MICROBIOLOGY SOCIETY

OPEN DATA · OPEN ACCESS

Surveillance of travel-associated isolates elucidates the diversity of non-pandemic *Vibrio cholerae*

Lia Bote[1], Alyce Taylor-Brown[1], Mailis Maes[1], Danielle J. Ingle[2], Mary Valcanis[3], Benjamin P. Howden[2,3,4,5,*] and Nicholas R. Thomson[1,6,*]

Abstract

Vibrio cholerae is a Gram-negative bacterium found in aquatic environments and is the aetiological agent of cholera, character-ized by acute watery diarrhoea and severe dehydration. Cholera presents a significant global health burden of an estimated 1.3–5 million annual cases, with the current pandemic caused by a toxigenic lineage of the O1 El Tor biotype called seventh pan-demic El Tor (7PET) that is still ongoing. Whilst it is known that non-7PET lineages can cause sporadic disease, little is known about the transmission of these non-epidemic lineages. Thirty-four *V. cholerae* isolates were obtained from travellers returning from Indonesia to Australia between 2005 and 2017. These were whole genome sequenced, placed into a global phylogenetic context with 883 isolates, and screened for known genes associated with antimicrobial resistance and virulence. This analysis revealed that 30 isolates fell within non-7PET lineages and four within the 7PET lineage. Both 7PET and non-7PET isolates carried genes for resistance to antibiotics that are commonly used in cholera treatment such as tetracyclines and fluoroqui-nolones. Diverse virulence factors were also present in non-7PET isolates, with two isolates notably carrying toxin-coregulated pilus genes, which are primarily responsible for intestinal colonization in 7PET *V. cholerae*. This study demonstrates the role of travel in long-range carriage of epidemic and non-epidemic lineages of *V. cholerae,* and how sentinel travel surveillance can enrich our knowledge of *V. cholerae* diversity, reveal new biology about the spread of diverse lineages with differing disease potential and illuminate disease presence in endemic regions with limited surveillance data.

Impact Statement

Vibrio cholerae remains one of the most important human pathogens globally. While surveillance efforts have focused on the current pandemic lineage, non-epidemic lineages have also been shown to cause sporadic disease. This study demonstrates the ability of diverse epidemic and non-epidemic lineages with virulence and antimicrobial resistance determinants to be carried across long distances, and highlights the importance of sentinel travel surveillance in understanding disease transmis-sion dynamics.

DATA SUMMARY

Sequencing reads were deposited in the NCBI Sequence Read Archive under Bioproject PRJNA856407. Individual accession IDs for each isolate are listed in the Supplementary Data. All other supporting data have been provided within the article or in the supplementary data files.

INTRODUCTION

Vibrio cholerae is a Gram-negative, curved or comma-shaped bacterium that is found in marine, coastal or brackish water. It is the causative agent of cholera, which is transmitted through the oral–faecal route and characterized by acute watery diarrhoea and severe dehydration. Cholera represents a significant global health burden of an estimated 1.3–5 million cases and 21000–143000 deaths annually [1]. Of the nearly 200 *V. cholerae* 'O' serogroups based on the O-antigen polysaccharide, only toxigenic strains of O1 and O139 have caused epidemics and global pandemics. Toxigenic *V. cholerae* strains carry the *ctxA* and *ctxB* genes encoding cholera toxin (CTX), which itself is encoded in the genome of the lysogenic CTXφ bacteriophage [2]. CTX is the primary virulence factor linked to secretory diarrhoea. The cell surface receptor for CTXφ is the toxin coregulated pilus (TCP), which is also a key virulence factor essential for intestinal colonization [3].

There have been seven recorded cholera pandemics in history since 1817. The classical biotype of the O1 serogroup was thought to be primarily responsible for the first six and was replaced by a toxigenic, highly clonal lineage of the O1 serogroup, El Tor biotype for the seventh, ongoing pandemic (7PET, seventh pandemic El Tor) [4]. The O139 serogroup is an El Tor derivative that rapidly emerged in 1992 but is no longer linked to outbreaks [5, 6]. The 7PET lineage was traced to a source in the Bay of Bengal and has since spread globally through three independent but overlapping waves of transmission [7]. The seventh pandemic itself was first documented in Indonesia in 1961 and continues globally to the present day [7, 8]. Despite large cholera outbreaks in the 1970s and 1980s [9, 10], and an estimated 0.5 per 1000 annual cases of both O1 and non-O1 *V. cholerae* between 2001 and 2003 [11], the World Health Organization has not reported cholera cases in Indonesia since 2011 [12, 13]. Nevertheless, diarrhoea remains a leading cause of mortality for children in Indonesia [1, 10].

As per WHO recommendations, while the primary mode of treatment for cholera is oral rehydration therapy, tetracycline is commonly used as the first-line antibiotic due to both efficacy and ease of administration, and fluoroquinolones such as ciprofloxacin are used as alternative treatment [14]. Multidrug resistance (MDR) in 7PET *V. cholerae* is often linked to the acquisition of various mobile genetic elements (MGEs), including the sulfamethoxazole–trimethoprim (SXT) family integrative conjugative element (ICE), which carries genes such as *dfrA1* and *sul1* conferring resistance to trimethoprim and sulfamethoxazole, respectively [15]. This SXT ICE was acquired in 7PET between 1978 and 1984, during the transition between wave 1 and wave 2 of 7PET transmission [7]. The SXT ICE has also been found in non-7PET lineages and can lose the MDR gene cassette as previously seen in O139 strains [4]. As such, the role of other MGEs, such as plasmids and other ICEs, in antimicrobial resistance in *V. cholerae* has also been described [15].

Isolates that do not belong to the 7PET lineage, which are collectively referred to as non-7PET *V. cholerae*, have also been associated with sporadic cases of diarrhoea, gastroenteritis and bacteraemia. These isolates often belong to non-O1, non-O139 serogroups and present with a wider range of symptoms than 7PET infection [16–18]. However, like 7PET, some non-7PET isolates also belong to the O1 serogroup and, although rarely found, CTXφ and TCP are present in some strains [6, 19]. In fact, recent cholera outbreaks in China have been attributed to CTX-negative, TCP-positive (CNTP), non-7PET lineages where CTXφ phage acquisition has been observed [20].

Type III secretion systems (T3SSs) have been linked to virulence in other pathogenic Gram-negative bacteria such as *Yersinia*, *Salmonella* and *Shigella*, allowing for the translocation of effectors into eukaryotic host cells [21, 22]. These effectors often result in cytoskeletal rearrangements and thus cell perturbation or death. In *Vibrio*, the T3SS was first detected in *V. parahaemolyticus*, which contains two T3SSs: T3SS1 and T3SS2. The *V. parahaemolyticus* T3SS2 is restricted to its pandemic serotype, O3:K6, and shares close sequence similarity to the T3SS first detected in *V. cholerae* [23]. T3SSs have been shown to be sufficient for intestinal colonization by non-7PET *V. cholerae* in infant mouse and rabbit models, even in the absence of TCP [23–25]. The presence and dynamic acquisition of these virulence factors demonstrate the disease-causing potential of non-7PET *V. cholerae*. However, due to the focus on surveillance for 7PET *V. cholerae*, little is known about the potential of non-7PET lineages as an emerging threat for future outbreaks [18, 26].

In this study, 34 isolates from travellers returning from Indonesia to Australia between 2005 and 2017 provided an opportunity to look at the potential for long-range dissemination of different *V. cholerae* lineages. The isolates were whole genome sequenced, placed in phylogenetic context, and screened for genetic elements of interest such as antimicrobial resistance (AMR) genes and virulence factors. This study provides insight into the potential of a range of *V. cholerae* lineages for long-range carriage.

METHODS

Biochemical testing and sequencing of bacterial isolates

V. cholerae isolates from patients returning from Indonesia to Australia presenting with symptomatic diarrhoea to primary pathology testing laboratories were referred to the Microbiological Diagnostic Unit Public Health Laboratory (MDU-PHL). Here, isolates were cultured for species confirmation, serotyping, biotyping and toxin gene detection.

The identity of these strains was first confirmed by growth on thiosulphate-citrate-bile salts-sucrose agar (TCBS, Oxoid) and a panel of biochemical tests [27]. The serogroup of these strains was subsequently confirmed using polyvalent O1, monovalent O1 (Inaba and Ogawa) and O139 agglutinating antisera (Mast Assure). All *V. cholerae* strains that serotyped as O1 were biotype tested for haemolysis of sheep erythrocytes and production of acetylmethylcarbinol (VP, Vogues–Proskauer), and submitted for PCR assays for the detection of the *V. cholerae* species-specific *ompW* and *ctxA* genes, as previously described [28, 29]. These isolates underwent whole genome sequencing (WGS) on the Illumina NextSeq 500 platform with 150 bp paired-end reads (Supplementary Data, available in the online version of this article). Reads are available in BioProject PRJNA856407 from the NCBI Sequence Read Archive and individual accession IDs are listed in the Supplementary Data.

Genome assembly and annotation

De novo sequence assembly was performed from paired-end short reads using Unicycler v0.5.0 [30] with default parameters. Contigs were then annotated with Prokka v1.14.5 [31]. Quality metrics for the generated assemblies and annotations were obtained using Quast v5.0.2 [32] and CheckM2 v1.0.1 [33], with thresholds of >99% completeness and <0.2% contamination. Pangenome analysis was performed with Panaroo v1.3.3 [34] from the annotated assemblies to generate a core gene alignment, with merged paralogues and a core gene threshold of 99%.

Phylogenetic analysis

Published *V. cholerae* genomes obtained from the collection in Pathogenwatch (https://pathogen.watch/genomes/all?organismId=666, accessed 3 August 2023) were used as contextual sequences (Supplementary Data). Five representative isolates from each of the three global waves of 7PET transmission from its origin in the Bay of Bengal, selected based on time and geographical region of collection, and 868 non-7PET isolates across diverse clades were included. A pangenome was generated using Panaroo, as described above, with 883 contextual genomes and the 34 isolates from this study. The polymorphic sites (*n*=760710 SNPs) from the resulting core gene alignment (*n*=2682 genes) were obtained using SNP-sites v2.5.1 [35] and used to build a maximum-likelihood (ML) phylogenetic tree using IQ-TREE v2.2.0 [36], with the GTR+F+ASC+R10 model (determined by first running model finder; -m MF) and 1000 ultrafast bootstraps.

An ML tree was also built for the 34 travel-associated isolates from this study based on polymorphic sites of the core gene alignment as described above. A mapped-SNP alignment to the N16961 7PET wave 1 reference (accession ID: GCA_000006745), made using SMALT v0.7.419 [37], was used to generate an ML tree for the 7PET isolates from this study and the five contextual 7PET isolates from each wave of transmission. All generated trees were visualized using iTOL v6.8.1 [38]. Pairwise SNP distances between isolates were calculated using SNP-dists v0.7.0 [39].

Characterizing genes of interest

The presence of AMR genes was determined through a BLASTn-based method with Abricate v.1.0.1 [40] with default parameters using the Comprehensive Antibiotic Resistance Database (CARD, v.3.2.4) [41] on the whole genome assemblies of all 34 isolates in this study. ARIBA v2.14.6 [42] was then used with the same database on the read sets of all 34 isolates. Matches were designated in ARIBA as 95% of the reference gene sequence matching to the assembly and the gene assembling into one contig and matches in Abricate were those with 80% identity and 80% coverage compared to the reference gene sequence. The results from these two methods were combined to generate a profile of AMR gene presence.

The same analysis was conducted with both Abricate and ARIBA using the Virulence Factor Database (VFDB, accessed 13 October 2023) [43]. Abricate was also used with the plasmidfinder database to find putative plasmid sequences. Structural proteins and putative effectors were also identified for different bacterial secretion systems through this VFDB search, and the HMMer Based UndeRstandinG of gene clustERs (hamburger) tool was used to find clusters of structural genes for the T3SS [44]. A custom database of Hidden Markov Models (HMMs) was built with HMMer v.3.2.1 [45] from known T3SS structural genes (Supplementary Data) in 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) for the three known types of T3SS in *V. cholerae*.

The *ctxB* genotype, which has been associated with different 7PET outbreaks, was identified by running BLASTn v.2.14.1 [46] on the whole genome assemblies for all 34 isolates with a custom database of representative sequences for each *ctxB* allele, using a 90% sequence identity and 70% gene coverage threshold. Presence of the SXT ICE was first explored through building a custom

ARIBA database with representative sequences in *V. cholerae* (Supplementary Data) and running a search on the read sets of all 34 isolates as described above. A BLASTn search was then undertaken with the same database on the assemblies for matches that were found in partially assembled contigs by ARIBA.

RESULTS

The majority of *V. cholerae* isolates from returning travellers to Australia come from non-7PET lineages

The 34 *V. cholerae* isolates obtained between 2005 and 2017 from returning travellers who presented with diarrhoea in Australia (Fig. 1a) underwent WGS, revealing this collection comprised seven *V. cholerae* O1 Ogawa and 27 *V. cholerae* non-O1, non-O139. *In silico* analysis for presence of the 7PET lineage-specific marker gene *VC2346* showed that four isolates carried this gene, indicating they belonged to the 7PET lineage (Fig. 1b), while the remaining 30 isolates belonged to non-7PET lineages.

Phylogenetic analysis reveals diversity of *V. cholerae* in Indonesia

Given that most isolates were from non-7PET lineages, we inferred the phylogeny of the isolates sequenced here in the context of a large collection of published non-7PET genomes (*n*=868), as well as five representatives from each wave of 7PET transmission, based on SNPs (*n*=760710 SNPs) of a core gene alignment (*n*=2682 genes). The contextual sequences included isolates from between 1916 and 2020 and across 49 countries in six continents, representing the known diversity of the species (Fig. 1c). Four of the isolates from this study were from the 7PET lineage and, of these, three (AUSMDU00028286, AUSMDU00028291, AUSMDU0028306) fell within wave 2 and one (AUSMDU00028282) within wave 3 of 7PET transmission (Fig. 1b). All four carried the *ctxB1* genotype of *ctxB*. The wave 2 isolates clustered with the Global 2B sublineage, and the wave 3 isolate with the Global 3B sublineage [26, 47]. Both sublineages have previously been described as having geographical signatures of predominantly Asian origin and dominant *ctxB1* genotypes (Fig. 1b) [26].

The 30 non-7PET isolates from this study fell across diverse phylogenetic branches (Fig. 1c). The mean number of SNPs among these non-7PET isolates was 52845 SNPs (minimum: 55, maximum: 120204), compared to 73 SNPs (minimum: 20, maximum: 129) among the 7PET isolates from this study, consistent with the knowledge that the 7PET lineage is clonal. None of these
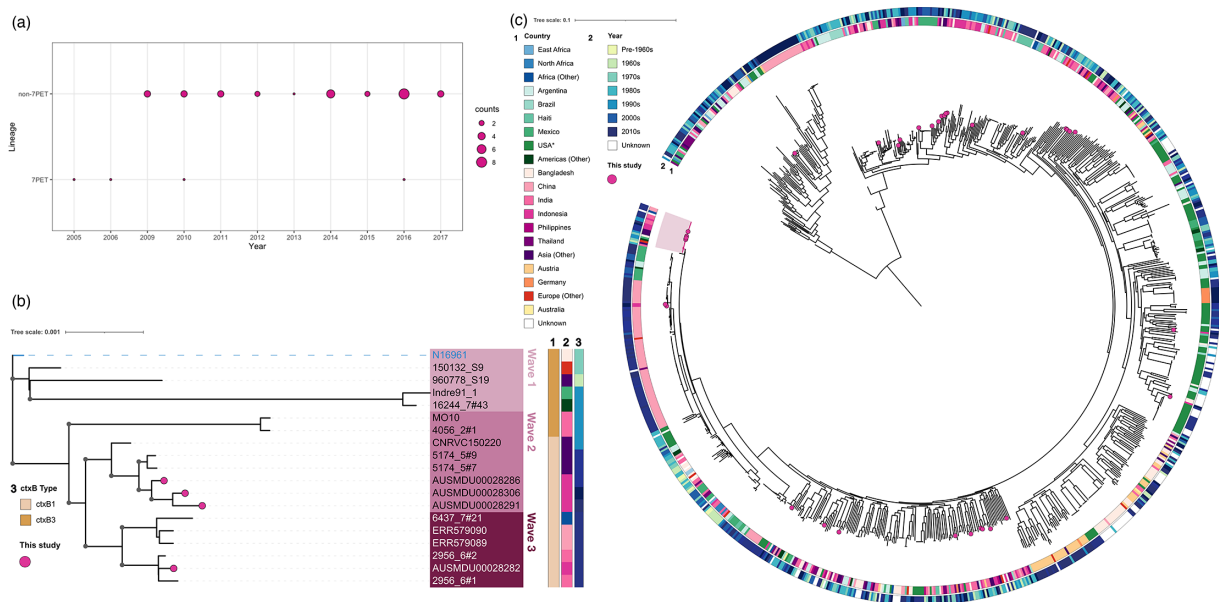

Fig. 1. Isolates from returning travellers to Australia included both 7PET and non-7PET *V. cholerae*. (a) Temporal distribution of non-7PET and 7PET *V. cholerae* isolates in this study, with the size of the circles relative to the number of genomes collected for the given lineage or group of lineages in each year. (b) ML phylogenetic tree with 1000 ultrafast bootstraps for 19 7PET *V. cholerae* genomes, based on SNPs (*n*=2603 SNPs) of a core gene alignment (*n*=3187 genes), rooted on the *V. cholerae* strain N16961 (highlighted in blue). Bootstraps are indicated by grey circles for nodes with bootstrap values ≥90. Metadata for country, year of isolation and *in silico* identification of *ctxB* type are shown in coloured blocks (1, 2 and 3 respectively; see key). Isolates from each wave of transmission are denoted by the three different pink blocks. Isolates from this study are indicated by pink circles on the tree tips. (c) ML phylogenetic tree with 1000 ultrafast bootstraps for non-7PET genomes (*n*=868) and five representatives from each wave of 7PET transmission based on SNPs (*n*=760710 SNPs) of a core gene alignment (*n*=2682 genes), rooted on a clade that contains isolates which have been suggested to belong to *V. paracholerae*. Metadata for country (ring 1) and year of collection (ring 2) are shown for each isolate. The clade shaded in pink denotes the 7PET lineage. Isolates from this study are indicated by pink circles on the tree tips. Trees were visualized and annotated with iTOL v6.8.1.
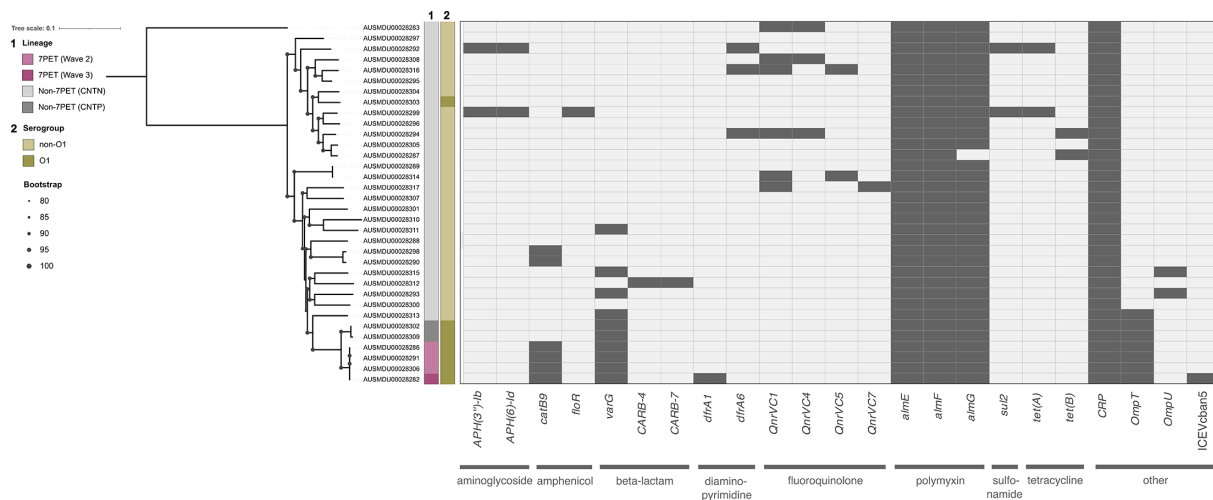
Fig. 2. AMR genes present in travel-associated *V. cholerae* isolates. ML phylogenetic tree with 1000 ultrafast bootstraps for the 34 *V. cholerae* genomes from this study, based on SNPs (*n*=271209 SNPs) of a core gene alignment (*n*=2833 genes), rooted on the AUSMDU00028283 isolate from this study. Metadata for the lineage and *in silico* identification of serogroup are shown. Presence (dark grey) and absence (light grey) of AMR genes for each isolate are shown, detected through a BLAST-based search of the CARD, with the antibiotic class that each gene confers resistance to indicated with grey bars underneath.

non-7PET genomes clustered with each other, except for two CNTP isolates (Fig. 2). These CNTP isolates differed from each other by 1585 SNPs. The mean SNP distance among these two isolates and the 28 other non-7PET isolates from this study was 54599 SNPs. In contrast, the mean SNP distance among these CNTP isolates and the four 7PET isolates from this study was 15343 SNPs, suggesting that they were more genetically similar to 7PET *V. cholerae*. The rest of the isolates did not cluster together, nor with any of the non-7PET clusters that have been previously linked to outbreaks. This suggests that *V. cholerae* isolated from returning travellers in this study were phylogenetically distinct both from each other and from previously observed non-7PET *V. cholerae*.

Travel-related *V. cholerae* isolates carried various AMR genes

Given that many of the travel-related isolates sequenced here were phylogenetically distinct, next we wanted to determine if they were also genotypically distinct. To do this, we first considered their AMR profiles. We looked for the SXT ICE to assess if presence of this element and its associated AMR genes were concordant with the waves of transmission in which our 7PET isolates fell. However, none of the wave 2 isolates in this study contained a complete SXT ICE. Only the wave 3 isolate (AUSMDU00028282) carried a sequence similar to the *V. cholerae* Ban5 ICE, ICEVchban5 (accession number GQ463140.1). This isolate was also the only one that encoded *dfrA1*, which is carried on ICEVchban5. One of the non-7PET isolates, AUSMDU00028299, carried *floR*, *sul2* and *tetA* genes, known to confer resistance to amphenicol, sulfonamide and tetracycline antibiotics, respectively. These were encoded on the same contig in a region that contained a partial match to the *V. cholerae* IDH_1986 strain SXT ICE. We also found one non-7PET isolate (AUSMDU00028290) with a partial match to the *Aeromonas hydrophila* IncU plasmid pRA3 (accession ID: DQ401103), spanning 565 bp with 99.8% nucleotide sequence identity.

We then looked for all other known AMR genes and found that these were sporadically present across both 7PET and non-7PET lineages. The *catB9* gene was detected in all four 7PET as well as two non-7PET isolates (Fig. 2). In contrast, some AMR genes were more conserved across the dataset. All but one of the isolates from this study carried the *almEFG* operon, while the other isolate carried *almE* and *almF* but not *almG*, which is the glycyltransferase primarily responsible for antimicrobial activity [48]. AlmEFG act together to modify lipid A with glycine or diglycine, conferring polymyxin B resistance. Polymyxin antibiotics are often used as a last line strategy against MDR Gram-negative bacteria [49]. Interestingly, the 7PET lineage, but not classical biotype O1, has previously been found to be polymyxin resistant [50]. Other AMR genes were sporadically present across the non-7PET isolates including *tetA* and *tetB*, which encode resistance to tetracycline antibiotics, and *qnrVC1*, *qnrVC4*, *qnrVC5* and *qnrVC7* genes for quinolone resistance (Fig. 2). These are of particular interest as tetracycline and fluoroquinolone antibiotics are recommended by the WHO for the treatment of cholera [51].

Virulence factors were detected in both 7PET and non-7PET isolates

Next, we screened for genetic determinants linked to virulence across different functional classes. The only genes that were present exclusively in 7PET isolates were those for the toxin genes *ctxAB*, accessory enterotoxin (*ace*) and zonula occludens toxin (*zot*) (Fig. 3a). These genes encode the primary toxins involved in 7PET *V. cholerae* pathogenicity and are all present together in a region
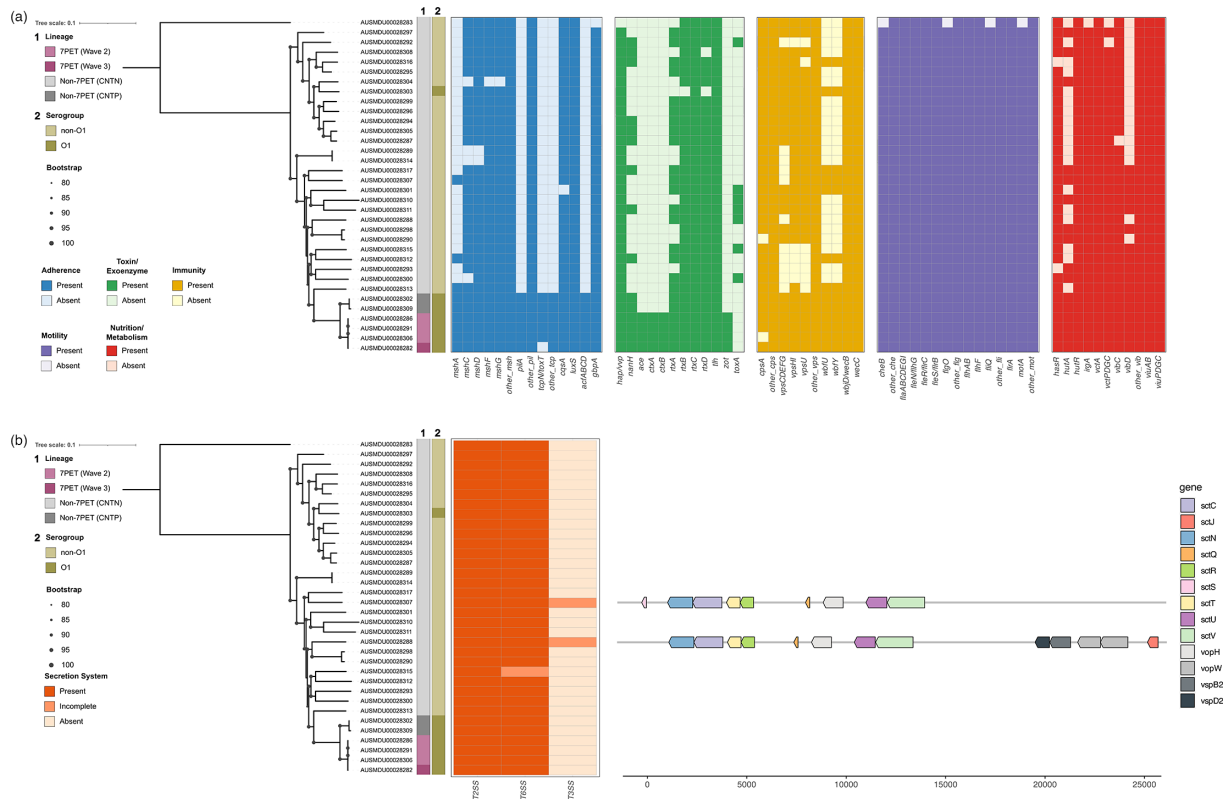
Fig. 3. Genes associated with virulence present in travel-associated *V. cholerae* isolates. (a) ML phylogenetic tree with 1000 ultrafast bootstraps for the 34 *V. cholerae* genomes from this study, based on SNPs (*n*=271209 SNPs) of a core gene alignment (*n*=2833 genes), rooted on AUSMDU00028283 from this study. Metadata for the lineage and *in silico* identification of serogroup are shown. Presence (dark) and absence (light) of virulence factor genes for each isolate are shown, detected through a BLAST-based search of the VFDB, where each block is coloured by functional category. Genes in an operon that are present in all or the same subset of isolates are denoted as 'other' (full list of genes in Supplementary Data). (b) The same ML phylogeny, with presence, absence or incomplete presence of secretion systems shown, where incomplete systems are those that lack at least one structural component. The operons of structural genes, detected through HMM-based methods for the two T3SS-positive isolates from this study, are shown.

in the core of the CTXφ phage called the virulence cassette. Zot modifies tight junctions to increase permeability of intestinal mucosa, while Ace potentially acts as an ion-transporting ATPase but its function is less well understood [52].

Genes for TCP biosynthesis and assembly, which are responsible for the primary mechanism of intestinal colonization by 7PET *V. cholerae*, were found in all four 7PET isolates as well as the two CNTP non-7PET isolates. These genes are clustered together in several transcriptional units: *tcpI*, *tcpP-H*, *tcpA-F* and *toxT-tcpJ* [53]. All six isolates contained the complete set of TCP genes, with the exception of *tcpN/toxT* that was found in five of these six isolates. The *tcpA* gene encodes the major structural subunit of TCP, and *tcpB* encodes the minor pilin gene [54]. These isolates also had other *tcp* genes such as: the putative outer membrane (OM) porin TcpF, which may form an OM channel where TcpA is extruded; the ATP-binding protein TcpT; TcpE that anchors TcpT to the cytoplasmic membrane; and the TcpJ pilin signal peptidase that processes TcpA [53]. *TcpN/toxT* is a regulatory gene that controls the transcription of TCP and other key virulence factors such as CTX.

From a BLAST-based search of the VFDB [43], most of the virulence factors detected were present across all of the isolates in this study. Genes for adherence (*msh*, *pil*), toxins and exoenzymes (*rtx*, *hap*/*vvp*), immunity (*cps*, *vps*, *wbf*) and for protective metabolism (*vctA/PDGC*, *vib*) were present in nearly all of the isolates. Operons encoding motility genes (*che*, *fla*, *fle*, *flg*, *flh*, *fli*, *mot*) were also conserved, and components of these clusters of motility genes, namely *cheB*, *flgO*, *fliQ* and *motA*, were absent from only one non-7PET isolate (Fig. 3a).

Structural components and effectors of type 2 and type 6 secretion systems (T2SS, T6SS) were found in all 34 isolates through the VFDB search. The T2SS has been previously observed to be constitutively expressed in *V. cholerae* at 37 °C, with a role in both virulence and environmental survival [55]. The T6SS has been found in all sequenced *V. cholerae* isolates and is responsible for the translocation of effectors into both eukaryotic and prokaryotic cells in a contact-dependent manner [56]. T3SS components were not found through the BLAST-based VFDB search. Hence, as these are difficult to detect through sequence identity-based methods due to close similarity of many structural components to flagellar proteins [57],

we performed an HMM-based search for T3SS clusters. This revealed that only two non-7PET isolates (AUSMDU00028288, AUSMDU00028307) contained structural components for the T3SS, consistent with previous studies that only found the T3SS is non-7PET *V. cholerae* [58]. Both isolates were non-toxigenic and TCP-negative, and both T3SSs found included secretin, which unlike the other components, does not have homologues in flagellar proteins [21]. Both structures also had conserved presence and order of structural gene components but were each predicted to be incomplete compared to the included references (Fig. 3b).

DISCUSSION

Here we report the nature and diversity of genomes generated from 34 *V. cholerae* isolates, taken from travellers returning to Australia from Indonesia. Four of these belong to the 7PET lineage. As there is little publicly available data on cholera incidence in Indonesia, and, of the years covered by our samples, the WHO only reported statistics in 2005 (1338 cases), it is difficult to determine if these 7PET isolates coincided with peaks in domestic cases. However, one isolate was obtained in 2010, when a cholera outbreak occurred in four Indonesian provinces and totalled 2160 cases [59]. The other 30 isolates were from diverse non-7PET lineages. Only the two CNTP non-7PET isolates clustered together. These were collected in 2014 and 2016, suggesting this lineage was present for at least these two years, and placed within a cluster that was associated with the previously described L3b lineage that also comprised CNTP isolates and caused an outbreak in Hangzhou, China, between 2001 and 2012 [18]. The other non-7PET isolates did not cluster with any previously observed lineages.

The observed trends in presence of AMR genes corresponded both with previous knowledge of AMR in *V. cholerae* and with antibiotic treatment recommendations by the WHO [51, 60]. Some of the AMR genes found are also known to be encoded on genomic islands. One of the non-7PET isolates in this study contained a partial match to the pRA3 plasmid, although only to a 565 bp region that did not map to any of the AMR genes (*sul1*, *aadA2*, *catA2*) previously described in pRA3 [61]. The *catB9* gene, which was found only in the four 7PET and two of the non-7PET isolates from this study, encodes a chloramphenicol acetyltransferase gene conferring low-level phenicol resistance. The *catB* cassette is encoded within the superintegron (SI) in chromosome 2 of *V. cholerae*, which itself allows for the capture and excision of different gene cassettes through site-specific recombination. The *catB9* gene is often silent in 7PET isolates due to its distance from a promoter [62, 63]. However, previous experimental studies have shown that integrase-mediated rearrangement of gene cassettes within the SI region can result in the relocation of *catB* closer to the promoter, leading to induction of chloramphenicol resistance [63]. Similarly, two members of the CARB family of carbenicillin-inactivating β-lactamases, CARB-4 and CARB-7, were found in one of the non-7PET isolates in this study. The bla_{CARB-7} gene cassette has previously been found in non-7PET *V. cholerae*, flanked by *V. cholerae* repeat regions contained within the SI, although was not able to be transferred through conjugation [64]. In contrast, the bla_{CARB-4} gene cassette is associated with the pUD12 plasmid in *Pseudomonas aeruginosa* and includes integrons in its flanking sequences [65, 66]. The *tetA* gene, which was found in two non-7PET isolates from this study, has also been found on MGEs such as in an IncA/C2 plasmid in *V. cholerae* [67] or in the *Salmonella* genomic island 1 [68]. Altogether, the results demonstrate the presence of diverse AMR genes in our sequenced isolates, with some sporadically distributed across non-7PET lineages and none of these genes unique to the 7PET lineage.

The presence of genes associated with virulence also highlighted the genetic diversity of these isolates. TCP was detected in the 7PET isolates and in two CNTP non-7PET isolates, but is typically associated with 7PET *V. cholerae*, where it drives intestinal colonization and thus pathogenesis. Genes for TCP biosynthesis have previously been found in non-7PET outbreak clades, even in isolates without CTXφ, such as the ELA-3, Gulf Coast and MX-2 lineages that have been associated with disease in the Americas between the 1970s and the 2010s [6]. Like the two CNTP isolates from this study, these lineages belong to the O1 serogroup, which is predominantly associated with pandemic cholera. The L3b (2001–2012) and L9 (2013–2018) non-7PET lineages responsible for cholera outbreaks in Hangzhou, China, were primarily CNTP, with CTXφ acquisition observed in 21% of collected isolates [18]. Moreover, the two CNTP isolates and the four 7PET isolates in this study had accessory colonization factor genes *acfABCD*, which like TCP are encoded on *V. cholerae* pathogenicity island 1, and have been shown in O1 classical biotype isolates to have a role in colonization of the mouse intestine [69].

Canonical *V. cholerae* toxin genes *ctxAB, zot* and *ace* were only detected in the 7PET isolates in this study, but other toxin and exoenzyme genes such as *rtx* and *nanH* were present in both 7PET and non-7PET isolates. The *rtxABCD* operon confers cytotoxic activity, with *rtxA* encoding a toxin that resembles repeats-in-toxin family exotoxins, while the *nanH* neuraminidase is able to unmask additional GM1 receptors for CTX, increasing binding and penetration and therefore severity of disease [70–72]. The T3SS is thought to be sufficient for both intestinal colonization and inflammatory diarrhoea even in the absence of TCP and CTX [23]. Here, the T3SS was not detected in 7PET isolates, consistent with previous literature [58], and in only two of 30 non-7PET isolates in this study, suggesting that this was not the primary virulence determinant of non-7PET *V. cholerae* found in this study.

Future research is required to assess the distribution and role of these virulence factors in establishing disease. This would require additional non-7PET *V. cholerae* genomes in order to undertake larger-scale comparative genomic analysis. For

example, genome-wide association studies could further elucidate relationships between the presence of known or novel virulence-associated genes across lineages and disease presentation or risk. Methods to accurately determine the serogroup *in silico* would also be essential for understanding the prevalence of different disease-causing non-O1, non-O139 lineages. Although the genetic context of the AMR genes found was not the focus of this study, one limitation of short-read sequencing data is that it is not possible to assemble complete plasmid or integron sequences. Notwithstanding this, we only found one plasmid linked to AMR genes by running the plasmidfinder database. Further work would be required to understand the detailed genomic context associated with such elements carrying AMR genes.

Nonetheless, through phylogenomic approaches, this study illustrates the potential for long-range carriage of a diverse selection of both 7PET and non-7PET *V. cholerae*. The carriage of *V. cholerae* through international travel has been previously described, for example to Australia (1972–2011) [73], Russia (2005–2012) [74] and the UK (2004–2018) [75]. In the UK, those with recent travel to cholera-endemic countries and who present with diarrhoea are recommended to test for *V. cholerae* according to the Investigation of Faecal Specimens for Enteric Pathogens, as per the UK Standards for Microbiology Investigations. Biochemical analysis of 836 *V. cholerae* isolates from the UK Health Security Agency collection revealed that 792 of these were associated with recent travel history abroad, and similar to this study, most of the isolates were non-O1, non-O139 (623/836, 74.5%) [75]. While we note that there are potential sampling biases in the referred Australian isolates, and travel was limited to country of reported travel, these data still enable the detailed exploration of the phylogenetic and genomic diversity of such travel-associated isolates. The distribution of the non-7PET isolates from this study across diverse branches of the global phylogeny demonstrates the highly diverse nature of these isolates, both from each other and from previously described non-7PET outbreak lineages. Most isolates collected in this study were from non-7PET lineages that carried multiple virulence factors and resistance genes against antibiotics that are commonly used against cholera, reinforcing the knowledge that non-7PET lineages are able to cause disease, although through different mechanisms than the 7PET lineage. Thus, WGS of travel-associated isolates from globally representative sources is important for generating a more complete picture of cholera diversity, pathogenicity and transmission dynamics.

References

1. **Ali M, Nelson AR, Lopez AL, Sack DA**. Updated global burden of cholera in endemic countries. *PLoS Negl Trop Dis* 2015;9:e0003832.

2. **Montero DA, Vidal RM, Velasco J, George S, Lucero Y**, *et al*. *Vibrio cholerae*, classification, pathogenesis, immune response, and trends in vaccine development. *Front Med* 2023;10:1155751.

3. **Ramamurthy T, Nandy RK, Mukhopadhyay AK, Dutta S, Mutreja A**, *et al*. Virulence regulation and innate host response in the pathogenicity of *Vibrio cholerae Front Cell Infect Microbiol* 2020;10:572096.

4. **Ramamurthy T, Pragasam AK, Taylor-Brown A, Will RC, Vasudevan K**, *et al*. *Vibrio cholerae* O139 genomes provide a clue to why it may have failed to usher in the eighth cholera pandemic. *Nat Commun* 2022;13:3864.

5. **Pradhan S, Baidya AK, Ghosh A, Paul K, Chowdhury R**. The El Tor biotype of *Vibrio cholerae* exhibits a growth advantage in the stationary phase in mixed cultures with the classical biotype. *J Bacteriol* 2010;192:955–963.

6. **Domman D, Quilici M-L, Dorman MJ, Njamkepo E, Mutreja A**, *et al*. Integrated view of *Vibrio cholerae* in the americas. *Science* 2017;358:789–793.

7. **Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH**, *et al*. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 2011;477:462–465.

8. **Grizhebovskiĭ GM, Akiev AK**. Epidemiology of El Tor cholera in Indonesia. *Zh Mikrobiol Epidemiol Immunobiol* 1975;0:36–40.

9. **Barua D, Burrows W**. *Cholera*. Philadelphia: WB Saunders; 1974.

10. **Deen JL, von Seidlein L, Sur D, Agtini M, Lucas MES**, *et al*. The high burden of cholera in children: comparison of incidence from endemic areas in Asia and Africa. *PLoS Negl Trop Dis* 2008;2:e173.

11. **Agtini MD, Soeharno R, Lesmana M, Punjabi NH, Simanjuntak C**, *et al*. Erratum To: the burden of diarrhoea, shigellosis, and cholera in North Jakarta, Indonesia: findings from 24 months surveillance. *BMC Infect Dis* 2007;7.

12. **Organization WH**. Cholera, 2015. *Wkly Epidemiol Rec* 2016;91:433–440.

13. **Organization WH**. Global Health Observatory data repository; 2017. https://apps.who.int/gho/data/node.main.175

14. **Deshayes S, Daurel C, Cattoir V, Parienti J-J, Quilici M-L**, *et al*. Non-O1, non-O139 Vibrio cholerae bacteraemia: case report and literature review. *Springer Plus* 2015;4:575.

15. **Das B, Verma J, Kumar P, Ghosh A, Ramamurthy T**. Antibiotic resistance in *Vibrio cholerae:* understanding the ecology of resistance genes and mechanisms. *Vaccine* 2020;38:A83–A92.

16. **Dorman MJ, Thomson NR**. *Vibrio cholerae* O37: one of the exceptions that prove the rule. *Microb Genom* 2023;9:mgen000980.

17. **Arteaga M, Velasco J, Rodriguez S, Vidal M, Arellano C**, *et al*. Genomic characterization of the non-O1/non-O139 *Vibrio cholerae* strain that caused a gastroenteritis outbreak in Santiago, Chile, 2018. *Microb Genom* 2020;6.

18. **Hao T, Zheng W, Wu Y, Yu H, Qian X**, *et al*. Population genomics implies potential public health risk of two non-toxigenic *Vibrio cholerae* lineages. *Infect Genet Evol* 2023;112:105441.

19. **Haley BJ, Choi SY, Grim CJ, Onifade TJ, Cinar HN**, *et al*. Genomic and phenotypic characterization of *Vibrio cholerae* non-O1 isolates from a US Gulf Coast cholera outbreak. *PLoS One* 2014;9:e86264.

20. **Miller KA, Tomberlin KF, Dziejman M**. *Vibrio* variations on a type three theme. *Curr Opin Microbiol* 2019;47:66–73.

21. **Abby SS, Rocha EPC**. The non-flagellar type iii secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet* 2012;8:e1002983.

22. Burkinshaw BJ, Strynadka NCJ. Assembly and structure of the T3SS. *Biochimica et Biophysica Acta (BBA)* 2014;1843:1649–1663.

23. Shin OS, Tam VC, Suzuki M, Ritchie JM, Bronson RT, *et al*. Type III secretion is essential for the rapidly fatal diarrheal disease caused by non-O1, non-O139 Vibrio cholerae. *mBio* 2011;2:e00106-11.

24. Dziejman M, Serruto D, Tam VC, Sturtevant D, Diraphat P, *et al*. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system . *Proc Natl Acad Sci USA* 2005;102:3465–3470.

25. Tam VC, Serruto D, Dziejman M, Brieher W, Mekalanos JJ. A type iii secretion system in *Vibrio cholerae* translocates a formin/spire hybrid-like actin nucleator to promote intestinal colonization. *Cell Host Microbe* 2007;1:95–107.

26. Taylor-Brown A, Afrad MH, Khan AI, Lassalle F, Islam MT, *et al*. Genomic epidemiology of *Vibrio cholerae* during a mass vacci-nation campaign of displaced communities in Bangladesh. *Nat Commun* 2023;14:3773.

27. Farmer JJ, Janda JM, Birkhead K. *Vibrio*. In: Murray PR (ed). *Manual of Clinical Microbiology*. Washington, DC: ASM Press; 2003. pp. 706–718.

28. Nandi B, Nandy RK, Sarkar A, Ghose AC. Structural features, prop-erties and regulation of the outer-membrane protein W (OmpW) of *Vibrio cholerae*. *Microbiology* 2005;151:2975–2986.

29. Lipp EK, Rivera ING, Gil AI, Espeland EM, Choopun N, *et al*. Direct detection of *Vibrio cholerae* and ctxA in Peruvian coastal water and plankton by PCR. *Appl Environ Microbiol* 2003;69:3676–3680.

30. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacte-rial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.

31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioin-formatics* 2014;30:2068–2069.

32. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.

33. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *bioRxiv*;2022:2022. DOI: 10.1101/2022.07.11.499243.

34. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, *et al*. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21:180.

35. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, *et al*. *SNP-sites*: rapid efficient extraction of SNPs from multi-FASTA align-ments. *Microb Genom* 2016;2:e000056.

36. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.

37. Ponstingl H, NingZ. SMALT -a new mapper for DNA sequencing reads. 2010.

38. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.

39. Seemann T, Klötzl F, Page AJ. snp-dists; 2021. https://github.com/tseemann/snp-dists

40. Seemann T. Abricate; 2020. https://github.com/tseemann/abricate

41. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, *et al*. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2023;51:D690–D699.

42. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, *et al*. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3:e000131.

43. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general clas-sification scheme for bacterial virulence factors. *Nucleic Acids Res* 2021;50:D912–D917.

44. Williams DJ. hamburger; 2022. https://github.com/djw533/hamburger

45. Eddy SR. *HMMER User's Guide*, 3.3.2 edn. Howard Hughes Medical Institute, 2020.

46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.

47. Ramamurthy T, Mutreja A, Weill F-X, Das B, Ghosh A, *et al*. Revis-iting the global epidemiology of cholera in conjuction with the genomics of *Vibrio cholerae*. *Front Public Health* 2019;7:203.

48. Henderson JC, Herrera CM, Trent MS. AlmG, responsible for poly-myxin resistance in pandemic *Vibrio cholerae*, is a glycyltrans-ferase distantly related to lipid A late acyltransferases. *J Biol Chem* 2017;292:21205–21215.

49. Mohapatra SS, Dwibedy SK, Padhy I. Polymyxins, the last-resort antibiotics: mode of action, resistance emergence, and potential solutions. *J Biosci* 2021;46:85.

50. Henderson JC, Fage CD, Cannon JR, Brodbelt JS, Keatinge-Clay AT, *et al*. Antimicrobial peptide resistance of *Vibrio cholerae* results from an LPS modification pathway related to nonribosomal peptide synthetases . *ACS Chem Biol* 2014;9:2382–2392.

51. Control GTFoC. Section 7: case management in treatment facilities; 2020. https://choleraoutbreak.org/book-page/section-7-case-management-treatment-facilities

52. Trucksis M, Galen JE, Michalski J, Fasano A, Kaper JB. Accessory cholera enterotoxin (Ace), the third toxin of a *Vibrio cholerae* viru-lence cassette. *Proc Natl Acad Sci U S A* 1993;90:5267–5271.

53. Manning PA. The TCP gene cluster of vibrio cholerae1presented at the workshop on `type 4 pili – biogenesis, adhesins, protein export, and DNA import', schloss ringberg, germany, 26–29 november 1995.12this article is dedicated to the memory of gunhild jonson, a dear colleague who contributed much to cholera research including her work on TCP, and who sadly passed away in march 1996.2. *Gene* 1997;192:63–70.

54. Gutierrez-Rodarte M, Kolappan S, Burrell BA, Craig L. The *Vibrio cholerae* minor pilin TcpB mediates uptake of the cholera toxin phage CTXϕ. *J Biol Chem* 2019;294:15698–15710.

55. Zielke RA, Simmons RS, Park BR, Nonogaki M, Emerson S, *et al*. The type II secretion pathway in Vibrio cholerae is characterized by growth phase-dependent expression of exoprotein genes and is positively regulated by σE. *Infect Immun* 2014;82:2788–2801.

56. Unterweger D, Miyata ST, Bachmann V, Brooks TM, Mullins T, *et al*. The *Vibrio cholerae* type VI secretion system employs diverse effector modules for intraspecific competition. *Nat Commun* 2014;5:3549.

57. Diepold A, Armitage JP. Type III secretion systems: the bacterial flagellum and the injectisome. *Philos Trans R Soc Lond B Biol Sci* 2015;370:20150020.

58. Dorman MJ, Domman D, Poklepovich T, Tolley C, Zolezzi G, *et al*. Genomics of the Argentinian cholera epidemic elucidate the contrasting dynamics of epidemic and endemic *Vibrio cholerae*. *Nat Commun* 2020;11:4918.

59. Kadim M. *Cholera Updates in Indonesia*. Indonesia: Gastrohepa-tology Division, Child Health Department, University of Indonesia; 2017.

60. Siahaan S, Herman MJ, Fitri N. Antimicrobial resistance situation in Indonesia: a challenge of multisector and global coordination. *J Trop Med* 2022;2022:2783300.

61. Kulinska A, Czeredys M, Hayes F, Jagura-Burdzy G. Genomic and functional characterization of the modular broad-host-range RA3 plasmid, the archetype of the IncU group. *Appl Environ Microbiol* 2008;74:4119–4132.

62. Weill FX, Domman D, Njamkepo E, Tarr C, Rauzier J, *et al*. Genomic history of the seventh pandemic of cholera in Africa. *Science* 2017;358:785–789.

63. Baharoglu Z, Bikard D, Mazel D. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet* 2010;6:e1001165.

64. Melano R, Petroni A, Garutti A, Saka HA, Mange L, *et al*. New carbenicillin-hydrolyzing beta-lactamase (CARB-7) from *Vibrio cholerae* non-O1, non-O139 strains encoded by the VCR region of the *V. cholerae* genome. *Antimicrob Agents Chemother* 2002;46:2162–2168.

65. Sanschagrin F, Bejaoui N, Levesque RC. Structure of CARB-4 and AER-1 carbenicillin hydrolyzing *β*-lactamases. *Antimicrob Agents Chemother* 1998;42:1966–1972.

66. Philippon AM, Paul GC, Thabaut AP, Jacoby GA. Properties of a novel carbenicillin-hydrolyzing beta-lactamase (CARB-4) specified by an IncP-2 plasmid from *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 1986;29:519–520.

67. Folster JP, Katz L, McCullough A, Parsons MB, Knipe K, *et al*. Multidrug-resistant IncA/C plasmid in *Vibrio cholerae* from Haiti. *Emerg Infect Dis* 2014;20:1951–1953.

68. Carattoli A, Filetici E, Villa L, Dionisi AM, Ricci A, *et al*. Antibiotic resistance genes and *Salmonella* genomic island 1 in *Salmonella enterica* serovar Typhimurium Isolated in Italy. *Antimicrob Agents Chemother*2002;46:2821–2828.

69. Valiente E, Davies C, Mills DC, Getino M, Ritchie JM, *et al*. *Vibrio cholerae* accessory colonisation factor AcfC: a chemotactic protein with a role in hyperinfectivity. *Sci Rep* 2018;8:8390.

70. Jermyn WS, Boyd EF. Characterization of a novel *Vibrio* pathogenicity island (VPI-2) encoding neuraminidase (nanH) among toxigenic *Vibrio cholerae* isolates. *Microbiology* 2002;148:3681–3693.

71. Lin W, Fullner KJ, Clayton R, Sexton JA, Rogers MB, *et al*. Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage . *Proc Natl Acad Sci USA* 1999;96:1071–1076.

72. Galen JE, Ketley JM, Fasano A, Richardson SH, Wasserman SS, *et al*. Role of *Vibrio cholerae* neuraminidase in the function of cholera toxin. *Infect Immun* 1992;60:406–415.

73. Bhandari M, Rathnayake IU, Huygens F, Nguyen S, Heron B, *et al*. Genomic and evolutionary insights into Australian toxigenic *Vibrio cholerae* O1 strains. *Microbiol Spectr* 2023;11:e03617-22.

74. Kuleshov KV, Vodop'ianov SO, Dedkov VG, Markelov ML, Deviatkin AA, *et al*. Travel-associated *Vibrio cholerae* O1 El Tor, Russia. *Emerg Infect Dis* 2016;22:2006–2008.

75. Greig DR, Schaefer U, Octavia S, Hunter E, Chattaway MA, *et al*. Evaluation of whole-genome sequencing for identification and typing of *Vibrio cholerae*. *J Clin Microbiol* 2018;56:00831–18.