

Genomic Epidemiology of 3,076 *Vibrio cholerae* Isolates Reveals ST69 Clonal Expansion and Multidrug Resistance across Africa

Ahmed Olowo-okere

`olowookere.ahmed@uniabuja.edu.ng`

University of Abuja

Mohammed Yahaya

Usmanu Danfodiyo University

Abdourahmane Yacouba

Abdou Moumouni University

Abubakar Jibril

Federal University of Health Sciences

Razaq Funsho Atata

University of Abuja

Busayo Olalekan Olayinka

Ahmadu Bello University

Research Article

Keywords: *Vibrio cholera*, Genomic Epidemiology, Antimicrobial resistance, Cholera, Africa

Posted Date: May 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9358362/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background Cholera remains a major public health threat across Africa, driven by complex interactions between environmental, socioeconomic, and microbial factors. However, the continental genomic epidemiology of *Vibrio cholerae* remains incompletely characterised. We aimed to conduct a comprehensive continental genomic analysis of *V. cholerae* to determine its spatiotemporal dynamics, virulence, and antimicrobial resistance (AMR) profiles.

Methods We conducted a large-scale genomic analysis of 3,076 *V. cholerae* isolates collected from 29 African countries between 1949 and 2025. Genomic data were retrieved from publicly available databases and subjected to standardized quality control, assembly, and annotation pipelines. Phylogenomic reconstruction was performed using core genome single-nucleotide polymorphism (SNP) analysis with recombination filtering. We conducted multilocus sequence typing (MLST), virulence profiling, AMR gene detection, and mobile genetic element characterization using established bioinformatics tools to assess associations between lineage distribution, geographic regions, temporal trends, and resistance patterns.

Results The population is highly clonal and dominated by sequence type 69 (ST69) (2,682 [87.2%] of 3,076 genomes). ST69 almost exclusively drove the 2022–2024 continent-wide outbreaks, representing 98.9% (463/468) of genomes in 2022, 99.3% (605/609) in 2023, and 97.3% (252/259) in 2024. Toxigenic O1 El Tor markers and *ctxA/ctxB* co-carriage exceeded 91% in clinical genomes but were significantly attenuated in environmental strains. We identified 101 distinct AMR genes, with 94.1% of isolates classified as multidrug-resistant. Ubiquitous chromosomal mutations, primarily *gyrA*_S83I (92.88%) and *parC*_S85L (84.82%), drove universal fluoroquinolone resistance. An SXT conjugative element conferring multidrug resistance was present in roughly 64% of the population. Clinical isolates exclusively harboured high-risk resistance determinants absent environmentally, including macrolide resistance genes (6.6%–6.8%) and extended-spectrum beta-lactamases such as *bla*PER-7 (6.4%).

Conclusion Cholera transmission across Africa is persistently driven by the clonal expansion of the ST69 lineage. Universal fluoroquinolone resistance and clinically restricted emergence of macrolide and beta-lactamase resistance highlight a critical therapeutic challenge. Sustained genomic surveillance is essential to monitor this multidrug-resistant clone and inform regional outbreak control.

Background

Vibrio cholerae is a Gram-negative bacterium, responsible for the acute secretory diarrheal disease known as cholera [1]. The disease remains a formidable and persistent threat to global public health, with the World Health Organization (WHO) estimating between 1.3 and 4.0 million cases and 21,000–143,000 deaths annually worldwide [2,3]. Sub-Saharan Africa bears a disproportionate fraction of this burden, accounting for over 30% of global cases and more than 80% of cholera deaths worldwide, with the region suffering 83% of global cholera mortality between 2000 and 2015 [4]. The scale of this crisis escalated dramatically in recent years: in 2023, Africa experienced a 125% increase in reported cholera

cases and a 62% rise in deaths compared with 2022, with 21 countries collectively reporting 225,857 cases and 3,167 deaths, yielding a case fatality ratio (CFR) of 1.4% [5]. By July 2024, a cumulative total of 399,508 cases and 7,023 deaths (CFR: 1.8%) had been reported across the continent since January 2022, with the Democratic Republic of the Congo (DRC), Ethiopia, Malawi, Mozambique, and Zimbabwe collectively accounting for 72.2% of all cases and 62.3% of all deaths [6]. These explosions of transmission are driven by complex intersections of inadequate water, sanitation, and hygiene (WASH) infrastructure, the exacerbating effects of climate change – including cyclones and El Niño-driven flooding – and persistent armed conflicts that fracture healthcare systems and displace populations [7–9]. Geospatial analyses further reveal that the annual incidence of cholera in Africa has dynamically shifted from western to central and eastern regions between 2011 and 2020, with an estimated 296 million people residing in high-incidence administrative units by 2020 [10].

The evolutionary success of *V. cholerae*, particularly the El Tor biotype responsible for the ongoing seventh pandemic (7PET), is intrinsically linked to its high genomic plasticity [11,12]. The pathogen possesses a bipartite genome consisting of a larger Chromosome I (~2.9 Mb), which encodes the majority of essential housekeeping genes and primary virulence determinants – including the toxin-coregulated pilus (TCP), encoded by the *VPI-1* pathogenicity island, and the core cholera toxin (CTX) prophage – and a smaller Chromosome II (~1.1 Mb) that houses accessory metabolic genes and the chromosomal superintegron [13]. The acquisition and exchange of mobile genetic elements (MGEs) play a significant role in driving the emergence of novel, highly virulent lineages and disseminating antimicrobial resistance genes [14]. Phylogenomic reconstructions have resolved the African 7PET population into a series of distinct sublineages, designated AFR10 through AFR15. Recent genomic analyses of 763 *V. cholerae* O1 isolates collected across sub-Saharan Africa between 2019 and 2024 identified the rapid continental circulation of the AFR15 lineage, which has been associated with unusually large outbreaks in Southern Africa and has expanded into new African Union member states, while cross-border transmission was found to be prevalent across both West and East Africa [15]. Multiple co-circulating lineages have been identified within individual countries, and what were previously classified as sporadic outbreak strains have, upon genomic re-evaluation, proven to be representatives of larger undersampled outbreaks, underscoring the persistent gaps in continental surveillance [16].

Compounding the threat of hypervirulent sublineages is the accelerating emergence of AMR within *V. cholerae* populations. The primary vehicle for multidrug resistance (MDR) in African 7PET isolates is the SXT/R391 family of integrating conjugative elements (ICEs), which are ~99–100 kb self-transmissible elements that integrate site-specifically into the 5' end of the chromosomal *prfC* gene [17]. These ICEs encode resistance to sulfamethoxazole, trimethoprim, streptomycin, chloramphenicol, nalidixic acid, and tetracycline, and carry well-characterised antimicrobial resistance genes (ARGs)[18].

Fluoroquinolone resistance, however, represents the most clinically alarming contemporary trend. Reduced ciprofloxacin susceptibility in 7PET strains arises primarily through sequential accumulation of point mutations in the quinolone resistance-determining regions (QRDRs) of chromosomal

topoisomerase genes, most notably the Ser83Ile substitution in *gyrA* (encoding DNA gyrase subunit A) and a corresponding Ser85Leu substitution in *parC* (encoding topoisomerase IV subunit C), which together confer high-level fluoroquinolone resistance [19]. A landmark example of the consequences of unchecked AMR accumulation was the 2018 cholera outbreak in Zimbabwe, in which the causative AFR13 sublineage strain acquired an approximately 160-kb IncA/C2 plasmid bearing 14 additional AMR genes; the isolates were intermediately resistant or resistant to tetracycline (*tetA*) and ciprofloxacin (*gyrA* and *parC* mutations and *aac(6')-Ib-cr*) and produced the extended-spectrum beta-lactamase CTX-M-15, rendering the outbreak strain resistant to two of the three WHO-recommended cholera antibiotics [20]. More recent PulseNet Africa genomic surveillance of isolates from Côte d'Ivoire, Ghana, Zambia, and South Africa (2010–2024) has confirmed the persistence of resistance genes associated with quinolones and trimethoprim, raising serious concerns about the continued efficacy of current treatment protocols [21].

Despite the severe and escalating impact of cholera across Africa, comprehensive genomic epidemiology of the pathogen remains fragmented. Between 2010 and 2019, sub-Saharan Africa documented 999 suspected cholera outbreaks with over 480,000 probable patients across 744 sub-national regions in 25 countries, yet molecular characterisation has largely been limited to localized, outbreak-specific investigations [22]. Consequently, the continental genomic diversity, pan-genome dynamics, AMR gene reservoir, and broad evolutionary relationships of African *V. cholerae* isolates remain incompletely characterised. The difficulty in containing cholera is compounded by poor understanding of how the pathogen circulates throughout the region; within-country variation in transmission dynamics and the finding that several formerly classified "sporadic" outbreaks represent undersampled continuous transmission events highlight the critical limitations of non-genomic surveillance [23]. Without a unified, large-scale genomic framework, tracking cross-border transmission events and identifying the emergence of high-risk MDR lineages in real time remains extremely challenging.

To address this critical knowledge gap, this study presents a comprehensive, large-scale comparative genomic analysis of 2,726 *V. cholerae* genomes sourced from across the African continent. Ultimately, this analysis provides a robust genomic framework to elucidate the evolutionary trajectory of the pathogen, characterise the continental distribution of AMR determinants, and inform highly targeted, real-time public health interventions consistent with the WHO Global Roadmap to End Cholera by 2030.

Materials and Methods

2.1 Data Identification and Genome Acquisition

Genomic data for *V. cholerae* were identified via the NCBI Pathogen Detection Isolates Browser (<https://www.ncbi.nlm.nih.gov/pathogens/isolates%23/search/%23/search/>), accessed on 19th March 2026. An initial global query yielded 19,040 isolate records, which were subsequently filtered to retain only 3185 originating from 29 African countries. These isolates were stratified into two cohorts: those

with pre-assembled genomes deposited in the NCBI Assembly database ($n=2733$), and those available exclusively as raw paired-end sequencing reads in the NCBI Sequence Read Archive (SRA/ENA) ($n=452$). These were subsequently retrieved using NCBI Datasets v18.9.0 [24] and *fasterq-dump* v3.2.1 [25], respectively. We verified the mutual exclusivity of the two cohorts by cross-referencing their BioSample identifiers, which confirmed that no single isolate was represented in both the assembly and raw read datasets. The genomes metadata, including isolation country, collection year, and sample source were programmatically retrieved via the NCBI E-utilities API [26].

2.2 Raw Read Processing, *De Novo* Genome Assembly, and Annotation

The downloaded raw paired-end reads were assembled *de novo* into contigs using the Shovill pipeline v1.1.0 (<https://github.com/tseemann/shovill>), which employs SPAdes v4.2.0 as the underlying assembler, with `-trim` option [27]. All genome assemblies were annotated in a standardised, reproducible manner using Prokka v1.13 [28].

2.4 Genome Quality Control and Taxonomic Verification

To ensure the integrity of downstream analyses, all assembled and downloaded genomes were subjected to rigorous quality control. Assembly contiguity was evaluated using QUAST v5.2.0, while genome completeness and contamination were assessed utilizing CheckM2 v1.0.2 [29,30]. Assemblies were systematically excluded from the dataset if they exhibited a completeness of less than 90% ($n = 6$) or a contamination level exceeding 10% ($n = 17$). Furthermore, genomes were discarded if their total assembly length fell outside the biologically expected range for *V. cholerae* (approximately 3.5 to 4.5 Mbp), or if they displayed high fragmentation, contig N50 lower than 10,000 bp or an L90 metric exceeding 100 contigs ($n = 86$). To preclude the inclusion of misclassified taxa, pairwise Average Nucleotide Identity (ANI) of all retained assemblies was calculated against the *V. cholerae* O1 biovar El Tor str. N16961_(GCF_900205735.1) using FastANI v1.3.4 [31]. Only assemblies exhibiting an ANI $\geq 95\%$ were definitively classified as *V. cholerae* and retained for downstream analyses [33]. All genomes passed this threshold, with pairwise ANI against the reference ranging from 96.01 to 100% (mean = 99.88%; median = 99.98%; SD = 0.45%), confirming the species-level identity of all retained assemblies.

Following the application of these exclusion filters, a final high-quality dataset comprising 3,076 genomes was retained (**Table S1**). Within this refined cohort, genome completeness ranged from 97.67% to 100% (mean = 100%, median = 100%), and contamination varied between 0% and 9.82% (mean = 2.25%, median = 0.05%) (**Table S2**). The retained assemblies contained between 2 and 281 contigs (mean = 75, median = 72), with N50 values spanning 39.0 to 3,141.1 Kb (mean = 278.8 Kb, median = 235.7 Kb). The overall GC content ranged from 46.93% to 48.01% (mean = 47.51%, median = 47.50%), and total genome sizes spanned 3.78 to 4.48 Mbp (mean = 4.10 Mbp, median = 4.05 Mbp) (**Table S3; Table S4**). These assembly metrics are highly consistent with the known genomic architecture and bipartite chromosomal structure of *V. cholerae* [32].

2.6 Core Genome SNP Alignment and Recombination Masking

Per-sample read mapping and variant calling were performed using Snippy v4.6.0 (<https://github.com/tseemann/snippy>) against the reference genome. To ensure high-quality variant detection, minimum thresholds were set for read depth (4×; --mincov 4) and variant allele frequency (0.75; --minfrac 0.75) [33]. A core genome alignment of all 3,076 isolates was subsequently generated using snippy-core, retaining only sites present across the entire dataset [33]. This strict core alignment spanned 4,033,501 bp of the reference genome, with a mean per-sample reference coverage of 96.14%, and yielded 245,973 variant sites.

To mitigate the confounding effects of homologous recombination and mobile genetic elements (MGEs) on phylogenetic inference, recombinant regions were identified and masked prior to tree reconstruction using Gubbins v3.4.3 [34]. Gubbins detects recombination under an iterative maximum-likelihood framework by identifying genomic windows with significantly elevated SNP densities relative to the background. The pipeline was configured to construct an initial guide tree using FastTree (--first-tree-builder fasttree), followed by iterative refinement using IQ-TREE (--tree-builder iqtree) under the GTRGAMMA substitution model. The process was capped at five iterations (--iterations 5), with convergence assessed via the weighted Robinson-Foulds metric. Recombinant blocks were defined by a minimum of three SNPs (--min-snps 3), and sequences containing more than 25% missing data were excluded from the analysis (--filter-percentage 25).

The resulting recombination-free polymorphic site alignment produced by Gubbins was utilized for all downstream phylogenetic analyses. Finally, to characterize within-dataset genomic diversity and delineate putative transmission clusters, pairwise SNP distances were calculated from this filtered alignment using snp-dists v0.8.2 [35].

2.7 Maximum Likelihood Phylogenetic Reconstruction and Temporal Analysis

A maximum-likelihood (ML) phylogeny was reconstructed from the Gubbins recombination-free core SNP alignment using IQ-TREE 2 v2.2.6 [36]. The optimal nucleotide substitution model was selected automatically using the integrated ModelFinder algorithm under the Bayesian Information Criterion (BIC) to avoid model misspecification [37]. Topological robustness was assessed by 1,000 ultrafast bootstrap (UFBoot2) replicates, with branch support values $\geq 95\%$ considered indicative of well-supported nodes [38]. The N16961 reference genome was included as an outgroup to root the phylogeny. Phylogenetic visualization and metadata annotation were performed using ggtree v3.16.3 and ggtreeExtra v1.18.1 [39,40]

Multilocus sequence typing

Multilocus sequence typing (MLST) was performed on all genomic assemblies using the command-line MLST tool v2.23.0 (<https://github.com/tseemann/mlst>), which performs automated allele matching against curated PubMLST databases. The Heidelberg seven-locus *V. cholerae* MLST scheme (vcholerae)

was applied, interrogating the following housekeeping gene loci: adenylate kinase (*adk*), DNA gyrase subunit B (*gyrB*), malate dehydrogenase (*mdh*), homocysteine synthase (*metE*), pyridine nucleotide transhydrogenase (*pntA*), phosphoribosylformylglycinamide synthetase (*purM*), and dihydroorotase (*pyrC*). Sequence types (STs) were assigned based on exact allele matches against the PubMLST *V. cholerae* reference database (<https://pubmlst.org/vcholerae>). Only assemblies yielding a PERFECT or GOOD call status, defined as 100% identity and full-length allele matches across all seven loci, were retained for downstream ST-based analyses.

2.8 In Silico Profiling of Virulence Factors, AMR Determinants, and Mobile Genetic Elements

Virulence, Serogrouping, and Biotyping.

Genomic assemblies were screened using CholeraeFinder v2.1 (Center for Genomic Epidemiology) to predict serogroup and biotype, and to identify major *V. cholerae*-specific virulence genes, resistance determinants, and mobile genetic elements [41]. Species identity was confirmed by detection of *ompW*, a highly conserved outer membrane protein gene specific to *V. cholerae*. Serogroup was assigned based on detection of the O-antigen biosynthesis genes *rfbV_01* (O1 serogroup) and *wbfZ_0139* (O139 serogroup), with a minimum nucleotide identity threshold of $\geq 98\%$ applied to the *wbfZ_0139* marker to exclude cross-reactive hits; isolates below this threshold were classified as Non-O1/Non-O139. Biotype was differentiated into El Tor, El Tor Variant, or Classical lineages using allele-specific markers for the transcriptional regulator (*rstR_et*, *rstR_cc*) and the toxin-coregulated pilus subunit (*tcpA_et3*, *tcpA_cc*), with El Tor Variant designation requiring co-detection of the *ctxB_7* allele. The cholera toxin B subunit (*ctxB*) was further genotyped into allelic variants (*ctxB_1*, *ctxB_3*, *ctxB_7*) to resolve lineage-level diversity.

To contextualise isolates within the ongoing seventh pandemic, the seventh pandemic El Tor (7PET) lineage-specific marker gene (*VC2346*) was assessed. Isolates were classified as toxigenic upon detection of the cholera enterotoxin A subunit gene (*ctxA*), the catalytic determinant of cholera toxin production; *ctxB* genotyping was used independently for lineage discrimination. The presence of *V. Pathogenicity Islands* (VPI-1, VPI-2) and *Seventh Pandemic Islands* (VSP-I, VSP-II) was assessed using the CholeraeFinder reference database, which employs embedded reference loci derived from the *V. cholerae* O1 El Tor N16961 reference genome (GenBank: AE003852). Mobile genetic elements were characterised by detection of the SXT/R391 integrating conjugative element (ICE) integrase gene (*intSXT*) and the Class 1 integron integrase (*intI1*).

Antimicrobial Resistance (AMR) Profiling and MDR Classification

The resistome of 3,076 *V. cholerae* isolates was characterized by screening genomic assemblies with AMRFinder plus v3.10.1 [42]. The abundance of AMR genes within the population was evaluated by calculating the Average Copy Number (ACN) per isolate, both overall and stratified by isolation source (clinical versus environmental) as previously described [43]. Gene prevalence was determined as the percentage of genomes harboring a specific resistance determinant. Isolates were classified as

multidrug-resistant (MDR) if they harbored acquired resistance genes spanning three or more distinct antibiotic classes as previously described [44]. To assess spatial and source-specific distributions, the diversity of unique AMR genes and their specific ACNs were compared between the clinical and environmental subsets.

Identification of Mobile Genetic Elements and Genomic Localization of Genes

Plasmid-derived contigs were identified using a consensus approach combining PlasForest and geNomad v1.8.0 [45,46]. PlasForest was run using the plasforest.sav model against the RefSeq plasmid database, while geNomad was executed in end-to-end mode with the --cleanup and --splits 8 flags to simultaneously detect plasmid and prophage sequences. As previously reported, contigs were classified as prophage where geNomad assigned a virus score exceeding both the chromosome and plasmid scores [43]. To determine the genomic context of the resistome, the coordinates of the identified AMR genes were intersected with the MGE predictions, allowing each gene occurrence to be localized to the chromosome, a plasmid, or a prophage region.

2.8 Statistical Analysis and Data Visualization

Descriptive statistics, including frequencies and percentages, were used to summarize categorical variables across isolates. All statistical analysis and visualizations were performed in R v4.5.1 and RStudio v2025.9.0.387 (R Core Team, 2023).

Results

A total of 3,076 sequenced *V. cholerae* genomes from the African continent were analyzed in this study. Geographically, the isolates originated from 29 countries, with a pronounced concentration in Central Africa ($n = 1,305$; 42.4%) and East Africa ($n = 1,174$; 38.2%), followed by West Africa ($n = 358$; 11.6%), Southern Africa ($n = 221$; 7.2%), and North Africa ($n = 18$; 0.6%) (Fig. 1A; Table S1). The Democratic Republic of the Congo (DRC) contributed the largest proportion of genomes ($n = 1,039$; 33.8%), followed by Cameroon ($n = 259$; 8.4%), Zambia ($n = 255$; 8.3%), Kenya ($n = 254$; 8.3%), and Mozambique ($n = 245$; 8.0%).

The majority of sequenced genomes were derived from clinical human samples ($n = 1,988$), while environmental sources accounted for a minor fraction of the dataset ($n = 96$; 3.1%), including water ($n = 101$), food/aquatic animal ($n = 17$), and other/unknown sources ($n = 12$).

Temporal analysis of collection dates revealed isolates spanning from 1949 to 2025, with 150 genomes lacking collection year metadata. The mean number of isolates per year was 84, ranging from 1 to 609 across years with available data (Table S1; Fig 1C). A marked escalation in genomic surveillance was observed in recent years, with the highest isolate counts recorded in 2023 ($n = 609$; 19.8%) and 2022 ($n = 468$; 15.2%), coinciding with intensified large-scale cholera outbreaks across the continent. Historical

isolates from earlier decades (≤ 2010) were characterised by lower sequencing volumes, with the earliest isolate in the dataset dating to 1949.

ST69 and its related lineage ST515 are the most dominant sequence types in Africa

Multi-locus sequence typing (MLST) of the 3,076 genome dataset revealed a highly clonal population structure characterised by the near-complete dominance of a single lineage across the continent (Fig. 3B, 3D). All isolates were typed using the *V. cholerae* MLST scheme, yielding 33 unique sequence types, with MLST assignments achieving perfect allelic concordance in 2,980 genomes (96.9%), while 35 genomes carried novel allelic profiles, 32 were partially assigned (OK), 27 were mixed, and 2 were missing data. MLST scores ranged from 74 to 100 (median = 100; mean = 99.6%), reflecting high-quality typing across the dataset (Figure 1; Table S5).

Sequence Type 69 (ST69) was the overwhelmingly prevalent sequence type, accounting for 2,682 genomes (87.2% of the entire dataset). Spatially, ST69 exhibited a pervasive pan-African distribution, driving epidemics across all five African regions – Central Africa (n = 1,117), East Africa (n = 1,113), West Africa (n = 314), Southern Africa (n = 137), and North Africa (n = 1). It was the dominant lineage in all heavily burdened nations, most notably the Democratic Republic of the Congo (DRC) (n = 869), Cameroon (n = 243), Mozambique (n = 243), Kenya (n = 250), Zambia (n = 247), Tanzania (n = 180), and Nigeria (n = 156). Longitudinally, ST69 has circulated persistently since at least 1970 and was almost exclusively responsible for the massive surge in clinical genomes sequenced during the recent 2022–2024 continent-wide outbreak waves, accounting for 463 of 468 genomes in 2022 (98.9%), 605 of 609 in 2023 (99.3%), and 252 of 259 in 2024 (97.3%).

The secondary lineage ST515 (n = 196; 6.4%) emerged as the only other numerically substantial sequence type and demonstrated restricted geographic mobility heavily localised to Central and East Africa. ST515 drove distinct temporal transmission peaks in 2015 (n = 76) and 2018 (n = 33), with the vast majority of isolates originating from the DRC (n = 150) and Tanzania (n = 18), alongside minor representation in Zambia (n = 4). Notably, ST515 was detected almost exclusively in clinical samples (n = 103) and was virtually absent from environmental sources.

The remaining sequence types collectively constituted a minor proportion of the dataset (n = 198; 6.4%), with 31 additional STs each accounting for $\leq 0.8\%$ of isolates. ST75 (n = 24; 0.8%) was predominantly observed in Southern Africa, particularly South Africa (n = 21), and showed a recent temporal resurgence between 2018 and 2025. ST1251 (n = 15; 0.5%) was geographically restricted to West Africa, almost entirely from Ghana (n = 15), suggesting a localised lineage. ST555 (n = 6; 0.2%) was exclusively detected in South Africa, while ST68 (n = 6; 0.2%) was confined to North Africa. Historically older isolates from 1949 were typed as ST73, representing one of the earliest sequenced African *V. cholerae* genomes in the dataset. Ninety-six genomes (3.1%) could not be assigned a definitive ST, disproportionately represented among environmental isolates (n = 34 of 96 environmental genomes), suggesting greater allelic diversity in non-clinical reservoir strains. Collectively, these findings underscore the profound geographic entrenchment and epidemiological dominance of the ST69 clone across contemporary

African cholera epidemics, with secondary lineages exhibiting spatially and temporally restricted circulation patterns.

Cross-tabulation of MLST sequence types with *ctxB* genotypes revealed distinct lineage-level patterns within the dataset. ST69, the dominant sequence type (n = 2,746), exhibited a near-equal distribution of *ctxB_1* (n = 1,276; 46.47%) and *ctxB_7* (n = 1,329; 48.40%) alleles, alongside minor proportions of *ctxB_3* (n = 31; 1.13%) and undetected *ctxB* (n = 62; 2.26%). The co-occurrence of both *ctxB* alleles within a single dominant sequence type is consistent with CTX prophage-mediated *ctxB* allele replacement within the 7PET genomic backbone, a well-documented mechanism of intra-lineage diversity in circulating *V. cholerae* [REF]. ST515, the second most prevalent sequence type (n = 197), was almost exclusively associated with *ctxB_1* (n = 193; 97.97%), indicating this lineage predates the post-2010 Haitian variant expansion. Notably, *ctxB_7* carriage was almost entirely confined to ST69 (n = 1,329; 98.7% of all *ctxB_7* isolates), confirming that the Haitian variant allele was acquired within the ST69 backbone rather than introduced via a distinct sequence type. Minor sequence types, collectively accounting for 314 isolates, were predominantly *ctxB*-negative, consistent with non-toxigenic environmental or pre-pandemic strains. These findings indicate that *ctxB* allele diversity in African *V. cholerae* reflects dynamic CTX prophage remodelling within a clonally dominant ST69 population rather than the co-circulation of genetically distinct pandemic lineages (Table S6).

Genomic Characterization and Reservoir-Specific Distribution

Genomic characterization of the *V. cholerae* isolates revealed a stark dichotomy in serogroup and biotype prevalence between clinical and environmental reservoirs (Table 1). The O1 serogroup marker (*rfbV*) was overwhelmingly dominant among clinical isolates across all African regions, ranging from 91.8% in Southern Africa to 99.2% in Central Africa. This trend was mirrored by biotype analysis, confirming the near-universal predominance of the El Tor biotype (*rstR_et*) in clinical settings (82.3% to 99.2%). The classical biotype marker (*rstR_cc*) was rare, detected only in minor clinical subsets (e.g., 8.9% in Southern Africa and 2.8% in East Africa). In sharp contrast, environmental isolates exhibited marked genomic heterogeneity and a distinct lack of canonical epidemic markers. The O1 serogroup and El Tor markers were entirely absent (0.0%) in the Southern African environmental cohort and severely reduced in Central Africa (31.6%) and West Africa (36.4% to 43.6%).

Regional Shifts in CTX Prophage Carriage and *ctxB* Allelic Variants

Intact CTX prophage carriage was a defining hallmark of clinical isolates. Co-carriage of the cholera toxin genes *ctxA* and *ctxB* exceeded 91% in clinical genomes from all regions, reaching 98.4% in Central Africa and 100% in North Africa. Conversely, environmental isolates demonstrated significant attenuation or complete absence of the CTX element, with carriage dropping to 36.4% in West Africa, 31.6% in Central Africa, and 0.0% in Southern Africa.

Analysis of *ctxB* allelic variants uncovered distinct geographic partitioning across the continent. In Central Africa, *ctxB* variant 1 (classical) dominated the clinical landscape, accounting for 92.0% of

isolates. However, a major allelic shift was observed elsewhere: *ctxB* variant 7 (El Tor) was the dominant allele in clinical isolates from West Africa (83.4%), Southern Africa (80.4%), and East Africa (70.2%). This geographic divergence in *ctxB* alleles suggests established, region-specific clonal endemicity rather than a homogenous continental population.

Conservation of Pandemic Markers and Pathogenicity Islands

The 7th pandemic El Tor (7PET) lineage-specific marker (VC2346) was highly conserved within the clinical subset (82.3% to 99.3%), confirming their pandemic lineage. In environmental subsets, the presence of VC2346 was substantially degraded, ranging from 43.6% in West Africa down to 0.0% in Southern Africa.

The four major *V. cholerae* pathogenicity islands (VPI-1, VPI-2, VSP-1, VSP-2) and the Type VI Secretion System (T6SS) core effector *vasX* followed identical trajectories of reservoir-specific conservation. Key anchoring loci for these islands were highly stable in clinical isolates, consistently detected in the vast majority of genomes across all regions (generally >91%). Environmental isolates, however, displayed systemic degradation of these crucial virulence loci. Notably, VPI-2 (VC1758) and the T6SS (*vasX*) appeared slightly more resilient in environmental subsets (e.g., VPI-2 was retained in 68.4% of Central African and 49.1% of West African environmental isolates) compared to the VSP elements, though all remained significantly diminished compared to their clinical counterparts.

Antimicrobial Resistance

Extensive AMR Burden and Ubiquitous Chromosomal Resistance

Genomic profiling of the 3,076 *V. cholerae* isolates revealed an extensive antimicrobial resistance repertoire, identifying 101 distinct AMR genes spanning 9 categories. The population exhibited a mean burden of 10.81 AMR genes per isolate (range: 3–24), with a staggering 94.1% of the collection classified as multidrug-resistant (MDR) (Figure 2).

Chromosomal resistance was ubiquitous across the dataset. Specifically, polymyxin resistance-associated lipid A modification genes (*almE*, *almF*, and *almG*) were detected in 100% of the isolates. Furthermore, the entire collection exhibited chromosomal mutations in the fluoroquinolone resistance-determining regions, driven predominantly by the highly prevalent *gyrA_S83I* (92.88%) and *parC_S85L* (84.82%) substitutions. This universal baseline resistance to fluoroquinolones—particularly ciprofloxacin, a WHO-recommended first-line therapy for severe cholera—highlights a critical and widespread therapeutic challenge across the continent.

Horizontally Acquired Determinants and the SXT/ICE Module

Beyond intrinsic chromosomal mutations, horizontally acquired AMR determinants were deeply entrenched in the population. A dominant SXT integrating conjugative element (ICE)-associated multidrug resistance module—comprising *sul2* (64.37%), *aph(3'')-Ib* (64.34%), *aph(6)-Id* (64.24%), and *floR*

(64.04%)—was identified in approximately 64% of the population, closely mirroring the anticipated carriage rates of the mobile element. Other highly prevalent acquired determinants included the chloramphenicol resistance gene *catB9* (95.61%) and the trimethoprim resistance gene *dfra1* (93.21%). The near-universal co-occurrence of chromosomal fluoroquinolone mutations with these acquired MDR determinants signals deeply consolidated multidrug resistance within the African *V. cholerae* population.

Diversity and Abundance Across Antibiotic Classes

Significant variation was observed in the diversity and abundance of genes within specific antibiotic categories. beta-lactams represented the most highly diverse resistance category, encompassing 30 unique AMR genes. Despite this striking diversity, the metallo-beta-lactamase *varG* emerged as the overwhelmingly dominant determinant (97.46%). Similar structural diversity was observed among aminoglycosides and fluoroquinolones, which featured 15 unique determinants each. In terms of genomic burden, the highest Average Copy Number (ACN) was observed in the MLS category, yielding 3.94 ± 0.37 copies per strain among those carrying them, significantly higher than the overall ACNs for aminoglycosides (1.88) or sulfonamides (1.00).

Ecological Discrepancy: Clinical Isolates Harbor Emerging Therapeutic Threats

A comparative analysis revealed a sharp discrepancy between isolation sources, with clinical isolates harboring a significantly broader and more concerning diversity of AMR determinants. Specifically, 38 resistance genes were completely restricted to the clinical population and absent from environmental counterparts, whereas only 7 genes (including *blaCARB-2* and *tet(G)*) were uniquely environmental (Figure 3).

This clinically restricted pool acts as a reservoir for emerging, high-risk resistance mechanisms. Of particular public health concern, macrolide resistance genes—including *mph(A)*, *mrx(A)*, *msr(E)*, and *mph(E)*—were restricted entirely to a clinical subset, present in 6.6% to 6.8% of the collection. Furthermore, a wide array of extended-spectrum beta-lactamases (ESBLs) were clinically exclusive, including *blaCTX-M-202*, *blaTEM-1*, and notably *blaPER-7* (detected in 6.4% of isolates).

Finally, clinical isolates exhibited an elevated resistance gene burden within shared categories, demonstrating higher ACNs for fluoroquinolones (1.92 ± 0.28 vs. 1.50 ± 0.59 in environmental strains) and phenicols (1.70 ± 0.47 vs. 1.33 ± 0.48). Together, the elevated genomic burden and the presence of clinically restricted MLS and ESBL determinants signal emerging resistance to azithromycin and broad-spectrum beta-lactams, threatening alternative and critically important therapeutic options for severe cholera cases (Figure 4).

Discussion

This study provides the most comprehensive genomic characterization to date of *V. cholerae* across the African continent. The population structure of *V. cholerae* in Africa is defined by the overwhelming

dominance of ST69, reflecting a pattern of extreme clonal expansion within the 7th pandemic El Tor (7PET) lineage. This aligns with the global paradigm of 7PET-driven cholera [47]. Recent studies in Asia, particularly in historically endemic regions such as Bangladesh, also reported that the majority of the seventh pandemic El Tor (7PET) *V. cholerae* O1 strains belong to ST69 [48]. In Europe, cholera is no longer endemic, and most cases are imported or travel-associated, with limited or no sustained local transmission. Genomic studies of European isolates typically reflect high diversity linked to multiple international sources, rather than local clonal expansion [49].

The secondary lineage ST515 in Africa further illustrates these continental contrasts. While Asia often harbors multiple co-circulating lineages with varying degrees of epidemic potential, ST515 remains geographically restricted and temporally episodic [50], largely confined to Central [51] and East Africa [52,53]. Its limited spread suggests that, unlike in Asia, where lineage competition and replacement are common, the African cholera ecosystem is dominated by competitive exclusion, with ST69 maintaining a strong selective advantage.

The genomic analysis of *V. cholerae* isolates across African regions reveals a pronounced divergence between clinical and environmental reservoirs, underscoring the ecological and epidemiological complexity of cholera transmission on the continent. A key finding is the overwhelming dominance of the O1 serogroup (rfbV) among clinical isolates, consistently exceeding 90% across all regions and reaching near fixation in Central Africa (99.2%). These genomic patterns observed in African *V. cholerae* isolates reveal strong similarities with, but also important distinctions from, other global regions, including Asia, Europe, and the Americas. Across all settings, a common theme emerges: epidemic cholera is driven by a highly conserved subset of pathogenic lineages, while environmental reservoirs harbor substantial genetic diversity. Asia differs in one critical aspect: the frequent detection of atypical El Tor variants and hybrid strains combining classical and El Tor traits [54,55]. These hybrid genotypes, often associated with increased virulence or altered toxin production, are more commonly reported in Asia than in Africa, where classical markers (rstR_cc) remain rare and geographically limited [54]. This suggests that Africa may experience less frequent recombination events or may have different selective pressures shaping strain evolution. In the Americas, particularly following the introduction of cholera in Haiti in 2010, genomic analyses have demonstrated a strong link to South Asian El Tor strains [56]. Much like in Africa, clinical isolates in the Americas are overwhelmingly dominated by O1 El Tor lineages, reflecting the global spread of a relatively uniform pandemic clone [57].

The present analysis highlights pronounced regional heterogeneity in CTX prophage carriage and ctxB allelic distribution among *V. cholerae* isolates across Africa, reinforcing the dynamic nature of cholera epidemiology on the continent. A central finding is the near-universal presence of intact CTX prophage elements in clinical isolates. The high co-carriage rates of ctxA and ctxB genes—exceeding 91% across all regions and reaching complete prevalence in North Africa—underscore the essential role of the CTX Φ prophage in pathogenicity. This observation aligns with established evidence that cholera toxin production, encoded by these genes, remains the primary virulence determinant driving epidemic

disease [51,58]. The particularly high prevalence in Central and North Africa suggests strong selective pressure favoring toxigenic strains [59].

The geographic partitioning of *ctxB* allelic variants provides further insight into cholera evolution and transmission dynamics. The dominance of *ctxB* variant 1 (classical biotype) in Central Africa contrasts sharply with the predominance of variant 7 (El Tor biotype) in West, East, and Southern Africa. This pattern is consistent with the global replacement of classical strains by El Tor variants, particularly the emergence of altered El Tor strains carrying classical toxin genes [60]. However, the persistence of classical *ctxB* alleles in Central Africa suggests either historical lineage retention or localized evolutionary trajectories.

The differential distribution of CTX prophage carriage and *ctxB* variants across Africa underscores the need for regionally tailored cholera control measures. Continued genomic surveillance will be essential to monitor lineage dynamics, detect emerging variants, and inform targeted interventions.

The high prevalence of intact CTX prophage in African clinical isolates mirrors observations from cholera-endemic regions in Asia, such as Bangladesh and India, and Russia, where toxigenic *V. cholerae* strains consistently dominate clinical cases [61–63]. In Asia, the epicenter of cholera evolution, a transition from classical to El Tor and subsequently to altered El Tor strains has been well documented [60]. In the Americas, cholera outbreaks have largely been attributed to imported El Tor strains, predominantly carrying *ctxB7* [56]. The geographic structuring of *ctxB* alleles observed in Africa reflects broader global evolutionary trends [51,55,64,65].

The major finding of this study is the extensive and deeply entrenched antimicrobial resistance across African *V. cholerae* populations. The genomic analysis of 3,076 *V. cholerae* isolates from Africa reveals an alarming antimicrobial resistance (AMR) landscape, characterized by both high gene diversity and widespread chromosomal resistance. With 101 distinct AMR genes across 9 categories and an average of 10.81 genes per isolate, the dataset reflects one of the most heavily burdened regional resistomes reported to date. The finding that 94.1% of isolates are multidrug-resistant (MDR) underscores a critical public health challenge for cholera management across the African continent. The near-universal presence of chromosomal fluoroquinolone resistance mutations, combined with widespread carriage of SXT/R391 integrative conjugative elements, indicates that multidrug resistance is now a defining feature of circulating epidemic strains. This has serious implications for treatment, particularly given the reliance on ciprofloxacin and azithromycin for severe cholera cases [66]. Fluoroquinolones, particularly ciprofloxacin, are recommended by the World Health Organization as first-line treatment for severe cholera [67]. The widespread baseline resistance observed here suggests that standard treatment protocols may be increasingly ineffective in African settings, raising concerns about treatment failure and prolonged transmission. Asia—particularly regions like Bangladesh and India—has historically been a hotspot for cholera and AMR [50,54]. The extensive AMR burden observed in African *V. cholerae* isolates signals a critical turning point in cholera management. Unlike other regions where resistance is still evolving, Africa shows a deeply entrenched resistance landscape, driven by both chromosomal

mutations and mobile genetic elements. Addressing this challenge will require coordinated global action, integrating antimicrobial stewardship, genomic surveillance, and strengthened public health infrastructure.

The dataset used in this study was derived from publicly available genomic sequences, which may be subject to sampling bias, with overrepresentation of certain countries and outbreak periods. There is also limited environmental sampling that restricts the ability to fully assess environmental reservoirs and transmission pathways. Moreover, phenotypic antimicrobial susceptibility testing was not performed, and resistance was inferred solely from genomic data. Finally, incomplete metadata for some isolates, including missing collection dates and clinical details, may limit the precision of temporal and epidemiological inferences.

Conclusion

This study demonstrates that cholera transmission across Africa is driven by the clonal expansion of a dominant multidrug-resistant *V. cholerae* lineage with sustained cross-border spread. The high prevalence of antimicrobial resistance and conserved virulence determinants underscores the growing challenge of managing cholera outbreaks in the region. Strengthening genomic surveillance and integrating real-time sequencing into public health systems will be critical for tracking transmission, guiding treatment strategies, and achieving the WHO target of cholera elimination by 2030.

Declarations

Funding

This study received no specific external funding.

Availability of data and materials

The dataset supporting the conclusions of this article is included within the article and its supplementary files. Genome sequence data of *Vibrio cholerae* isolates are available in NCBI Sequence Read Archive and NCBI GenBank. All accession numbers are listed in supplementary Table S1.

Acknowledgement

The authors acknowledge the use of publicly available genomic datasets and computational resources that supported this work.

Ethics approval and consent to participate

This study involve analysis of publicly available datasets. No patient specimens were used and patient-protected health information was not collected. Therefore, informed consent was not required.

Consent for publication

Not applicable.

Conflict of Interest

The authors declare no conflicts of interest.

Authors Contribution

AO and YM contributed equally to this work. AO: Conceptualization, Methodology, Formal analysis, Data curation, Writing - Original draft preparation. YM: Conceptualization, Methodology, Formal analysis, Data curation, Writing - Original draft preparation. AY: Software, Validation, Visualization, Writing - Review & editing. AJ: Software, Validation, Investigation, Writing - Review & editing. RFA: Investigation, Resources, Writing - Review & editing. BOO: Supervision, Project administration, Conceptualization, Writing - Review & editing. All authors read and approved the final manuscript.

References

1. Chowdhury F, Ross AG, Islam MT, McMillan NAJ, Qadri F. Diagnosis, management, and future control of cholera. *Clin Microbiol Rev* 2022; 35: e00211-21.
2. Ali M, Nelson AR, Lopez AL, Sack DA. Updated global burden of cholera in endemic countries. *PLoS Negl Trop Dis* 2015; 9: e0003832.
3. Kapaya F, et al. An assessment of the progress made in the implementation of the regional framework for cholera prevention and control in the WHO African region. *BMJ Glob Health* 2025; 10. doi:10.1136/bmjgh-2024-017xxx.
4. Enserink M, Ochoa TJ, Pham NTK, et al. Global patterns of trends in cholera mortality. *Trop Med Infect Dis* 2023; 8: 169.
5. Global Alliance Against Cholera. WHO reports surge in cholera cases in 2023, surpassing the numbers recorded in 2022. 2024. <https://www.choleraalliance.org/en/news/who-reports-surge-cholera-cases-2023-surpassing-numbers-recorded-2022> (accessed April 6, 2026).
6. World Health Organization Regional Office for Africa. Cholera in the WHO African region. 2026. <https://www.afro.who.int/health-topics/disease-outbreaks/cholera-who-african-region> (accessed April 6, 2026).
7. Bekele BK, et al. Cholera in Africa: a climate change crisis. *J Epidemiol Glob Health* 2025; 15: 68.
8. Idoga PE, Toyacan M, Zayyad MA. Analysis of factors contributing to the spread of cholera in developing countries. *Eurasian J Med* 2019; 51: 121–27.
9. Giroto CD, et al. Analysis of environmental factors influencing endemic cholera risks in sub-Saharan Africa. *Sci Total Environ* 2024; 926: 171896.
10. Perez-Saez J, et al. Geographical shifting of cholera burden in Africa and its implications for disease control. *Nat Med* 2025; 31: 3380–87.

11. Smirnova NI, et al. Genomic diversity of toxigenic strains of *V. cholerae* O1 biovar El Tor isolated during three waves of the 7th cholera pandemic. Dokl Biochem Biophys 2025; 525: 593–604.
12. Mukhopadhyay AK, Takeda Y, Balakrish Nair G. Cholera outbreaks in the El Tor biotype era and the impact of the new El Tor variants. In: Nair GB, Takeda Y, eds. Cholera outbreaks. Berlin: Springer, 2014: 17–47.
13. Heidelberg JF, et al. DNA sequence of both chromosomes of the cholera pathogen *V. cholerae*. Nature 2000; 406: 477–83.
14. Frontiers in Microbiology. Genomic characterization of *V. cholerae* isolated from clinical and environmental sources during the 2022–2023 cholera outbreak in Kenya. Front Microbiol 2025. doi:10.3389/fmicb.2025.1603736.
15. Mboowa G, et al. Multicountry genomic analysis underscores regional cholera spread in Africa. Nat Commun 2026; 17: 2539.
16. Xiao S, et al. Whole genome sequencing and transmission analysis of *V. cholerae* isolates from eastern and southern Africa: a genomic epidemiology study. medRxiv 2025. doi:10.1101/2024.03.28.24302717.
17. Hochhut B, Waldor MK. Site-specific integration of the conjugal *V. cholerae* SXT element into prfC. Mol Microbiol 1999; 32: 99–110.
18. De R. Mobile genetic elements of *V. cholerae* and the evolution of its antimicrobial resistance. Front Trop Dis 2021; 2. doi:10.3389/ftd.2021.691604.
19. Weill F-X, et al. Genomic history of the seventh pandemic of cholera in Africa. Science 2017; 358: 785–89.
20. Rouard C, et al. Long-distance spread of a highly drug-resistant epidemic cholera strain. N Engl J Med 2024; 391: 2271–73.
21. Foster-Nyarko E, et al. Genomic diversity and antimicrobial resistance of *V. cholerae* isolates from Africa: a PulseNet Africa initiative using nanopore sequencing to enhance genomic surveillance. Microb Genomics 2025; 11: 001586.
22. Amisu BO, et al. Cholera resurgence in Africa: assessing progress, challenges, and public health response towards the 2030 global elimination target. Infez Med 2024; 32: 148–56.
23. Xiao S, et al. Whole genome sequencing and transmission analysis of *V. cholerae* isolates from eastern and southern Africa: a genomic epidemiology study. medRxiv 2025. doi:10.1101/2024.03.28.24302717.
24. Schoch CL, et al. NCBI taxonomy: enhanced access via NCBI datasets. Nucleic Acids Res 2025; 53: D1711–17.
25. Heldenbrand J, Ren Y, Asmann Y, Mainzer LS. Step-by-step guide for downloading very large datasets to a supercomputer using the SRA Toolkit. protocols.io 2017. doi:10.17504/protocols.io.kb6csre.

26. Sayers E. A general introduction to the E-utilities. In: Entrez programming utilities help. Bethesda: National Center for Biotechnology Information, 2010.
27. Bankevich A, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; 19: 455–77.
28. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; 30: 2068–69.
29. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; 29: 1072–75.
30. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023; 20: 1203–12.
31. Chun J, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018; 68: 461–66.
32. Heidelberg JF, et al. DNA sequence of both chromosomes of the cholera pathogen *V. cholerae*. *Nature* 2000; 406: 477–83.
33. Connor CH, et al. Rapid, reference-free identification of bacterial pathogen transmission using optimized split k-mer analysis. *Microb Genomics* 2025; 11: 001347.
34. Croucher NJ, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015; 43: e15.
35. Seemann T. snp-dists: pairwise SNP distance matrix from a FASTA sequence alignment. Zenodo 2018. doi:10.5281/zenodo.1411986.
36. Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020; 37: 1530–34.
37. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; 14: 587–89.
38. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018; 35: 518–22.
39. Xu S, et al. ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol Biol Evol* 2021; 38: 4039–42.
40. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 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.
41. Siriphap A, et al. Characterization and genetic variation of *V. cholerae* isolated from clinical and environmental sources in Thailand. *PLoS One* 2017; 12: e0169324.
42. Feldgarden M, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 2019; 63: e00483-19.
43. Abdulgader SM, et al. Genomic epidemiology of antimicrobial resistance in *Proteus mirabilis*: core genome and plasmid-mediated drivers. *BMC Microbiol* 2025; 25. doi:10.1186/s12866-025-04651-8.

44. Magiorakos A-P, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect 2012; 18: 268–81.
45. Pfeifer E, et al. PlasForest: a homology-based random forest classifier for plasmid detection in genomic datasets. BMC Bioinformatics 2021; 22: 349.
46. Camargo AP, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol 2023; 41: 1723–31.
47. Ramamurthy T, et al. *V. cholerae* O139 genomes provide a clue to why it may have failed to usher in the eighth cholera pandemic. Nat Commun 2022; 13: 3864.
48. Jubyda FT, et al. *V. cholerae* O1 associated with recent endemic cholera shows temporal changes in serotype, genotype, and drug-resistance patterns in Bangladesh. Gut Pathog 2023; 15: 17.
49. Bhandari M, et al. Genomic and evolutionary insights into Australian toxigenic *V. cholerae* O1 strains. Microbiol Spectr 2022; 11: e03617-22.
50. Ayyappan MV, et al. Emergence of multidrug resistant, ctx negative seventh pandemic *V. cholerae* O1 El Tor sequence type (ST) 69 in coastal water of Kerala, India. Sci Rep 2024; 14: 2031.
51. Ireng LM, et al. Genomic analysis of pathogenic isolates of *V. cholerae* from eastern Democratic Republic of the Congo (2014–2017). PLoS Negl Trop Dis 2020; 14: e0007642.
52. Hounmanou YMG, et al. Genomic insights into *V. cholerae* O1 responsible for cholera epidemics in Tanzania between 1993 and 2017. PLoS Negl Trop Dis 2019; 13: e0007934.
53. Hounmanou YMG, et al. Genomic microevolution of *V. cholerae* O1, Lake Tanganyika basin, Africa. Emerg Infect Dis 2023; 29: 149–53.
54. Monir MM, et al. Genomic characteristics of recently recognized *V. cholerae* El Tor lineages associated with cholera in Bangladesh, 1991 to 2017. Microbiol Spectr 2022; 10: e00391-22.
55. Jubyda FT, et al. *V. cholerae* O1 associated with recent endemic cholera shows temporal changes in serotype, genotype, and drug-resistance patterns in Bangladesh. Gut Pathog 2023; 15: 17.
56. Ayyappan MV, et al. Presence of unique clone of seventh pandemic *V. cholerae* O1 El Tor with Haitian cholera toxin (7PET-HCT) in seafood. EMI Anim Environ 2025; 1: 2469919.
57. Johura F-T, et al. *V. cholerae* O1 El Tor strains linked to global cholera show region-specific patterns by pulsed-field gel electrophoresis. Infect Genet Evol 2022; 105: 105363.
58. Li X, et al. Diversity and complexity of CTX Φ and pre-CTX Φ families in *V. cholerae* from seventh pandemic. Microorganisms 2024; 12: 1935.
59. Foster-Nyarko E, et al. Genomic diversity and antimicrobial resistance of *V. cholerae* isolates from Africa: a PulseNet Africa initiative using nanopore sequencing to enhance genomic surveillance. Microb Genomics 2025; 11: 001586.
60. Smirnova NI, et al. Genomic diversity of toxigenic strains of *V. cholerae* O1 biovar El Tor isolated during three waves of the 7th cholera pandemic. Dokl Biochem Biophys 2025; 525: 593–604.

61. Islam MT, et al. *V. cholerae* O47 associated with a cholera-like diarrheal outbreak concurrent with seasonal cholera in Bangladesh. *mSphere* 2025; 10: e00831-24.

62. Havelikar U, et al. Recent epidemiological and management approaches for cholera in India. *Proc (Bayl Univ Med Cent)* 2026; published online ahead of print. doi:10.1080/08998280.2026.xxxxxxx.

63. Smirnova NI, et al. New genetic variants of the cholera agent and their distribution in endemic countries and Russia. *Mol Genet Microbiol Virol* 2023; 38: 8–15.

64. Rouard C, et al. Genomic analysis of *V. cholerae* O1 isolates from cholera cases, Europe, 2022. *Euro Surveill* 2024; 29: 2400069.

65. Smirnova NI, Rybal'chenko DA, Lozovsky YuV, Kutyrev VV. Genome instability of the cholera agent: role in the evolution of pathogenicity, drug resistance, and adaptation to a changing environment. *Mol Genet Microbiol Virol* 2025; 40: 91–104.

66. Chowdhury F, Ross AG, Islam MT, McMillan NAJ, Qadri F. Diagnosis, management, and future control of cholera. *Clin Microbiol Rev* 2022; 35: e00211-21.

67. Global Task Force on Cholera Control. Technical note on the use of antibiotics for the treatment of cholera. Geneva: World Health Organization, 2018.

Table 1

Table 1 is available in the Supplementary Files section.

Figures

Image not available with this version

Figure 1

Geographic distribution, temporal trends, and sequence type dynamics of *V. cholerae* isolates across Africa, 1949–2025. **(A)** Geographic distribution of *V. cholerae* isolates across Africa. The choropleth map displays the total number of genomes per country, with colour intensity proportional to isolate count. Overlaid pie charts represent the proportion of isolates by source category (clinical, environmental, and unknown). **(B)** Geographic distribution of isolates stratified by multilocus sequence types (MLST). Pie charts illustrate the relative contribution of dominant sequence types across African regions, highlighting the widespread distribution of ST69. **(C)** Temporal distribution of sequenced isolates from 1949 to 2025. Bars represent the annual number of isolates, stacked by source category (clinical, environmental, and unknown), demonstrating increased sequencing activity in recent outbreak years. **(D)**

Temporal distribution of MLST sequence types. Stacked bar plots show the yearly distribution of major sequence types, illustrating the sustained dominance of ST69 over time, particularly during the 2022–2024 outbreak period.

Image not available with this version

Figure 2

Genomic location, profile, and prevalence of antimicrobial resistance (AMR) genes in African *V. cholerae* isolates. (A) Ranked prevalence of detected AMR genes across the analyzed isolate population. Bars are colored according to their respective resistance class, corresponding to the legend in Panel B. (B) Circular representation of the identified AMR gene repertoire. The inner ring represents the overarching antimicrobial resistance class, while the outer ring visually differentiates individual AMR genes and point mutations. (C) Genomic location of AMR genes across chromosomal, plasmid, and prophage contexts.

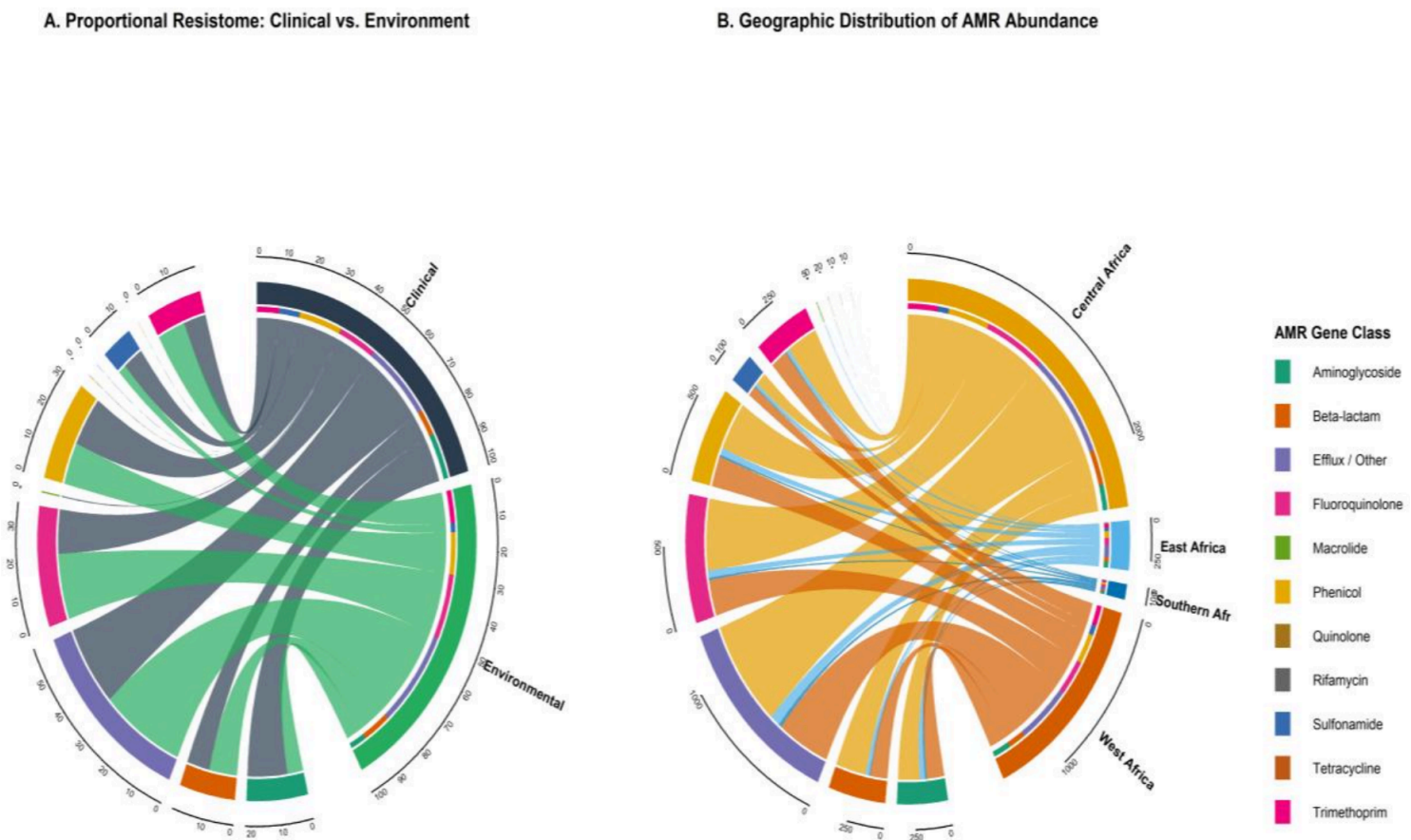


Figure 3

Structure and geographic distribution of the *V. cholerae* resistome. (A) Proportional representation of antibiotic resistance gene (ARG) classes across clinical and environmental reservoirs. Ribbon width represents the percentage contribution of each ARG class to the total reservoir resistome. (B) Absolute ARG counts distributed across five African sub-regions. Ribbon width is proportional to raw gene counts. Sector colours denote antibiotic class (see legend), reservoir origin (Panel A), or geographic region (Panel B). Directionality (height difference at chord ends) indicates asymmetric ARG sharing. Colour palette validated for colour-vision deficiency.

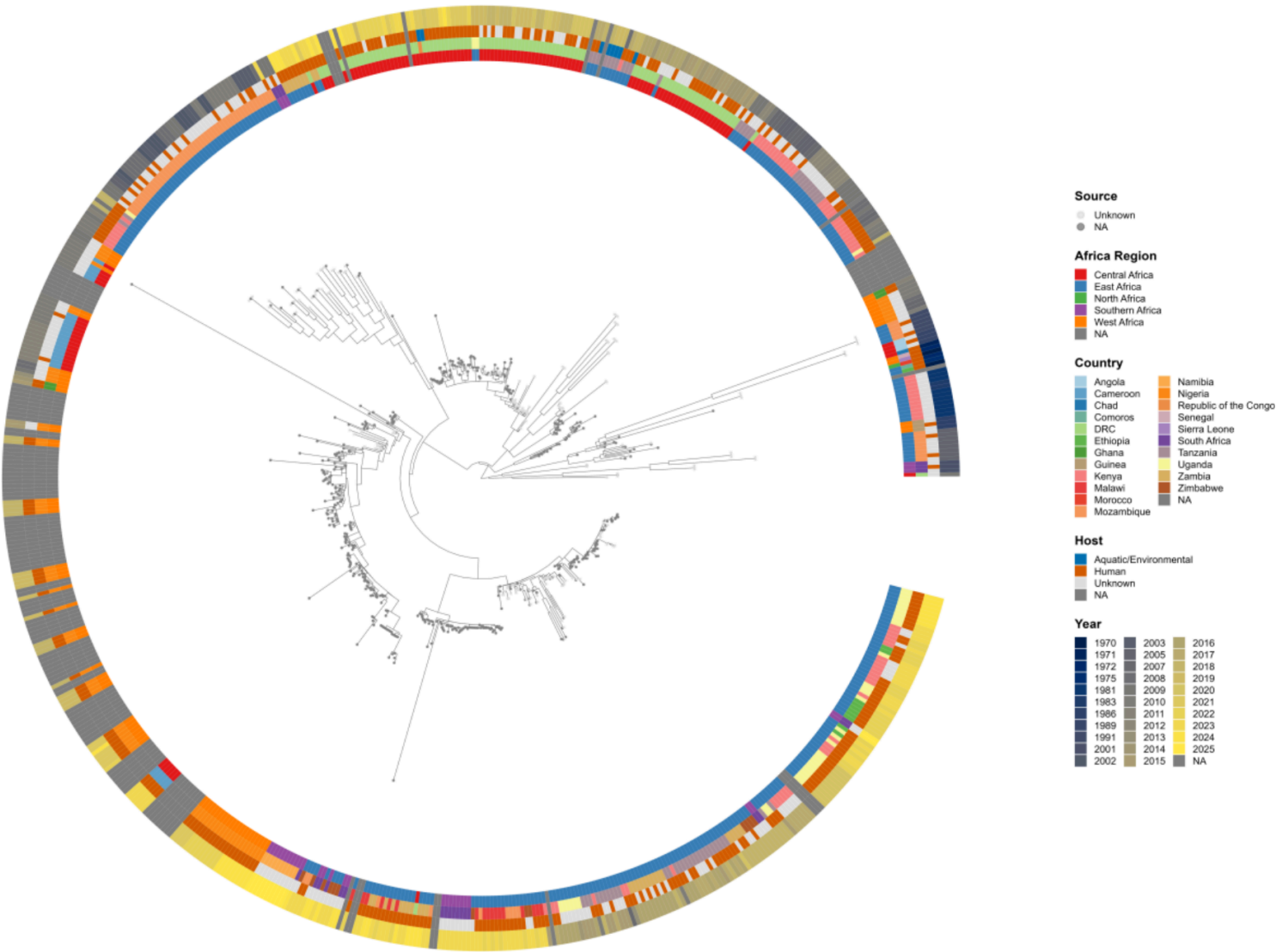


Figure 4

Genomic epidemiology of *V. cholerae* across the African continent. Midpoint-rooted maximum-likelihood phylogenetic tree based on core single nucleotide polymorphisms. Metadata rings (innermost to outermost) represent: Africa Region, Country, Host, and Collection Year. Missing or unavailable metadata is designated in light gray.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplMat.xlsx](#)
- [Table1.docx](#)