RESEARCH ARTICLE

MICROBIOLOGY SOCIETY

OPEN DATA   OPEN ACCESS

Population genomics of the food-borne pathogen *Vibrio fluvialis* reveals lineage associated pathogenicity-related genetic elements

Hongyuan Zheng[1]†, Yuanming Huang[2]†, Ping Liu[2], Lin Yan[3], Yanyan Zhou[4], Chao Yang[1], Yarong Wu[1], Jingliang Qin[1], Yan Guo[1], Xiaoyan Pei[3], Yunchang Guo[3], Yujun Cui[1,*] and Weili Liang[2,*]

Abstract

Vibrio fluvialis is a food-borne pathogen with epidemic potential that causes cholera-like acute gastroenteritis and sometimes extraintestinal infections in humans. However, research on its genetic diversity and pathogenicity-related genetic elements based on whole genome sequences is lacking. In this study, we collected and sequenced 130 strains of *V. fluvialis* from 14 provinces of China, and also determined the susceptibility of 35 of the strains to 30 different antibiotics. Combined with 52 publicly available *V. fluvialis* genomes, we inferred the population structure and investigated the characteristics of pathogenicity-related factors. The *V. fluvialis* strains exhibited high levels of homologous recombination and were assigned to two major populations, VflPop1 and VflPop2, according to the different compositions of their gene pools. VflPop2 was subdivided into groups 2.1 and 2.2. Except for VflPop2.2, which consisted only of Asian strains, the strains in VflPop1 and VflPop2.1 were distributed in the Americas, Asia and Europe. Analysis of the pathogenicity potential of *V. fluvialis* showed that most of the identified virulence-related genes or gene clusters showed high prevalence in *V. fluvialis*, except for three mobile genetic elements: pBD146, ICE*Vfl*Ind1 and MGI*Vfl*Ind1, which were scattered in only a few strains. A total of 21 antimicrobial resistance genes were identified in the genomes of the 182 strains analysed in this study, and 19 (90%) of them were exclusively present in VflPop2. Notably, the tetracycline resistance-related gene *tet*(35) was present in 150 (95%) of the strains in VflPop2, and in only one (4%) strain in VflPop1, indicating it was population-specific. In total, 91% of the 35 selected strains showed resistance to cefazolin, indicating *V. fluvialis* has a high resistance rate to cefazolin. Among the 15 genomes that carried the previously reported drug resistance-related plasmid pBD146, 11 (73%) showed resistance to trimethoprim-sulfamethoxazole, which we inferred was related to the presence of the *dfr6* gene in the plasmid. On the basis of the population genomics analysis, the genetic diversity, population structure and distribution of pathogenicity-related factors of *V. fluvialis* were delineated in this study. The results will provide further clues regarding the evolution and pathogenic mechanisms of *V. fluvialis*, and improve our knowledge for the prevention and control of this pathogen.

Impact Statement

Vibrio fluvialis is a food-borne pathogen frequently found in marine environments or products, which causes watery diarrhoea (with or without the presence of blood in stools) in humans, and causes enormous economic losses to the aquaculture industry because it is also pathogenic to cultured fish and lobsters. This is the first study of the genetic diversity, population structure and pathogenicity-related elements in *V. fluvialis* based on population genomics analysis. We found that the *V. fluvialis* strains analysed in this study could be divided into two major populations, both with wide geographical distribution. The identification of virulence factors and antimicrobial resistance genes, and the structure, prevalence and distribution of important mobile genetic elements (e.g. pBD146, ICE*Vfl*Ind1, MGI*Vfl*Ind1) in the *V. fluvialis* strains have broadened our knowledge of the pathogenicity of *V. fluvialis*, and laid a good foundation for further genomic epidemiological studies and the improvement of clinical treatments. This population genomics study has deepened our understanding of *V. fluvialis* and facilitated future prevention and control efforts.

DATA SUMMARY

All sequence data for the 130 newly sequenced *V. fluvialis* strains have been submitted to the National Center for Biotechnology Information (NCBI) BioProject database under number PRJNA744458. The strain metadata are listed in Supplementary Data S1 (available in the online version of this article). Other supplementary data have been uploaded to figshare (https://doi.org/10.6084/m9.figshare.17054564.v2).

INTRODUCTION

Vibrio fluvialis, once called 'group F Vibrio' and 'EF-6 Vibrio', is a halophilic Gram-negative bacterium that is widely distributed in marine and estuarine environments [1–4]. The emergence of *V. fluvialis* is thought to have been associated with global changes impacting on social and natural environments, although this is still of debate [3]. One speculation is that *V. fluvialis* strains originated from highly polluted waters under intense population pressure in large urban concentrations [3]. *V. fluvialis*-associated infections or outbreaks occur through the faecal–oral route by ingesting contaminated food or water and have been reported worldwide, especially in regions with poor sanitation [1, 2, 5–7]. Increased isolation rates of *V. fluvialis* from patients with cholera-like diarrhoea and of multidrug-resistant clinical strains have been reported over the years [6, 8–11]. Drug-resistance genes harboured by plasmids, Class I integrons and the SXT element, MATE family efflux pumps, and mutations in DNA gyrase and topoisomerase IV account for the multiple drug-resistance phenotypes of *V. fluvialis* strains [9, 10, 12–14]. Furthermore, azithromycin-resistant *V. fluvialis* [15] and NDM-1-producing multidrug-resistant *V. fluvialis* strains have been detected, and the large plasmid harbouring blaNDM-1 was found to be easily transferred to other enteric pathogens [16]. Considering the prevalence of *V. fluvialis* infections in humans, its ubiquitous presence in the environment and the emergence of multidrug-resistant strains, *V. fluvialis* poses a potentially serious threat to public health. The prevalence of *V. fluvialis* and its related disease burden has probably been overlooked because of the challenge of distinguishing it from *Vibrio cholerae* and *Aeromonas* species, which have a close phenotypic resemblance to *V. fluvialis* [4], and the lack of systematic surveillance [7] and commercial diagnostic sera. Therefore, it is important to clarify the evolutionary and pathogenic characteristics of *V. fluvialis* to improve the surveillance and control of this pathogen.

In a previous study, we primarily investigated the virulence genes and molecular epidemiological characteristics of *V. fluvialis* in China using traditional molecular biological methods, such as PFGE and PCR [7]. We also studied the composition of the quorum sensing (QS) system and its regulated pathogenesis in *V. fluvialis* [17], determined the genetic bases of *V. fluvialis* species-specific biochemical pathways [18], and further delineated the genetic content and organization, functional operation and regulation of the type VI secretion system (T6SS) in *V. fluvialis* that favours its environmental fitness [19, 20]. Recently, we showed that a cholerae autoinducer 1 and autoinducer 2-based CqsA/LuxS-HapR QS cascade modulated T6SS2 in *V. fluvialis*, further enhancing our understanding of crosstalk between T6SS and QS in microbes [21]. Other researchers have investigated the structure and functional characteristics of pathogenic- or antibiotic-related mobile genetic elements (MGEs), including plasmids, integrons and integrating conjugative elements (ICEs), in *V. fluvialis* [12, 22, 23].

Although *V. fluvialis* is now recognized as an important pathogen, its genetic diversity, pathogenicity and genomic epidemiological characteristics are still poorly understood. A few whole genome sequence studies of *V. fluvialis* have been reported, which focused on the genetic components of MGEs such as ICE*Vfl*Ind1 in certain *V. fluvialis* strain [24], characterized genomic features of two bile-isolated *V. fluvialis* strains [25] and genetically explained biological phenomena in specific strains [26]. In this study, we used large-scale genomic epidemiological techniques to investigate the population structure, virulence factors (VFs), antimicrobial resistance genes (ARGs), MGEs and their distribution in strains of *V. fluvialis*, to further characterize the genomic epidemiology of *V. fluvialis* and improve our knowledge for the prevention and control of this pathogen.

METHODS

Preparation of genome dataset

For this study, we collected and sequenced a total of 130 *V. fluvialis* strains; 129 were isolated from China and one was collected from Bangladesh; 61 were isolated from patients with diarrhoea and 68 were from environment samples, while the source of the remaining one strain was unclear. These strains were collected between 1963 and 2018 from 14 provinces in China. All the *V. fluvialis* strains were grown in Luria-Bertani (LB) medium containing 2% (w/v) NaCl at 37 °C. Genomic DNA was extracted with overnight culture using a Wizard Genomic DNA Purification Kit (Promega). DNA libraries of the 130 strains were prepared according to the Illumina protocol and sequenced on an Illumina NovaSeq 6000 system, with paired-end runs and read length of 150 bp. An average of 1.2 Gb clean data were generated for each strain and the average sequencing depth was 279-fold (123–451-fold). We also downloaded 52 publicly available *V. fluvialis* genome sequences (assembly and SRA) from the National Center for Biotechnology Information (NCBI) database (up to November 2020). These strains were isolated mainly from the USA, England and Bangladesh between 1970 and 2018. Details of the genomes of all 182 strains analysed in this study are given in Supplementary Data S1.

Pan-genome analysis, variant calling and GWAS analysis

The genomes were assembled *de novo* using SOAPdenovo v2.04 [27], as previously described [28]. The genes and protein coding sequences (CDSs) were predicted for each genome using Prokka [29], and the GFF3 files produced by Prokka were used for pan-genome analysis with Roary [30]. We identified SNPs according to a previously described method [31]. First, all the assemblies were aligned against the reference genome ATCC 33809 (CP014034.2, CP014035.2) using MUMmer v3.1 [32] to generate a whole-genome alignment and identify SNPs in the core genome (regions present in all 182 strains). Second, the clean sequencing reads of newly sequenced genomes were mapped to their respective draft genome assemblies to evaluate the SNP accuracy with SOAPaligner [33]. Only high-quality SNPs (supported by >10 reads, quality value >30) were retained in the analysis. The repetitive regions in the reference genome were identified using TRF v4 [34] and a BLASTN search against itself. SNPs in the repetitive regions were excluded and only biallelic SNPs (92%) were used in the subsequent analysis. A total of 382629 SNPs were identified in the 182 strains, which were then used to construct a maximum-likelihood (ML) tree with IQ-Tree v1.6.5 [35]. The main parameters were '-bb 1000 -m GTR +G -nt AUTO', which meant IQ-Tree performed ultrafast bootstrap for 1000 replicates, the models were GTR +Gamma, and the number of threads was set to auto. The phylogenetic tree was visualized with iToL [36]. We also tried to identify the genetic differences between clinical and environmental strains based on a unitigs-based genome-wide association study (GWAS) through the software pyseer v1.2.0 [37, 38] (https://pyseer.readthedocs.io/en/master/index.html). We followed the manual of pyseer and conducted the analysis based on unitigs with default parameters. A strain from a diarrhoea sample was regarded as the positive phenotype and accordingly 72 clinical strains were clustered into positive group as phenotype 1, and the 80 environmental strains were clustered into the negative/control group as phenotype 0.

Calculation of linkage disequilibrium (LD) decay

LD decay was calculated using the Haploview software [39] based on previously reported SNP sets of eight species (*Burkholderia pseudomallei*, *Escherichia coli*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Bacillus thuringiensis*, *Legionella pneumophila*, *Vibrio cholerae*, *Vibrio parahaemolyticus*) [40] and newly identified SNP sets of *V. fluvialis* in the current study. The main parameters were '-maxdistance 30 -minGeno 0.6 -minMAF 0 -hwcutoff 0', which meant the maximum intermarker distance for LD comparisons was 30 kb, markers with <60% valid data were excluded, and the threshold of the minor allele frequency and the Hardy–Weinberg *P* value were set to 0 to include all the identified SNPs in the calculation. A sliding window of 100 bp was used to summarize the pairwise LD (r^2) values.

Population structure identification

The fineSTRUCTURE algorithm v2.0.7 [41] was used to determine the population structure, as previously described [42]. First, we converted the matrix of core genome SNPs to a Fasta format file and then calculated the pairwise SNP distances (PSDs) of the 182 genomes through he software snp-dists v0.6 (https://github.com/tseemann/snp-dists). To avoid interference from the clonal signal, we identified 131 semi-clonal groups (SCGs) based on the threshold of 5000 PSDs, and randomly selected one strain from each SCG for population clustering. The preliminary clustering result indicated that some small groups comprised only two or three strains and still showed strong clonal signal, so we removed all but one strain from these small groups, and finally a total of 104 representative genomes were obtained. Then, we reran fineSTRUCTURE with the SNPs of the 104 representative genomes using the default settings, to infer the population structure.

Antimicrobial susceptibility testing

The antimicrobial susceptibility phenotype of the 15 strains with plasmid pBD146 and 20 randomly selected strains without pBD146 was tested using the microbroth dilution method in accordance with the instructions of the Clinical and Laboratory Standards Institute (CLSI) [43]. Thirty antimicrobial agents, namely ampicillin, ampicillin-sulbactam, tetracycline,

chloramphenicol, trimethoprim-sulfamethoxazole, cefazolin, cefotaxime, ceftazidime, cefoxitin, gentamicin, imipenem, nalidixic acid, azithromycin, sulfisoxazole, ciprofloxacin, amoxicillin-clavulanic acid, cefotaxime-clavulanic acid, ceftazidime-clavulanic acid, colistin, polymyxin B, minocycline, amikacin, aztreonam, cefepime, meropenem, levofloxacin, doxycycline, kanamycin, streptomycin and gemifloxacin, were tested and interpreted according to the CLSI minimum inhibitory concentration (MIC) interpretive standards for *Enterobacteriaceae* and *Vibrio* to determine susceptibility and resistance [44, 45]. Because no interpretive criteria for *V. fluvialis* were available based on CLSI guidelines, the antimicrobial MICs were determined by referring to the CLSI standards for *V. cholerae* if available; otherwise breakpoints for *Enterobacteriaceae* were used. Multidrug resistance was defined as an absence of susceptibility to two or more classes of antibiotics. *E. coli* strain ATCC 25922 was used as the quality control for the antibiotic susceptibility tests.

Identification of VFs, ARGs and MGEs

The assembled genomes of all 182 strains were compared with the core dataset (setA) of the Virulence Factors Database (VFDB) [46] to detect VFs using BLASTN. We also selected VFs that have been reported and/or experimentally verified to be involved in the pathogenesis in *Vibrio* species, *E. coli* and *Salmonella* spcies, and searched for the associated encoding genes in the whole genome dataset, including fimbrial low-molecular-weight protein (Flp) pili biogenesis gene clusters, PilA pili biogenesis gene clusters, flagellar-related genes, haemolysin, haem utilization protein HupO, metalloprotease HapA, and transcriptional regulator coding genes, QS-related genes, potential T2SS gene cluster, mannose-sensitive haemagglutinin (MSHA) synthesis genes and curli pili biogenesis gene cluster. The ID or accession numbers of these virulence-related genes are listed in Supplementary Data S2. ARGs were identified using ResFinder [47] by comparing the 182 genomes to the ResFinder database (https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/). We also compared all the assembled genomes with the core genome sequences of VflT6SS (VflT6SS1: KY319184, VflT6SS2: KY319183, three 'orphan' *hcp-vgrG* modules: *hcp-vgrG_a*: KY319185, *hcp-vgrG_b*: KY319186, *hcp-vgrG_c*: KY319187 [19]), T3SS1 gene cluster (KF990997) and three MGEs (pBD146: EU574928, ICE*Vfl*Ind1: GQ463144, MGI*Vfl*Ind1: KC117176) in *V. fluvialis*, using BLASTN to reveal the structure and prevalence of VflT6SS and MGEs in the strains of *V. fluvialis*. A gene/segment was considered present if the overall hit coverage and identity were both ≥70%, as previously described [48]. Because of the high genetic diversity of ICEs and mobilizable genomic islands (MGIs) in the evolutionary process, the coverage threshold was relaxed to 40%. Sequence alignment of VflT6SS and MGEs was performed and visualized using Geneious v4.8.3 (www.geneious.com/) and BRIG v0.95 [49], respectively. Default parameters were used for all the software settings, unless stated otherwise.

RESULTS

Genome features of *V. fluvialis*

We collected and sequenced 130 *V. fluvialis* strains from patients with diarrhoea and from environmental aquatic products from 14 provinces in China, between 1963 and 2018. The 130 sequenced genomes together with 52 publicly available *V. fluvialis* genome sequences from the NCBI that were mainly collected from the Americas, Europe and Asia between 1970 and 2018 (a total of 182 genomes) were included in the analysis (Data S1). The number of contigs and average sizes of the assemblies were 44 (16–91) and 4.7 Mb (4.4–5.0 Mb), respectively, with an average GC content of 50% (49.56–50.33%). Pan-genome analysis of the whole dataset identified 18135 genes, including 2786 core genes (genes present in all of strains) and 15349 accessory genes. The relationship between the numbers of genes and *V. fluvialis* genomes (Fig. S1) showed that the number of pan-genes of all the *V. fluvialis* strains rose rapidly along with the increase of analysed strains, whereas the number of core genes tended to be stable. This indicates that the increasing number of accessory genes and the 'open' pan-genome of *V. fluvialis* may mean that this pathogen is able to continuously acquire new genes through horizontal gene transfer in the process of evolution.

Genetic diversity and population structure of *V. fluvialis*

Genome-wide coverage ranged from 83.7 to 93.4% when aligning draft genome assemblies against the reference genome, and the core genome of *V. fluvialis* was 3.2 Mb based on the current dataset. We reconstructed an ML phylogenetic tree of the whole genome dataset based on the 382629 SNPs identified from the core genome (Fig. 1a). First, to identify the differences between clinical and environmental strains, we observed the genealogical relationship between the clinical and environmental strains based on the phylogenetic tree of all the 182 samples (Fig. S2). We found that strains from different sources were scattered along the tree, with no lineage revealing a source-specific trend. In particular, strains from different sources could appear in the same clone, suggesting no phylogenetic signal related to the source of strains. Then, by using the software pyseer v1.2.0 [37, 38], we conducted a GWAS analysis of *V. fluvialis* to identify the non-phylogenetic signal possibly related to the different isolation sources. The pyseer program scanned the 152 genomes of phenotype 1 and 0 to search for specific genetic markers (unitigs) that can be used to predict the source of the strains. However, no specific unitigs could satisfy the significant threshold, which suggested that no genome fragments in our dataset were specifically relevant to the different sources of the samples. So, no specific differences were observed between clinical and environmental strains of the current dataset. Similar to *Vibrio alginolyticus*, *V. parahaemolyticus* and other high-frequency recombination bacteria [42, 50], the inferred phylogenetic tree of *V. fluvialis* had long terminal
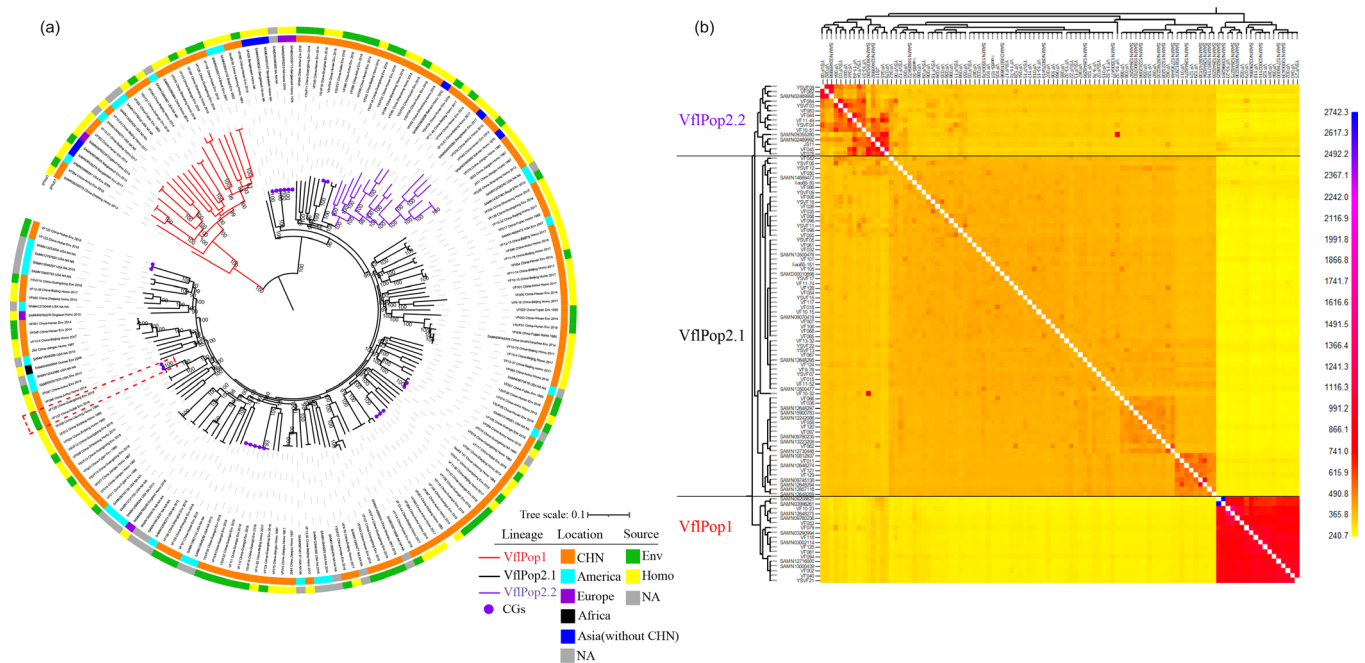
Fig. 1. Phylogenic tree and population structure of *V. fluvialis* strains. (a) Phylogenic tree of 182 *V. fluvialis* strains based on the core genome SNPs. Red, black and purple branches represent strains in VflPop1, VflPop2.1 and VflPop2.2, respectively. Purple dots on top of branches mark the seven identified clonal groups. The red dotted boxes indicate the two strains involved in the cross-provinces spread in China (lower left). Numbers at nodes are bootstrap values, and only values greater than 80 are displayed. The ring colours from inner to outer indicate isolation location and sample type, respectively. Bar, 0.1 nucleotide substitutions per site. (b) Population structure of the 104 representative strains based on fineSTRUCTURE. Columns and rows represent the donor and recipient strains, respectively. The colour of the boxes indicates the number of genomic fragments introduced from the donor to the recipient. The black lines represent the boundary between the three populations.

branches, indicating that frequent homologous recombination events among strains had disrupted the vertical genetic signal. We compared the LD decay of *V. fluvialis* with eight previously reported species [40]. The curve of LD decay of *V. fluvialis* was very close to that of *Burkholderia pseudomallei* and *E. coli* (Fig. S3), further confirming the high recombination level of *V. fluvialis*.

According to the protocol used previously for other high recombination bacterial species, such as *V. parahaemolyticus* and *Helicobacter pylori* [28, 42, 51], we reconstructed the population structure across all the 182 genomes using fineSTRUCTURE based on genome-wide SNPs, which divided *V. fluvialis* into two main populations, VflPop1 and VflPop2, and VflPop2 was further divided into two subpopulations, VflPop2.1 and VflPop2.2 (Fig. 1). The fineSTRUCTURE software infers the population structure and clusters strains into different populations by calculating the pairwise co-ancestry rate, which is intuitively reflected by the number of genomic fragments (chunk) provided by the donor to the recipient [41]. In the present study, the co-ancestry matrix of *V. fluvialis* was composed of two main boxes shown in red and orange in Fig. 1(b), indicating the strains in each of them shared more co-ancestry chunks, whereas the strains within the yellow boundary shared fewer co-ancestry chunks, suggesting a relatively small shared gene pool between them. The statistics for VflPop1 and VflPop2.1 showed that they both contained strains from the Americas, Asia and Europe (Table 1), the only African strain belonged to VflPop2.1, and VflPop2.2 contained only Asian strains (Fisher's exact test, *P*=0.0082), indicating that strains of VflPop2.1 spread more widely around the word.

Table 1. Number of strains in different populations according to their isolation locations

	China	America	Europe	Asia (without China)	Africa	Others	Total
VflPop1	14 (58.3%)	7 (29.2%)	1 (4.2%)	2 (8.3%)	0	0	24
VflPop2.1	104 (78.4%)	22 (15.8%)	4 (2.9%)	2 (1.4%)	1 (0.7%)	6 (4.3%)	139
VflPop2.2	16 (84.2%)	0	0	3 (15.8%)	0	0	19

Numbers and percentages indicate the number of strains and percentages in each population.
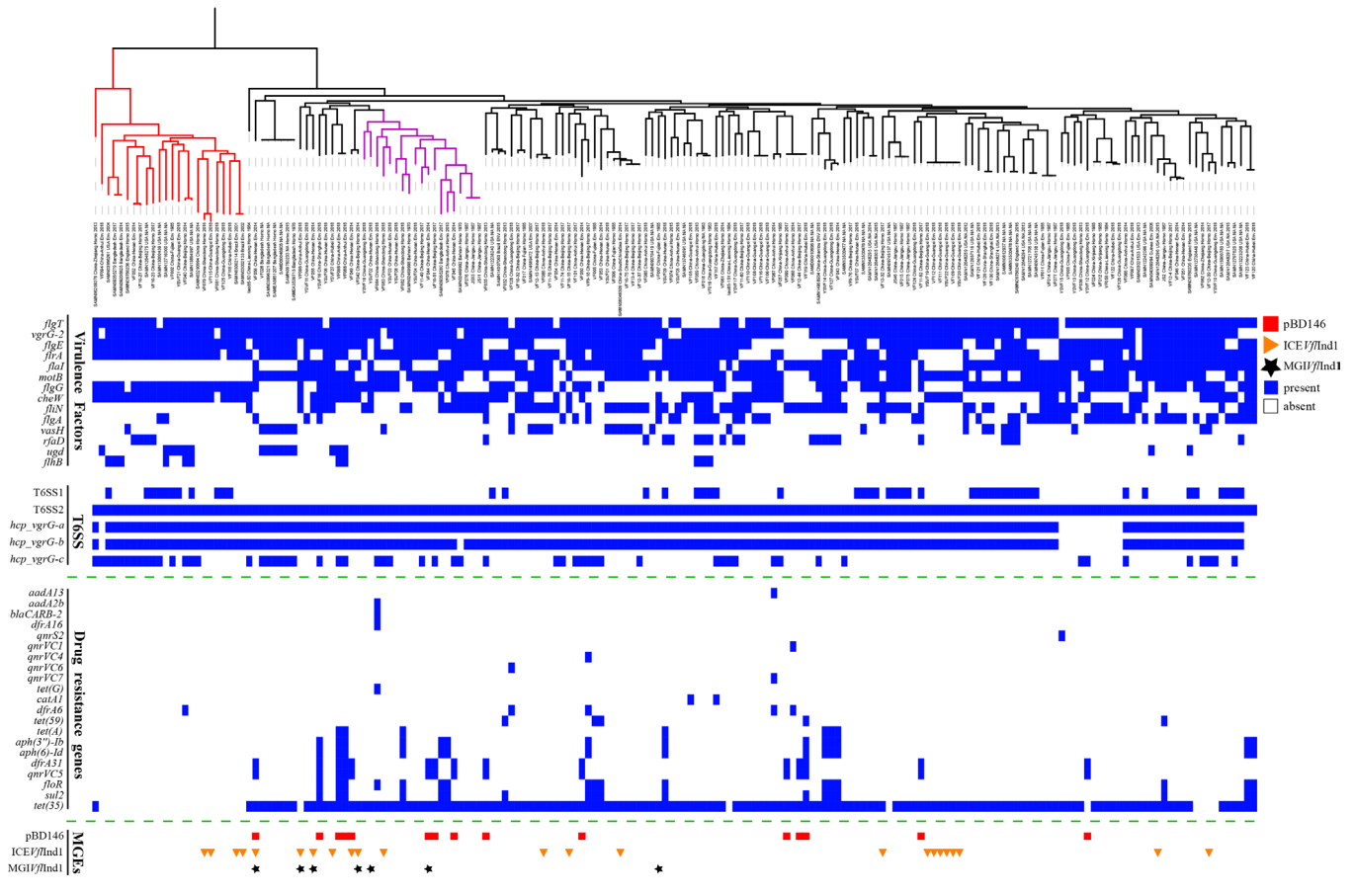
Fig. 2. Heatmap of VFs, T6SS, drug resistance genes and MGEs in *V. fluvialis*. Blue boxes indicate the presence of VFs, T6SS and drug resistance genes; red boxes indicate the presence of the MGE pBD146; orange triangles and black pentacles indicate the presence of ICE*Vfl*Ind1 and MGI*Vfl*Ind1, respectively; white boxes indicate absence. The VFs were the 14 variably present VFs identified in the VFDB.

The presence of a clonal group (CG) usually suggests the recent spread of a specific bacterial lineage. We identified seven CGs (Fig. 1a), involving 22 strains, according to the threshold of PSD of <10 SNPs [31]. However, no CG was observed that contained strains isolated from different countries. Therefore, the genome dataset in the current study could not prove a recent transoceanic spread event. The strains in each of the other six CGs were all isolated from the same province in China, except for one CG of the lower left dashed box in Fig. 1(a), which was composed of two strains from two different provinces (VF126, Guangdong, 2018; and VF122, Hubei, 2018). The PSD between them was nine SNPs, indicating the occurrence of cross-province spread events in China.

Identification and distribution of VFs

We mapped the assembled genomes of all 182 strains against the core dataset of the VFDB to assess the virulence potential of *V. fluvialis*. A total of 52 VF genes were detected, among which the flagellum-related gene cluster and T6SS-related gene cluster were predominant, involving 34 and 11 related genes, respectively, accounting for 45 (87%) of the 52 VFs (Data S3). The flagellum determines the motility and chemotaxis of microbes that are generally considered as virulence determinants in bacterial pathogenesis, and it also functions as an adhesin factor, mediating the adhesion of microbes to the cell surface of the host [52]. T6SS is involved in the virulence and environmental fitness of various bacterial species [53, 54]. Additionally, all the 182 genomes carried the immunogenic lipoprotein A encoding gene *IlpA*, which triggers cytokine production in human monocytes by activating the toll-like receptor 2 (TLR2) and functions as an adhesion factor in *Vibrio vulnificus* [55, 56]. The lipooligosaccharide synthesis gene *rfaD*, capsule-related gene *ugd*, and RTX toxin-related genes *rtxB* and *rtxD* were found in 1–13 % of all the genomes (Data S3).

Among the 52 VFs identified in VFDB, 31 were present in >95% of all genomes, seven were present in only a few strains (<5%) and the remaining 14 were variably present among 5–95% of the genomes. By observing the distribution of the 14 variably present VFs in the phylogenetic tree (Fig. 2) and in the three populations (Fig. 3), we found that seven of them (*flgT*, *vgrG-2*, *flgE*, *flrA*, *flgG*, *cheW*, *vasH*) were scattered in all three populations, although the percentage varied. However, three flagellum-related VFs
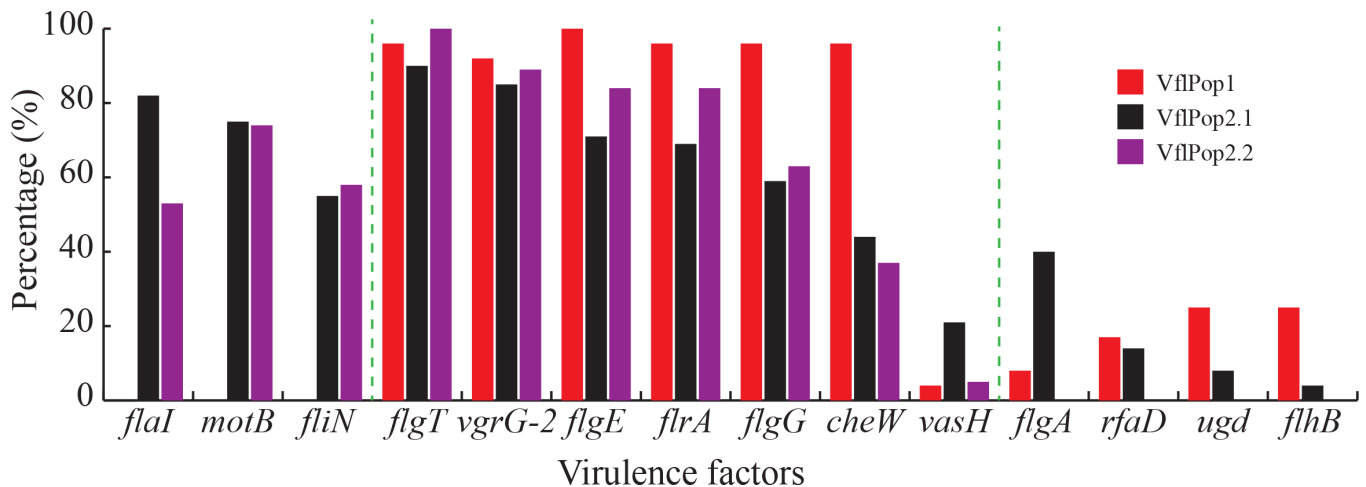
Fig. 3. Proportions of the 14 variably distributed VFs in three populations. The 14 variably present VFs are on the horizontal axis; the percentages of the VFs in three different populations are shown on the vertical axis.

(*flaI*, *motB*, *fliN*) were present only in VflPop2 and absent in VflPop1. Another four VFs (*flgA*, *flhB*, *rfaD*, *ugd*) were present only in VflPop1 and VflPop2.1, and absent in VflPop2.2.

We also investigated the prevalence of selected VFs by searching the whole genome with the sequences of virulence-related genes annotated in *V. fluvialis* ATCC 33809 and *V. fluvialis* XJ85003 (Data S2). We found that nearly all strains were equipped with VFs (Fig. S4), including flagellar, HapA metalloprotease, haem utilization protein HupO, virulence transcriptional regulators ToxR/S, CqsA/LuxS-HapR and acyl-homoserine lactone-based VfqI-VfqR QS systems, T2SS, and curli pili. Haemolysin, PilA pili and MSHA synthesis gene clusters were also highly prevalent in the *V. fluvialis* genomes in the current dataset except for a few specific loci in these three gene clusters. HupO is a haem utilization protein identified in *V. fluvialis*, which has high amino acid sequence homology to bacterial outer membrane haem receptors in *Vibrio* species and contains motifs that are common to bacterial haem receptors, including a consensus TonB box, a FRAP domain and an NPNL domain. HupO has been shown to be induced under iron-restricted conditions and mediate the acquisition of iron from hemin in *V. fluvialis* [57]. However, the direct interaction of HupO with possible TonB transport proteins remains to be examined. The QS system has been demonstrated to positively regulate the production of two potential VFs (extracellular protease and haemolysin) in *V. fluvialis*, and it affects cytotoxic activity against epithelial tissue cultures [17]. The T2SS is used to selectively translocate VFs from the periplasmic space into the extracellular environment by many Gram-negative bacteria, such as cholera toxin, and several enzymes in *V. cholerae* [58]. Curli promote biofilm formation in *E. coli* and *Salmonella* species. and thus is an important determinant in microbial adaptation to different environments [59]. Haemolysin is the most widely distributed exotoxin with pore-forming activity and has various roles in the infection process [60]. Several predicated haemolysins were detected in *V. fluvialis*, among which VFH encoded by *hlyA* (AL536_RS29720) is classified as HlyA or El Tor haemolysin [61, 62]. VFH damages the erythrocyte membrane by acting as a pore-forming toxin, with the pores estimated to be approximately 2.8–3.7 nm in diameter on rabbit erythrocytes [61]. We also demonstrated that VFH induces IL-1β secretion by activating the NLRP3 inflammasome and contributes to the pathogenicity of *V. fluvialis* [63]. The haemolysin encoded by AL536_RS29725 is similar to the thermolabile haemolysin (TLH) of *V. parahaemolyticus*. AL536_RS29725 is 67 % homologous to the TLH coding gene *tlh*. TLH of *V. parahaemolyticus* was heat labile (60°C for 10 min) and had phospholipase A2/lysophospholipase activity [60]. AL536_RS43070 encodes a putative haemolysin III with 82 % nucleotide sequence identity to *hlyIII* from *V. vulnificus,* which has proven haemolytic activity [64]. The other putative haemolysins or precursors remain to be verified. As surface filamentous organelles of bacteria, PilA plays key roles in microbe–microbe interactions and in the formation of bacterial aggregates and microcolonies, thus enabling pathogens to adhere to host cells [65]. MSHA was shown to be associated with the adherence and biofilm formation of *V. cholerae* [66, 67]. Notably, the prevalence of Flp pili-associated gene clusters dwas highly variable among the strains and the three populations. Two predicted Flp pili coding gene clusters were found in *V. fluvialis*. The Flp pili gene cluster I (AL536_RS40405–AL536_RS40340), which was located on the large chromosome of *V. fluvialis* and displayed relatively high homology with the *tad* III locus of *V. vulnificus* that encodes CMCP6 Flp pili [68], was scattered in VflPop1 and VflPop2.1 while being almost absent in VflPop2.2. The Flp pili gene cluster II (AL536_RS28745–AL536_RS28710) on the small chromosome was scattered among the strains in VflPop2.1 and VflPop2.2, but was present in all the strains in VflPop1. Flp pili play essential roles in adherence, colonization, biofilm formation and pathogenicity in a number of genera and were recently demonstrated to confer successful invasion of *V. vulnificus* into host

deeper tissue, survive in the bloodstream and show resistance to complement activation [68]. The correlation of the two Flp pili gene clusters with strain virulence in different *V. fluvialis* populations requires further study.

The T6SS plays an important role in interbacterial or intraspecies antagonism, biofilm formation, host–microbe interactions and competition in the host [69]. Previously, we characterized the genetic organization of two T6SS gene clusters (VflT6SS1 and VflT6SS2) in the clinical *V. fluvialis* strain 85003, and PCR screening of selected genes in the two VflT6SS loci showed that VflT6SS2 was more prevalent than VflT6SS1 [19]. In strain 85003, VflT6SS2 contained three 'orphan' *hcp-vgrG* modules (*hcp-vgrG_a*, *hcp-vgrG_b*, *hcp-vgrG_c*) that were functionally expressed to mediate the antibacterial activity of strain 85003 [18–20]. These studies used a limited number of strains and traditional molecular biological methods and did not provide a full description of the distribution of the two VflT6SS clusters and three *hcp-vgrG* modules in the strains of *V. fluvialis*. In the present study, the genomes of all the strains (100%) contained VflT6SS2, but VflT6SS1 was present in only 48 (26%) of the genomes (Fig. 2), confirming the previous finding that VflT6SS2 was more prevalent than VflT6SS1 in *V. fluvialis*. Furthermore, the 48 genomes carrying VflT6SS1 were all from VflPop1 and VflPop2.1, and VflT6SS1 was absent in the genomes from VflPop2.2. Additionally, *hcp-vgrG_a*, *hcp-vgrG_b* and *hcp-vgrG_c* were present in 169 (93%), 168 (93%), and 57 (31%) genomes, respectively (Fig. 2), indicating that *hcp-vgrG_a* and *hcp-vgrG_b* were much more prevalent than *hcp-vgrG_c* in *V. fluvialis*. Most of the genomes that contained *hcp-vgrG_c* also contained *hcp-vgrG_a* and *hcp-vgrG_b*, whereas several strains contained only *hcp-vgrG_c*, the least prevalent of the three. Whether *hcp-vgrG_c* can support VflT6SS2 assembly remains unclear considering its lowest expression level [20]. Multiple sequence alignments of the 'core' gene clusters of the 48 identified VflT6SS1 clusters showed that the pairwise identity was 98.3%, indicating the internal structure of VflT6SS1 was highly conserved in *V. fluvialis* (Fig. S5a). VflT6SS2 had higher diversity than VflT6SS1 and could be split into three groups based on coverage of the 'core' gene clusters: 85% (41 strains), 96% (61 strains) and 100% (80 strains) (Fig. S5b). When VflT6SS2 was aligned against the reference genome, we found that most of the strains with lower coverage had incomplete deletions of *tssI2*, a gene that encodes VgrG protein, in their genomes.

The T3SS was first discovered in *Yersinia* species and then on the surface of many other Gram-negative bacteria, including *Vibrio* species [70]. It plays a critical role in cytotoxicity in mammalian epithelial cells and leukocytes [71]. Two distinct T3SSs (T3SS1 and T3SS2) were originally identified in *V. parahaemolyticus* (serotype O3:K6, strain RIMD2210633) [72]. First, we compared all the genomes in the current study to the VFDB core dataset (setA) based on BLASTN, but none of the T3SS-related genes were identified. Then, we downloaded the sequence of the T3SS1 gene cluster (KF990997) of *V. parahaemolyticus* strain VP49, and compared it to all the 182 genomes through BLASTN. Coverage ranged from 0 to 3.6%, which also indicated the absence of T3SS1 in the current genome dataset. We thus consider that *V. fluvialis* does not possess the T3SS.

Identification and distribution of ARGs

To determine the distribution of antimicrobial ARGs in *V. fluvialis*, we searched for ARGs in the 182 genomes using ResFinder and identified a total of 21 ARGs, which were related to resistance mainly to tetracycline, aminoglycoside and quinolone (Table 2). According to the distribution of the ARGs on the phylogenetic tree, we found that 19 (90%) of them were concentrated in VflPop2, and the other two were concentrated in VflPop1 [*drfA6* and *tet*(35)]. In particular, *tet*(35), the tetracycline resistance-related gene, was present in 95% of the strains in VflPop2, but in only 4% of the strains in VflPop1 (Fig. 2).

Antibiotic resistance phenotype of plasmid pBD146-carrying strains

Previous researches reported a clinical *V. fluvialis* strain (BD146) with multiple drug resistance, which was introduced by a high-copy plasmid pBD146 (GenBank: EU574928) and a low-copy plasmid carrying a class I integron through horizontal gene transfer [12, 22]. Antimicrobial susceptibility of pBD146 transformants showed that the 7.5 kb long pBD146 conferred resistance

Table 2. ARGs, related antibiotics, and resistance mechanisms of *V. fluvialis*

Antimicrobial-resistance genes (ARGs)	Related antibiotics	Resistance mechanisms
tet(35)(83%) *tet*(A)(4%) *tet*(59)(3%) *tet*(G)(1%)	Tetracycline resistance	Antibiotic efflux
aph(3″)-Ib(8%) *aph(6)-Id*(8%) *aadA13*(1%) *aadA2b*(1%)	Aminoglycoside resistance	Antibiotic inactivation
qnrVC5(8%) *qnrS2*(1%) *qnrVC1*(1%) *qnrVC4*(1%) *qnrVC6*(1%) *qnrVC7*(1%)	Quinolone resistance	Antibiotic target protection
sul2(10%) *floR*(9%) *catA1*(1%)	Sulphonamide resistance, phenicol resistance	Antibiotic target replacement (*sul2*), antibiotic efflux (*floR*), antibiotic inactivation (*catA1*)
dfrA31(8%) *dfrA6*(3%) *dfrA16*(1%)	Trimethoprim resistance	Antibiotic target replacement
blaCARB-2(1%)	Beta-lactam resistance	Antibiotic inactivation

Percentages in parentheses are the percentage of strains carrying the corresponding genes in the whole dataset.
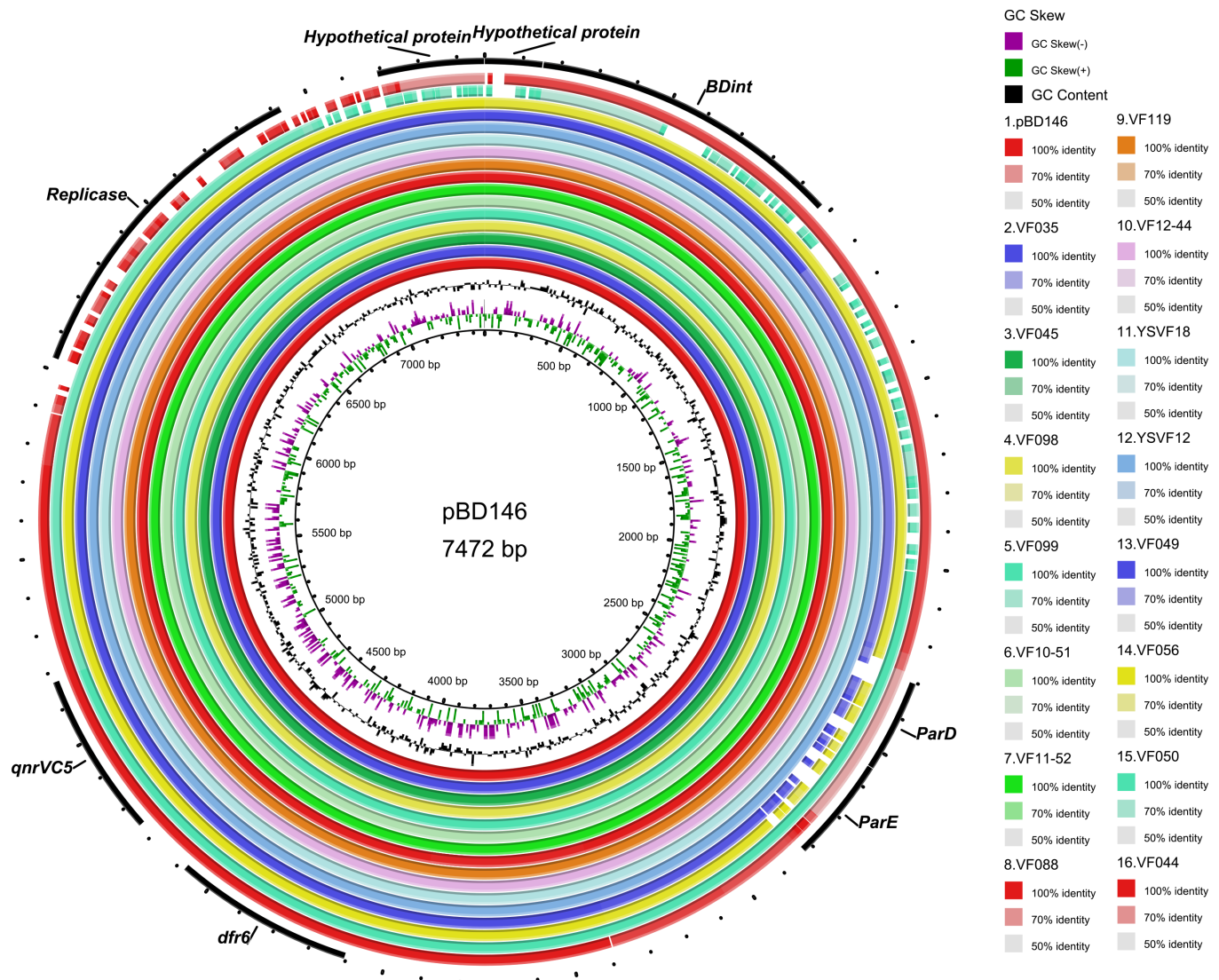
Fig. 4. Comparison of the genetic structure of pBD146. The 15 newly identified pBD146 structures are shown. The outer ring indicates genes based on the annotation of pBD146 (GenBank: EU574928).

only to trimethoprim, and the resistance to the other antibiotics was possibly due to the low-copy plasmid with a class I integron [12]. For the low-copy plasmid, only the sequence of the 4.0 kb variable region amplified from the class I integron (GenBank: KY883670) was available. However, this region was absent in all 182 genomes in this study, suggesting that the low-copy plasmid is rare in *V. fluvialis*.

By aligning reference sequences of pBD146 with the 182 genome sequences, we found that pBD146 was present in 15 of the genomes from China; 11 were complete plasmids and four had fragments missing (coverage on reference plasmid sequences was 66–100%, Fig. 4). Strains VF049 and VF056 had lost the important toxin–antitoxin genes *parD*/*parE*, which are genetic elements that promote stable plasmid inheritance [73], and strains VF044 and VF050 had lost genetic fragments involving genes that encode replicase and integrase (BDint), respectively. The distribution of pBD146 on the phylogenic tree (Fig. 2) showed that all the pBD146-carrying strains were in VflPop2.

To verify the antimicrobial resistance phenotypes of the pBD146-carrying strains, we determined the antimicrobial susceptibility of the 15 strains with plasmid pBD146 and 20 randomly selected strains without pBD146 (Table 3). The pBD146-carrying strains were susceptible to most of the antibiotics tested except for three common clinical antibiotics, trimethoprim-sulfamethoxazole (SXT), cefazolin (CFZ) and streptomycin, with resistance rates of 73, 87 and 27%, respectively (Fig. 5). Similar to the Indian multiple drug resistance strain BD146, the 15 pBD146-carrying strains also had much higher resistance rates (73%) to SXT than the 20 strains without pBD146 (10%). This phenotype was supposed be caused by the *dfr6* gene in pBD146, according to Rajpara

Table 3. Antimicrobial susceptibility of *V. fluvialis* strains with and without pBD146

The MIC values filled in light blue indicate resistance to different kinds of antibiotics based on the breakpoints for *Enterobacteriaceae*. The colour-coded top indicates different kinds of antibiotics: yellow, β-lactams; blue, chloramphenicols; orange, quinolones; red, aminoglycosides; green, tetracyclines; grey, chloramphenicols; cyan, macrolides; purple, peptides. Cefazolin (CFZ), ampicillin (AMP), ampicillin-sulbactam (AMS), amoxicillin-clavulanic acid (AMC), cefotaxime (CTX), aztreonam (AZM), cefepime (FEP), ceftazidime (CAZ), cefoxitin (CFX), gentamicin (GEN), imipenem (IMI), meropenem (MEM), cefotaxime-clavulanic acid (CTX-C), ceftazidime-clavulanic acid (CAZ-C), trimethoprim-sulfamethoxazole (SXT), sulfisoxazole (Sul), nalidixic acid (NAL), gemifloxacin (GEM), levofloxacin (LEV), ciprofloxacin (CIP), streptomycin (STR), amikacin (AMI), kanamycin (KAN), tetracycline (TET), minocycline (MIN), doxycycline (DOX), chloramphenicol (CHL), azithromycin (AZI), colistin (CT), polymyxin B (PB).

		CFZ	AMP	AMS	AMC	CTX	AZM	FEP	CAZ	CFX	GEN	IMI	MEM	CTX-C	CAZ-C	SXT	Sul	NAL	GEM	LEV	CIP	STR	AMI	KAN	TET	MIN	DOX	CHL	AZI	CT	PB
15 strains with pBD146	VF035	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.25	0.25	0.125	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF044	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	1	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.12	0.25	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF045	>16	16	8/4	16/8	0.5	≤1	≤0.25	≤0.5	4	≤1	1	0.125	0.5/4	≤0.25/4	8/152	≤32	≤4	0.12	0.25	0.125	8	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF049	4	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	4	≤1	1	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.12	0.25	0.125	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF050	>16	64	32/16	64/32	2	≤1	≤0.25	≤0.5	8	≤1	0.5	0.125	1/4	≤0.25/4	4/76	≤32	≤4	0.12	0.25	0.125	8	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
	VF056	≤0.5	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	0.5	0.125	≤0.12/4	≤0.25/4	4/76	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF088	8	4	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	8	≤1	0.5	0.125	≤0.12/4	≤0.25/4	>8/152	256	≤4	0.12	0.25	0.125	32	≤4	≤8	8	≤1	≤0.5	8	≤2	≤0.5	≤0.5
	VF098	8	≤2	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	4	≤1	0.5	0.125	≤0.12/4	≤0.25/4	>8/152	256	≤4	0.12	0.25	0.125	16	≤4	≤8	8	≤1	≤0.5	8	≤2	≤0.5	≤0.5
	VF099	8	4	8/4	4/2	≤0.25	≤1	≤0.25	≤0.5	4	≤1	1	0.125	≤0.12/4	≤0.25/4	8/152	128	≤4	0.12	0.25	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF119	8	4	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	8	≤1	0.5	0.125	≤0.12/4	≤0.25/4	>8/152	128	≤4	0.25	0.25	0.25	>32	≤4	≤8	16	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	YSVF12	>16	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	≤0.25	≤0.06	≤0.12/4	≤0.25/4	4/76	≤32	≤4	0.25	0.25	0.125	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	YSVF18	8	4	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	4	≤1	1	0.125	≤0.12/4	≤0.25/4	>8/152	256	>64	2	8	2	>32	8	≤8	≤1	≤1	≤0.5	8	≤2	≤0.5	1
	VF10-51	8	4	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	8	≤1	0.5	0.25	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	8	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF11-52	>16	8	8/4	16/8	0.5	≤1	≤0.25	≤0.5	4	≤1	1	0.125	0.5/4	≤0.25/4	4/76	≤32	≤4	0.12	0.25	0.125	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
	VF12-44	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	0.5	0.125	≤0.12/4	≤0.25/4	4/76	≤32	≤4	0.06	≤0.125	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5

Continued

Table 3. Continued

|	CFZ	AMP	AMS	AMC	CTX	AZM	FEP	CAZ	CFX	GEN	IMI	MEM	CTX-C	CAZ-C	SXT	Sul	NAL	GEM	LEV	CIP	STR	AMI	KAN	TET	MIN	DOX	CHL	AZI	CT	PB
20 representative strains without pBD146																														
VF096	>16	32	32/16	64/32	2	≤1	≤0.25	≤0.5	8	≤1	0.5	0.125	2/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF097	>16	32	16/8	32/16	2	≤1	≤0.25	≤0.5	16	≤1	0.5	0.125	0.5/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF10-27	16	8	8/4	8/4	≤0.25	32	16	≤0.5	16	4	0.5	0.125	0.5/4	8/4	≤0.25/4.75	128	≤4	0.25	0.25	0.125	16	32	16	≤1	≤1	1	≤2	≤2	≤0.5	≤0.5
VF10-32	>16	4	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	4	2	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF002	>16	16	16/8	≤2/1	≤0.25	≤1	≤0.25	≤0.5	8	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	8	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF003	>16	16	16/8	16/8	≤0.25	≤1	≤0.25	≤0.5	4	≤1	1	0.125	0.25/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF126	>16	8	8/4	8/4	≤0.25	≤1	≤0.25	≤0.5	8	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF127	4	≤2	≤2/1	≤2/1	≤0.25	≤1	≤0.25	≤0.5	≤2	2	0.5	0.125	≤0.12/4	≤0.25/4	>8/152	128	>64	1	1	0.5	>32	8	≤8	4	≤1	≤0.5	8	≤2	≤0.5	1
VF106	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	8	≤1	1	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.06	≤0.125	0.125	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
VF108	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.03	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
VF051	>16	32	32/16	64/32	4	≤1	≤0.25	≤0.5	8	≤1	1	0.125	2/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
VF052	16	16	8/4	16/8	≤0.25	2	≤0.25	≤0.5	16	≤1	0.5	0.125	≤0.12/4	0.5/4	≤0.25/4.75	≤32	≤4	0.06	≤0.125	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
VF121	>16	16	16/8	16/8	≤0.25	≤1	≤0.25	≤0.5	4	≤1	0.5	0.125	0.25/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	8	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
VF122	>16	8	8/4	8/4	≤0.25	≤1	≤0.25	≤0.5	4	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	64	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
YSVF02	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	4	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.25	0.25	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
YSVF03	16	4	4/2	4/2	≤0.25	≤1	≤0.25	1	4	≤1	1	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF081	>16	8	8/4	8/4	≤0.25	≤1	≤0.25	≤0.5	4	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	128	≤4	0.06	≤0.125	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF082	8	≤2	≤2/1	4/2	≤0.25	≤1	≤0.25	≤0.5	≤2	≤1	0.5	0.125	≤0.12/4	≤0.25/4	>8/152	256	>64	0.25	1	0.5	8	≤4	≤8	4	≤1	≤0.5	8	≤2	≤0.5	≤0.5
VF061	8	≤2	4/2	4/2	≤0.25	≤1	≤0.25	≤0.5	4	≤1	0.5	0.125	≤0.12/4	≤0.25/4	≤0.25/4.75	≤32	≤4	0.03	≤0.125	0.06	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	≤0.5
VF062	>16	8	8/4	16/8	≤0.25	≤1	≤0.25	≤0.5	4	≤1	1	0.125	0.25/4	≤0.25/4	≤0.25/4.75	≤32	≤4	≤0.015	≤0.125	≤0.03	≤4	≤4	≤8	≤1	≤1	≤0.5	≤2	≤2	≤0.5	1
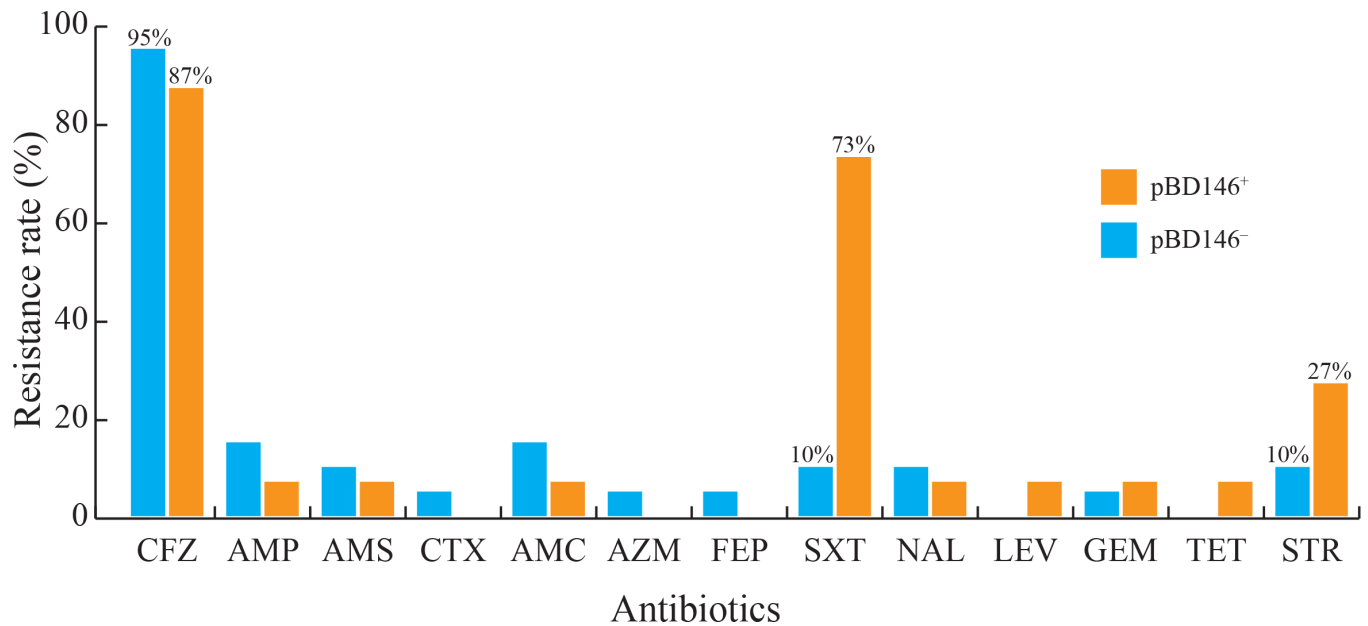
Fig. 5. Resistance rates of pBD146-carrying and pBD146-absent strains to 13 antibiotics. The average percentage resistance rates of 15 pBD146-carrying strains (orange) and 20 pBD146-absent strains (blue) to each antibiotic are shown. See Table 3 for antibiotic abbreviations.

et al. [12], who reported that the presence of pBD146 conferred resistance to trimethoprim, a component of SXT. However, four strains (VF035, VF044, VF049, VF10-51) that carried the *dfr6* gene, including two with a complete pBD146 (VF035 and VF10-51), showed susceptibility to SXT (Table 3). This may be because the *dfr6* genes in these *V. fluvialis* strains were lowly or not expressed for some unknown reason, and therefore the acquisition of *dfr6* or even the full pBD146 did not provide resistance to SXT in *V. fluvialis*. Notably, the pBD146-carrying strains had higher resistance rates (27%) to streptomycin than the pBD146-absent strains (10%); however, this difference was not statistically significant (Fisher's exact test, $P=0.37$). Furthermore, we tried to identify the possible streptomycin resistance-related ARGs or homologues, but none of the identified ARGs (Table 2) were related to resistance to streptomycin according to available annotations. Additionally, we extracted the *strA* and *strB* sequences from the sequence of the antibiotic resistance gene cluster of SXT[MO10] (accession no. AY034138) [74], and compared them to the whole genome dataset through BLASTN. The result showed that only two strains carried *strAB* genes: VF127 and YSVF09. Although strain VF127 indeed showed resistance to streptomycin (Table 3), the extremely low detection rate of *strAB* genes in the current dataset hardly explained the streptomycin resistance of all related strains. Interestingly, both the pBD146-carrying and pBD146-absent strains showed high resistance rates (87 and 95%, respectively) to CFZ, a *β*-lactam antibiotic, but the mechanism of resistance is still unknown.

Identification and structure of MGEs

ICEs are important MGEs that spread drug resistance and virulence among bacteria by horizontal gene transfer [75]. Although many studies have tried to identify the possible ICEs in *V. fluvialis* [9, 23–25], to date, only one *V. fluvialis* ICE named ICE*Vfl*Ind1, which was first recognized in strain H-08942, has been reported. The sequence of ICE*Vfl*Ind1 has been submitted to the ICEberg [76] (ICEberg ID 36) and GenBank (GQ463144) databases. By aligning the ICE*Vfl*Ind1 sequence to our 182 genome sequences, we identified ICE*Vfl*Ind1 in 23 genomes; four (17.4%) were from strains in VflPop1, 18 (78.3%) were from strains in VflPop2.1, and one (4.3%) was from a strain in VflPop2.2 (Fig. 2). Comparison of these ICE*Vfl*Ind1 sequences identified eight unique genomic regions (GR1–8, Fig. 6) that were absent in the 23 newly identified ICE*Vfl*Ind1, except for GR5, which was in strains SAMN12648296, VF083 and VF115, and GR6, which was in strains SAMN12648296, VF13–32 and YSVF27. The eight GRs contained a total of 42 genes (ranging from one to 17 genes) (Data S4). The impact of the absence of these GRs on the drug resistance and virulence of *V. fluvialis* needs further analysis.

MGIs are small genomic islands associated with drug resistance and virulence in bacteria [77]. Few studies have focused on MGIs in *Vibrio* species. The most well-studied *V. fluvialis* MGI is MGI*Vfl*Ind1 (GenBank: KC117176), which is 24 kb long [77]. In the current dataset, MGI*Vfl*Ind1 was identified in the genomes of seven strains in VflPop2 (four in VflPop2.1 and three in VflPop2.2; Fig. 2) with coverage of 42–73%. Comparison of these MGI*Vfl*Ind1 sequences (Fig. 7) showed that they all contained the key MGI integrase gene, although four in strains VF055, VF056, YSVF02, and VF007 contained short fragment deletions. Whether these deletions only reflect sequence diversity or affect the function of integrase needs to be confirmed. Furthermore, the coverage of MGI*Vfl*Ind1 in four strains (VF056, VF042, YSVF02, VF007) was lower than 50%, indicating the high diversity of MGIs in *V.*
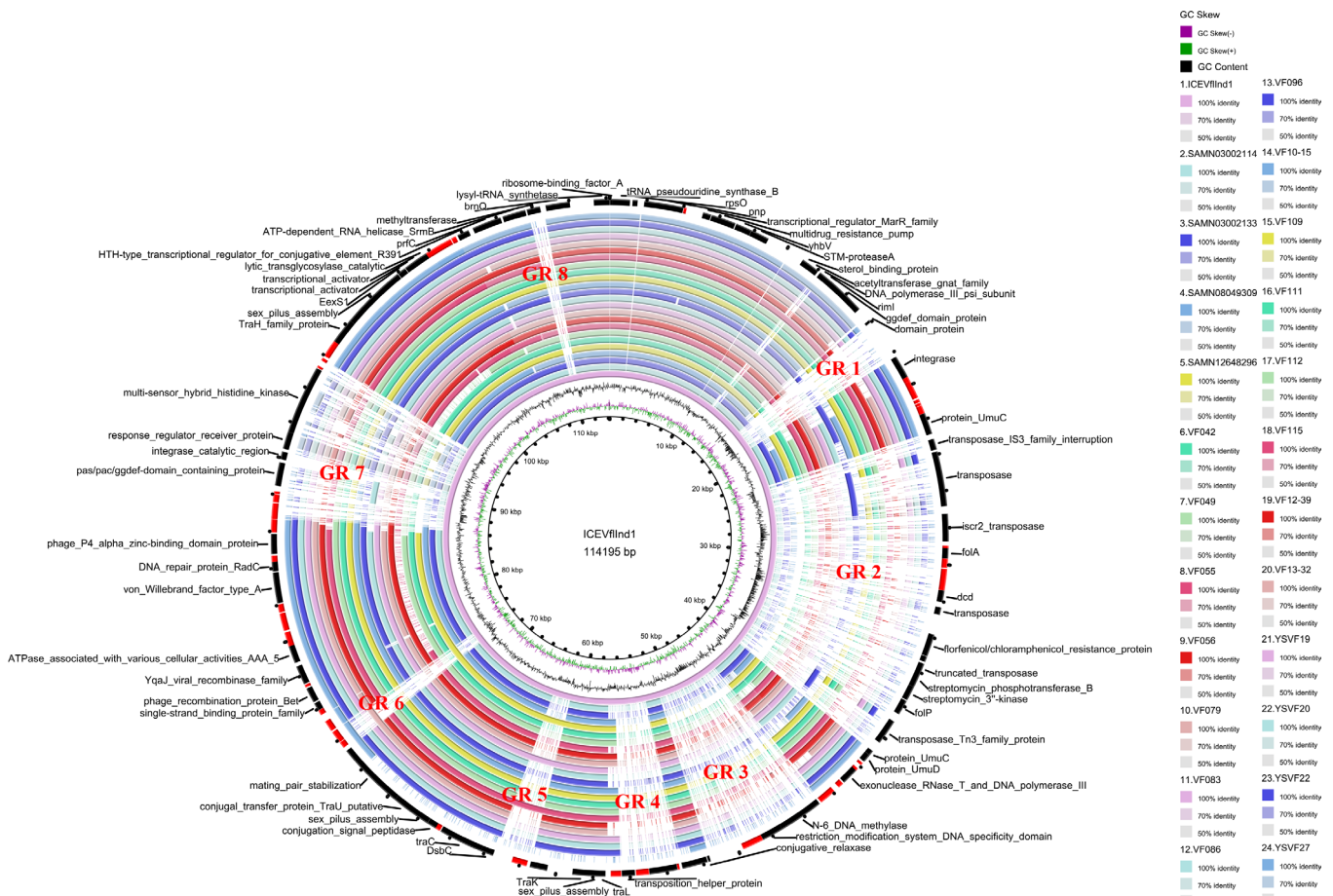
Fig. 6. Comparison of the genetic structure of ICE*Vfl*Ind1. The 23 newly identified ICE*Vfl*Ind1 sequences are shown. The outer ring indicates the genes based on the annotation of ICE*Vfl*Ind1 (GenBank: GQ463144); the genes marked in red encode hypothetical proteins. The eight unique genomic regions (GR) are marked GR1–8.

fluvialis. Notably, one of the hypervariable regions harboured the host specificity determinant (*hsd*) genes (*hsdR*, *hsdM*, *hsdS*) and the gene that codes for DEAD/DEAH box helicase. The *hsd* locus encodes the type I restriction–modification systems that are composed of multisubunit enzymes with modification methyltransferase, restriction endonuclease or translocase activities [78]. Phages, viruses and MGEs have evolved multiple strategies, including acquiring modifications from their hosts or synthesizing their own modification methyltransferase, to evade host restriction endonuclease, thus favouring their own growth or spread [78]. DEAD/DEAH box helicases are essential in numerous RNA metabolic processes, and some also act as sensors of cytosolic DNA/RNA, as adaptor proteins, or as regulators of signalling and gene expression [79]. Bacterial DEAD/DEAH box helicases involved in motility and biofilm formation [80] and stress tolerance [81] have been reported. The biological significance or MGI*Vfl*Ind1 phenotypes with/without the *hsd* locus and DEAD/DEAH box helicase to *V. fluvialis* deserve further study.

DISCUSSION

The population structure of *V. parahaemolyticus* was shown to be closely related to different oceans [42]. However, in the present study, the *V. fluvialis* strains from different sampling locations did not show regional aggregation on the phylogenetic tree and among different populations; instead, they showed a 'dispersed distribution' similar to that of *V. alginolyticus* [50]. According to the theory of population genetics, population differentiation only occurs when the migration rate among different populations is very low [82]. When migrants transmit to a new niche and coexist with the local population of the same species, they could possibly share a common gene pool, allowing the exchange of genetic elements between each other, and finally leading to the merging of the migrants with the local population. In the current study, the *V. fluvialis* strains were divided into three populations. However, except for VflPop2.2, which only contained Asian strains, VflPop1 and VflPop2.1 both contained strains from different continents, such as the Americas, Asia and Europe. This may be because there was inadequate exchange of genetic elements between the local strains and the migrants that still retained the genetic characteristics of the 'hometown' populations. Possible reasons for this include: (1) the transoceanic
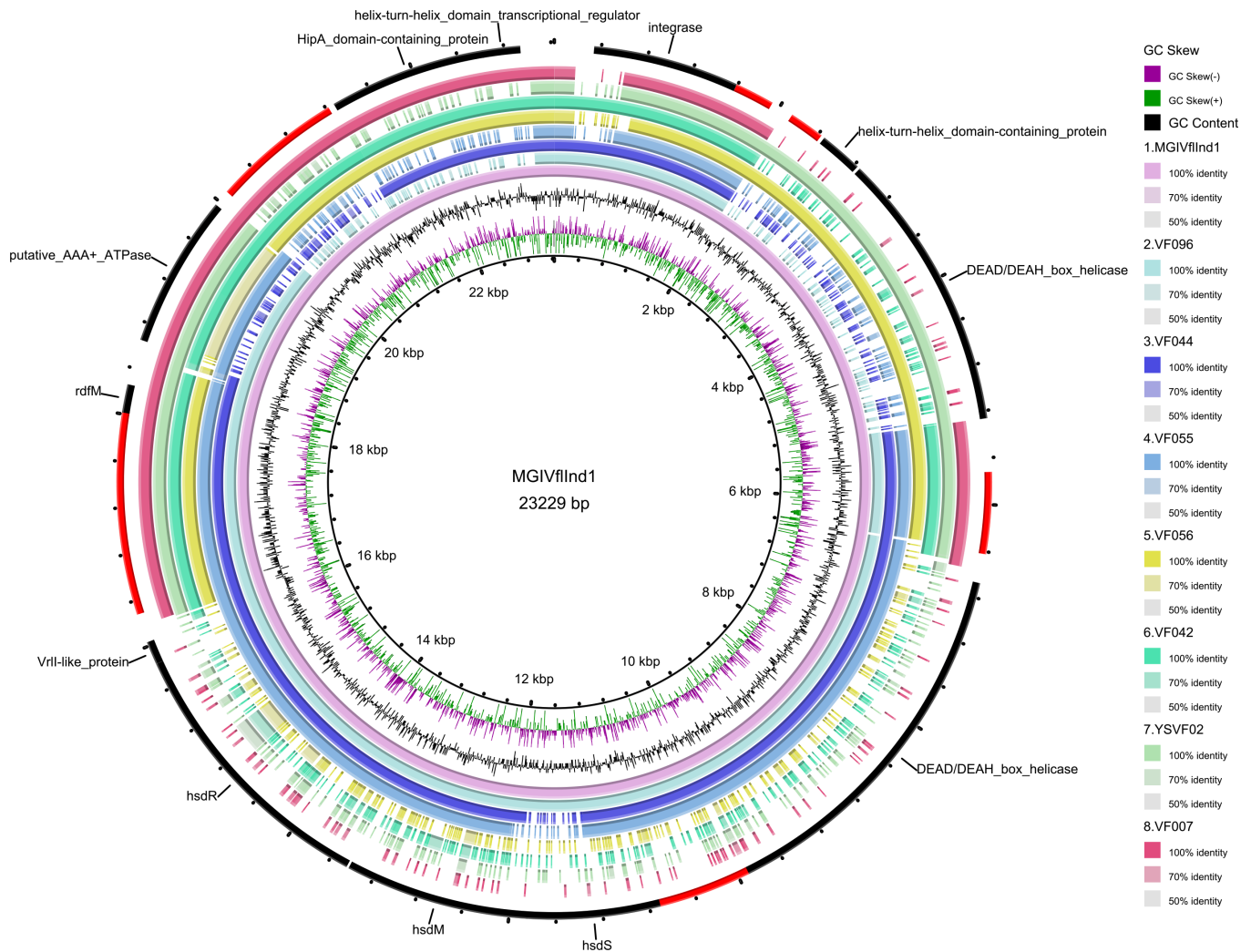
Fig. 7. Comparison of the genetic structure of MGI*Vfl*Ind1. The seven newly identified MGI*Vfl*Ind1 sequences are shown. The outer ring indicates the genes based on the annotation of MGI*Vfl*Ind1 (GenBank: KC117176); the genes marked in red encode hypothetical proteins.

spread occurred recently and the evolution time was too short; (2) the inherited genetic characteristics of *V. fluvialis* restricted the exchange of genetic elements between strains in different populations; and (3) *V. fluvialis* strains in different populations existed in different microhabitats in the same geographical area, leading to *de facto* isolation of ecological conditions. These conjectures need to be verified by ecological and laboratory investigations in the future.

The scattering of strains from different continents on the phylogenetic tree and their distribution in VflPop1 and VflPop2.1 indicate multiple historical transmission events. However, we did not identify a recent transoceanic transmission event, and only one cross-provincial transmission event was identified in the strains from China. This may be caused by sampling bias because the relatively small number of available *V. fluvialis* genomes in NCBI may not represent the global prevalence and epidemiological characteristics of this species in other continents. Potential transoceanic spread and the confirmation of local spread in China has further characterized the phylogeography and epidemiological characteristics of *V. fluvialis* worldwide and this new information will help to improve prevention strategies, making them more comprehensive and integrated.

Here, we investigated the distribution of pathogenicity-related factors in *V. fluvialis*, including VFs, ARGs and MGEs. Similar in other *Vibrio* species [50], flagellum VFs were predominant in *V. fluvialis*. Consistent with previous studies [19], VflT6SS2 showed a high prevalence in the strains in the current study. Additionally, we showed that genes encoding virulence-associated systems and protein factors, such as QS, T2SS, curli pili, PilA pili, MSHA, haemolysin, HapA, HupO, IlpA and ToxR/S, were prevalent in *V. fluvialis*. The broad distribution of these VFs indicated their important roles in the life cycle of *V. fluvialis*. MSHA and PilA pili are type IVa pili commonly found in *Vibrio* genomes and play roles in the colonization of *Vibrio* species in the environment and/or host tissue [83]. Interestingly, within the PilA pili and MSHA biogenesis gene clusters, *pilA* (AL536_RS30945) and *mshA* (AL536_RS41655) displayed

obvious coverage variations among the strains (Fig. S4), suggesting selective pressure may have caused changes in the sequences of the MSHA and PilA pili subunits. These findings are consistent to those in other *Vibrio* species, including *V. parahaemolyticus*, *V. vulnificus* and *V. cholerae* strains without the toxin co-regulated pilus, a type IVb pilus where the *mshA* and *pilA* sequences also have polymorphisms [83]. Most importantly, we found that the distribution of two potential clusters of Flp pili coding genes exhibited the most striking difference. Considering the involvement of Flp pili in microbial adherence, colonization, biofilm formation and thereafter pathogenesis [68], we inferred that strains that harbour the two Flp pili clusters may have better survival advantage and be more virulent in infected human hosts than those without or with only one Flp pili cluster. Identification of ARGs showed that the population-specific characteristics of *tet*(35), which was present in 150 (95%) of the strains in VflPop2 but only in one (4%) of the strains in VflPop1, requires further investigation to confirm its biological role. Additionally, a few VFs, most of the annotated ARGs, including *ugd, rtxB, rtxD, aph(3″)-Ib, qnrVC5, sul2, floR* and *blaCARB-2*, and the MGEs, including pBD146, ICE*Vfl*Ind1 and MGI*Vfl*Ind1, had low allele frequencies (<10%) and were scattered among the different populations, suggesting such elements may have a minor role in the life cycle of *V. fluvialis*.

Although the RTX (repeat in toxin) family of toxins are produced by several pathogenic gram-negative bacteria, our VFDB-based searches revealed that only 1% of the studied *V. fluvialis* genomes contain *rtxB* and *rtxD* genes, which are generally believed to encode an RTX toxin-associated ATP-binding cassette transporter system. We reasoned that the *rtx* gene cluster is not widely disseminated in *V. fluvialis*. This is not surprising because the *rtx* gene cluster is absent from the *V. cholerae* classical O1 serogroup strain, but is frequent in the *V. cholerae* El Tor biotype strain. Vigil *et al.* reported that in uropathogenic *E. coli*, the RTX homologue gene *tosA* is present in 11% of faecal strains but 25 % of urinary isolates [84]. These findings suggest that the distribution of the *rtx* gene cluster varies in microbial species and groups. Overall, currently there is little definite information available on the virulence factors associated with *V. fluvialis* infections in humans and less on the mechanism of pathogenicity of this organism. We consider that the pathogenesis of *V. fluvialis* involves multiple virulence factors determining its colonization, nutrition uptake, haem utilization, proliferation, etc., in the human intestine and the underlying pathogenic mechanisms remain to be investigated. VFH haemolysin and the QS system are two relatively well-demonstrated VFs in *V. fluvialis* [17, 21, 63]. The T6SS has also been shown to contribute to the virulence of *V. fluvialis* by providing an enhanced competitive fitness in the marine environment [19]. VFH haemolysin seems to be of most importance among the VFs in *V. fluvialis*. According to the literature, pores formed by VFH on the erythrocyte membrane are even larger than those formed by other *Vibrio* haemolysins such as in *V. cholerae*, *V. parahaemolyticus* and *V. vulnificus* [3]. This phenomenon seems to correlate well with the bloody diarrhoea in *V. fluvialis* infection.

Similar to the Indian multiple drug resistance strain BD146, 11 of the 15 newly identified pBD146-carrying strains had high resistance rates to SXT, which should be related to the *dfr6* gene in pBD146. In 2018, Rajpara *et al.* [22] reported that strain BD146 had high extended-spectrum beta-lactamase (ESBL) and ampC beta-lactamase activities, leading to resistance to ceftazidime, cefotaxime, cefepime and cloxacillin, but we found that the pBD146-carrying strains in our study were sensitive to all these antibiotics (except cloxacillin, which was not included in our antimicrobial susceptibility testing) (Table 3). This phenotype is consistent with the finding that the low-copy plasmid with a class I integron and the *blaOXA10* gene, which are present in strain BD146, were missing in the 182 genomes in our study. As shown in Fig. 5, the pBD146-carrying and pBD146-absent strains both had a high resistance rate to CFZ, implying that CFZ is not a suitable antibiotic for the clinical treatment of *V. fluvialis* infections. We inferred that CFZ resistance may be conferred by chromosomal genes, which could be verified in future studies by larger-scale antibiotic susceptibility testing and GWAS to locate the specific genetic determinants. Furthermore, except for strain YSVF18, the other 14 pBD146-carrying strains were still sensitive to nalidixic acid, ciprofloxacin, levofloxacin and gemifloxacin (Table 3), even though they contained *qnrVC5*, which was considered to confer resistance to quinolone antibiotics. These findings indicate that the function of *qnrVC5* may be affected by a variety of factors, and that the presence of ARGs in the genome does not necessarily mediate resistance to the corresponding antibiotics.

In conclusion, on the basis of the whole genome sequencing and population genomics analysis results, we identified the genetic diversity and pathogenicity-related characteristics of *V. fluvialis*. Our results show that the *V. fluvialis* strains formed three populations, among which population fusion was not observed in VflPop1 and VflPop2.1. Furthermore, local spread in China of *V. fluvialis* was identified based on the phylogenetic tree and analysis of CGs. Identification of pathogenicity-related factors provided information about the distribution, prevalence and genetic structure of VFs, ARGs and MGEs in *V. fluvialis*, which broadened our understanding of the virulence and multiple drug resistance potential of this species and provided further clues for clinical therapy and vaccine development. Our results also provide fundamental data for further studies into the genomics, characteristics and pathogenic mechanisms of *V. fluvialis*, which will help to improve the surveillance and control of this pathogen. However, because the sample size in this study was small and there was an uneven distribution of strains from the different continents, the genetic characteristics of *V. fluvialis* in the Americas, Europe, Africa and other countries in Asia need to be further explored.

Author contributions

Conflicts of interest

Ethical statement

This study did not involve human materials, human data or animal experiments. No ethics or consent approval was required for the research in this study.

References

1. Huq MI, Alam AK, Brenner DJ, Morris GK. Isolation of Vibrio-like group, EF-6, from patients with diarrhea. *J Clin Microbiol* 1980;11:621–624.

2. Seidler RJ, Allen DA, Colwell RR, Joseph SW, Daily OP. Biochemical characteristics and virulence of environmental group F bacteria isolated in the United States. *Appl Environ Microbiol* 1980;40:715–720.

3. Igbinosa EO, Okoh AI. *Vibrio fluvialis*: an unusual enteric pathogen of increasing public health concern. *Int J Environ Res Public Health* 2010;7:3628–3643.

4. Ramamurthy T, Chowdhury G, Pazhani GP, Shinoda S. *Vibrio fluvialis*: an emerging human pathogen. *Front Microbiol* 2014;5:91.

5. Bhattacharjee S, Bhattacharjee S, Bal B, Pal R, Niyogi SK, *et al*. Is *Vibrio fluvialis* emerging as a pathogen with epidemic potential in coastal region of eastern India following cyclone Aila? *J Health Popul Nutr* 2010;28:311–317.

6. Chowdhury G, Pazhani GP, Dutta D, Guin S, Dutta S, *et al*. *Vibrio fluvialis* in patients with diarrhea, Kolkata, India. *Emerg Infect Dis* 2012;18:1868–1871.

7. Liang P, Cui X, Du X, Kan B, Liang W. The virulence phenotypes and molecular epidemiological characteristics of *Vibrio fluvialis* in China. *Gut Pathog* 2013;5:6.

8. Ahmed AM, Nakagawa T, Arakawa E, Ramamurthy T, Shinoda S, *et al*. New aminoglycoside acetyltransferase gene, *aac(3)-Id*, in a class 1 integron from a multiresistant strain of *Vibrio fluvialis* isolated from an infant aged 6 months. *J Antimicrob Chemother* 2004;53:947–951.

9. Ahmed AM, Shinoda S, Shimamoto T. A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of Vibrio fluvialis. *FEMS Microbiol Lett* 2005;242:241–247.

10. Chowdhury G, Pazhani GP, Nair GB, Ghosh A, Ramamurthy T. Transferable plasmid-mediated quinolone resistance in association with extended-spectrum β-lactamases and fluoroquinolone-acetylating aminoglycoside-6′-N-acetyltransferase in clinical isolates of *Vibrio fluvialis*. *Int J Antimicrob Agents* 2011;38:169–173.

11. Srinivasan VB, Virk RK, Kaundal A, Chakraborty R, Datta B, *et al*. Mechanism of drug resistance in clonally related clinical isolates of *Vibrio fluvialis* isolated in Kolkata, India. *Antimicrob Agents Chemother* 2006;50:2428–2432.

12. Rajpara N, Patel A, Tiwari N, Bahuguna J, Antony A, *et al*. Mechanism of drug resistance in a clinical isolate of *Vibrio fluvialis*: involvement of multiple plasmids and integrons. *Int J Antimicrob Agents* 2009;34:220–225.

13. Singh R, Rajpara N, Tak J, Patel A, Mohanty P, *et al*. Clinical isolates of *Vibrio fluvialis* from Kolkata, India, obtained during 2006: plasmids, the *qnr* gene and a mutation in gyrase A as mechanisms of multidrug resistance. *J Med Microbiol* 2012;61:369–374.

14. Mohanty P, Patel A, Kushwaha Bhardwaj A. Role of H- and D-MATE-type transporters from multidrug resistant clinical isolates of *Vibrio fluvialis* in conferring fluoroquinolone resistance. *PLoS One* 2012;7:e35752.

15. Chowdhury G, Ramamurthy T, Ghosh A, Dutta S, Takahashi E, *et al*. Emergence of azithromycin resistance mediated by phosphotransferase-encoding *mph*(A) in diarrheagenic *Vibrio fluvialis*. *mSphere* 2019;4:e00215-19.

16. Chowdhury G, Pazhani GP, Sarkar A, Rajendran K, Mukhopadhyay AK, *et al*. Carbapenem resistance in clonally distinct clinical strains of *Vibrio fluvialis* isolated from diarrheal samples. *Emerg Infect Dis* 2016;22:1754–1761.

17. Wang Y, Wang H, Liang W, Hay AJ, Zhong Z, *et al*. Quorum sensing regulatory cascades control *Vibrio fluvialis* pathogenesis. *J Bacteriol* 2013;195:3583–3589.

18. Lu X, Liang W, Wang Y, Xu J, Zhu J, *et al*. Identification of genetic bases of *Vibrio fluvialis* species-specific biochemical pathways and potential virulence factors by comparative genomic analysis. *Appl Environ Microbiol* 2014;80:2029–2037.

19. Huang Y, Du P, Zhao M, Liu W, Du Y, *et al*. Functional Characterization and Conditional Regulation of the Type VI Secretion System in *Vibrio fluvialis. Front Microbiol 2017*;8:528. 10.3389/fmicb.2017.00528

20. Pan J, Zhao M, Huang Y, Li J, Liu X, *et al*. Integration Host Factor Modulates the Expression and Function of T6SS2 in *Vibrio fluvialis. Front Microbiol* 2018;9:962. 10.3389/fmicb.2018.00962

21. Liu X, Pan J, Gao H, Han Y, Zhang A, *et al*. CqsA/LuxS-HapR Quorum sensing circuit modulates type VI secretion system VflT6SS2 in *Vibrio fluvialis. Emerg Microbes Infect* 2021;10:589–601. 10.1080/22221751.2021.1902244

22. Rajpara N, Nair M, Bhardwaj AK. A Highly Promiscuous Integron, Plasmids, Extended Spectrum Beta Lactamases and Efflux Pumps as Factors Governing Multidrug Resistance in a Highly Drug Resistant *Vibrio fluvialis* Isolate BD146 from Kolkata, India. *Indian J Microbiol* 2018;58:60–67.

23. Daccord A, Ceccarelli D, Burrus V. Integrating conjugative elements of the SXT/R391 family trigger the excision and drive the mobilization of a new class of Vibrio genomic islands. *Mol Microbiol* 2010;78:576–588.

24. Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, *et al*. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet* 2009;5:12.

25. Zheng B, Jiang X, Cheng H, Guo L, Zhang J, *et al*. Genome characterization of two bile-isolated *Vibrio fluvialis* strains: an insight into pathogenicity and bile salt adaption. *Sci Rep* 2017;7:11827.

26. Li C, Li C, Li L, Yang X, Chen S, *et al*. Comparative genomic and secretomic analysis provide insights into unique agar degradation function of marine bacterium *Vibrio fluvialis* A8 through horizontal gene transfer. *Front Microbiol* 2020;11:1934.

27. Luo R, Liu B, Xie Y, Li Z, Huang W, *et al*. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 2012;1:18.

28. Cui Y, Yang X, Didelot X, Guo C, Li D, *et al*. Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. *Mol Biol Evol* 2015;32:1396–1410.

29. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.

30. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.

31. Yang C, Zhang X, Fan H, Li Y, Hu Q, *et al*. Genetic diversity, virulence factors and farm-to-table spread pattern of Vibrio parahaemolyticus food-associated isolates. *Food Microbiol* 2019;84:103270.

32. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* 2003;Chapter 10:Unit .

33. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, *et al*. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966–1967.

34. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999;27:573–580.

35. 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.

36. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242-5.

37. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* 2018;34:4310–4312.

38. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, *et al*. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet* 2018;14:11.

39. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harb Protoc* 2009;2009:pdb.ip71.

40. Yang C, Cui Y, Didelot X, Yang R, Falush D. Why panmictic bacteriaare rare. *BioRxiv* 2018:385336.

41. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet* 2012;8:e1002453.

42. Yang C, Pei X, Wu Y, Yan L, Yan Y, *et al*. Recent mixing of Vibrio para-haemolyticus populations. *ISME J* 2019;13:2578–2588.

43. CLSI. *Performance Standards for Antimicrobial Susceptibility Testing*. 30th ed. CLSI supplement M100. Wayne, PA: Clinical and Laboratory Standards Institute; 2020.

44. CLSI. *Methods for Antimicrobial Dilution and Disk Susceptibility Testing of Infrequently Isolated or Fastidious Bacteria*. 3rd ed. CLSI guideline M45. Wayne, PA: Clinical and Laboratory Standards Institute; 2015.

45. CLSI. *Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Second Informational Supplement. CLSI document M100-S22*. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

46. Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 2019;47:D687–D692.

47. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, *et al*. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–3500.

48. Wang H, Yang C, Sun Z, Zheng W, Zhang W, *et al*. Genomic epidemiology of Vibrio cholerae reveals the regional and global spread of two epidemic non-toxigenic lineages. *PLoS Negl Trop Dis* 2020;14:e0008046.

49. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.

50. Zheng HY, Yan L, Yang C, Wu YR, Qin JL, *et al*. Population genomics study of *Vibrio alginolyticus*. *Yi Chuan* 2021;43:350–361.

51. Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, *et al*. Chromosome painting *in silico* in a bacterial species reveals fine population structure. *Mol Biol Evol* 2013;30:1454–1464.

52. Syed KA, Beyhan S, Correa N, Queen J, Liu J, *et al*. The *Vibrio cholerae* flagellar regulatory hierarchy controls expression of virulence factors. *J Bacteriol* 2009;191:6555–6570.

53. Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, *et al*. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol Microbiol* 2007;64:1466–1485.

54. Yu Y, Yang H, Li J, Zhang P, Wu B, *et al*. Putative type VI secretion systems of *Vibrio parahaemolyticus* contribute to adhesion to cultured cell monolayers. *Arch Microbiol* 2012;194:827–835.

55. Goo SY, Han YS, Kim WH, Lee K-H, Park S-J. *Vibrio vulnificus* IlpA-induced cytokine production is mediated by Toll-like receptor 2. *J Biol Chem* 2007;282:27647–27658.

56. Lee K-J, Lee NY, Han Y-S, Kim J, Lee K-H, *et al*. Functional characterization of the IlpA protein of *Vibrio vulnificus* as an adhesin and its role in bacterial pathogenesis. *Infect Immun* 2010;78:2408–2417.

57. Ahn SH, Han JH, Lee JH, Park KJ, Kong IS. Identification of an iron-regulated hemin-binding outer membrane protein, HupO, in *Vibrio fluvialis*: effects on hemolytic activity and the oxidative stress response. *Infect Immun* 2005;73:722–729.

58. Yanez ME, Korotkov KV, Abendroth J, Hol WGJ. Structure of the minor pseudopilin EpsH from the Type 2 secretion system of *Vibrio cholerae*. *J Mol Biol* 2008;377:91–103.

59. Soo VWC, Wood TK. Antitoxin MqsA represses curli formation through the master biofilm regulator CsgD. *Sci Rep* 2013;3:3186.

60. Zhang XH, Austin B. Haemolysins in *Vibrio* species. *J Appl Microbiol* 2005;98:1011–1019.

61. Han J-H, Lee J-H, Choi Y-H, Park J-H, Choi T-J, *et al*. Purification, characterization and molecular cloning of *Vibrio fluvialis* hemolysin. *Biochim Biophys Acta* 2002;1599:106–114.

62. Kothary MH, Lowman H, McCardell BA, Tall BD. Purification and characterization of enterotoxigenic El Tor-like hemolysin produced by *Vibrio fluvialis*. *Infect Immun* 2003;71:3213–3220.

63. Song L, Huang Y, Zhao M, Wang Z, Wang S, *et al*. A critical role for hemolysin in *Vibrio fluvialis*-induced IL-1β secretion mediated by the NLRP3 inflammasome in macrophages. *Front Microbiol* 2015;6:510.

64. Chen YC, Chang MC, Chuang YC, Jeang CL. Characterization and virulence of hemolysin III from *Vibrio vulnificus*. *Curr Microbiol* 2004;49:175–179.

65. Helaine S, Dyer DH, Nassif X, Pelicic V, Forest KT. 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. *Proc Natl Acad Sci U S A* 2007;104:15888–15893.

66. Marsh JW, Taylor RK. Genetic and transcriptional analyses of the *Vibrio cholerae* mannose-sensitive hemagglutinin type 4 pilus gene locus. *J Bacteriol* 1999;181:1110–1117.

67. Chiavelli DA, Marsh JW, Taylor RK. The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl Environ Microbiol* 2001;67:3220–3225.

68. Duong-Nu T-M, Jeong K, Hong SH, Puth S, Kim SY, *et al*. A stealth adhesion factor contributes to *Vibrio vulnificus* pathogenicity: Flp pili play roles in host invasion, survival in the blood stream and resistance to complement activation. *PLoS Pathog* 2019;15:e1007767.

69. Alteri CJ, Mobley HLT. The versatile Type VI secretion system. *Microbiol Spectr* 2016;4.

70. Zhao Z, Liu J, Deng Y, Huang W, Ren C, *et al*. The *Vibrio alginolyticus* T3SS effectors, Val1686 and Val1680, induce cell rounding, apoptosis and lysis of fish epithelial cells. *Virulence* 2018;9:318–330.

71. Zhou X, Konkel ME, Call DR. Type III secretion system 1 of *Vibrio parahaemolyticus* induces oncosis in both epithelial and monocytic cell lines. *Microbiology (Reading)* 2009;155:837–851.

72. Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, *et al*. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of V cholerae. *Lancet* 2003;361:743–749.

73. Dalton KM, Crosson S. A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. *Biochemistry* 2010;49:2205–2215.

74. Hochhut B, Lotfi Y, Mazel D, Faruque SM, Woodgate R, *et al*. Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob Agents Chemother* 2001;45:2991–3000.

75. Burrus V, Pavlovic G, Decaris B, Guédon G. Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 2002;46:601–610.

76. **Bi D**, **Xu Z**, **Harrison EM**, **Tai C**, **Wei Y**, *et al*. ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res* 2012;40:D621-6.

77. **Daccord A**, **Ceccarelli D**, **Rodrigue S**, **Burrus V**. Comparative analysis of mobilizable genomic islands. *J Bacteriol* 2013;195:606–614.

78. **Loenen WAM**, **Dryden DTF**, **Raleigh EA**, **Wilson GG**. Type I restriction enzymes and their relatives. *Nucleic Acids Res* 2014;42:20–44.

79. **Perčulija V**, **Ouyang S**. Diverse roles of dead/deah-box helicases in innate immunity and diseases. *Helicases from All Domains of Life: Elsevier* 2019:141–171.

80. **Granato LM**, **Picchi SC**, **AndradeMO**, **Takita MA**, **de Souza AA**, *et al*. The ATP-dependent RNA helicase HrpB plays an important role in motility and biofilm formation in *Xanthomonas citri* subsp. *citri*. *BMC Microbiol* 2016;16:55.

81. **Illakkiam D**, **Shankar M**, **Ponraj P**, **Rajendhran J**, **Gunasekaran P**. Genome sequencing of a mung bean plant growth promoting strain of *P. aeruginosa* with biocontrol ability. *Int J Genomics* 2014;2014:123058.

82. **Wright S**. Evolution in mendelian populations. *Genetics* 1931;16:97–159.

83. **Aagesen AM**, **Häse CC**. Sequence analyses of type IV pili from *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus*. *Microb Ecol* 2012;64:509–524.

84. **Vigil PD**, **Stapleton AE**, **Johnson JR**, **Hooton TM**, **Hodges AP**, *et al*. Presence of putative repeat-in-toxin gene *tosA* in *Escherichia coli* predicts successful colonization of the urinary tract. *mBio* 2011;2:e00066-11.