

RESEARCH

Open Access



Epidemiological and Genomic analysis of *Vibrio parahaemolyticus* isolated from imported travelers at the port of Shanghai, China (2017-2019)

Danlei Liu^{1,2†}, Lei Zhou^{2†}, Zilei Zhang^{3†}, Ying Zhang⁴, Zhiyi Wang¹, Shenwei Li¹, Yongqiang Zhu², Huajun Zheng^{2*}, Zilong Zhang^{1*} and Zhengan Tian^{1*}

Abstract

Background *Vibrio parahaemolyticus* is the predominant etiological agent of seafood-associated foodborne illnesses on a global scale. It is essential to elucidate the mechanisms by which this pathogen disseminates. Given the existing research predominantly concentrates on localized outbreaks, there is a pressing necessity for a comprehensive investigation to capture strains of *V. parahaemolyticus* cross borders.

Results This study examined the frequency and genetic attributes of imported *V. parahaemolyticus* strains among travelers entering Shanghai Port, China, between 2017 and 2019. Through the collection of 21 strains from diverse countries and regions, Southeast Asia was pinpointed as a significant source for the emergence of *V. parahaemolyticus*. Phylogenetic analysis revealed clear delineation between strains originating from human and environmental sources, emphasizing that underlying genome data of foodborne pathogens is essential for environmental monitoring, food safety and early diagnosis of diseases. Furthermore, our study identified the presence of virulence genes (*tdh* and *tlh*) and approximately 120 antibiotic resistance-related genes in the majority of isolates, highlighting their crucial involvement in the pathogenesis of *V. parahaemolyticus*.

Conclusions This research enhanced our comprehension of the worldwide transmission of *V. parahaemolyticus* and its antimicrobial resistance patterns. The findings have important implications for public health interventions and antimicrobial stewardship strategies, underscoring the necessity for epidemiological surveillance of pathogen at international travel hubs.

Keywords Port surveillance, *Vibrio parahaemolyticus*, Whole genome sequencing, Virulence factor, Antibiotic resistance

[†]Danlei Liu, Lei Zhou, and Zilei Zhang contributed equally to this work.

*Correspondence:

Huajun Zheng
zhenghj@chgc.sh.cn
Zilong Zhang
heilong82@163.com
Zhengan Tian
tianzhenganciq@163.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Vibrio parahaemolyticus, a Gram-negative bacterium with halophilic characteristics, inhabits in coastal estuarine ecosystems and stands as a significant worldwide pathogen accountable for foodborne infections [1, 2]. The escalating occurrence of outbreaks of *V. parahaemolyticus* are closely tied to climate change, posing a significant public health risk [3]. Investigating the potential transmission pathways of *V. parahaemolyticus*, from its environmental reservoir to human exposure through retail channels, emerges as an imperative stride in mitigating the global prevalence of acute gastroenteritis.

Nations across Asia have reported significant impacts of *Vibrio* infections on human health, thereby detrimentally affecting the seafood industry [4]. To facilitate research in this field, the Foodborne *Vibrio parahaemolyticus* Genome Database (FVPGD) has been established, encompassing a compilation of 643 *V. parahaemolyticus* strains obtained from 39 cities in China [5]. Moreover, from 2013 to 2017, there was a comprehensive examination in southeastern China where 1,220 *V. parahaemolyticus* strains were isolated and analyzed [6]. The escalating incidence of *V. parahaemolyticus* infections in China highlights the exigent requirement to assess the prevalence and genetic heterogeneity of this virulent bacterium [5].

The global spread of *V. parahaemolyticus* underscores the imperative to enhance our comprehension of its heightened propagation. As most of the previous studies were conducted independently and focused on local outbreaks, there is a need to perform comprehensive investigations to capture cross-border strains of *V. parahaemolyticus* at the whole-genome level. As the most densely populated urban center in China, Shanghai is witnessing an escalating connectivity within the global city network [7]. Notably, the city boasts two commercial airports, Shanghai Pudong International Airport and Shanghai Hongqiao International Airport, rendering it the world's fifth-busiest gateway. This confluence of factors amplifies Shanghai's role as an ideal sentinel for monitoring the global transmission dynamics of *V. parahaemolyticus*.

To trace the origin of *V. parahaemolyticus*, serological typing, pulsed-field gel electrophoresis, multilocus sequence typing (MLST), and whole genome sequencing (WGS) have been widely employed [6]. Nowadays, 16 O serotypes and 71 K serotypes can be identified using commercial antisera [8]. Nonetheless, this method is naturally limited by challenges including its costly, intricate protocols, potential for cross-immunoreactivity, and need for subjective interpretation [9]. WGS of *V. parahaemolyticus* isolates and comparative genomics analyses have revealed various mutations, chromosomal

rearrangements and gene gain or loss events caused by duplication or horizontal gene transfer [10]. Bioinformatics methods such as the VPsero [11] and Kaptive database [12] have been developed to identify gene clusters associated with O-loci (lipopolysaccharide, LPS) and K-loci (capsular polysaccharide, CPS) for serotype determination of *V. parahaemolyticus*. Clinical isolates of *V. parahaemolyticus* also contain numerous virulence factors, including the thermostable direct hemolysin (*tdh*) and the *tdh*-related hemolysin (*trh*) [1, 13].

Our research conducted a pangenome analysis to genetically characterize isolates of 21 *V. parahaemolyticus* strains that were collected from imported travelers in Shanghai port, China during 2017-2019. Furthermore, this study enhanced our comprehension of the worldwide spread of *V. parahaemolyticus* and the genetic factors that affect its resistance to antimicrobial substances. Results had implications for public health interventions and strategies to preserve the effectiveness of antimicrobial drugs, highlighting the importance of ongoing monitoring and control of this pathogen.

Methods

Sample collection

Shanghai Customs adopted the principle of voluntary reporting and sample collection paradigm to facilitate its disease surveillance efforts. In instances where diarrheal-like symptoms were suspected among inbound travelers, a proactive approach was adopted whereby passengers were requested to self-disclose any abnormal health manifestations. Following the attainment of passengers' informed consent, the process entailed the collection of anal swab samples by medical personnel stationed at the airport's customs facility [14, 15]. Subsequently, these samples were meticulously stored under refrigeration conditions and expeditiously conveyed to the Shanghai International Travel Healthcare Center for comprehensive testing and characterization.

Pathogen detection and isolation of *V. parahaemolyticus*

DNA/RNA of swab samples was extracted using a magnetic beads nucleic acid isolation kit (BioPerfectus technologies, China, Catalog No. SDK60104) following the manufacturer's instructions. The presence of foodborne pathogens including *Salmonella* spp., *Shigella* spp., *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Escherichia coli* O157:H7, Rotavirus A and GI/GII Norovirus were detected using commercial qPCR kit (Bioperfectus Technologies, Shanghai, China) according to the voluntary border screening surveillance strategy of entry points. The qPCR was performed using the QuantStudio 5 Real Time PCR system (Thermo Fisher, USA) and data were analyzed using the QuantStudio Design and Analysis

software. In adherence to the established criterion, samples recording a Ct value below 35 were categorized as positive outcomes for *V. parahaemolyticus* detection.

V. parahaemolyticus strains were isolated according to the GB 4789.7-2013 food microbiological examination (National Food Safety Standards of China) [16]. Samples were streaked onto thiosulfate citrate bile-salt sucrose (TCBS) agar plates and incubated at 37 °C for 18-24 h. Presumptive colonies (green or blue green colonies, 2-3 mm in diameter) were then transferred to Chromogenic Vibrio Medium and incubated at 37 °C for 24 h. Genomic DNA was extracted using Bacteria Genomic DNA Extraction kit with magnetic beads (Bioperfectus technologies, Catalog No. SDK60108). Species identification was further confirmed through 16s rRNA gene sequencing. All positive *V. parahaemolyticus* colonies were preserved in Brucella broth with 17.5% glycerol and stored in Biological Sample Bank of Shanghai International Travel Healthcare Center.

DNA extraction and shotgun genome sequencing

All isolates were cultivated in tryptic soy broth medium (HuanKai Microbial, China) overnight at 37 °C. Genomic DNA was extracted using Bacteria Genomic DNA Extraction kit with magnetic beads (Bioperfectus technologies, Catalog No. SDK60108). Subsequently, DNA integrity and quantity were assessed on a Nanodrop UV-vis spectrophotometer (Thermo Fisher, USA) and 1% agarose gel electrophoresis visualized with ethidium bromide in a ChemiDoc XRS photodocumenter (BioRad, USA). Paired-end libraries were constructed using the TrueSeq DNA Sample Prep Kit (Illumina, USA) according to the manufacturer's instructions. The libraries were subjected to purification using AMPure XP Beads (Beckman Coulter, USA). For the sequencing phase, the Illumina X-10 platform was employed, generating 2 × 150 bp paired-end reads, consequently accumulating a total of 2 GB data per individual sample.

De Novo assembly and annotation

Raw reads were trimmed using Trimmomatic v0.36 [17]. The trimmed reads were subjected to assembly using Velvet v1.2.03. Predictions of putative open reading frames (ORFs) within each strain were accomplished using Glimmer3 [18]. Subsequently, the functionality of these ORFs was inferred using BLASTP against the NR protein database of the National Center for Biotechnology Information (NCBI) [19]. To further enhance the comprehension of the gene functions, a multifaceted annotation process was undertaken including pathway annotation and COG functional annotation, all of which contributed to unraveling the intricate functional attributes embedded within the genes of interest. To identify the conservation

of *V. parahaemolyticus* strains, average nucleotide identity (ANI) through alignment-free approximate sequence mapping was calculated using Fast Average Nucleotide Identity (FastANI) [20]. Reads were aligned against the *V. parahaemolyticus* reference genome, RIMD 2210633 (assembly GCF_000196095.1).

The prokaryotic Pan-Genome Analysis Pipeline (PGAP) was implemented to perform a pan-genome analysis of 21 *V. parahaemolyticus* genomes with a threshold of 95% identity and coverage [21]. Antimicrobial resistance (AMR) genes were identified using the Comprehensive Antibiotic Resistance Database (CARD), and virulence genes were identified using the virulence finder database (VFDB) [22], with a stringent threshold of 1e-50.

Serotyping and phylogenetic analysis

Comparison of the serotype-specific genes in VPsero database [9] with 21 human-derived *V. parahaemolyticus* genomes was performed by BLASTN to identify potential serotypes. Meanwhile, the 21 genomes were uploaded to the Kaptive website (<https://kaptive-web.erc.monash.edu/>) for review and supplementation [10], aiming to obtain a more comprehensive serotype classification. In Kaptive database, the nomenclature for the O- and K- loci was adopted based on the *Klebsiella* capsule synthesis loci, each specific O- and K- locus was separately denoted as OL (O-locus) and KL (K-locus), followed by a unique numerical identifier.

To investigate the potential avenues of cross-transmission of *V. parahaemolyticus*, 74 environmental or foodborne genomes from diverse countries and collection dates were selected and downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/assembly/?term=vibrio>). A phylogenetic tree of the 21 human-derived and 74 environmental or foodborne *V. parahaemolyticus* isolate was built using IQ-TREE based on the concatenation of the core genomes [23]. The evolutionary distances were computed using the Maximum Likelihood method. The phylogenetic tree was visualized and annotated using Interactive tree of Life (iTOL) [24].

Results

Epidemiological features of *V. parahaemolyticus* isolates

In this study, a total of 1,435 anal swab samples were collected from imported travelers ranged from 2017 to 2019, with 35 positive cases being detected, for a positive rate of 2.44% for the *V. parahaemolyticus*. Twenty-one (21/35, 60.0%) human-origin strains were successfully isolated and purified for further analysis (Table 1). According to statistics, these strains were traced back to seven Asian countries and regions (Fig. 1A). Notably, Thailand accounted for the largest proportion of 57.14%,

Table 1 Detailed information and serotyping analysis of 21 isolates collected in Shanghai Port, China

ID	Information of epidemiological survey				Result of genome sequencing					ST	VPsero serotype			Kaptive serotype		
	Date	Region	Gender	Age	Gene	Protein	RNA	Scaffold	Length (bp)		O-loci	K-loci	Ox:Ky	O-loci	K-loci	Ox:Ky
SHP-1108	8/31/2017	Philippines	Male	24	4,748	4,748	116	17	5,190,976	3	7	68	07:K68	OL4-1	KL68-1	OL4-1:KL68-1
SHP-1117	10/10/2017	Viet Nam	Female	-	4,581	4,581	119	17	5,078,856	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1177	10/11/2017	Philippines	Female	-	4,666	4,666	111	33	5,157,286	17	12	4	012:K4	OL4	KL4	OL4:KL4
SHP-1179	6/20/2017	Thailand	Female	56	4,602	4,602	110	22	5,038,963	2516	4	8	04:K8	OL4	KL8	OL4:KL8
SHP-1269	8/6/2017	Hong Kong, China	Female	11	4,603	4,603	114	13	5,104,622	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1361	4/8/2017	Japan	Male	22	4,729	4,729	106	36	5,108,367	-	2	-	-	OL2	KL22	OL2:KL22
SHP-1429	9/13/2017	Thailand	Female	40	4,669	4,669	102	18	5,153,566	-	12	55	012:K55	OL4	KL55	OL4:KL55
SHP-1434	8/12/2017	Thailand	Female	28	4,625	4,625	119	24	5,075,979	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1678	8/19/2017	Thailand	-	-	4,652	4,652	121	20	5,127,586	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1433	2/26/2018	Philippines	Female	34	4,615	4,615	120	21	5,093,560	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1611	5/21/2018	Thailand	Male	32	4,606	4,606	104	29	4,990,142	635	-	-	-	OL1	KL118	OL1:KL118
SHP-1675	8/12/2018	Thailand	Female	10	4,548	4,548	112	25	4,991,106	-	12	-	-	OL10	KL24	OL10:KL24
SHP-1847	3/17/2018	Thailand	Male	36	4,576	4,576	126	22	5,077,773	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1926	3/21/2018	Thailand	Male	32	4,688	4,688	122	24	5,128,246	3	3	6	03:K6	OL3_or_OL13	KL6	OL3_or_OL13:KL6
SHP-1949	6/21/2018	Thailand	Female	44	4,647	4,647	115	15	5,056,207	-	12	9	012:K9	OL4	KL9	OL4:KL9
SHP-1780	9/14/2018	Malaysia	Female	11	4,763	4,763	115	15	5,190,902	3	4	68	04:K68	OL4	KL68-1	OL4:KL68-1
SHP-3796	12/23/2018	Thailand	Male	66	4,629	4,629	99	42	5,068,935	199	-	-	-	OL1	KL123	OL1:KL123
SHP-8214	12/19/2018	Malaysia	Female	35	4,604	4,604	98	36	5,015,880	2516	12	8	012:K8	OL4	KL8	OL4:KL8
SHP-8216	12/19/2018	Indonesia	Female	40	4,614	4,614	95	34	5,021,272	2516	12	8	012:K8	OL4	KL8	OL4:KL8
SHP-3690	7/4/2019	Thailand	Female	45	4,622	4,622	98	51	5,079,348	-	1	-	-	OL1	KL123	OL1:KL123
SHP-8502	8/14/2019	Thailand	Female	33	4,721	4,721	98	40	5,107,560	2516	12	-	-	OL4	KLUT5	OL4:KLUT5

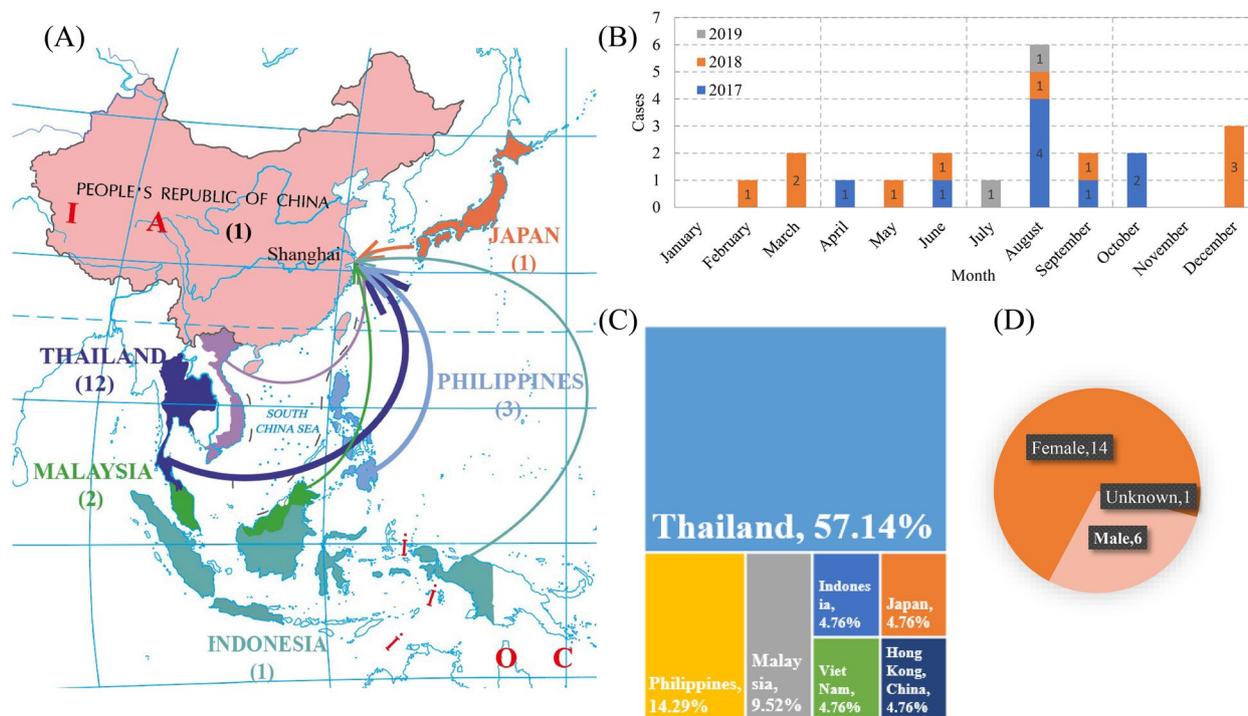


Fig. 1 The information of epidemiological and genomic features of 21 human-origin *V. parahaemolyticus* isolates. **A** The regional distribution of the strains. Arrows in different colors represent diverse countries where samples isolated from. The width of arrows reflects the number of the strains. **B** The number of strains isolated in different months from 2017 to 2019. Different years are marked in diverse colors. **C** The proportion of strains isolated from different countries. The size of the rectangle represents the number of isolates collected in corresponding locations. **D** Percentage of different sample genders is shown by pie chart

followed by the Philippines accounted for 14.29%, and Malaysia accounted for 9.52% (Fig. 1C). To gain insights into the temporal dynamics of *V. parahaemolyticus* incidents, a comprehensive analysis of the distribution pattern across quarters was undertaken. The distribution of *V. parahaemolyticus* cases exhibited noteworthy variations across quarters. The third quarter stood out with a peak incidence rate of 42.86%, closely followed by the fourth quarter which accounted for 23.81% of the cases. A detailed examination of the data unveiled that the highest surge in cases was recorded during the month of August, accounting for 28.57% of the total cases over the course of three years (Fig. 1B). Upon delving into the demographic profile of the affected individuals, it was ascertained that the mean age of the cases was approximately 33.27 years old. Gender distribution revealed that females constituted a significant majority, comprising 66.67% of the cases (Fig. 1D).

Genomic features of 21 human-origin *V. parahaemolyticus* isolates

Genome sequencing of the 21 *V. parahaemolyticus* isolates revealed an average genome size of 5.08 ± 0.058 Mb, with variations spanning from 4.99 Mb (SHP-1611)

to 5.19 Mb (SHP-1108). The gene content analysis illustrated that these genomes, in aggregate, consisted of an average of 4,643 protein-encoding genes (Table 1). Evidencing the degree of genomic similarity, the average nucleotide identity (ANI) was evaluated among the 21 isolates and the reference genome. The presence and absence of genes was visualized in the highest and lowest ANI value genomes compared with the rest of isolates, and found that most regions of the strain SHP-1611 with the lowest ANI value (98.30%) are consistent with the reference genome RIMD 221063. The pan-genome of the 21 *V. parahaemolyticus* genomes consisted of 97,508 protein-coding genes, of which, 2,461 core genes were identified. Rarefaction analysis demonstrated an open pan-genome, indicative of the diversity limitation within the gene pool.

Serotyping analysis and phylogenetic relatedness

Combining the serotype classification from the VPsero database and Kaptive database (Table 1), it was observed that the main serotype of these 21 strains was O3:K6 or OL3:KL6 (7/21, mainly detected in travelers originating in Thailand), consistent with the prevailing trend of *V. parahaemolyticus* in the worldwide. For example, strains

SHP-1611 and SHP-3796, which lacked serotype classification in the VPsero database, were categorized as serotype OL1:KL118, while SHP3796 was categorized as OL1:KL123 in the Kaptive database, respectively. Therefore, these results could be a useful basic for clinical treatment and basic research investigation.

Seventy-four strains downloaded from NCBI were distributed in fourteen countries, including China, Mexico, Viet Nam, Colombia, Venezuela, Malaysia, Nigeria, Thailand, Philippines, South Korea, India, Chile, Peru and the USA, with samples from environmental or foodborne sources ranging from 2010 to 2019 (Additional file 1). According to the phylogenetic

tree based on 96 concatenated genomes of 1,675 core genes, there was a close genetic relationship between the isolates detected in Thailand, Indonesia, Malaysia and Philippines (e.g., SHP-8502, SHP-8214, SHP-8216 and SHP-1179) (Fig. 2). Strains of Thailand isolated in different years still shared close genetic relationship (e.g., SHP-3690 and SHP-3796, SHP-1675 and SHP-1429). Only three strains separated in Thailand were found to have a close phylogenetic relationship with strains collected from the Chinese environment, such as strains SHP-1611 (2018, Thailand) had a tight phylogenetic connection with YK20 isolated from shrimp bond in 2019, China.

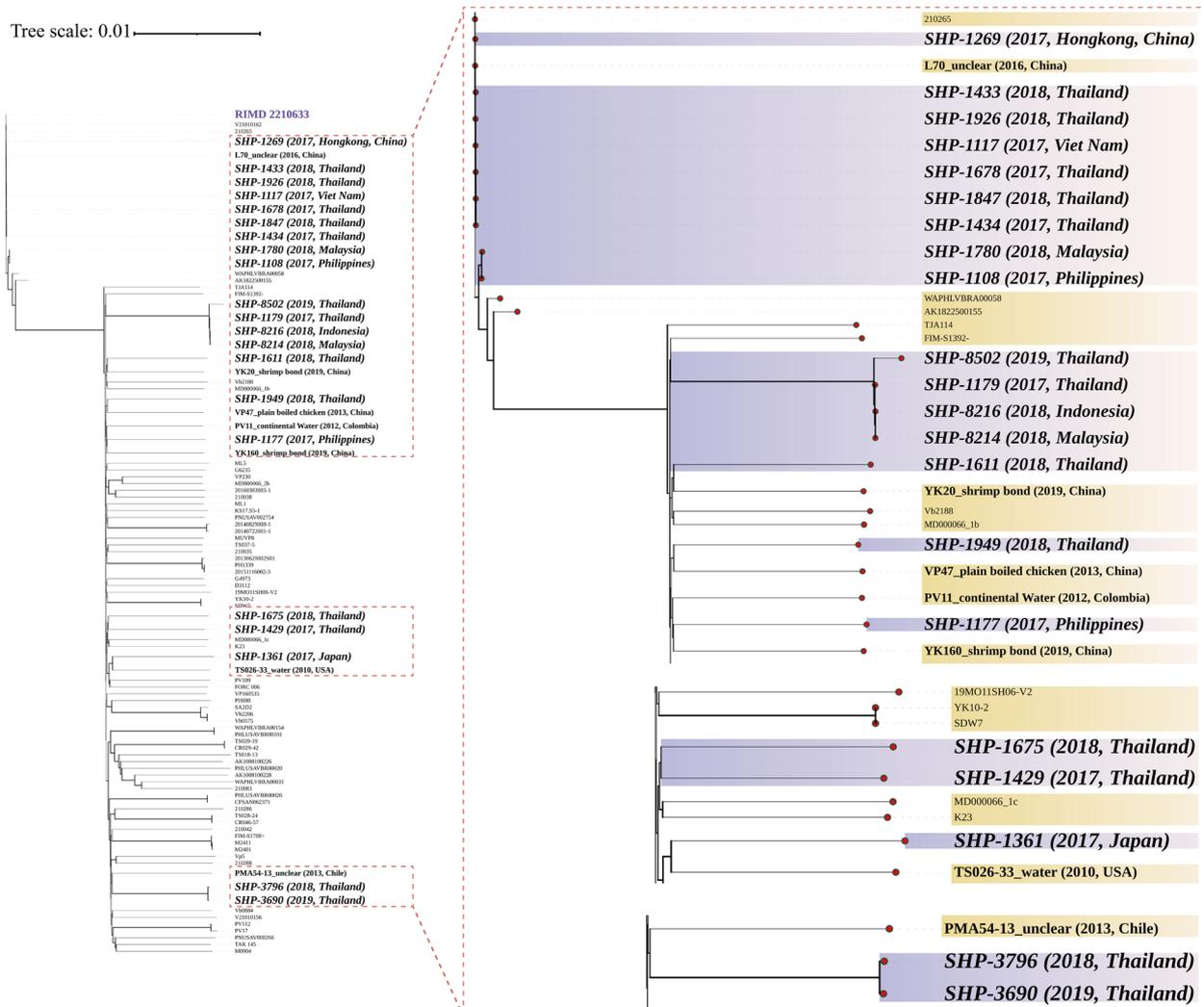


Fig. 2 The phylogenetic tree of 96 *V. parahaemolyticus* isolates. The tree was constructed based on the alignment of concatenated core genome of 96 strains, including 21 isolates detected in this study and 74 genomes downloaded from NCBI collected in different years and geographical regions. The reference genome, RIMD 2210633 was also added. The comprehensive evolutionary tree is shown on the left, while the right side provides an enlarged view of the relationship between 21 human-derived *V. parahaemolyticus* isolates (purple strip) and other strains derived from environmental other sources (yellow strip)

Distribution of potential virulence-associated genes and antimicrobial resistance genes

The exploration of evolutionary relationships was further deepened by conducting a phylogenetic analysis based on the 2,461 core genes encompassing a collection of 22 *V. parahaemolyticus* strains, including reference genome RIMD 2210633 (Fig. 3). Isolates sharing the same Sequence Type (ST) were perceptibly grouped together within the Maximum Likelihood (ML) tree. Nine of the 21 *V. parahaemolyticus* isolates were attributed to ST3, four isolates attributed to ST2516, and three isolates separately assigned to ST17, ST635, and ST199. Five additional isolates remained untyped. It's noteworthy that despite the variance in both geographic origins and the temporal context of collection, a discernible pattern emerged. Regional clustering was evident, with strains originating from Southeast Asian countries and regions displaying a marked similarity.

Turning attention to potential virulence-associated gene detection, the examination spotlighted 15 virulent mechanisms governed by the predicted genes (Additional file 2). These genes predominantly participated in adhesion, exotoxin production, and effector delivery systems, constituting pivotal virulence attributes. Remarkably, the presence of the *tlh* gene, encoding thermolabile hemolysin, a specific marker of *V. parahaemolyticus*, was identified in 20 out of 21 strains

(95.24%). In stark contrast, the occurrence of the *tdh* and *trh* genes was observed in 16 out of 21 strains (76.19%) and two out of 21 strains (9.52%). However, only SHP-3796 contained both the *tdh* and *trh* genes.

Through a comprehensive analysis based on the CARD, a robust assortment of 147 drug resistance genes was identified (Additional file 3). On average, each strain was found to harbor approximately 120 antibiotic resistance-related genes, spanning a spectrum of 25 distinct antibiotics. Remarkably, a substantial proportion of these resistance genes encoded efflux pumps, effectively conferring resistance against an array of antibiotics encompassing the top five predicted antibiotics including peptides, disinfecting agents and antiseptics, tetracycline, aminoglycoside, glycopeptide, and more (Fig. 3). The categories of drug-resistant genes contained in these strains indicated the possibility of the presence multidrug resistance. Of the 21 *V. parahaemolyticus* genomes, 20 (95.24%) strains contained the gene *tet* (W/N/W), *tet* (W/32/O) and *smeS*. Moreover, within the context of STs, the gene *vanH_in_vanM_cl* was solely identified in strains (SHP-8216, SHP-8214, SHP-8502 and SHP-1179) designated as ST2516. This congruence between the ST types and the associated resistance patterns provided intriguing insights into the genetic determinants influencing antibiotic resistance within the *V. parahaemolyticus* population.

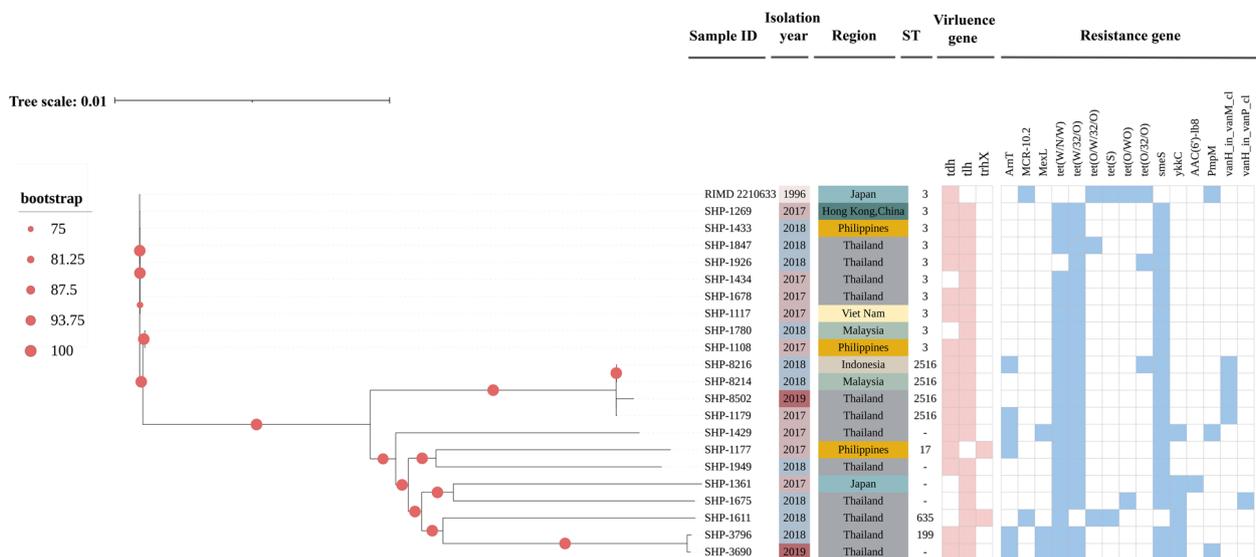


Fig. 3 Maximum-likelihood phylogeny using core genome alignment of 21 *V. parahaemolyticus* isolates. Sample ID represents the nomenclature for the strains, Isolation year means the concrete year of isolation, ST describes the sequence types, the presence of the three most common virulence genes in each strain is labeled with red stripe and the blue represents the distribution of resistance genes of the top five antibiotics contained the most predicted resistance genes (peptide, disinfecting agents and antiseptics, tetracycline, aminoglycoside and glycopeptide). The phylogenetic tree was rooted to a reference genome, *V. parahaemolyticus* RIMD 2210633

Discussion

The true global burden of illness caused by *Vibrio* exposure is currently underestimated due to the limited detection and surveillance of vibriosis [11]. This study represents a comprehensive genomic investigation of *V. parahaemolyticus* isolates and provides insight into the diversity and epidemiology of cross-border spread pathogens. The use of WGS provides a high level of resolution, making it crucial for public health and safety to mitigate the risk of cross-border transmission and spread of the pathogen in China. Cross border surveillance is a key issue in understanding their epidemiology and health risks to human [25].

With the increasing accessibility of genome-wide sequencing technologies, numerous pathogen genomes have become available. These genomes have facilitated the identification of effector proteins with diverse functions in various pathogens [26]. The study benefits from the utilization of network techniques and the construction of databases, which overcome the barriers associated with cross-border pathogen surveillance. Additionally, pangenome analyses represent an adequate comparative genomics approach to investigate the intraspecific diversity of *V. parahaemolyticus* by analyzing the core and accessory genomes of different isolates of species.

The findings indicated the utilization of WGS data from foodborne bacteria is crucial for monitoring the environment, ensuring food safety, and diagnosing diseases early. Advances in sequencing technology and reduced costs have contributed to an abundance of bacterial genomic data, enabling a clearer understanding of the transmission patterns of foodborne pathogens.

Travel, including medical tourism, is an important risk factor for the spread of pathogens. Southeast Asia has been identified as a hotspot for the emergence of both extra-intestinal bacteria and gastrointestinal pathogens [27]. Our previous investigation of human norovirus at Shanghai Port has revealed a concentration of cases originating from Southeast Asia, particularly from Thailand [14, 15]. Climate is known to have influence on the distribution of human pathogens [28]. Thailand is predominantly an agricultural country, locates in a tropical zone, which has a hot and humid climate with different epidemiology from other countries [29]. Meanwhile, it is one of the most visited countries globally, improving opportunities for pathogens transmission all around the world. These factors contribute to the formation of a 'melting pot' of diverse pathogens. In this study, half of the samples were collected from Thailand (57.14%) during the three years. Therefore, it is essential to intensify *V. parahaemolyticus* surveillance in Thailand and other Southeast Asian countries to detect emerging strains with epidemic potential at an early stage.

For the serotype identification of 21 *V. parahaemolyticus* isolates, the Kaptive database provided a relatively comprehensive serological typing when serotypes cannot be conclusively determined in the VPsero database, such as SHP-1611 previously described (Table 1). But the O-serogroup genetic determinant region (OGD region) in this database, serotype O13 was found to share 100% identity with corresponding region of O3 serotype, for which the whole genome has been published previously [30]. Therefore, in Kaptive O-database, the serotype O3 or O13 genomes were categorized as OL3_or_OL13. For example, in the VPsero database, the serotype of SHP-1611 was identified as O3:K6, but in the Kaptive website, it was identified as OL3_OL13:KL6. Notably, these two classification methods exhibit a high degree of consistency in the analysis of the K antigen.

In terms of pathogen risk assessment, virulence genes (*tdh*, *trh*) were found in most of isolates in this study. The presence of these two virulence genes in *V. parahaemolyticus* is commonly used in clinical, surveillance, and research testing as an indicator of pathogenicity [1, 12]. The *tlh* gene, as the unique marker for *V. parahaemolyticus* identification, revealed in all the 74 isolates obtained from environmental or other sources that were included in our study [31] (Additional file 1). Furthermore, the presence of the *tdh* gene was nearly negligible in environmental strains, while it was extensively found in clinical strains, consistent with findings from prior reports [32]. It is known that horizontal gene transfer plays a role in the transfer of pathogenicity [33]. Therefore, it is important to continue monitoring and investigating other pathogenicity markers in these products.

There is strong evidence to suggest that AMR can disseminate globally across borders [34]. Multi-resistant strains are being distributed worldwide through air travel and other forms of travel across borders [35]. Our AMR isolates demonstrated resistance to three or more antimicrobial agents. This relatively high prevalence of these strains and their AMR profiles reveal the potential public health problems associated with illegal import of food in passengers' luggage. Therefore, it is imperative for nations to collaborate in order to mitigate the emergence of global AMR. Also, appropriate measures should be taken to prevent the spread of these resistant-bacterial isolates by increasing awareness about health and hygiene and by restricting the random use of antibiotics and antiseptics [36].

Conclusions

Through the examination of 21 strains sourced from diverse countries and regions, our research has pinpointed Southeast Asia as a significant hotspot for the prevalence of *V. parahaemolyticus*. Furthermore,

phylogenetic analysis has delineated clear delineations between strains originating from human and environmental sources, underscoring the restricted transmission dynamics between these two reservoirs. Additionally, the detection of virulence genes and antibiotic resistance-related genes in the majority of isolates underscores their pivotal contributions to the pathogenicity of *V. parahaemolyticus*-induced illnesses. The findings of this research enhanced our comprehension of the worldwide spread of *V. parahaemolyticus* and carried substantial implications for future public health prevention efforts.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-024-03303-7>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgments

We thank the editor and the anonymous reviewers for their comments and suggestions. We would like to thank Xiaoyun Ma, Sunmei Nie and Leiping Zhang for invaluable help in sample collection.

Authors' contributions

Huajun Zheng, Zilong Zhang and Zhengan Tian designed the project. Danlei Liu, Lei Zhou and Zilei Zhang analyzed the data and prepared the primary manuscript. Zhiyi Wang, Shiwen Chen, Ying Zhang and Yongqiang Zhu participated in sample collection, bioinformatics analysis and figures illustration. Zilong Zhang and Huajun Zheng revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Key R&D Program of China under grant 2022YFC2302800; China Postdoctoral Science Foundation under grant 2022M722152; Shanghai Sailing Program under grant 22YF1415600; Starting Research Fund from Shanghai Customs College (kyqd202212) and General Administration of Customs Project (2022HK136).

Availability of data and materials

The raw sequencing data reported in this study have been deposited in the Genome Sequence Archive in the National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformatics), Chinese Academy of Sciences, under accession number CRA014330 and are publicly accessible at <https://bigd.big.ac.cn/gsa>.

Declarations

Ethics approval and consent to participate

No applicable.

Consent for publication

No applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Shanghai International Travel Healthcare Center, Shanghai Customs District P. R. China, Shanghai 200136, China. ²Shanghai-MOST Key Laboratory of Health and Disease Genomics, NHC Key Lab of Reproduction Regulation, Shanghai Institute for Biomedical and Pharmaceutical Technologies, School of Life

Sciences, Fudan University, Shanghai 200237, China. ³Inspection and Quarantine Technology Communication Department, Shanghai Customs College, Shanghai 200136, China. ⁴State Key Laboratory of Applied Microbiology Southern China, Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, Guangdong Open Laboratory of Applied Microbiology, Guangdong Institute of Microbiology, Guangzhou 510070, China.

Received: 2 January 2024 Accepted: 12 April 2024

Published online: 26 April 2024

References

- Hartnell RE, Stockley L, Keay W, Rosec JP, Hervio-Heath D, Van den Berg H, et al. A pan-European ring trial to validate an international standard for detection of *Vibrio cholerae*, *Vibrio parahaemolyticus* and *Vibrio vulnificus* in seafoods. *Int J Food Microbiol*. 2019;2(288):58–65.
- Plaza N, Pérez-Reytor D, Ramírez-Araya S, Pavón A, Corsini G, Loyola DE, et al. Conservation of small regulatory RNAs in *Vibrio parahaemolyticus*: Possible role of RNA-OUT encoded by the pathogenicity island (VPal-7) of pandemic strains. *Int J Mol Sci*. 2019;20(11):2827.
- Vezzulli L, Grande C, Reid PC, Hélaouët P, Edwards M, Höfle MG, et al. Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic. *Proc Natl Acad Sci USA*. 2016;113(34):E5062–5071.
- Mok JS, Ryu A, Kwon JY, Kim B, Park K. Distribution of *Vibrio* species isolated from bivalves and bivalve culture environments along the Gyeongnam coast in Korea: Virulence and antimicrobial resistance of *Vibrio parahaemolyticus* isolates. *Food Control*. 2019;106:106697.
- Pang R, Li Y, Chen M, Zeng H, Lei T, Zhang J, et al. A database for risk assessment and comparative genomic analysis of foodborne *Vibrio parahaemolyticus* in China. *Sci Data*. 2020;7(1):321.
- Chen X, Zhu Q, Liu Y, Wang R, Xie H, Chen J, et al. Pathogenic characteristics of and variation in *Vibrio parahaemolyticus* isolated from acute diarrhoeal patients in southeastern China from 2013 to 2017. *Infect Drug Resist*. 2020;13:1307–18.
- Zhao X, Zhou Q. Path of building a global city in Shanghai based on the "Space of Flows" theory. *J Donghua University (English Edition)*. 2019;36(1):88–95.
- Guo X, Liu B, Chen M, Wang Y, Wang L, Chen H, et al. Genetic and serological identification of three *Vibrio parahaemolyticus* strains as candidates for novel provisional O serotypes. *Int J Food Microbiol*. 2017;20(245):53–8.
- Twedt RM, Spaulding PL, Johnson HM. Antigenic relationships among strains of *Vibrio parahaemolyticus*. *Appl Microbiol*. 1972;23(5):966–71.
- Janecko N, Bloomfield SJ, Palau R, Mather AE. Whole genome sequencing reveals great diversity of *Vibrio spp* in prawns at retail. *Microbial Genomics*. 2021;7(9):000647.
- Bian S, Jia Y, Zhan Q, Wong NK, Hu Q, Zhang W, et al. VPsero: Rapid serotyping of *Vibrio parahaemolyticus* using serogroup-specific genes based on whole-genome sequencing data. *Front Microbiol*. 2021;30:12.
- van der Graaf-van Bloois L, Chen H, Wagenaar JA, Zomer AL. Development of Kaptive databases for *Vibrio parahaemolyticus* O- and K-antigen genotyping. *Microbial Genomics*. 2023;9(5):mgen001007.
- O'Boyle N, Boyd A. Manipulation of intestinal epithelial cell function by the cell contact-dependent type III secretion systems of *Vibrio parahaemolyticus*. *Front Cell Infect Microbiol*. 2014;3:114.
- Liu D, Zhang Z, Li S, Wu Q, Tian P, Zhang Z, et al. Fingerprinting of human noroviruses co-infections in a possible foodborne outbreak by metagenomics. *Int J Food Microbiol*. 2020;16(333):108787.
- Zhang Z, Liu D, Li S, Zhang Z, Hou J, Wang D, et al. Imported human norovirus in travelers, Shanghai port, China 2018: An epidemiological and whole genome sequencing study. *Travel Med Infect Dis*. 2021;43:102140.
- Li Y, Xie T, Pang R, Wu Q, Zhang J, Lei T, et al. Food-borne *Vibrio parahaemolyticus* in China: prevalence, antibiotic susceptibility, and genetic characterization. *Front Microbiol*. 2020;11:1670.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*. 2014;30(15):2114–20.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*. 2007;23(6):673–9.

19. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;15(10):421.
20. Jain C, Rodríguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114.
21. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. PGAP: pan-genomes analysis pipeline. *Bioinformatics*. 2012;28(3):416–8.
22. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2023;51(D1):D690–9.
23. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic Era. *Mol Biol Evol*. 2020;37(5):1530–4.
24. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
25. Xu Q, Li ZW, Zhang XA, Liu MY, Wang JL, Zhang HY, et al. The imported infections among foreign travelers in China: an observational study. *Int Libr Eth Law New*. 2022;18(1):97.
26. Huang Z, Yu K, Xiao Y, Wang Y, Xiao D, Wang D. Comparative genomic analysis reveals potential pathogenicity and slow-growth characteristics of genus *Brevundimonas* and description of *Brevundimonas pishanensis* sp. nov. *Microbiol Spectr*. 2022;10(2):e0246821.
27. Molton JS, Tambyah PA, Ang BSP, Ling ML, Fisher DA. The global spread of healthcare-associated multidrug-resistant bacteria: a perspective from Asia. *Clin Infect Dis*. 2013;56(9):1310–8.
28. Guernier V, Hochberg ME, Guégan JF. Ecology drives the worldwide distribution of human diseases. *PLoS biology*. 2004;2(6):e141.
29. Pongpan S, Thanatrakolsri P, Vittaporn S, Khamnuan P, Daraswang P. Prognostic factors for leptospirosis infection severity. *Trop Med Infect Dis*. 2023;8(2):112.
30. Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, et al. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet*. 2003;361(9359):743–9.
31. Park JY, Jeon S, Kim JY, Park M, Kim S. Multiplex real-time polymerase chain reaction assays for simultaneous detection of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus*. *Osong Public Health Res Perspect*. 2013;4(3):133–9.
32. Raghunath P. Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in *Vibrio parahaemolyticus*. *Front Microbiol*. 2015;22(5):805.
33. Le Roux F, Blokesch M. Eco-evolutionary dynamics linked to horizontal gene transfer in *Vibrios*. *Annu Rev Microbiol*. 2018;8(72):89–110.
34. Barlam TF, Gupta K. Antibiotic resistance spreads internationally across borders. *J Law Med Ethics*. 2015;43(Suppl 3):12–6.
35. Rodríguez-Lázaro D, Oniciuc EA, García PG, Gallego D, Fernández-Natal I, Domínguez-Gil M, et al. Detection and characterization of *Staphylococcus aureus* and methicillin-resistant *S. aureus* in foods confiscated in EU borders. *Front Microbiol*. 2017;8:1344.
36. Safain KS, Islam MS, Amatullah J, Mahmud-Un-Nabi MA, Bhuyan GS, Rahman J, et al. Prevalence of silver resistance determinants and extended-spectrum β -lactamases in bacterial species causing wound infection: First report from Bangladesh. *New Microbes New Infect*. 2023;52:101104.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.