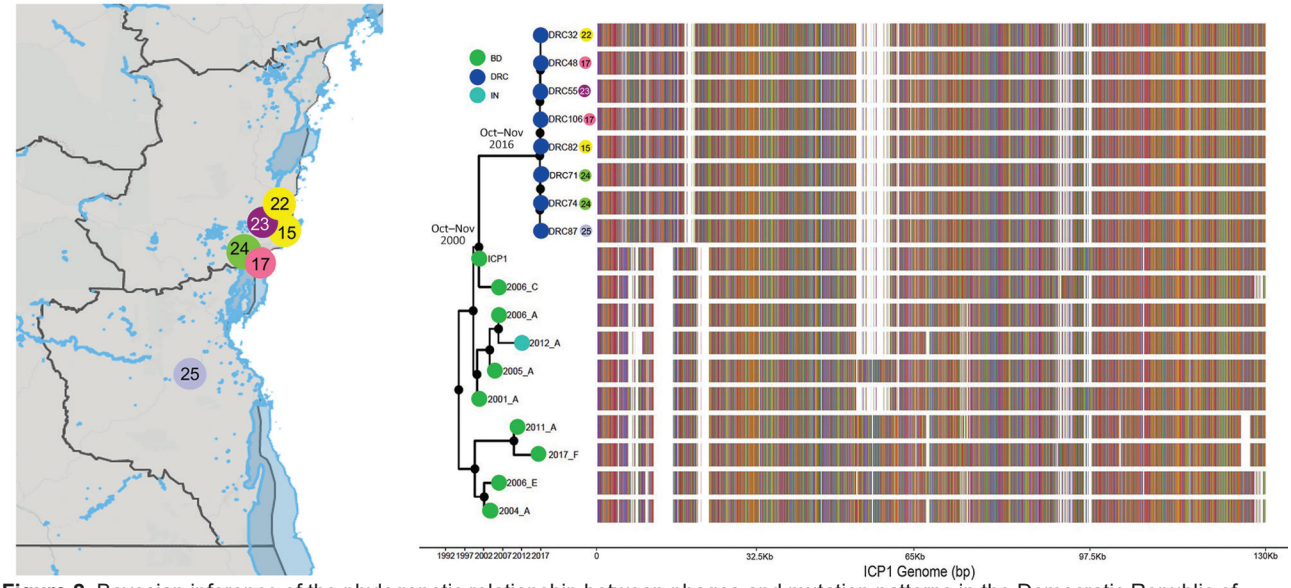


Figure 2. Bayesian inference of the phylogenetic relationship between phages and mutation patterns in the Democratic Republic of the Congo and ICP1 (Bangladesh cholera phage 1) patterns from Asia. Sampling locations of ICP1 strains from DRC are shown on the map. Each sampling location is coded by color and number, also indicated at the tip of the maximum clade credibility tree (exact locations in Appendix Table 2, https://wwwnc.cdc.gov/EID/article/28/12/22-0572-App1.pdf). The tree branches are scaled in time, and the circle tip points are colored by the location of origin, as indicated in the key. Circles in internal node indicate posterior probability support >0.9. To the right of the MCC tree, the genomic composition of each isolate is displayed: red, adenine; green, cytosine; yellow, guanine; and blue, thymine. White spaces indicate gaps at that location in the genome.

sublineages (IIa and IIb in Figure 1) that diverged in November 2004 (95% HPD September 2002–March 2007). Overall, molecular clock and phylogeographic reconstruction suggests circulation of cholera lineages in DRC years before the first reported cholera outbreaks in 2008–2009.

Phages were isolated from 17/41 (41.5%) fecal samples screened, on the basis of formation of plaques on strain AGC_15_CD_2017. Whole-genome sequencing of a subset (n = 8) of these phages showed that they shared high similarity in sequence (hqSNPs = 114) and diverged substantially (hqSNPs = 8,441) from the ICP1 phage isolated from Bangladesh and India (Figure 2). Previously, a total of 185 core open reading frames were identified as being conserved in ICP1 isolates collected over a 12-year period in Bangladesh and India (*15*). Uniquely, the DRC ICP1 lacks 15 of these core open reading frames and also has 10.6 kb of novel sequence in the first third of the genome. Of particular note, genomes do have the anti-*Vch*ind5 factor OrbA (*12*). However, like most ICP1-encoded gene products, most genes unique to DRC ICP1 are classified as hypothetical proteins because of a lack of an informative BLAST identification.

We screened our 24 DRC *V. cholerae* O1 strains for susceptibility to DRC ICP1 by plaque assays using ICP1_2017_A_DRC as a reference phage. Eighteen (75%) of the 24 *V. cholerae* strains were susceptible to ICP1_2017_A_DRC (Table); we observed that 2 strains had turbid plaques on plaque assay, as described elsewhere (*28*). At the genome level, all resistant strains, and 3 of 6 sensitive strains, including the 2 strains that produced turbid plaques, carried ≥1 mutation in genes that belong to the O1-antigen biosynthetic gene cluster. ICP1 uses the O1 antigen as its receptor, and *V. cholerae* is known to undergo phase variation to decrease or produce modified forms of the O1 antigen to evade ICP1 infection (*7*). However, the fact that resistant strains gave a positive serologic response when tested for the O1 antigen suggests that there are mechanisms for resistance to ICP1 that lie elsewhere in the genome.

SXT-ICE has been reported to have 5 hotspots within the accessory region, including hotspot 5 (*Vch*Ind5), which confers resistance to ICP1 phage infection (*12*). When we evaluated the SXT-ICE sequence in all 24 DRC *V. cholerae* genomes, we found that all harbored genes identical to the wild-type SXT-ICE, which should make the strains phage-resistant. However, as already noted, the DRC ICP1 phage that we identified encodes the anti-BREX factor OrbA, which protects against the host *Vch*Ind5 (*12*). We did not detect other mechanisms usually associated with resistance of host cells to phage, such as acquisition and expression of a family of phage-inducible chromosomal island–like elements (*17*), and none of the DRC ICP1 isolates encoded a

previously described CRISPR-cas system specifically targeting phage-inducible chromosomal island–like elements for destruction, which might enable the phage to evade host immunity (*29*). At this point we cannot comment further on the mechanisms underlying resistance of our DRC *V. cholerae* strains to the regional DRC ICP1 phage other than to note the complexity of these regional phage/host interactions.

Comparison of genome-wide weighted averages of *dS* and *dN* along the internal branches of the cholera phylogeny showed a *dN*/*dS* ratio significantly >1 ($p<0.001$) (Appendix Figure 3, panel A). Moreover, the difference between *dN* and *dS* divergence accumulating over time along the internal branches of the phylogeny also appears to be increasing (Appendix Figure 3, panel B). In other words, the mixed presence of susceptible and resistant *V. cholerae* phenotypes, at least in the DRC Goma region where samples were collected, together with *dN*/*dS* patterns suggest that *V. cholerae* has been evolving under pressure of increasing diversifying selection, possibly driven by the co-circulation of predatory phages. Indeed, the map of sampling locations shows that phage-resistant or phage-susceptible *V. cholerae* strains, as well as independently sampled phages, have tended to co-circulate in the DRC Goma region and surrounding locales (Figure 3).

To examine in more detail the adaptive fitness landscape that might confer either resistance or sensitivity to phage predation, we optimized an MCC tree for the subset of *V. cholerae* sequences including all strains in the $\phi_{R/S}$ clade, as well as 2 outgroup strains, AGC-2-CD-2015 and AGC-8-CD-2015, that clustered outside the clade (Figure 4). We used a Bayesian phylogeographic model with phage resistance or susceptibility as discrete phenotypic characters to infer the most likely phenotype of the ancestral (internal) nodes of the tree. The analysis clearly shows that the backbone path (trunk) of the $\phi_{R/S}$ clade, which represents the surviving lineage successfully propagating through time (*28*), is dominated by isolates with the phage-sensitive phenotype and connects phage-sensitive ancestral sequences that first generated a subcluster of strains circulating in 2015–2016 and then a subcluster including 2017 strains.

We cannot say which mutations are responsible for acquisition of phage resistance, but mutations in genes belonging to the O1-antigen biosynthetic gene cluster appear to have emerged, independently, along 3 distinct evolutionary lineages. The first lineage, leading to strain AGC-6-2015-DRC sampled in 2015, is characterized by amino acid substitutions in the *rfbV*, *rpIE*, *phrA*, and *fliD* genes. The second lineage resulted in a monophyletic clade of phage-resistant strains with mutations in either *rfbN* (strains sampled in 2015) or *rfbD* (sampled in 2016). The third lineage, leading to strain AGC-23-2017-DRC (sampled in 2017), was characterized again by an amino acid substitution in the *rfbB* gene.

Discussion

Cholera continues to be a major public health problem in the Great Lakes region of Africa (*1*–*4*). To optimize

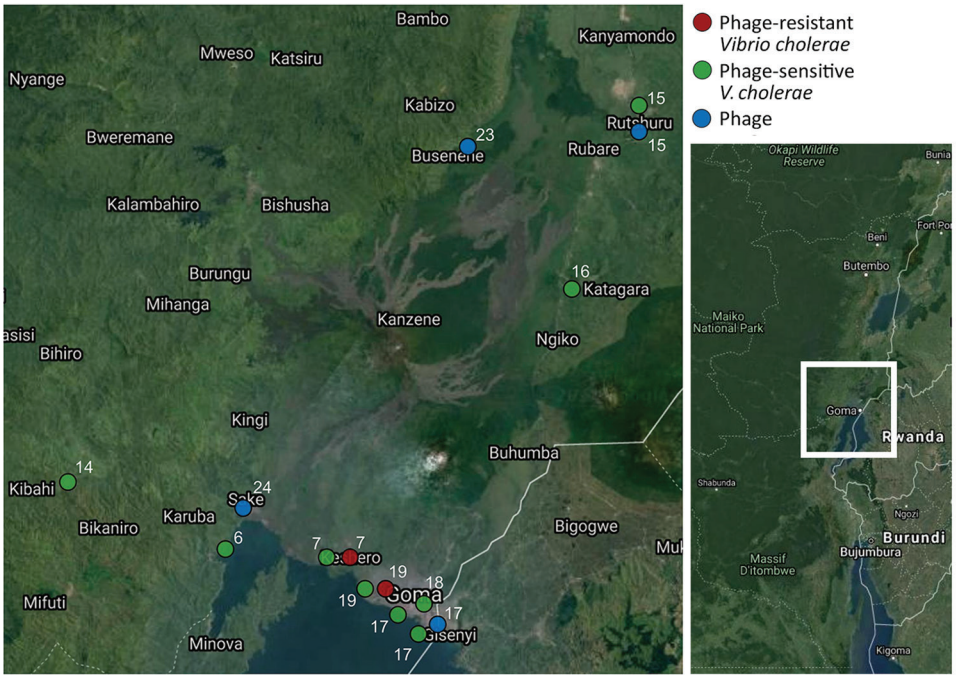


Figure 3. Sampling locations of phages and phage-resistant or sensitive *Vibrio cholerae* isolates in the Democratic Republic of the Congo. Each sampling location is coded by color (key) and number, which also appear at the tips of the maximum clade credibility tree in Figure 4 for comparison. Inset shows location of sampling area in the Democratic Republic of the Congo.

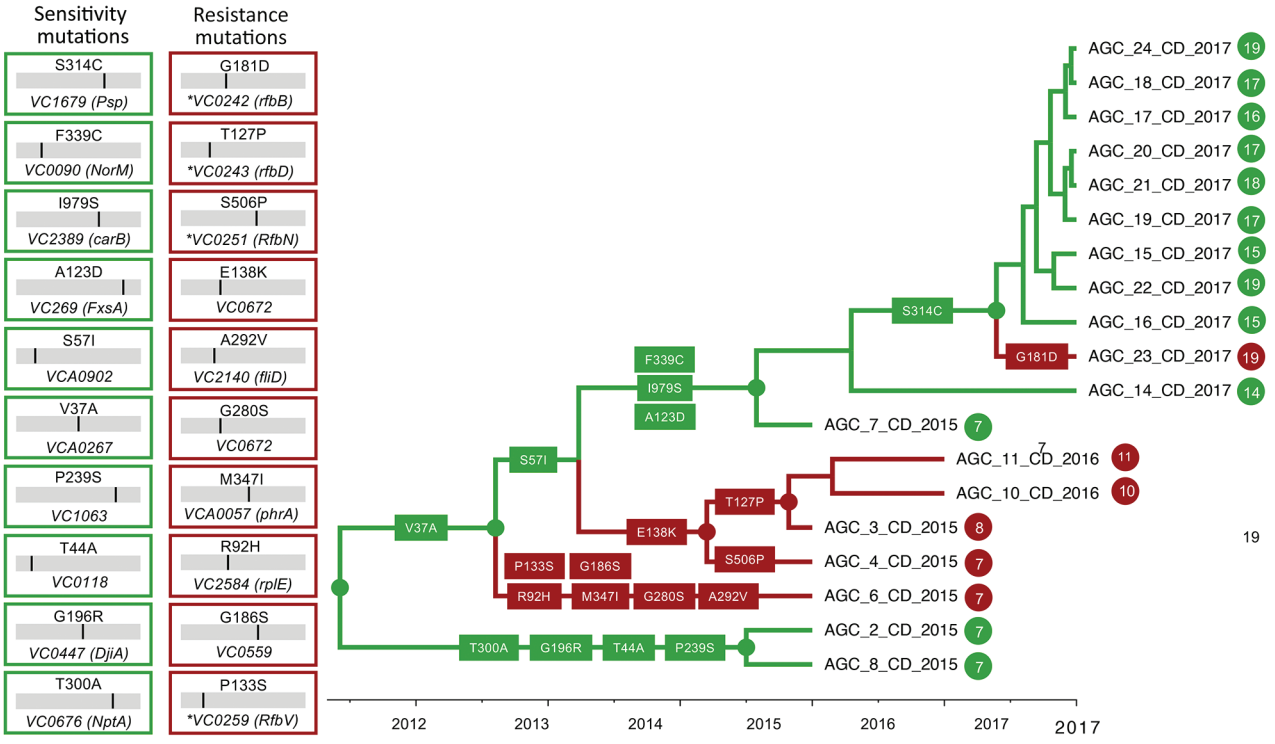


Figure 4. Phage-resistant or phase-sensitive dynamics and mutational patterns of *Vibrio cholerae* isolates in the Democratic Republic of the Congo. The boxes show the mutations that have been found along the backbone, internal, or external branches of the maximum clade credibility tree. Each number is the amino acid position of the protein where the mutation was mapped in the tree. The branches are scaled in time and colored on the basis of resistance (red) or sensitivity (green), matching colors in Figure 3. Circles and colors in internal node indicate posterior probability support >0.9 for an ancestor to be resistant or sensitive.

cholera control and appropriately target public health interventions, evolutionary drivers for *V. cholerae* in this area need to be determined, including reasons why certain *V. cholerae* strains emerge and persist while others fail to propagate. Our data provide further information on sources and subsequent development of endemic *V. cholerae* O1 in eastern DRC. Our work also highlights the effects a novel regional bacteriophage can have on cholera evolution. As reflected in the trunk of the phylogeny of the $\phi_{R/S}$ clade, *V. cholerae* isolates displaying the phage-sensitive phenotype appear to be successfully propagating, with every branch that leads to phage-resistant phenotypes in the phylogeny eventually dying out. Our findings are somewhat counterintuitive: phage resistance, rather than encouraging expansion of the epidemic clone, led to evolutionary dead ends; however, our data highlight the ability of *V. cholerae* to explore and quickly abandon different evolutionary pathways during epidemic spread. This finding is not surprising considering the potential fitness cost of phage resistance, particularly if resistance results in mutants highly attenuated for virulence (*30*,*31*). Further work will be needed to determine the exact mechanism by which the *V. cholerae* strains isolated in this study were either completely or partially resistant (turbid plaque) to ICP1_2017_A_DRC and whether the phage can mutate to regain virulence (*12*).

In summary, our study documents a complex coevolutionary dynamic involving *V. cholerae* and predatory phages (*30*) in the Great Lakes region of DRC. Phage-sensitive and highly infectious strains co-circulate with phage-resistant ones that occasionally emerge and eventually die out along different evolutionary pathways in response to the presence or absence of predatory phages in the environment, although the main phage-sensitive evolutionary lineage continued to propagate over time. The ability of *V. cholerae* to explore multiple mutational pathways in different genes and achieve phage resistance provides a substantial evolutionary advantage in terms of quick adaptive response to a changing environment, leading to emergence of new strains. Continuous monitoring of toxigenic *V. cholerae* and predator ICP1 phages in both patient fecal samples and aquatic environments in DRC and elsewhere could provide invaluable epidemiologic data for monitoring the spread of cholera, identifying environmental actors driving successful dissemination, and assessing the potential for new outbreaks.

R scripts and XML files are available from the authors upon request.

This work was supported by a National Institutes of Health/National Institute of Allergies and Infectious Disease–sponsored grant (R01 AI138554 to J.G.M., Jr.)

A. Ali, A.C., C.M., J.G.M., and M.S. conceived and designed the experiments; A. Angermeyer, A.M., C.M., K.K.S., M.N.C., and T.K.P. performed the experiments; A.M., C.M., F.M.M., K.D.S., M.S.T., and T.K.P. analyzed the data; A. Ali, A.C., F.M.M., J.G.M., and R.K.K.S. contributed materials/analysis tools; A. Ali, A.C., C.M., J.G.M., and M.S. wrote the paper.

About the Author

Dr. Alam is a biological scientist in the University of Florida College of Public Health and Health Professions and the Emerging Pathogens Institute. His research focuses on the factors and processes promoting persistence of *Vibrio cholerae* in aquatic reservoirs and evolution resulting from the lytic bacteriophage–*V. cholerae* in both clinical and environmental settings. Dr. Mavian is on the faculty at the University of Florida College of Medicine in the Department of Pathology, Immunology, and Laboratory Medicine. Her research focuses on intrahost evolution of HIV, population dynamics of cholera, and evolution and global spread of arboviruses and, since the beginning of the pandemic, the dynamics of SARS-CoV-2.

References

1. World Health Organization. Cholera annual report 2020. Wkly Epidemiol Rec. 2021;96:445–60.
2. 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
3. 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
4. Okeke IN. Africa in the time of cholera: a history of pandemics from 1817 to the present [book review]. Emerg Infect Dis. 2012;18:362. https://doi.org/10.3201/eid1802.111535
5. Moore S, Miwanda B, Sadji AY, Thefenne H, Jeddi F, Rebaudet S, et al. Relationship between distinct African cholera epidemics revealed via MLVA haplotyping of 337 *Vibrio cholerae* isolates. PLoS Negl Trop Dis. 2015;9:e0003817–0003817. https://doi.org/10.1371/journal.pntd.0003817
6. 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
7. Seed KD. Battling phages: how bacteria defend against viral attack. PLoS Pathog. 2015;11:e1004847. https://doi.org/10.1371/journal.ppat.1004847
8. Faruque SM, Naser IB, Islam MJ, Faruque AS, Ghosh AN, Nair GB, et al. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proc Natl Acad Sci U S A. 2005;102:1702–7. https://doi.org/10.1073/pnas.0408992102
9. Huq A, Sack RB, Nizam A, Longini IM, Nair GB, Ali A, et al. Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. Appl Environ Microbiol. 2005;71:4645–54. https://doi.org/10.1128/AEM.71.8.4645-4654.2005
10. Silva-Valenzuela CA, Camilli A. Niche adaptation limits bacteriophage predation of *Vibrio cholerae* in a nutrient-poor aquatic environment. Proc Natl Acad Sci U S A. 2019;116:1627–32. https://doi.org/10.1073/pnas.1810138116
11. Seed KD, Yen M, Shapiro BJ, Hilaire IJ, Charles RC, Teng JE, et al. Evolutionary consequences of intra-patient phage predation on microbial populations. eLife. 2014;3:e03497. https://doi.org/10.7554/eLife.03497
12. LeGault KN, Hays SG, Angermeyer A, McKitterick AC, Johura FT, Sultana M, et al. Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. Science. 2021;373:eabg2166. https://doi.org/10.1126/science.abg2166
13. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, et al. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. Science. 2021;374:488–92. 10.1126/science.abb1083 https://doi.org/10.1126/science.abb1083
14. Seed KD, Bodi KL, Kropinski AM, Ackermann HW, Calderwood SB, Qadri F, et al. Evidence of a dominant lineage of *Vibrio cholerae*–specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. MBio. 2011;2:e00334–10. https://doi.org/10.1128/mBio.00334-10
15. Angermeyer A, Das MM, Singh DV, Seed KD. Analysis of 19 highly conserved *Vibrio cholerae* bacteriophages isolated from environmental and patient sources over a twelve-year period. Viruses. 2018;10:10. https://doi.org/10.3390/v10060299
16. Ali A, Chen Y, Johnson JA, Redden E, Mayette Y, Rashid MH, et al. Recent clonal origin of cholera in Haiti. Emerg Infect Dis. 2011;17:699–701. https://doi.org/10.3201/eid1704.101973
17. O'Hara BJ, Barth ZK, McKitterick AC, Seed KD. A highly specific phage defense system is a conserved feature of the *Vibrio cholerae* mobilome. PLoS Genet. 2017;13:e1006838. https://doi.org/10.1371/journal.pgen.1006838
18. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLOS Comput Biol. 2009;5:e1000520. https://doi.org/10.1371/journal.pcbi.1000520
19. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303:327–32. https://doi.org/10.1126/science.1090727
20. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214. https://doi.org/10.1186/1471-2148-7-214
21. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22:160–74. https://doi.org/10.1007/BF02101694
22. Leaché AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. Short tree, long tree, right tree, wrong tree:

new acquisition bias corrections for inferring SNP phylogenies. Syst Biol. 2015;64:1032–47. https://doi.org/10.1093/sysbio/syv053
23. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008;25:1459–71. https://doi.org/10.1093/molbev/msn090
24. Strimmer K, Pybus OG. Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol Biol Evol. 2001;18:2298–305. https://doi.org/10.1093/oxfordjournals.molbev.a003776
25. Hall MD, Woolhouse ME, Rambaut A. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. Virus Evol. 2016;2:vew003. https://doi.org/10.1093/ve/vew003
26. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLOS Comput Biol. 2007;3:e29. https://doi.org/10.1371/journal.pcbi.0030029
27. Mavian C, Paisie TK, Alam MT, Browne C, Beau De Rochars VM, Nembrini S, et al. Toxigenic *Vibrio cholerae* evolution and establishment of reservoirs in aquatic ecosystems. Proc Natl Acad Sci U S A. 2020;117:7897–904. https://doi.org/10.1073/pnas.1918763117
28. Seed KD, Faruque SM, Mekalanos JJ, Calderwood SB, Qadri F, Camilli A. Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in *Vibrio cholerae* O1. PLoS Pathog. 2012;8:e1002917. https://doi.org/10.1371/journal.ppat.1002917
29. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature. 2013;494:489–91. https://doi.org/10.1038/nature11927
30. Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A. Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. PLoS Pathog. 2013;9:e1003800. https://doi.org/10.1371/journal.ppat.1003800
31. Oechslin F. Resistance development to bacteriophages occurring during bacteriophage therapy. Viruses. 2018;10:10. https://doi.org/10.3390/v10070351

Address for correspondence: Afsar Ali and Carla Mavian, Emerging Pathogens Institute, 2055 Mowry Rd, University of Florida, Gainesville, FL 32601, USA; email: afsarali@epi.ufl.edu and cmavian@ufl.edu

Article DOI: https://doi.org/10.3201/eid2812.220572

Emergence and Evolutionary Response of *Vibrio cholerae* to a Novel Bacteriophage in the Democratic Republic of the Congo

Appendix

Materials and Methods

Isolation and Characterization of Toxigenic *Vibrio cholerae* O1 and Virulent Phages

During 2015–2017, fecal samples from suspected cholera patients admitted in different cholera treatment centers around Goma, Democratic Republic of the Congo (DRC) (Table) were placed on Cary-Blair transport media and brought to the Laboratoire Provincial de Sante Publique du Nord-Kivu in Goma for microbiological and serologic analysis. *V. cholerae* in fecal samples were enriched in alkaline peptone water as described elsewhere (*1*). Following enrichment, a loopful of culture was streaked onto thiosulfate citrate bile salts (TCBS) agar and the plates were incubated overnight at 37°C. Bacterial colonies grown as yellow color on TCBS agar were subcultured onto Luria-Bertani, Miller (LB) agar and the culture plates were incubated overnight at 37°C. To determine serogroup, translucent colonies grown on LB-agar were tested against polyvalent antiserum specific for *V. cholerae* O1 and O139 by slide agglutination tests; each O1 positive strain was further typed for serotype using antiserum specific for Ogawa, Inaba, or Hikojima serotype. The toxigenic *V. cholerae* were stored in soft LB-agar (0.7% agar) and sent to the Emerging Pathogens Institute (EPI) at University of Florida for further analysis.

For isolation of potential virulent phages, cholera rice-water fecal samples collected between 2016 and 2017 were centrifuged at 5,000 x g for 10 min in a microfuge. Fecal samples sused to detect and characterize virulent phages were different than the fecal samples used for detection of 24 *V. cholerae* O1 strains described above. The resultant supernatant was filtered through a 0.22 µm syringe filter, stored at 4°C in a sterile microfuge tube, and sent to EPI for further analysis. For further characterization of potential phages in cholera-confirmed patients' fecal samples, 41 fecal sample filtrates were brought to EPI.

Virulent phage plaque assay: For detection and characterization of potential virulent phages, each filtered fecal sample was tested by plaque assay using a host toxigenic *V. cholera*e O1 Inaba strain, AGC-15 (Appendix Table 1). Isolated in DRC, AGC-15 has a wild-type *ompU* sequence, which encodes the receptor for ICP2. AGC-15 also has the O1-antigen receptor for ICP1 and ICP3 (*2*) and lacks any PLE elements mediating immunity to ICP1. Briefly, a sterile glass tube was inoculated with 100 µl of filtered-sterilized fecal sample (potential source of virulent phage), 9.8 ml of LB-broth, and 100 µl of host *V. cholerae* AGC-15 culture (freshly grown to mid-exponential phase). The culture mixture was incubated overnight at 37°C with aeration to enrich any phages capable of infecting AGC-15. Following incubation, the culture was transferred to a 15 ml conical tube and centrifuged for 10 min at 5,000 x g at 4°C. To eliminate residual bacterial cells, the supernatant was filtered through a 0.22 µm syringe filter and the filtrate was stored at 4°C. The filtered supernatant was serially diluted (10-fold) in LB-broth to reach a dilution of 10^{-5} in a 96-well microtiter plate. One hundred µl of AGC-15 culture (grown to a mid-exponential phase) were added to the undiluted and each serially diluted filtrate in the microtiter plate and was incubated at room temperature for 10 min to enable phage adsorption. All 190 µl of each mixture from the microtiter plate was transferred to wells of six-well tissue culture plates.

Three ml of soft LB-agar (0.35% agar) kept at 55°C in a water bath were added to each well of the six-well plate and the plate was gently swirled to mix the bacteria and phage evenly. To allow the agar to solidify, the plate was left for 30 min at room temperature followed by incubation at 37°C for 3–4 hours. After incubation, the plate was visually observed for plaque formation. If no plaques were observed after 4 hours of incubation, the plate was incubated overnight at room temperature to determine if any plaques were formed following extended incubation time. For virulent phage purification, a single clear plaque was picked using a Pasteur pipette into 1 ml of LB-broth and incubated overnight at 4°C to allow phage to diffuse out of the soft agar piece. High titer stocks of purified plaques were made by multiplication in broth culture as described above.

Whole genome mapping and hqSNP calling: Toxigenic *V. cholerae* O1 samples collected were confirmed by serology and PCR (*1*). After subculture, gDNA extraction was performed using the Qiagen DNeasy Blood and Tissue kit. Genomic DNA from all isolates were cultured and extracted from a bacterial pellet. Sample library construction was performed using the

Nextera XT DNA Library Preparation Kit (Illumina, https://www.illumina.com). Whole-genome sequencing on all isolates was performed with the Illumina MiSeq for 500 cycles. Adaptor and raw sequence reads were filtered by length and quality by using the program Trimmomatic (*3*). After quality filtering, Bowtie2 (*4*) was used to map the sequence reads to the reference genome, *V. cholerae* O1 str. N16961 (GenBank Accessions: NC_002505.1 and NC_002506.1) (*5*). After reads were mapped to the reference genome, duplicate reads were marked and realigned using Picard (http://broadinstitute.github.io/picard). The reference-based mapping alignment was then verified and fixed accordingly. Freebayes (Garrison MG, unpub data; https://arxiv.org/abs/1207.3907) was used to create a custom genome-wide SNP calling database (dbSNP) from all isolates in the dataset to perform base quality score recalibration (BQSR), as outlined in GATK's best practice guidelines for germline variation (https://software.broadinstitute.org/gatk/). The newly created variant call format (VCF) file obtained from Freebayes was subsequently filtered only for SNPs. The VCF file was used as a dbSNP for BQSR of the reference-based mapped alignment files; alignment files were then recalibrated and variants calling on the newly recalibrated files performed with Freebayes. The newly created VCF file was filtered only for SNPs and normalized using the program BCFtools (http://www.htslib.org/doc/bcftools.html). Normalization simplifies the represented variants in the VCF file by showing as few bases as possible at particular SNP sites in the genome. SNPs were filtered by depth of coverage, quality, and genotype likelihood, as described in Azarian *et al.* (*6*). Finally, the SNP FASTA alignment was extracted by a custom python script from the VCF file. The SNP alignment was filtered site-by-site, leaving sites with only greater than 75% of SNPs at that particular site, making a high-quality SNP (hqSNP) alignment. Our hqSNP alignment was then annotated using the program SnpEff (*7*). The final genome-wide SNP alignment included 120 T10 sublineage *V. cholerae* genome sequences (Appendix Table 1): 24 strains were collected in DRC (eight collected in 2015, five in 2016, 11 in 2017) and sequenced in this study; 71 were publicly available genomes from outbreaks in eastern DRC between 2014 and 2016 (*8*); six archival and publicly available DRC genomes collected between 2001–2013; 17 genomes collected across Africa between 1998 and 2014; and two publicly available genomes from India, ancestor of T10 sublineage (*9*). While the previously available DRC samples spanned 2014–2016 (*10*), we expanded the temporal dimension of the DRC collection as we sequenced mainly strains collected in 2017. The MLST analysis was performed on the online

tool PubMLST (Jolley K, unpub data; https://doi.org/10.12688/wellcomeopenres.14826.1); results are shown in Appendix Table 2.

Phylogenetic and temporal signal: All datasets used in this study passed phylogenetic quality checks such as evaluating the presence of phylogenetic signal, to resolve the phylogenetic relationship among the *V. cholerae* isolates, and temporal signal for a robust calibration of the molecular clock (Appendix Figure 1). We performed likelihood mapping analysis using IQ-TREE (*10*), which enables the report likelihood values of the three possible unrooted trees, inferred using the best-fitting nucleotide substitution model, of each possible quartet (set of four sequences) on an equilateral triangle (likelihood map). In a likelihood map, dots (likelihood values) in the center of the triangle represent phylogenetic noise and simulation have shown that datasets with <35% nose (as it was in our case) can reliably be used for phylogeny inference (*11*). For each dataset, the presence of temporal signal was assessed by calculating the tree root-to-tip divergence regression plot with TempEst v1.5 (http://tree.bio.ed.ac.uk/software/tempest) (*12*), using maximum-likelihood (ML) phylogenies inferred with IQ-TREE (*10*) and the best-fitting nucleotide substitution model according to Bayesian Information Criterion (BIC), and ultrafast bootstrap (BB) approximation (1000 replicates) to assess robustness of the phylogeny internal branches.

Bayesian Phylogeography of DRC isolates: To test the hypothesis of whether cholera outbreaks in the DRC were caused by endemic *V. cholerae* O1 strains, or strains recently introduced from other African countries surrounding the Great Lakes region, we used the Bayesian phylogeographic (*13*) coalescent-based method (*14*) implemented in the BEAST v1.10.4 (*15*) software package. The reconstruction of *V. cholerae* O1 spatiotemporal spread from different locations through Bayesian phylogeography requires the calibration of a molecular clock. Evolutionary rates were estimated implementing a HKY nucleotide substitution model (*16*) with empirical base frequencies, gamma distribution of site-specific rate heterogeneity, and ascertainment bias correction (*17*), testing a constant demographic prior against non-parametric demographic models – Gaussian Markov randomfield Skyride (BSR) (*18*) and Bayesian Skyline Plot (BSP) (*19*) – to rule out spurious changes in effective population size inferred by a non-parametric model that would in turn effect timing of divergence events (*20*). Additionally, for each demographic model, we compared a strict and relaxed uncorrelated (lognormal distribution among branches) molecular clock (*21*). The best fitting molecular clock and demographic model

were chosen by estimating the marginal likelihood of each model by using path sampling (PS) and stepping-stone (SS) methods, followed by Bayes Factor comparison test (*11*,*15*). A Markov Chain Monte Carlo (MCMC) sampler was run for 500 million generations, sampling every 50,000 generations. Proper mixing of the Markov chain was evaluated by the effective population size (ESS) of each parameter estimate under a specific model. ESS values >200 for all parameter estimates are considered as evidence of proper mixing in the analysis. The sampling location for each isolate was used as a discreet trait to reconstruct likely locations of ancestral sequences (internal nodes in the tree) and infer migration events (bacterial flow) that took place in the DRC and throughout the Great Lakes Region. Phylogeographic analysis was performed with the BEAST package v1.10.4 (*15*). Transitions between discrete states (location of where the isolate was collected) were estimated using the continuous-time Markov chain model operating the asymmetric migration model with Bayesian Stochastic Search Variable Selection (*13*). In our reconstruction of ancestral states, we assume migration occurs along branches connecting the tree nodes. The maximum clade credibility (MCC) tree chosen from the posterior distribution of trees using TreeAnnotator v1.10.4 after 10% burn-in. The MCC tree was annotated in R using the package ggtree (*22*) for publishing purposes.

Calculation of weighted average of nonsynonmous (*dN*) and synonymous substitution rates (*dS*), and selection analysis: A codon alignment from the DRC clade was generated to analyze other mutations in the *V. cholerae* genome from the isolates in the phylogeny, and a subset of 200 Bayesian MCC genealogies randomly was obtained from the posterior distribution of trees for each subsampled dataset. The weighted average of synonymous substitution rates (*dS*) and non-synonymous substitution rates (*dN*) in the protein-coding regions of the *V. cholerae* O1 genome for all, internal and external branches were obtained from a subset of 200 Bayesian MCC trees randomly obtained from the posterior distribution of trees, as described by Lemey et al. (*23*). The subset of trees and in-house java scripts were then used to calculate *dN/dS* rates and divergence in the isolates located within the DRC clade and plotted using the package ggplot2 in R.

Whole genome sequencing, genome assembly and annotation of DRC phages

To characterize DRC phages, we plaque-purified phages from eight different patient samples, and prepared high titer stocks of each of these phages. Sample library construction for whole-genome sequencing was performed using the Nextera XT DNA Library Preparation Kit

(Illumina). Whole-genome sequencing was performed with the Illumina MiSeq for 50 cycles. We obtained over 200-fold coverage facilitating de novo assembly of each phage genome into one complete contig using CLC Genomics Workbench (QIAGEN, https://www.qiagen.com). Manual confirmation/correction of low coverage areas and/or problem areas were performed to ensure authentic genome assembly. Annotation of phage genomes was performed as described elsewhere (*24*). Briefly, open reading frames from existing annotated ICP1 phages were compared against the DRC phage genomes using BLASTn to find homologs. Additional new putative open reading frames were discovered with de novo prediction software and added to the annotation. This was performed for each of the eight phages and due to very high similarity between them, one representative was randomly chosen and designated as ICP1_2017_A_DRC. This representative was aligned with existing ICP1 phage genomes using Mauve (*25*) and a maximum-likelihood, bootstrapped phylogenetic tree was generated with PhyML (-s BEST 92–rand_start–n_rand_starts 10 -b 100) (*26*).

References

1. Ali A, Chen Y, Johnson JA, Redden E, Mayette Y, Rashid MH, et al. Recent clonal origin of cholera in Haiti. Emerg Infect Dis. 2011;17:699–701. PubMed https://doi.org/10.3201/eid1704.101973

2. Seed KD, Bodi KL, Kropinski AM, Ackermann HW, Calderwood SB, Qadri F, et al. Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. MBio. 2011;2:e00334–10. PubMed https://doi.org/10.1128/mBio.00334-10

3. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. PubMed https://doi.org/10.1093/bioinformatics/btu170

4. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. PubMed https://doi.org/10.1038/nmeth.1923

5. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae.* Nature. 2000;406:477–83. PubMed https://doi.org/10.1038/35020000

6. Azarian T, Ali A, Johnson JA, Mohr D, Prosperi M, Veras NM, et al. Phylodynamic analysis of clinical and environmental *Vibrio cholerae* isolates from Haiti reveals diversification driven by positive selection. MBio. 2014;5:5. PubMed https://doi.org/10.1128/mBio.01824-14

7. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80–92. PubMed https://doi.org/10.4161/fly.19695

8. 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. PubMed https://doi.org/10.1371/journal.pntd.0007642

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. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. PubMed https://doi.org/10.1093/molbev/msu300

11. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 2002;18:502–4. PubMed https://doi.org/10.1093/bioinformatics/18.3.502

12. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2:vew007. PubMed https://doi.org/10.1093/ve/vew007

13. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLOS Comput Biol. 2009;5:e1000520. PubMed https://doi.org/10.1371/journal.pcbi.1000520

14. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303:327–32. PubMed https://doi.org/10.1126/science.1090727

15. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214. PubMed https://doi.org/10.1186/1471-2148-7-214

16. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22:160–74. PubMed https://doi.org/10.1007/BF02101694

17. Leaché AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst Biol. 2015;64:1032–47. PubMed https://doi.org/10.1093/sysbio/syv053

18. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008;25:1459–71. PubMed https://doi.org/10.1093/molbev/msn090

19. Strimmer K, Pybus OG. Exploring the demographic history of DNA sequences using the generalized skyline plot. Mol Biol Evol. 2001;18:2298–305. PubMed https://doi.org/10.1093/oxfordjournals.molbev.a003776

20. Hall MD, Woolhouse ME, Rambaut A. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. Virus Evol. 2016;2:vew003. PubMed https://doi.org/10.1093/ve/vew003

21. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22:1185–92. PubMed https://doi.org/10.1093/molbev/msi103

22. Yu GC, Smith DK, Zhu HC, Guan Y, Lam TTY. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. https://doi.org/10.1111/2041-210X.12628

23. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLOS Comput Biol. 2007;3:e29. PubMed https://doi.org/10.1371/journal.pcbi.0030029

24. Angermeyer A, Das MM, Singh DV, Seed KD. Analysis of 19 highly conserved *Vibrio cholerae* bacteriophages isolated from environmental and patient sources over a twelve-year period. Viruses. 2018;10:10. PubMed https://doi.org/10.3390/v10060299

25. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5:e11147. PubMed https://doi.org/10.1371/journal.pone.0011147

26. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21. PubMed https://doi.org/10.1093/sysbio/syq010

Appendix Table 1. *Vibrio cholera* strains included in phylogenetic studies in Democratic Republic of the Congo*

V. cholera name	Year	Location, country or sea	Location, region/ locality	Lineage	Latitude	Longitude	Location on map†	SRA ID
AGC_1_CD_2015	2015	DRC	North Kivu/ Kirotshe	ST515	−1.613051	29.03132	6	SRR15192533
AGC_2_CD_2015	2015	DRC	Goma/ Buhimba	ST69	−1.621214	29.156623	7	SRR15192532
AGC_3_CD_2015	2015	DRC	Mutwanga	ST515	0.514939	25.191932	8	SRR15192521
AGC_4_CD_2015	2015	DRC	Goma/ Buhimba	ST515	−1.621214	29.156623	7	SRR15192516
AGC_5_CD_2015	2015	DRC	Goma/ Buhimba	ST515	−1.621214	29.156623	7	SRR15192515
AGC_6_CD_2015	2015	DRC	Goma/ Buhimba	ST515	−1.621214	29.156623	7	SRR15192514
AGC_7_CD_2015	2015	DRC	Goma/ Buhimba	ST515	−1.621214	29.156623	7	SRR15192513
AGC_8_CD_2015	2015	DRC	Goma/ Buhimba	ST515	−1.621214	29.156623	7	SRR15192512
AGC_9_CD_2016	2016	DRC	Maniema/ Kabambare	ST515	−4.400901	27.765835	9	SRR15192511
AGC_10_CD_2016	2016	DRC	Karisimbi/ Hop Millitaire	ST515	1.5064	29.4508	10	SRR15192510
AGC_11_CD_2016	2016	DRC	Alimbongo	ST515	−0.36879	29.156179	11	SRR15192531
AGC_12_CD_2016	2016	DRC	South Kivu/ Fizi	ST515	−4.30058	28.94212	12	SRR15192530
AGC_13_CD_2016	2016	DRC	South Kivu/ Kimbilulenge	ST515	−3.21838	28.25855	13	SRR15192529
AGC_14_CD_2017	2017	DRC	Kirotshe/ Rubaya	ST515	−1.546277	28.873122	14	SRR15192528
AGC_15_CD_2017	2017	DRC	Rutshuru/ Hgr	ST515	−1.188054595	29.4459123	15	SRR15192527
AGC_16_CD_2017	2017	DRC	Rutshuru/ Hgr	ST515	−1.188054595	29.4459123	15	SRR15192526
AGC_17_CD_2017	2017	DRC	Nyiragongo/ Turunga	ST515	−1.3527161	29.37873	16	SRR15192525
AGC_18_CD_2017	2017	DRC	Goma/Hop Provincial	ST515	−1.678865426	29.8	17	SRR15192524
AGC_19_CD_2017	2017	DRC	Goma/Hop Provincial	ST515	−1.678865426	29.8	17	SRR15192523
AGC_20_CD_2017	2017	DRC	Goma/Hop Provincial	ST515	−1.678865426	29.8	17	SRR15192522
AGC_21_CD_2017	2017	DRC	Karisimbi/ Prison centrale	ST69	−1.9	29	18	SRR15192520
AGC_22_CD_2017	2017	DRC	Karisimbi/ Majengo	ST69	−1.65388	29.5	19	SRR15192519
AGC_23_CD_2017	2017	DRC	Karisimbi/ Majengo	ST515	−1.65388	29.5	19	SRR15192518
AGC_24_CD_2017	2017	DRC	Karisimbi/ Majengo	ST515	−1.65388	29.5	19	SRR15192517
ERR019292_KE_2007	2007	Kenya						ERR019292
ERR037738_KE_2010	2010	Kenya						ERR037738
ERR044795_ZM_2003	2003	Zambia						ERR044795
ERR1877642_RW_2000	2000	Rwanda						ERR1877642
ERR1878097_CD_2003	2003	DRC						ERR1878097
ERR1878101_CD_2002	2002	DRC	Congo/ Zaire		−11.64112	27.51818	3	ERR1878101
ERR1878103_KM_2003	2003	Comoros						ERR1878103
ERR1878154_KE_2006	2006	Kenya						ERR1878154
ERR1878551_DJ_2007	2007	Djibouti						ERR1878551
ERR1879386_TZ_1998	1998	Tanzania						ERR1879386
ERR1879540_KE_1998	1998	Kenya						ERR1879540

V. cholera name	Year	Location, country or sea	Location, region/ locality	Lineage	Latitude	Longitude	Location on map†	SRA ID
ERR1880767_TZ_1998	1998	Tanzania						ERR1880767
ERR1880801_IN_1997	1997	India						ERR1880801
ERR1880812_IN_1998	1998	India						ERR1880812
ERR2265670_NE_2014	2014	Niger						ERR2265670
ERR3268992_CD_2017	2017	DRC	Minova		−4.32153	15.31185	3	ERR3268992
ERR3268993_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268993
ERR3268994_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268994
ERR3268995_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268995
ERR3268996_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268996
ERR3268997_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268997
ERR3268998_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268998
ERR3268999_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3268999
ERR3269000_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3269000
ERR3269001_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3269001
ERR3269002_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3269002
ERR3269003_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3269003
ERR3269004_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3269004
ERR3269005_CD_2015	2015	DRC	Goma		−1.6835	29.2356	17	ERR3269005
ERR3269006_CD_2015	2015	DRC	Goma	ST515	−1.6835	29.2356	17	ERR3269006
ERR3269007_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269007
ERR3269008_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269008
ERR3269009_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269009
ERR3269010_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269010
ERR3269011_CD_2015	2015	DRC	Bukavu	ST515	−2.50316	28.85309	21	ERR3269011
ERR3269012_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269012
ERR3269013_CD_2015	2015	DRC	Minova	ST515	−4.32153	15.31185	3	ERR3269013
ERR3269014_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269014
ERR3269015_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269015
ERR3269016_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269016
ERR3269017_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269017
ERR3269018_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269018
ERR3269019_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269019
ERR3269020_CD_2016	2016	DRC	Uvira	ST515	−3.38413	29.1415	20	ERR3269020
ERR3269021_CD_2016	2016	DRC	Uvira	ST515	−3.38413	29.1415	20	ERR3269021

V. cholera name	Year	Location, country or sea	Location, region/ locality	Lineage	Latitude	Longitude	Location on map†	SRA ID
ERR3269022_CD_2016	2016	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269022
ERR3269023_CD_2016	2016	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269023
ERR3269024_CD_2016	2016	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269024
ERR3269025_CD_2016	2016	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269025
ERR3269026_CD_2016	2016	DRC	Uvira	ST515	−3.38413	29.1415	20	ERR3269026
ERR3269027_CD_2016	2016	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269027
ERR3269028_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269028
ERR3269029_CD_2015	2015	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269029
ERR3269031_CD_2016	2016	DRC	Alimbongo	ST515	−0.3692	29.155569	8	ERR3269031
ERR3269033_CD_2016	2016	DRC	Kabambare	ST69	−4.68967	27.69298	9	ERR3269033
ERR3269034_CD_2016	2016	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269034
ERR3269035_CD_2016	2016	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269035
ERR3269036_CD_2016	2016	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269036
ERR3269037_CD_2016	2016	DRC	Kimbilulenge	ST69	−4.32153	15.31185	3	ERR3269037
ERR3269038_CD_2016	2016	DRC	Kimbilulenge	ST69	−4.32153	15.31185	3	ERR3269038
ERR3269039_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269039
ERR3269040_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269040
ERR3269041_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269041
ERR3269042_CD_2014	2014	DRC	Alimbongo	ST69	−0.3692	29.155569	8	ERR3269042
ERR3269043_CD_2015	2015	DRC	Masisi	ST69	−1.3527161	29.37873	16	ERR3269043
ERR3269044_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269044
ERR3269045_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269045
ERR3269047_CD_2015	2015	DRC	Goma	ST69	−1.6835	29.2356	17	ERR3269047
ERR3269048_CD_2015	2015	DRC	Goma	ST515	−1.6835	29.2356	17	ERR3269048
ERR3269049_CD_2015	2015	DRC	Goma	ST69	−1.6835	29.2356	17	ERR3269049
ERR3269050_CD_2015	2015	DRC	Goma	ST69	−1.6835	29.2356	17	ERR3269050
ERR3269051_CD_2014	2014	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269051
ERR3269052_CD_2014	2014	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269052
ERR3269053_CD_2014	2014	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269053
ERR3269054_CD_2014	2014	DRC	Uvira	ST69	−3.38413	29.1415	20	ERR3269054
ERR3269056_CD_2014	2014	DRC	Uvira	ST515	−3.38413	29.1415	20	ERR3269056
ERR3269057_CD_2014	2014	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269057
ERR3269058_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269058

V. cholera name	Year	Location, country or sea	Location, region/ locality	Lineage	Latitude	Longitude	Location on map†	SRA ID
ERR3269059_CD_2014	2014	DRC	Goma	ST515	−1.6835	29.2356	17	ERR3269059
ERR3269060_CD_2014	2014	DRC	Fizi-Baraka	ST515	−4.30058	28.94212	12	ERR3269060
ERR3269061_CD_2014	2014	DRC	Fizi-Baraka	ST69	−4.30058	28.94212	12	ERR3269061
ERR3269062_CD_2014	2014	DRC	Goma	ST69	−1.6835	29.2356	17	ERR3269062
ERR3269063_CD_2014	2014	DRC	Fizi-Baraka	ST613	−4.30058	28.94212	12	ERR3269063
ERR3269064_CD_2014	2014	DRC	Fizi-Baraka	ST612	−4.30058	28.94212	12	ERR3269064
ERR3269065_CD_2015	2015	DRC	Goma	ST515	−1.6835	29.2356	17	ERR3269065
ERR3269066_CD_2015	2015	DRC	Goma	ST515	−1.6835	29.2356	17	ERR3269066
ERR386629_CD_2009	2009	DRC			−5.91312	29.20005	4	ERR386629
ERR386661_ZM_2012	2012	DRC						ERR386661
ERR386712_CD_2012	2012	DRC	Biyela	ST515	−4.32153	15.31185	5	ERR386712
ERR572559_CD_2013	2013	DRC	Nyemba	ST69	−8.16966	25.38507	1	ERR572559
ERR572810_CD_2001	2001	DRC	Ankoro		−6.75053	26.94274	2	ERR572810
ERR976553_UG_1998	1998	Uganda						ERR976553
ERR976558_KM_1998	1998	Comoros						ERR976558
ERR976569_RW_1998	1998	Rwanda						ERR976569
ERR976575_SD_1998	1998	Sudan						ERR976575
ERR976593_MG_2000	2000	Madagascar						ERR976593

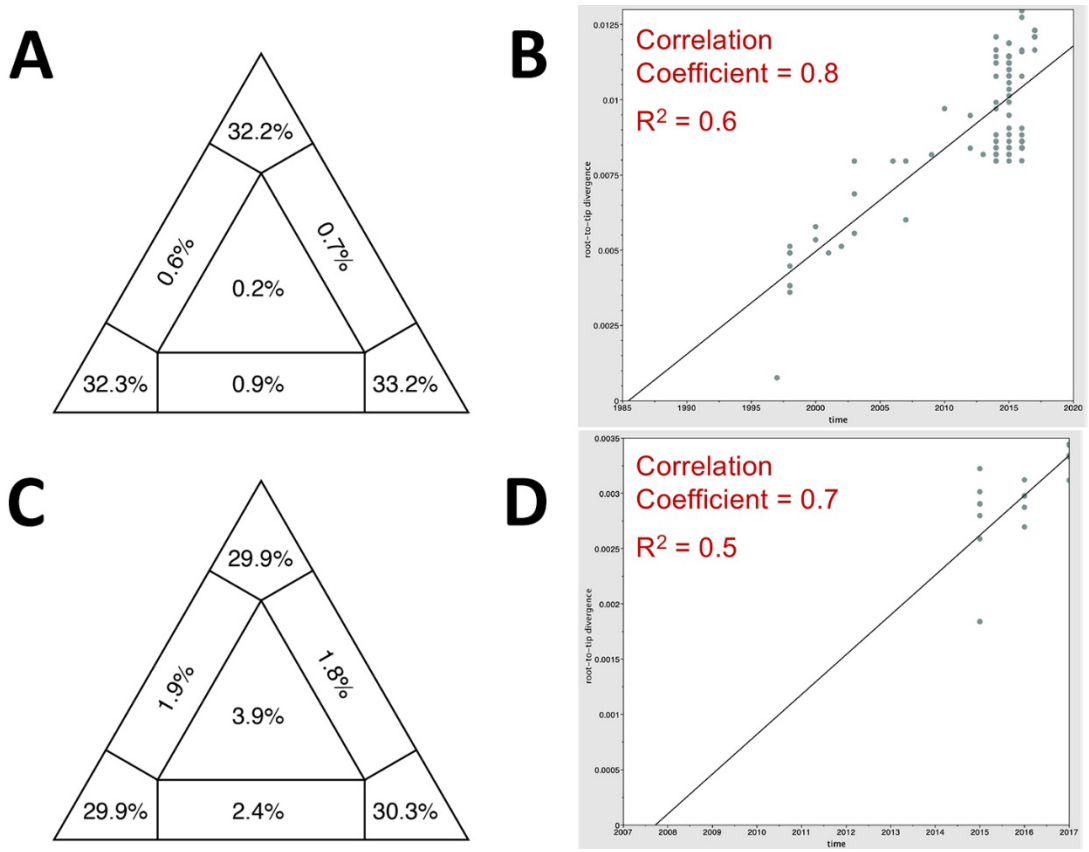
*DRC, Democratic Republic of the Congo.
†Map in Figure 1.

Appendix Table 2. Results of MLST analysis of DRC *V. cholerae* strains

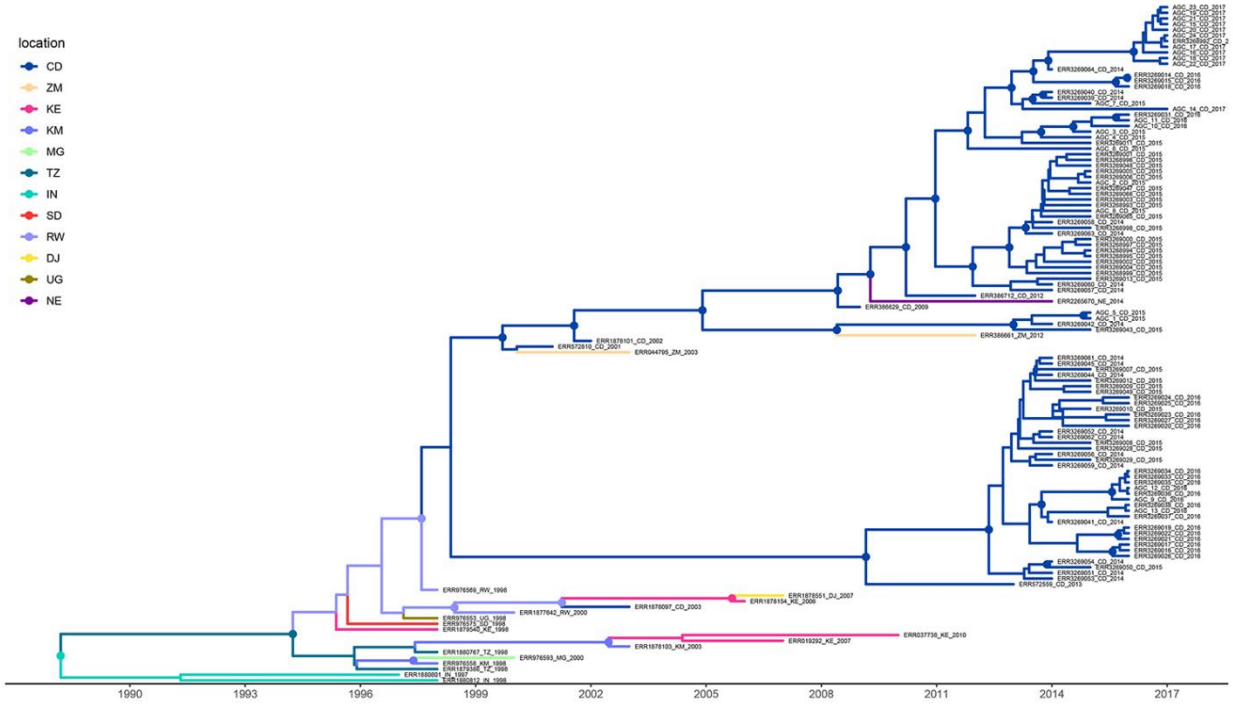
Sample name	MLST profile (ST)
AGC_12_CD_2016	69
AGC_13_CD_2016	69
AGC_9_CD_2016	69
AGC_1_CD_2015	515
AGC_10_CD_2016	515
AGC_11_CD_2016	515
AGC_14_CD_2017	515
AGC_15_CD_2017	515
AGC_16_CD_2017	515
AGC_17_CD_2017	515
AGC_18_CD_2017	515
AGC_19_CD_2017	515
AGC_2_CD_2015	515
AGC_20_CD_2017	515
AGC_21_CD_2017	515
AGC_22_CD_2017	515
AGC_23_CD_2017	515
AGC_24_CD_2017	515
AGC_3_CD_2015	515
AGC_4_CD_2015	515
AGC_5_CD_2015	515
AGC_6_CD_2015	515
AGC_7_CD_2015	515
AGC_8_CD_2015	515

Appendix Table 3. ICP1 phage which were sequenced

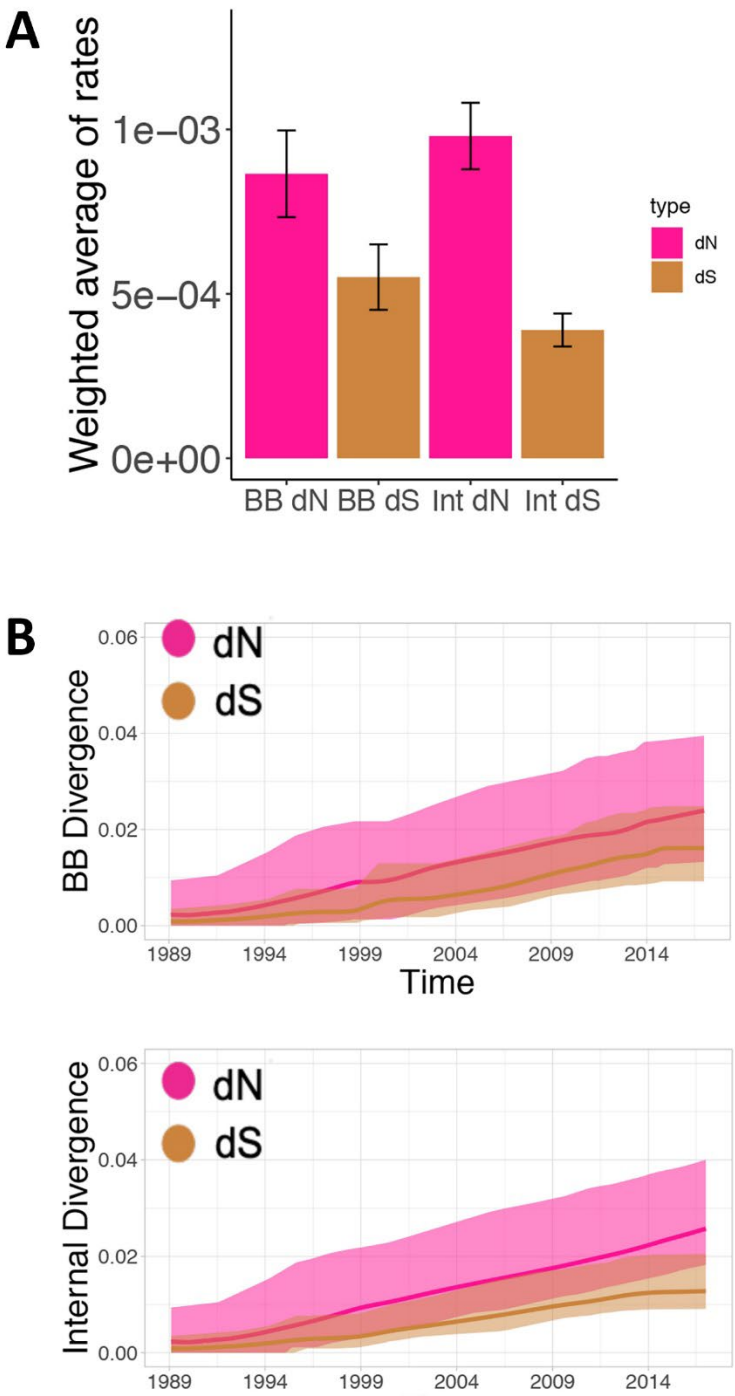
Strain	Isolation date	Province/Location	SRA ID
DRC32	1/11/2017	Rutshuru/Hgr	SRR18305671
DRC48	3/18/2017	Goma/Hgr	SRR18305670
DRC55	3/30/2017	Rutshuru/Tongo	SRR18305669
DRC72	3/2/2017	Kirotshe/Sake	SRR18305668
DRC74	3/2/2017	Kirotshe/Sake	SRR18305667
DRC82	4/11/2017	Rutshuru/Ntamugenga	SRR18305666
DRC87	4/15/2017	Nyiragongo/Kibumba	SRR18305665
DRC106	4/22/2017	Kibumba	SRR18305664



Appendix Figure 1. Estimations of phylogenetic and temporal signal from the DRC phylogenies. Presence of phylogenetic signal in the dataset of the *V. cholerae* A) dataset displayed in Figure 1 (isolates collected by our group and downloaded from NCBI), as well as C) the dataset displayed in Figure 3 (isolates collected by our group only), was evaluated by likelihood mapping check for alternative topologies (tips), unresolved quartets (center) and partly resolved quartets (edges) for each dataset. Linear regressions of root-to-tip genetic distance within the ML phylogeny (tree in Figure 1) against sampling time for each taxon. Temporal resolution for B) dataset displayed Figure 1 or D) dataset displayed in Figure 3 was assessed using the slope of the regression, with positive slope indicating sufficient temporal signal. Correlation coefficient, r, and R^2 are reported in the temporal signal plot.



Appendix Figure 2. MCC in Figure 1 tree with tips. Phylogeny reported in Figure 1 with tips. Branches of the phylogeny are scaled in time and colored by country of origin as shown in the legend (location). Circles in internal node indicate posterior probability support >0.9 and the colors indicate the ancestral country inferred by Bayesian phylogeography reconstruction. CD, Democratic Republic of Congo; ZM, Zambia; KE, Kenya; KM, Comoros; MG, Kyrgyzstan; TZ, United Republic of Tanzania; IN, India; SD, Sudan; RW, Rwanda; DJ, Djibouti; UG, Uganda; NE, Niger



Appendix Figure 3. Tree, synonymous and nonsynonymous substitution rates. A) Weighted average of synonymous and nonsynonymous substitution rates of backbone and internal branches based on the Bayesian phylogeography tree in Figure 1. B) Absolute synonymous (tan) and nonsynonymous (pink) divergence rates (y-axis) against time in years (x-axis) of backbone (top) and internal branches (bottom) of the phylogeny. BB, backbone; dN, nonsynonymous substitution rates; dS, synonymous substitution rates