Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome dynamics of Vibrio mimicus in cholera-endemic regions of Bangladesh

Lia Bote

lb34@sanger.ac.uk

Wellcome Sanger Institute https://orcid.org/0009-0003-2261-4249

Stephanie McGimpsey

Wellcome Sanger Institute

Mohammad Tarequl Islam

icddr,b

Abdus Sadique

icddr,b (International Center for Diarrhoeal Diseases Research, Bangladesh)

Md Mamun Monir

icddr,b

Marzia Sultana

icddr,b

Nisha Singh

Wellcome Sanger Institute https://orcid.org/0000-0002-4140-5560

Chetan Thakur

Post Graduate Institute of Medical Education & Research (PGIMER)

Neelam Taneja

Postgraduate Institute of Medical Education & Research, Chandigarh https://orcid.org/0000-0003-1198-6138

Nur A. Hasan

icddr,b (International Center for Diarrhoeal Diseases Research, Bangladesh)

Anwar Huq

University of Maryland, College Park

Rita Colwell

University of Maryland Centre for Bioinformatics & Computational Biology https://orcid.org/0000-0001-5432-1502

Munirul Alam

icddr,b (International Center for Diarrhoeal Diseases Research, Bangladesh) https://orcid.org/0000-0001-9157-5560

Nicholas Thomson

The Wellcome Trust Sanger Institute https://orcid.org/0000-0002-4432-8505

Article

Keywords:

Posted Date: June 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9904606/v1

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

Additional Declarations: There is **NO** Competing Interest.

Abstract

Vibrio mimicus is a Gram-negative, marine bacterium that can cause diarrhoea and gastroenteritis. It is closely related to *Vibrio cholerae*, the aetiological agent of cholera. *V. mimicus* is considered an emerging human pathogen, with evidence of it carrying the canonical virulence genes in pandemic *V. cholerae*: cholera toxin (CTX) and toxin-coregulated pilus (TCP). Despite this, the presence, persistence, and evolution of *V. mimicus* in a cholera-endemic setting is unknown. Here, we present genomic analysis of *V. mimicus* isolated from environmental sources between 2005–2016 across cholera-endemic regions of Bangladesh. We contextualise these genomes with the global *V. mimicus* phylogeny and uncover *V. mimicus* as a diverse environmental species, with Bangladeshi genomes forming many distinct, deeply branching clades. CTX and TCP were absent from *V. mimicus* in Bangladesh and instead acquired independently in two clades encompassing closely-related, spatiotemporally dispersed genomes. The genomic islands carrying CTX and TCP were nearly identical in *V. mimicus* and *V. cholerae*, suggesting recent acquisition, and absence of these elements in the Bangladeshi genomes raises questions about niche sharing between these species. Taken together, this study sheds light into the genome dynamics of *V. mimicus* in a cholera-endemic context, and the continued evolutionary success of pandemic *V. cholerae*.

Introduction

Vibrio mimicus is a Gram-negative, aquatic bacterium that is a member of the *Vibrionaceae* family within *Gammaproteobacteria*. It was first identified as an atypical member of *Vibrio cholerae*, but redefined, using biochemical tests, as a novel species and differentiated by negative sucrose, Voges-Proskauer, corn oil, and Jordan tartrate reactions[1]. It was confirmed with advancements in whole genome sequencing as a novel species closely related to *V. cholerae*[2]. *V. mimicus* has been isolated in environmental sources, clinical cases of diseases such as diarrhoea and otitis media, and in seafood and shellfish infections. Disease encompasses a broader range of symptoms than cholera, such as diarrhoea occasionally accompanied by fever, cramps, vomiting, nausea, and bloody stool[3–6].

The canonical virulence factors of pandemic cholera caused by *V. cholerae*, namely cholera toxin (CTX) and toxin-coregulated pilus (TCP), have also been described in a small number of clinical and environmental *V. mimicus* isolates[3,7]. CTX and TCP are responsible for secretory diarrhoea and intestinal colonisation, respectively. These genes were found to be carried on the same genomic islands in *V. mimicus* as in *V. cholerae*, with some variation in sequence identity[2,8]. CTX is encoded by *ctxAB* genes, which are carried on the lysogenic bacteriophage CTXφ that includes a core and a Repeat Sequence 2 (RS2) module[9,10]. Variations in CTXφ, and notably in the *ctxB* gene, are linked to different cholera epidemics[11]. Similarly, TCP is encoded by an operon of genes that includes the major pilin subunit gene *tcpA*, which is carried on a genomic island called *Vibrio* pathogenicity island 1 (VPI-1). VPI-1 can be transferred between *V. cholerae* strains and has been shown to excise itself from the bacterial chromosome[12,13]. Other genomic islands in *V. cholerae* include VPI-2, which encodes genes for sialic

acid metabolism, and *Vibrio* seventh pandemic islands I and II (VSP-I, VSP-II), which include phage defence genes. These genomic islands are largely confined to isolates of the seventh, ongoing pandemic of *V. cholerae*, although variants of VPI-2 are also found throughout the species[14,15]. Previous work has detailed the assembly of these genomic islands in one clinical and one environmental *V. mimicus* isolate[2]. *V. mimicus* has also been found to carry other *V. cholerae* virulence factors, such as haemolysins, type III secretion systems, accessory cholera enterotoxin (*ace*), and zonula occludens toxin (*zot*)[2].

Sporadic disease associated with both toxigenic and non-toxigenic strains of *V. mimicus* have been recorded in some countries, including Thailand, China, and the United States[4–6]. Whilst *V. mimicus* has never been associated with large-scale epidemics, there have been a small number of reported outbreaks in humans associated with ingestion of seafood and shellfish[2,3,16,17] and in animals linked to intensive aquaculture[5,6]. Previous studies have suggested *V. mimicus* and other environmental *Vibrio* species to be important emerging human pathogens, linked with climatic factors[18–21]. For example, the USA has seen a latitude increase for *V. mimicus* of ~ 42 km/year, influenced largely by temperature and salinity[22]. In contrast, the current cholera pandemic has been reported through whole genome sequencing to be dominated by a single, toxigenic lineage of *V. cholerae*, referred to as seventh pandemic El Tor (7PET)[23]. The population structure of *V. mimicus*, including distribution and mobility of key virulence genes and emergence of pathogenicity within the species, remains to be fully understood.

Here, we present genome sequence data of *V. mimicus* isolates collected from the Bakergonj, Chhatak, and Mathbaria subdistricts in Bangladesh between 2005 and 2016. We compared this with all publicly available *V. mimicus* genomes to describe the phylogenetic structure and the distribution of genes of interest across the species. Understanding the population diversity of *V. mimicus* in the cholera-endemic niche of Bangladesh allows us to explore both the evolution of pandemic *V. cholerae* and the potential emergence of other pathogenic *Vibrio* species.

Results

Vibrio mimicus species phylogeny

We performed whole genome sequencing of 53 *V. mimicus* isolates obtained from various environmental sources including water, sediment, and plankton. The isolates were collected between 2005 and 2016 across three subdistricts of Bangladesh: Bakergonj, Chhatak, and Mathbaria. Most of the isolates were collected from Mathbaria (42), followed by Chhatak (6), and Bakergonj (5) (Supplementary Data). We inferred a phylogeny from these 53 genomes alongside 108 publicly available *V. mimicus* genomes, based on single nucleotide polymorphisms (n = 454147 SNPs) in core genes (n = 3049 genes) (Fig. 1) present in 95% or more of included genomes. The publicly available *V. mimicus* genomes were from Bangladesh, Brazil, USA, China, and Japan, collected between 1977 and 2025. Of these, 28 were from a variety of human clinical sources, namely diarrhoea, wound, and ear swabs. Twenty-two were from non-

clinical sources, e.g., environmental water, seafood, fish tanks, “ice display”, “utensils”, and “laboratory”. Here, we refer to these diverse non-clinical isolates as “environmental”. Fifty-seven publicly available genomes had no specified source of isolation.

The 161 assembled genomes displayed considerable genetic diversity across the species, with deep branches in the tree and genomes differing by a mean of 67,000 core gene SNPs. The resulting phylogenetic tree was partitioned through hierarchical Bayesian analysis of population structure (BAPS), which uses nucleotide frequencies to infer genetic structure and histories within a population, resulting in 22 phylogenetic clades. The genomes from this study fell within 13 clades, with four of these only containing genomes from this study [clades 3 (n = 3), 11 (n = 4), 12 (n = 4), and 16 (n = 4)]. The remaining 38 genomes from this study were distributed across nine clades (5, 7, 9, 10, 13, 14, 15, 17, 19) that also included genomes from diverse geographic and temporal sources. Interestingly, where metadata were available, five of the clades contained genomes from both clinical and environmental samples. For example, clade 13 was composed of three genomes from this study, as well as eight genomes isolated from patient stool (USA, 2024), marine shellfish (Brazil, 1998), or unknown sources. Clade 15 included 10 genomes from this study and 17 isolated from wound specimens, fish, etc.

Within the *V. mimicus* phylogeny, we observed a few clonal clades: clades 1, 2, and 6 with intra-clade diversity of 482, 15, and 32 mean SNPs, respectively. Moreover, clade 1 and 6 were geographically and temporally dispersed, both comprising genomes from China and the USA, with isolates in clade 1 collected between 1980 and 2023, and those in clade 6 collected between 1977 and 2020. The genomes for which metadata were available had been isolated from patient diarrhoea, ear infection, and laboratory strains. Clade 2 comprised 13 genomes from the USA between 2013 and 2025 from patient samples. Clade 8 contained seven closely related genomes sequenced in the USA from 2019 to 2023, which varied by a mean of 12 SNPs.

Presence and distribution of antimicrobial resistance and virulence genes

We searched for antimicrobial resistance (AMR) and virulence genes across these 161 *V. mimicus* genomes. We only found two AMR genes, *QnrVC4* and *QnrVC7*, which both confer fluoroquinolone resistance (Fig. 1). *QnrVC4* was present in 65/161 (40.4%) of the genomes, including in 28 from Bangladesh, whereas *QnrVC7* was present in only 2/161 genomes - one sequenced from a water sample collected in Japan in 1990 and the other isolated in China in 2019 - both belonging to clade 18. None of the *V. mimicus* genomes were found to contain plasmids.

We found two common virulence genes in *V. mimicus: cqsA* (160/161, 99.4%) and *luxS* (161/161, 100%), both of which are involved in quorum sensing and associated with swarming and biofilm formation. Type III secretion system (T3SS) genes were present in 44/161 (27.3%) of the *V. mimicus* genomes, with 33 of these containing all of the core-encoded structural genes previously identified in *V. cholerae* T3SS[24]. None of the genomes containing a T3SS encoded CTX or TCP. In contrast, type VI secretion system

(T6SS) genes, linked to inter-bacterial competition[25], were present in 159/161 (98.8%) of the *V. mimicus* genomes, with 143 encoding at least one of the effector-immunity auxiliary gene clusters previously described in *V. cholerae*[26].

A significant finding was that the genes encoding CTX and TCP were detected in 22 and 23 of the genomes, respectively, with 20/161 (12.4%) genomes containing both CTX and TCP genes, 2/161 with CTX only, and 3/161 with TCP only. CTX is encoded by two genes, *ctxA* and *ctxB*, that are both present in all of the toxigenic isolates identified here. CTX and TCP are canonical virulence determinants, conserved in both the classical and 7PET epidemic lineages of *V. cholerae*. In *V. mimicus*, these toxigenic isolates all belonged to BAPS clades 1 and 8, collected from China and the USA. We determined the *ctxB* allele carried by these genomes and found that nine genomes had an identical sequence to *V. cholerae ctxB1* and 11 to *ctxB2*. Two genomes had a *ctxB* allele that varied from *ctxB2* by a single-base deletion within a homopolymeric region in the 23rd nucleotide position, introducing a premature stop codon likely resulting in ablated function (Fig. S1). All genomes containing *ctxB2* belonged to clade 1, while genomes with *ctxB1* were in both clades 1 and 8 (Fig. 2).

Where metadata were available, the genomes carrying genes for CTX and TCP were all collected from diarrhoeal patients or human sources. No toxigenic genomes were detected in environmental samples. In contrast, 2/3 non-toxigenic genomes with TCP operon genes were isolated from water sources: one from Japan in 1990 and one from this study, collected in Mathbaria in 2013. Neither belonged to clonal clades and instead were in clade 14, which comprised genomes from this study as well as from environmental isolates collected in Brazil, Japan, and the USA.

Mobility of pathogenicity islands in V. mimicus

CTXφ bacteriophage

Since CTX and TCP genes were detected in two different, spatiotemporally dispersed clades, we next wanted to determine the genomic context of *ctxAB* and TCP genes to shed light on their mechanism of mobility. All of the *ctxAB* genes in *V. mimicus* were found to be carried on CTXφ, which shared 97.2% mean nucleotide identity across the entire CTXφ carried by 7PET reference strain N16961 (Fig. 2). Two additional toxins, *ace* and *zot*, were present on CTXφ and were found in all of the toxigenic *V. mimicus* isolates described here. The two environmental isolates that encoded TCP genes contained partial matches to CTXφ, which carried only *ace* and not *ctxAB*.

Analysis of the complete prophage elements revealed six of the genomes contained two copies of CTXφ tandemly repeated, while two others contained three tandem repeats of CTXφ. All of the CTXφ regions we found had both RS2 and CTX core modules, but not RS1 satellite phage. A phylogenetic tree generated from these sequences revealed three major clusters of CTXφ in *V. mimicus* (Fig. S2). CTXφ from genomes in clades 1 and 8 were distinct from each other. The third was present in two genomes from clade 1, which encoded *ctxB1* and a variant of *rstA* in the RS2 region. While CTXφ in *V. cholerae* is

inserted at the *rtxA* gene, we did not find *rtx* genes in any of the *V. mimicus* genomes described here. Furthermore, sequencing reads from toxigenic *V. mimicus* genomes did not map to the region of *rtx* genes flanking CTXφ in *V. cholerae*. From Fig. 2, it is evident that where there were repeat prophage copies carried in the same genome, the copies encoded the same *ctxB* allele, suggesting intraspecific duplication rather than multiple independent infections.

TCP operon

In *V. cholerae*, the TCP gene cluster is carried on the genomic island denoted *V. cholerae* pathogenicity island 1 (VPI-1)[27]. Fifteen *V. mimicus* genomes had TCP genes assembled in one contig with no overlapping gene annotations. We analysed these further to determine their genomic context (Fig. 3). All of the TCP operons were located within a region similar to VPI-1, which contained the 15 TCP genes, as well as the upstream and downstream regions encoding *aldA*, *tagA*, and accessory colonisation factor (*acf*) genes. These genes are flanked by a transposase and tRNA-Ser in *V. cholerae*. Here, we found different gene modules inserted between these sites and the TCP operon, suggesting this region to be a hotspot for horizontal gene transfer and insertion into the chromosome in both *Vibrio* species, as seen in other bacteria[28].

Like CTXφ, TCP operons in *V. mimicus* are highly conserved (> 96% nucleotide identity compared to *V. cholerae* reference strain N16961), and the phylogenetic tree constructed from the operon sequences showed that TCP operons were different between genomes in clades 1 and 8 (Fig. S3). What is unusual is the *tcpA* gene, which encodes the major structural subunit of TCP, differed in *V. mimicus* by 158 SNPs from *V. cholerae tcpA*, and yet the *tcpA* sequences appeared homoplastic, being identical in 13 isolates across clades 1 and 8. Moreover, site-heterogeneous nonsynonymous/ synonymous substitution rate ratios (dN/dS) suggested that 12 frequently mutated residues in TcpA were under positive selection (Fig. S4). The other two TCP clusters were from non-toxigenic, environmental isolate genomes from clade 14, which both contained partial matches to CTXφ. One carried *tcpA* (ERS17183197, this study) identical to that of *V. cholerae tcpA* reference, and the other varied by 155 SNPs (GCA_036553415.1). These two operons also possessed highly divergent *tcpF* genes, sharing < 39% amino acid sequence identity with the protein product of the *V. cholerae tcpF* gene, essential for efficient colonisation of the host intestine[29].

Gene sharing between V. mimicus and V. cholerae

The conserved CTXφ and TCP operon sequences between *V. mimicus* and *V. cholerae* suggest horizontal gene transfer between the two species. Given this, we also searched for key *V. cholerae* genomic islands in *V. mimicus*. Partial VPI-2 fragments were present in 151 genomes, although all were a mean of 14kb in length, compared to the entire 57 kb element previously described in 7PET *V. cholerae*. These fragments encoded the *nan-nag* cluster of genes involved in sialic acid metabolism. 7PET-specific genomic islands, VSP-I and VSP-II, were rare in *V. mimicus*. Four genomes shared partial matches to VSP-I, all of which were from the USA, and two to VSP-II, both of which were from environmental samples from Brazil

collected in the 1990s. Matches to VSP-I covered the entire cyclic oligonucleotide based anti-phage signaling system (CBASS) region, which encodes *capV*, *dncV*, *cap2*, and *cap3* genes involved in phage defence through abortive infection[30,31]. Fragments mapping to ~6kb of VSP-II encoded an AraC-like transcriptional activator, a methyl-accepting chemotaxis protein, an EAL domain protein, and a phage integrase near the tRNA integration site of VSP-II[32].

Given evidence of horizontal gene transfer, we then assessed gene sharing by calculating the combined pangenome of the 161 *V. mimicus* genomes included here and 144 *V. cholerae* genomes that represented the known diversity of the species. We found 8778 genes present only in *V. mimicus*, 5945 genes only in *V. cholerae*, and 6946 genes in both species. Of those unique to one species, 29 AMR genes were unique to *V. cholerae* and absent from *V. mimicus*, including genes for resistance to tetracyclines, macrolides, and aminoglycosides. None of the genes unique to *V. mimicus* had predicted virulence and AMR functions. While *V. mimicus* genomes were on average 9% larger and with 10% more CDSs than *V. cholerae* genomes, *V. mimicus* was not enriched across all functional gene ontologies compared to *V. cholerae*.

Instead, there were differences in the distribution of genes across COG categories between the species ($p = 0.0005$, Fisher's exact test). Namely, *V. mimicus* was enriched for genes belonging to COG categories G, L, V, and K (corresponding to Carbohydrate transport and metabolism; Replication, recombination and repair; Defence mechanisms; Transcription), while *V. cholerae* was enriched for genes belonging to COG categories T and P (Signal transduction mechanisms; Inorganic ion transport and metabolism), based on standardised residuals. When we reconstructed metabolic pathways based on these genes, we found that genes from both species spanned the same 32 unique KEGG pathway classes. Nonetheless, key genes were found to be core across both *V. mimicus* and *V. cholerae*. These include known virulence determinants such *cqsA* and *luxS*, involved in quorum sensing, biofilm formation and swarming, as well as T6SS clusters for bacterial competition.

Discussion

In this study, we provide a comprehensive population genomic analysis of *Vibrio mimicus*, integrating 53 newly sequenced environmental isolates from Bangladesh with publicly available genomes. Our findings bring together multiple different genomic studies to reveal a species characterised by deep phylogenetic structure, extensive environmental diversity, and the sporadic emergence of clonal toxigenic lineages. By placing genomes from Bangladesh in context of all publicly available *V. mimicus* genome data, we showed the species comprises 22 clades separated by notable core genome divergence. The Bangladeshi environmental isolates were found to be widely distributed across the phylogeny, occupying 13 distinct clades, four of which are unique to the region and this study. This pattern, together with the high mean core SNP distance (mean 67,000 SNPs or ~ 1.5% sequence variation across the whole genome), supports *V. mimicus* being a long-established, naturally diverse aquatic species rather than a recently expanding, globally distributed pathogen. Clades frequently contained isolates from multiple

continents and decades, and, based on available metadata, both derive from both environmental and clinical sources.

Interestingly, toxigenic *V. mimicus* isolates were strikingly restricted, as reported previously[2,3,8,33]. CTX and TCP genes were present in only 12.4% of genomes and were confined to clonal clades 1 and 8. Both the CTXφ and TCP operons in *V. mimicus* were very similar to those in pandemic *V. cholerae*, indicating they were recently acquired through horizontal gene transfer. However, as previously observed, we only found CTXφ in isolates from the USA and China[3], where cholera is not endemic. In contrast to previous molecular-based studies[34], no environmental isolate from Bangladesh carried *ctxAB*, a complete CTXφ prophage, or both CTX and TCP, despite the longitudinal sampling in this cholera-endemic region. This restricted distribution appears to mirror the clonal structure reported in epidemic *V. cholerae*, where toxigenicity is similarly concentrated largely within a specific lineage linked to human disease. However, unlike the 7PET lineage of *V. cholerae*, which is dominated by a single globally disseminated clone[23], clades of *V. mimicus* have not been associated with major outbreaks or shown evidence of global epidemic spread, suggesting possible ecological or genetic constraints on transmission.

CTXφ sequences were largely concordant with the whole genome phylogeny, with clades 1 and 8 encoding different CTXφ types that carried *ctxB2* and *ctxB1* alleles, respectively. The *ctxB1* allele is characteristic of the classical biotype *V. cholerae*, responsible for the previous six cholera pandemics, as well as the more recent 7PET cholera outbreaks in India (1995–2016) and linked to the expansion of BD-1 sublineage in Bangladesh (1999–2007)[35,36]. On the other hand, *ctxB2* is a rare variant previously detected in O1 serogroup, El Tor *V. cholerae* isolates from Australia[36,37], in contrast with *V. cholerae* where the *ctxB7* allele has largely replaced *ctxB1* in 7PET over the last two decades[38–40]. Notably, insertion occurred in the absence of the canonical *rtxA* locus described for *V. cholerae* CTXφ, implying alternative *attB* and *attP* sites for CTXφ in *V. mimicus*. The absence of the RS1 satellite phage further distinguishes *V. mimicus* CTXφ from that of the typical 7PET *V. cholerae* CTXφ.

Like CTXφ, the phylogeny of the TCP operon was largely concordant with that of their host genome. The exception being the *tcpA* gene, which displayed an unusual evolutionary pattern. That is, although divergent from the *V. cholerae* reference (~ 158 SNPs), identical *tcpA* sequences were observed across toxigenic isolates in the two distinct clades, consistent with homoplasy or recent shared acquisition. We found no evidence of this *V. mimicus tcpA* allele elsewhere in *V. cholerae*. Signals of positive selection in multiple TcpA residues suggest adaptive modification, potentially influencing host interaction or phage susceptibility. Of the SNPs differentiating *V. cholerae* from *V. mimicus* TcpA, only a few fell within the α1 and αβ-loop domains, known to be involved in the interactions between the individual TcpA monomers necessary for pilin filament formation[41], while most were present within the protruding, solvent exposed D-region[41]. Six of these residues also showed signals of being under positive selection. The TcpA D-region in *V. cholerae* promotes interaction between pilin monomers, leading to aggregation and microcolony formation, the latter being essential for attaching to the intestinal epithelium and establishing infection[42]. On the other hand, TcpB, which is present at the end of the TCP filament in *V.*

cholerae and directly mediates uptake of the CTXφ via interaction with pIII, was identical in all *V. mimicus* and the reference *V. cholerae* sequence[43]. These results suggest that *V. mimicus* TCP retains the ability to act as a receptor for CTXφ but may cause less efficient intestinal colonisation upon human infection. This interpretation agrees with previous work which found that toxigenic, TCP-carrying *V. mimicus* is severely attenuated for colonisation in the suckling mouse model compared to *V. cholerae*[4].

Beyond CTX and TCP, *V. mimicus* can host mobile genetic elements found in *V. cholerae*, evidenced by the mosaic presence of discrete genetic modules from VPI-2, VSP-I, and VSP-II, often at the same insertion sites as in *V. cholerae*. For example, the VPI-2 *nan-nag* sialic acid metabolism cluster was nearly ubiquitous, whereas VSP-I and VSP-II fragments were rare and incomplete. These patterns mirror observations in non-pandemic *V. cholerae*[15], and underscore the modular, recombinogenic nature of *Vibrio* genomes. More broadly, comparative pangenome analysis revealed notable gene sharing (6,946 genes) as well as large species-specific gene pools, for which 8,778 genes were detected only in *V. mimicus*, consistent with *V. mimicus* genomes being larger, encoding ~ 10% more CDSs. However, pathway reconstruction showed both species spanned the same KEGG classes, thus rather than expanded metabolic capability, *V. mimicus* appears to carry a higher degree of functional redundancy. This is often associated with the ability to thrive in more complex environments.[44]

The genomes from Bangladesh shed light on the pathogenic potential of environmental *V. mimicus*. Despite longitudinal sampling in a cholera endemic region, we found no complete CTXφ in isolates from Bangladesh. Instead, a single genome isolated from Bangladesh carried a partial CTXφ with *ace* and *zot* but lacking *ctxAB*. This resembled pre-CTXφ sequences previously seen in environmental, non-toxigenic *V. cholerae* and *V. mimicus* strains[2]. This genome also carried a complete TCP operon and was the only one across *V. mimicus* with a *tcpA* sequence identical to that found in 7PET *V. cholerae*. Only one other environmental isolate, collected from Japan in 1990, contained TCP genes and a partial CTXφ region that carried *zot* but not *ace* or *ctxAB*. *V. mimicus* from Bangladesh also lacked the 7PET islands, VSP-I and VSP-II. Thus, we did not observe the acquisition of complete 7PET-specific genomic islands in environmental *V. mimicus* genomes from Bangladesh, which may suggest that environmental *V. mimicus* does not share the same niche as 7PET *V. cholerae*.

Collectively, our findings position *V. mimicus* as a genetically diverse environmental species with recurrent, independent acquisition of canonical cholera virulence determinants. Other *Vibrio* species more closely related to *V. cholerae* have been recently described from clinical and environmental sampling[18,45]. These carry other virulence determinants but lack CTX and TCP, highlighting the unique ability of *V. mimicus* to acquire these elements. Yet, unlike *V. cholerae*, toxigenic lineages in *V. mimicus* remain geographically restricted and clinically sporadic. Differences in genomic context, allelic variation in *ctxB* and *tcpA*, absence of key accessory elements, and broader ecological constraints may limit epidemic potential. Understanding the mechanisms governing gene flow between *V. mimicus* and *V. cholerae*, and selective forces shaping retention or loss of virulence-associated loci, may give us key insights into both the continual evolution of epidemic *V. cholerae*, and will be critical for anticipating

emergence of novel pathogenic *Vibrio* lineages, likely from multiple species, in the context of the changing climate.

Methods

Sampling and whole genome sequencing

Fifty-three *V. mimicus* isolates from Bangladesh were obtained from environmental sources, including water, sediment, and plankton. These were collected across three subdistricts (Bakergonj, Chhatak, and Mathbaria), between 2005 and 2016. DNA was extracted from bacterial isolates, Illumina sequencing libraries were prepared, and these were sequenced on the NovaSeq 6000 platform at the Wellcome Sanger Institute.

Genome assembly and phylogenetic analysis

De novo sequence assembly was performed from paired-end short reads using Unicycler v0.5.0[46] with default parameters, and the generated assemblies were annotated with Bakta v1.11.0[47]. All 115 publicly available *V. mimicus* genomes were retrieved from NCBI GenBank (accessed 30 May 2025). We selected 108 of these as contextual genomes that passed the quality control parameters described below. Additionally, published *V. cholerae* genomes were also collected from the collection in Pathogenwatch (accessed 30 May 2025)[48]. We selected 144 genomes encompassing representative references from Pathogenwatch for all sublineages across the *V. cholerae* phylogeny, as well as genomes capturing diverse virulence profiles. Quality control of all genomes was conducted using Quast v5.0.2[48,49] and CheckM2 v1.0.1[50], and assemblies were filtered based on thresholds of > 99% completeness and < 5% contamination. Taxonomic classification of assemblies was performed using GTDB-Tk v.2.3.0[51].

Pangenome analysis was performed to generate a core genome alignment separately for each species, with Panaroo v1.3.3[52] in sensitive mode and removing invalid genes, then combined using panaroo-merge. The pangenome reference sequences were annotated using eggnog-mapper v2.1.11[53] to assign functions based on Clusters of Orthologous Genes (COGs). Species-aware pangenomics was performed with Twilight v1[54] to classify genes based on presence and prevalence in *V. cholerae* and *V. mimicus*. The polymorphic sites from the resulting *V. mimicus* core gene alignment were obtained using SNP-sites v2.5.1[55] and used to build a maximum-likelihood (ML) phylogenetic tree using IQ-TREE v2.2.0[56], with the GTR + F+R10 model (determined by first running model finder; -m MF) and 1000 ultrafast bootstraps. Bayesian Analysis of Population Structure (BAPS) using fastBAPS v1.0.6[57] was used to define phylogenetic clusters within the species.

Detecting genetic elements of interest in V. mimicus

We searched for known genomic islands that carry key virulence determinants in *V. cholerae* using BLAST v2.14[57,58]. A BLAST database was generated from representative sequences obtained for VPI-1, VPI-2, VSP-I and VSP-II from *V. cholerae* strain N16961 (accession ID: GCA_003063785.1), and for the full CTXφ bacteriophage (accession ID: KF664572.1). The *V. mimicus* genomes in this study were then searched against this database using BLASTn, and matches were filtered based on having 90% nucleotide identity and contiguous coverage of 50% of the reference.

AMR and virulence-associated genes were identified using Abricate v1.0.1[59] with default parameters to search the Comprehensive Antibiotic Resistance Database (CARD, v.3.2.4)[60] and the Virulence Factor Database (VFDB, accessed 13 October 2023)[61], respectively. Default parameters were used, with matches designated as hits with 80% identity and 80% coverage compared to the reference sequence. Presence of key virulence genes was also confirmed by comparing with the Bakta annotations. Signatures of positive selection were estimated by calculating the ratio of synonymous and non-synonymous mutations using PAML v4.10.9 with site-heterogenous Model 8[62].

The presence of Type III and Type VI secretion system gene clusters was detected using the HMmer Based UndeRstandinG of gene clustERs (Hamburger) tool[63]. The large cluster was found by searching the inbuilt T6SS database in Hamburger using the -t flag and default parameters. A custom HMM database was generated for T3SS structural genes, and effectors encoded within the structural gene core region, from representative isolates of *V. cholerae* (accession IDs: GCA_000153785.3, GCA_900538065.1, ERS2493980), and close relatives *Vibrio parahaemolyticus* (accession IDs: GCA_000196095.1, AB455531.1) and *Vibrio anguillarum* (accession ID: GCA_002813795.1), previously described to encode each of the three types of T3SS in *V. cholerae.* A gap of 15 non-T3SS genes in between T3SS genes was permitted based on these reference sequences. Due to homology of T3SS and flagellar genes, HMMs for flagellum-specific genes *flgB*, *flgC*, and *fliE* were included. T3SS clusters were thus designated as those that did not include flagellar genes but did contain secretin (*sctC*), which is a T3SS gene that has no flagellar homologues[64]. Sequence structure for *V. mimicus* TcpA was obtained from the AlphaFold Protein Structure Database and visualised with Pymol and Jalview[62–64].

References

1. Davis BR, Fanning GR, Madden JM, Steigerwalt AG, Bradford HB Jr, Smith HL Jr, Brenner DJ (1981) Characterization of biochemically atypical Vibrio cholerae strains and designation of a new pathogenic species, Vibrio mimicus. J Clin Microbiol 14:631–639
2. Hasan NA, Grim CJ, Haley BJ, Chun J, Alam M, Taviani E, Hoq M, Munk AC, Saunders E, Brettin TS et al (2010) Comparative genomics of clinical and environmental Vibrio mimicus. Proc Natl Acad Sci U S A 107:21134–21139
3. Morgado SM, Dos Santos Freitas F, da Fonseca Lourenço, E., and, Vicente ACP (2024) Vibrio mimicus Lineage Carrying Cholera Toxin and Vibrio Pathogenicity Island, United States and China. Emerg Infect Dis 30:1729–1732

4. Chitov T, Kirikaew P, Yungyune P, Ruengprapan N, Sontikun K (2009) An incidence of large foodborne outbreak associated with Vibrio mimicus. Eur J Clin Microbiol Infect Dis 28:421–424
5. Kay MK, Cartwright EJ, Maceachern D, McCullough J, Barzilay E, Mintz E, Duchin JS, Macdonald K, Turnsek M, Tarr C et al (2012) Vibrio mimicus infection associated with crayfish consumption, Spokane, Washington, 2010. J Food Prot 75:762–764
6. Alam MT, Stern SR, Frison D, Taylor K, Tagliamonte MS, Nazmus SS, Paisie T, Hilliard NB, Jones RG, Iovine NM et al (2023) Seafood-Associated Outbreak of ctx-Negative Vibrio mimicus Causing Cholera-Like Illness, Florida, USA. Emerg Infect Dis 29:2141–2144
7. Chowdhury MA, Aziz KM, Kay BA, Rahim Z (1987) Toxin production by Vibrio mimicus strains isolated from human and environmental sources in Bangladesh. J Clin Microbiol 25:2200–2203
8. Boyd EF, Moyer KE, Shi L, Waldor MK (2000) Infectious CTXPhi and the vibrio pathogenicity island prophage in Vibrio mimicus: evidence for recent horizontal transfer between V. mimicus and V. cholerae. Infect Immun 68:1507–1513
9. Pant A, Das B, Bhadra RK (2020) CTX phage of Vibrio cholerae: Genomics and applications. Vaccine 38(Suppl 1):A7–A12
10. Waldor MK, Rubin EJ, Pearson GD, Kimsey H, Mekalanos JJ (1997) Regulation, replication, and integration functions of the Vibrio cholerae CTXphi are encoded by region RS2. Mol Microbiol 24:917–926
11. Kim EJ, Lee D, Moon SH, Lee CH, Kim DW (2014) CTX prophages in Vibrio cholerae O1 strains. J Microbiol Biotechnol 24:725–731
12. Rajanna C, Wang J, Zhang D, Xu Z, Ali A, Hou Y-M, Karaolis DKR (2003) The vibrio pathogenicity island of epidemic Vibrio cholerae forms precise extrachromosomal circular excision products. J Bacteriol 185:6893–6901
13. O'Shea YA, Boyd EF (2002) Mobilization of theVibriopathogenicity island betweenVibrio choleraeisolates mediated by CP-T1 generalized transduction. FEMS Microbiol Lett 214:153–157
14. Murphy RA, Boyd EF (2008) Three pathogenicity islands of Vibrio cholerae can excise from the chromosome and form circular intermediates. J Bacteriol 190:636–647
15. Drebes Dörr NC, Lemopoulos A, Blokesch M (2025) Exploring Mobile Genetic Elements in Vibrio cholerae. Genome Biol Evol 17. 10.1093/gbe/evaf079
16. Shandera WX, Johnston JM, Davis BR, Blake PA (1983) Disease from infection with Vibrio mimicus, a newly recognized Vibrio species. Clinical characteristics and edipemiology. Ann Intern Med 99:169–171
17. Eyisi OAL, Nwodo UU, Iroegbu CU (2013) Distribution of vibrio species in shellfish and water samples collected from the atlantic coastline of south-east Nigeria. J Health Popul Nutr 31:314–320
18. Islam MT, Liang K, Orata FD, Im MS, Alam M, Lee CC, Boucher YF (2022) Vibrio tarriae sp. nov., a novel member of the Cholerae clade. Int J Syst Evol Microbiol 72. 10.1099/ijsem.0.005571

19. Baker-Austin C, Oliver JD, Alam M, Ali A, Waldor MK, Qadri F, Martinez-Urtaza J (2018) Vibrio spp. infections. Nat Rev Dis Primers 4:8
20. Tapia-Poza S, Jurado-Tarifa E, Guerrero-Lozano I, Trujillo-Soto T, Galán-Sánchez F, Sánchez-Calvo JM, Triñanes-Fernández JA, Martínez-Urtaza J, Rodríguez-Iglesias MA (2025) Emergence of Vibrio and related genera infections in a hotspot of climate risks, southern Spain, 2010–2023. One Health 21:101267
21. Bote L, Maes M (2024) Tracking pathogen evolution through climate change. Nat Rev Microbiol 22:390
22. Magers BM, Brumfield KD, Kumar S, Colwell RR, Jutla AS (2025) Machine learning in understanding environmental variability of vibriosis in coastal waters. Appl Environ Microbiol 91, e0071625
23. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M et al (2011) Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477:462–U111
24. Miller KA, Tomberlin KF, Dziejman M (2019) Vibrio variations on a type three theme. Curr Opin Microbiol 47:66–73
25. Russell AB, Peterson SB, Mougous JD (2014) Type VI secretion system effectors: poisons with a purpose. Nat Rev Microbiol 12:137–148
26. Unterweger D, Miyata ST, Bachmann V, Brooks TM, Mullins T, Kostiuk B, Provenzano D, Pukatzki S (2014) The Vibrio cholerae type VI secretion system employs diverse effector modules for intraspecific competition. Nat Commun 5:3549
27. Kumar A, Das B, Kumar N (2020) Vibrio Pathogenicity Island-1: The Master Determinant of Cholera Pathogenesis. Front Cell Infect Microbiol 10:561296
28. Mirold S, Ehrbar K, Weissmüller A, Prager R, Tschäpe H, Rüssmann H, Hardt WD (2001) Salmonella host cell invasion emerged by acquisition of a mosaic of separate genetic elements, including Salmonella pathogenicity island 1 (SPI1), SPI5, and sopE2. J Bacteriol 183:2348–2358
29. Kirn TJ, Taylor RK (2005) TcpF is a soluble colonization factor and protective antigen secreted by El Tor and classical O1 and O139 Vibrio cholerae serogroups. Infect Immun 73:4461–4470
30. Severin GB, Ramliden MS, Ford KC, Van Alst AJ, Sanath-Kumar R, Decker KA, Hsueh BY, Chen G, Yoon SH, Demey LM et al (2023) Activation of a Vibrio cholerae CBASS anti-phage system by quorum sensing and folate depletion. MBio 14, e0087523
31. Blokesch M (2025) Defence systems encoded by core genomic islands of seventh pandemic Vibrio cholerae. Philos Trans R Soc Lond B Biol Sci 380:20240083
32. Taviani E, Grim CJ, Choi J, Chun J, Haley B, Hasan NA, Huq A, Colwell RR (2010) Discovery of novel Vibrio cholerae VSP-II genomic islands using comparative genomic analysis: Vibrio cholerae VSP-II genomic islands. FEMS Microbiol Lett 308:130–137
33. Campos E, Bolaños H, Acuña MT, Díaz G, Matamoros MC, Raventós H, Sánchez LM, Sánchez O, Barquero C (1996) Vibrio mimicus diarrhea following ingestion of raw turtle eggs. Appl Environ Microbiol 62:1141–1144

34. Colombara DV (2010) Comparison of Cholera Risk Factors. in Urban & Rural Bangladesh, pp 2000–2008
35. Emergence (2021) spread of different ctxB alleles of Vibrio cholerae O1 in Odisha, India (. Int J Infect Dis 105:730–732
36. Kim EJ, Lee CH, Nair GB, Kim DW (2015) Whole-genome sequence comparisons reveal the evolution of Vibrio cholerae O1. Trends Microbiol 23:479–489
37. Safa A, Bhuiyan NA, Murphy D, Bates J, Nusrin S, Kong RYC, Chongsanguan M, Chaicumpa W, Nair GB (2009) Multilocus genetic analysis reveals that the Australian strains of Vibrio cholerae O1 are similar to the pre-seventh pandemic strains of the El Tor biotype. J Med Microbiol 58:105–111
38. Naha A, Mandal RS, Samanta P, Saha RN, Shaw S, Ghosh A, Chatterjee NS, Dutta P, Okamoto K, Dutta S et al (2020) Deciphering the possible role of ctxB7 allele on higher production of cholera toxin by Haitian variant Vibrio cholerae O1. PLoS Negl Trop Dis 14:e0008128
39. Barton A, Afrad MH, Taylor-Brown A, Singh N, Thakur C, Islam T, Rahman SIA, Akhtar M, Begum YA, Bhuiyan TR et al (2025) Evolution of Pandemic Cholera at its Global Source. medRxiv, 2025.02.03.25321585. 10.1101/2025.02.03.25321585
40. Monir MM, Islam MT, Mazumder R, Mondal D, Nahar KS, Sultana M, Morita M, Ohnishi M, Huq A, Watanabe H et al (2023) Genomic attributes of Vibrio cholerae O1 responsible for 2022 massive cholera outbreak in Bangladesh. Nat Commun 14:1154
41. Li J, Lim MS, Li S, Brock M, Pique ME, Woods VL Jr, Craig L (2008) Vibrio cholerae toxin-coregulated pilus structure analyzed by hydrogen/deuterium exchange mass spectrometry. Structure 16:137–148
42. Kirn TJ, Lafferty MJ, Sandoe CM, Taylor RK (2000) Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by Vibrio cholerae. Mol Microbiol 35:896–910
43. Gutierrez-Rodarte M, Kolappan S, Burrell BA, Craig L (2019) The Vibrio cholerae minor pilin TcpB mediates uptake of the cholera toxin phage CTXφ. J Biol Chem 294:15698–15710
44. Konstantinidis KT, Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. Proc. Natl. Acad. Sci. U. S. A. 101, 3160–3165
45. Boyd EF, Carpenter MR, Chowdhury N, Cohen AL, Haines-Menges BL, Kalburge SS, Kingston JJ, Lubin JB, Ongagna-Yhombi SY, Whitaker WB (2015) Post-genomic analysis of members of the family Vibrionaceae. Microbiol Spectr 3. 10.1128/microbiolspec.VE-0009-2014
46. Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595
47. Beyvers S, Jelonek L, Goesmann A, Schwengers O (2025) Bakta Web - rapid and standardized genome annotation on scalable infrastructures. Nucleic Acids Res 53:W51–W56
48. The Centre for Genomic Pathogen Surveillance Pathogenwatch https://pathogen.watch/genomes/all?organismId=666

49. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075
50. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW (2023) CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat Methods 20:1203–1212
51. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927
52. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA et al (2020) Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol 21:180
53. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol 38:5825–5829
54. Horesh G, Taylor-Brown A, McGimpsey S, Lassalle F, Corander J, Heinz E, Thomson NR (2021) Different evolutionary trends form the twilight zone of the bacterial pan-genome. Microb Genom 7. 10.1099/mgen.0.000670
55. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR (2016) : rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2, e000056
56. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534
57. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J (2019) Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res 47:5539–5549
58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
59. GitHub - tseemann /abricate::mag_right: Mass screening of contigs for antimicrobial and virulence genes GitHub. https://github.com/tseemann/abricate
60. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN et al (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573
61. Chen L, Zheng D, Liu B, Yang J, Jin Q (2016) VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. Nucleic Acids Res 44:D694–D697
62. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591
63. Mariano G, Trunk K, Williams DJ, Monlezun L, Strahl H, Pitt SJ, Coulthurst SJ (2019) A family of Type VI secretion system effector proteins that form ion-selective pores. Nat Commun 10:5484
64. Bote L, Taylor-Brown A, Maes M, Ingle DJ, Valcanis M, Howden BP, Thomson NR (2024) Surveillance of travel-associated isolates elucidates the diversity of non-pandemic. Microb Genom 10. 10.1099/mgen.0.001307

65. Jumper J et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. 10.1038/s41586-021-03819-2
66. Varadi M et al (2023) AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. Nucleic Acids Res gkad1011. 10.1093/nar/gkad1011
67. Varadi M et al (2021) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439–D444. 10.1093/nar/gkab1061

Figures

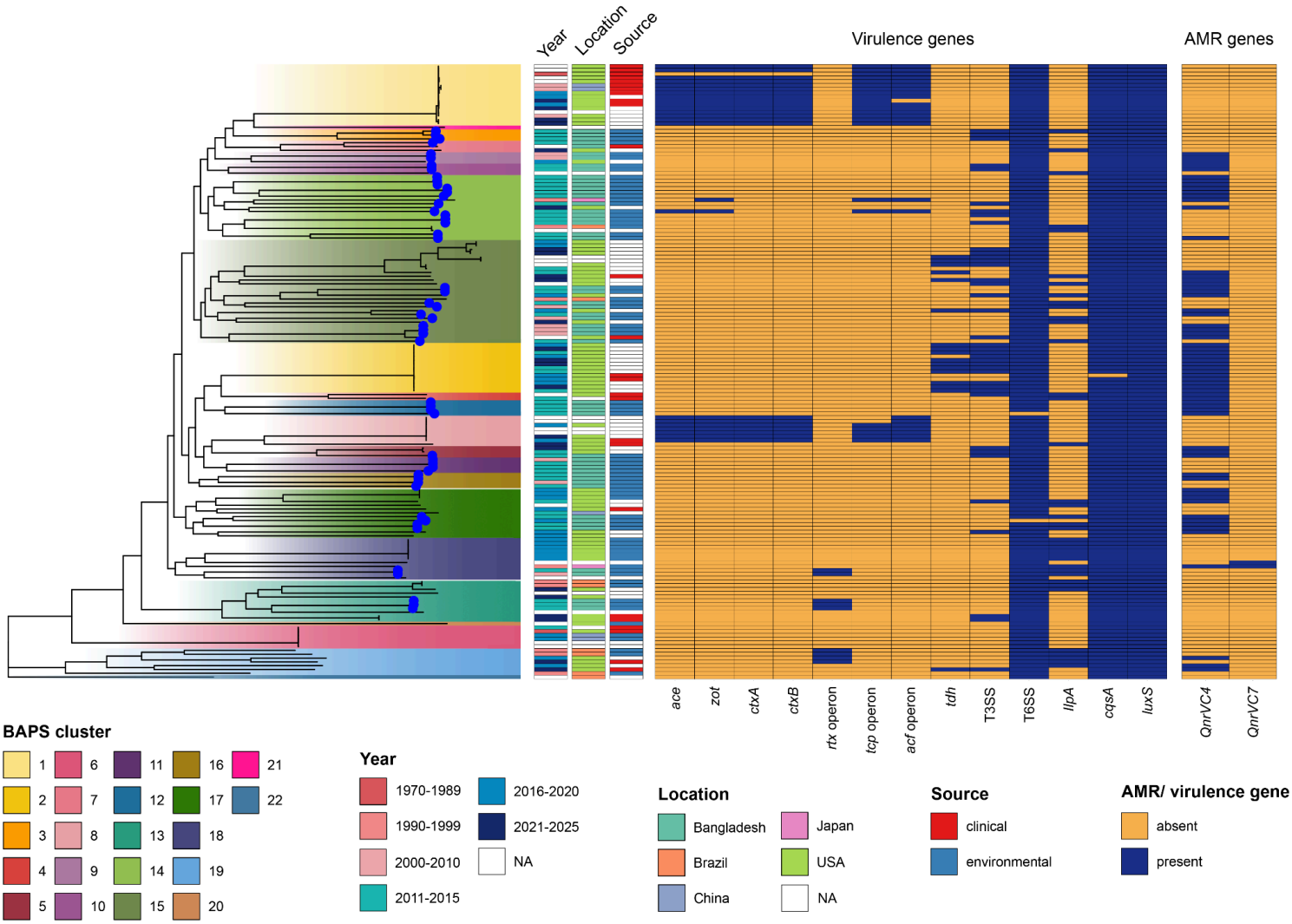


Figure 1

Maximum likelihood phylogenetic tree of 161 *V. mimicus* genomes, with novel genomes from Bangladesh indicated with blue tips. Clades are coloured by BAPS clusters. Corresponding metadata is shown for year, location, source of isolation, and AMR and virulence gene presence for each genome.

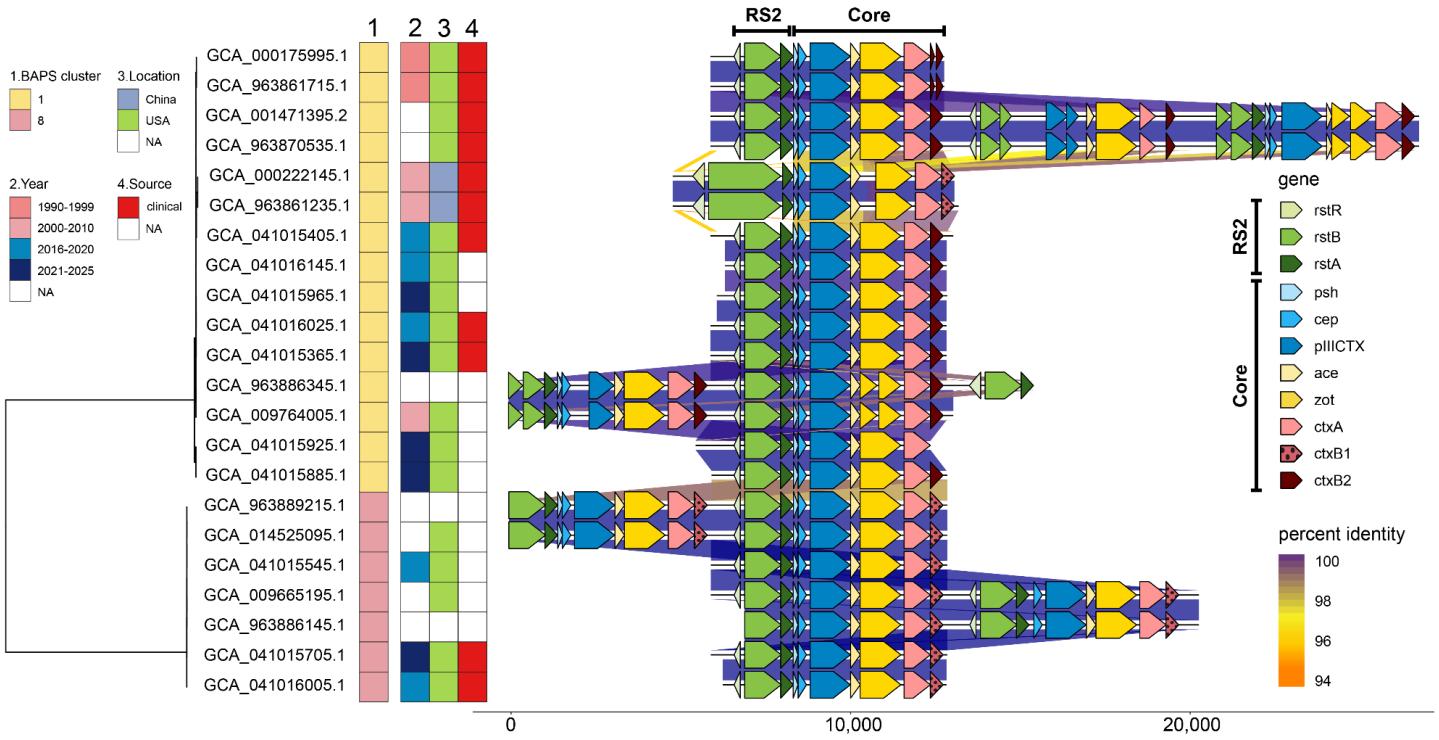


Figure 2

Gene-based synteny alignment for CTX phage in V. mimicus, with the corresponding maximum likelihood phylogenetic tree for the host genomes carrying CTX phage sequences, and associated metadata for BAPS cluster, year, location, and source of isolation. Genes belonging to RS2 and core regions are indicated, and links between gene tracks are coloured by % nucleotide identity.

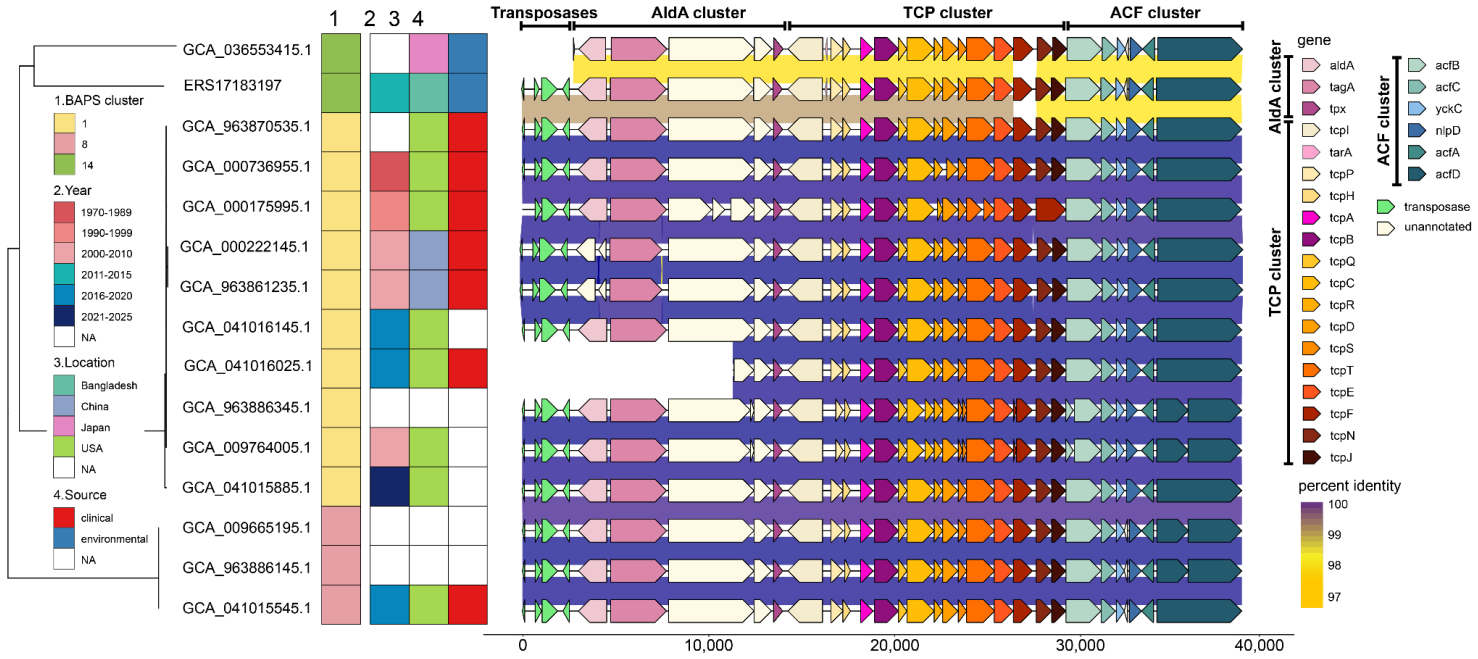


Figure 3

Gene-based synteny alignment for TCP operon phage in V. mimicus, with the corresponding maximum likelihood phylogenetic tree for the host genomes, and metadata for BAPS cluster, year, location, and source of isolation. Transposases and genes belonging to AldA, TCP, and ACF clusters regions are indicated, and links between gene tracks are coloured by % nucleotide identity.

Supplementary Files

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

- supplementaryfigures.docx
- supplementarydata.csv