BMC Microbiology

RESEARCH

Open Access

MtlA gene sequencing based on 3rd -generation sequencing for *Vibrio* typing

Rensi Zhou[1,2†], Yingju Qi[3†], Lei Zhou[2†], Danlei Liu[4], Yaozhong Zou[3], Ting Xian[3], Zilong Zhang[4*], Huan Hu[3*] and Huajun Zheng[2*]

Abstract

Background Traditional serotyping and multilocus sequence typing (MLST) for *Vibrio parahaemolyticus* have limitations in resolution and accuracy, particularly for strains with close genetic relationships. The long read length and high sequencing quality of a new type of nanopore sequencer G-seq500 make it possible to type *V. parahaemolyticus* strains using *mtlA*, a new identified typing gene for *Vibrio*. This gene not only offers advantages in cost-effectiveness and speed relative to MLST but also demonstrates comparable resolution.

Results We sequenced the full-length *mtlA* gene (1,953 bp) of 33 *V. parahaemolyticus* strains using Geneus G-seq500. A total of 199,658 long reads with an average length of 17,306 bp were produced by one sequencing chip. After recognizing adaptor and barcode, each long read can be split into 12 subreads with mean accuracy of 93.25%, which were then aligned into 64,086 consensus sequences with a mean quality Q35 (sequence accuracy 99.88–99.93%). The final *mtlA* gene was generated by clustering all the consensus sequencing from same strain, and its accuracy reached 100% with over four consensus sequences clustered together. Then we developed a naming rule based on *mtlA* clustering of 7,573 *mtlA* sequences from NCBI together with the 33 *mtlA* gene sequenced by G-seq500, with the largest cluster composed of 2,321 strains designated as mtlA0001.

Conclusions Our study confirms the high sequencing accuracy of Geneus G-seq500, and the effectiveness of using the full-length *mtlA* gene for *V. parahaemolyticus* strain typing and epidemiological surveillance.

Keywords *Vibrio parahaemolyticus*, Genotyping, *MtlA*, Third-Generation sequencing, Epidemiological surveillance

†Rensi Zhou, Yingju Qi, Lei Zhou these authors contributed equally to this work.

*Correspondence:
Zilong Zhang
heilong82@163.com
Huan Hu
huan.hu@geneus-tech.com
Huajun Zheng
zhenghj@chgc.sh.cn
[1]College of Life Sciences, Shanghai Normal University, Shanghai 200234, China
[2]Shanghai-MOST Key Laboratory of Health and Disease Genomics, NHC Key Lab of Reproduction Regulation, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai 200237, China
[3]Geneus Technologies (Chengdu) Co.,Ltd, Chengdu 610041, China
[4]Shanghai International Travel Healthcare Center, Shanghai Customs District P. R. China, Shanghai 200335, China

Background

Vibrio parahaemolyticus is a species of halophilic gram-negative bacteria found in water, sediment, aerosols, plankton, fish and shellfish. *V. parahaemolyticus* not only seriously endangers the mariculture industry [1] but can also cause gastrointestinal diseases in humans. Eating food contaminated with *V. parahaemolyticus* will cause nausea, vomiting, abdominal pain, diarrhea, headache and fever and, in severe cases, may induce septicemia [2], and its global incidence is increasing [3].

Rapid and accurate genotyping of *V. parahaemolyticus* and understanding its prevalence are highly important for quickly proposing solutions and controlling the

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

spread of diseases and pathogens [4]. For instance, in a study analyzing the genetic and evolutionary relationships of 162 *V. parahaemolyticus* strains isolated from Guangdong Province, China, researchers utilized multilocus sequence typing (MLST) technology to trace the sources of clinical isolates and prevent the outbreaks of *V. parahaemolyticus* infection [5]. In another study, researchers performed a genome wide phylogenetic analysis of a global collection of 111 *V. parahaemolyticus* ST36 isolates and revealed the trend of its transcontinental expansion, which reflects the crucial role of genotyping in monitoring and controlling the spread of pathogens [6]. At present, serotyping and MLST are the main methods for genotyping *V. parahaemolyticus*.

The serotype of *V. parahaemolyticus* is detected mainly by mixing the cell antigen O-antigen and the envelope antigen K-antigen with polyvalent antiserum to observe whether agglutination occurs [7]. However, the resolution of the serotype is poor, subjective judgment has a large impact, most strains have the same serotype, or once the corresponding antigen site is mutated, a new serotype may appear, and the real difference between strains cannot be reflected. This series of problems present great challenges for the epidemiological surveillance and tracing of *V. parahaemolyticus* [8].

MLST is a typing method for amplification analysis of specific gene sequences in the genome. By uploading the sequences of typing genes (generally 6 to 8) to the PubMLST database, the STs of a strain can be obtained through multisequence comparisons with the sequences of existing alleles in the database [9]. The MLST database of *V. parahaemolyticus* contains seven housekeeping genes, namely, *dnaE* (557 bp), *gyrB* (592 bp), *recA* (729 bp), *didS* (458 bp), *pntA* (430 bp), *pyrC* (493 bp) and *tnaA* (423 bp). However, for strains with close relatives, it is still difficult to use MLST to achieve accurate traceability [3]. Interspecies recombination events have been reported in the *recA* gene, leading to undefined MLST types. Furthermore, such recombination in *recA* can alter the topology of the MLST phylogenetic tree, potentially obscuring true evolutionary relationships [10].

To screen new genes that are easier to standardize, reduce time and cost and can be carried out by grassroots laboratories on their own, we identified *mtlA*, the core gene of 1,953 bp with the highest variation rate among the genomes of *V. parahaemolyticus* [4]. This gene encodes PTS mannitol transporter subunit IICBA and can achieve a resolution comparable to or even higher than that of traditional MLST. Furthermore, *mtlA* shows broad applicability, being present not only in *V. parahaemolyticus* but also universally detected in 13 pathogenic *Vibrio* species (e.g., *V. cholerae*, *V. alginolyticus*) [11]. Notably, *mtlA* undergoes positive selection pressure, indicating adaptive evolution associated with environmental survival and pathogenicity. Owing to its economic and computational advantages and universality as a single gene, the *mtlA* gene may replace MLST as a new gene for typing and tracing with high efficiency.

The MLST scheme was first developed for the species *Neisseria meningitides* in 1998 [12], and it has been relatively expensive to execute, mainly because of the laborious process of DNA sequencing by Sanger technology at that time [13].

In genotyping, the most central step is gene sequencing. Sanger sequencing, also known as first-generation sequencing technology, has successfully promoted the human genome project, but the shortcomings of low throughput and high cost limit its further large-scale application. Second-generation sequencing technology, also known as next-generation sequencing technology, has been widely used in basic research and clinical diagnosis and treatment because of its advantages of high throughput and low cost. However, the short length of sequencing reads limits their application in genotyping [14].

Third-generation sequencing technology, owing to its advantages of long reading length, fast speed and real-time single-molecule sequencing, has gradually demonstrated clinical application value in tumors, immunity, reproduction and other related fields [15]. Over the past decade, nanopore sequencers, exemplified by the products of Oxford Nanopore Technologies (ONT), have drawn significant attention within sequencing communities due to their abilities to provide long-read sequencing results at a relatively low cost and in a timely manner.

The G-seq500 developed by Geneus Technologies is a new type of nanopore sequencer commercially available recently. Different from the sequencers by ONT that utilizes nanopore as the biosensor to directly read bases of DNA single strand, G-seq500 employs the approach of "sequencing by synthesizing (SBS)". *CW. Fuller* et al. and *PB Stranges* et al. introduced prototype platforms that use nanopores to detect nucleotides one by one during DNA synthesis to determine the sequence of the template DNA [16, 17]. In their work, a DNA polymerase bound with primed DNA template was covalently attached to a nanopore, and the nucleotides were tagged with distinct polymers specific to the A, T, C, G bases. When a nucleotide was captured by the polymerase for synthesis, the tag interacted with the attached nanopore, producing base-specific electronic signal that could be detected by a chip array for real time sequencing. The G-seq500 builds upon a similar strategy, using a nanopore to identify nucleotide bases during DNA synthesis. With extensive optimization of the biochemical system, including a novel nanopore and polymerase, uniquely modified nucleotides, and an improved design of the nanopore-polymerase conjugation, G-seq500 shows significantly

enhanced performance compared to the prototype platform, including the read length at the range of tens of kilobase pairs and a remarkably high consensus accuracy exceeding Q50 at the coverage of 50X.

As a third-generation sequencer with superior performance and a viable alternative to products from Pacific Biosciences or ONT, G-seq500 draws our great interest in exploring how its high-accuracy long reads can facilitate genotyping for *V. parahaemolyticus*. In this study, we tested the efficiency of sequencing the full-length *mtlA* gene (1,953 bp) via G-seq500, verifying its application in genotyping.

Methods

Samples

Thirty-three *Vibrio parahaemolyticus* strains were isolated from the anal swabs of imported travelers from Shanghai Customs from 2017 to 2019, and their genomes were sequenced through 2 × 150 bp paired-end sequencing in the Illumina X-10 platform [18]. In this study, all strains 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).

PCR amplification of the *MtlA* gene

The primers for the *mtlA* genes were designed with Primer 3.0 [19] and synthesized by Tsingke Biotech Co., Ltd. (Shanghai, China).

PCRs were performed in a volume of 25 μl containing 5 μl of DNA, 12.5 μl of 2 × PCR mix (Yeasen, Shanghai, China), 1 μl of the forward and reverse primers for the *mtlA* genes, and 5.5 μl of sterile distilled water. PCRs were performed under the following conditions: initial denaturation at 94 °C for 5 min; 35 cycles of denaturation at 94 °C for 30 s, annealing at 63 °C for 30 s, and elongation at 72 °C for 2 min; and elongation at 72 °C for 5 min.

G-seq500 library Preparation and sequencing

Library preparation was performed via library preparation kits, complex incubation kits, and magnetic bead kits (Cat.No: 030.001.001, Geneus Technologies (Chengdu) Co.,Ltd, China). In brief, equal amounts of amplicons from 33 samples were first mixed and purified. One microgram of DNA, which was modified to create blunt ends suitable for ligation with sequencing adapters, was then subjected to end repair at 20 °C for 30 min. After end repair, the DNA fragments are ligated to the sequencing adapters at 20 °C for 15 min, followed by digestion at 37 °C for 30 min, the addition of sequencing primers and annealing at 25 °C for 20 min. Finally, sequencing complexes were obtained after protein complexes were combined with the library at 25 °C for 10 min, followed by magnetic bead purification. Sequencing was performed on the G-seq500 sequencer using one cell, and consensus sequences were finally generated via local algorithms.

Amplicon assembly

A total of 109,877 consensus sequences were obtained from the 33 samples and were filtered by read length and sample barcode sequences via local scripts, which resulted in 64,860 consensus sequences. All reads with identical indices were assembled into a consensus sequence.

Sequence accuracy evaluation

We assessed errors in the assemblies via comparison with the *mtlA* gene of the genome sequence of each strain, which had been reported previously [18]. To assess the sequencing accuracy, we aligned each consensus sequence to the final *mtlA* gene sequence via BLASR [20] with the following settings: --bestn 1 -m 5 --header --nproc 100. We analyzed the identity of each consensus sequence and the average identity of all sample consensus sequences via local scripts.

Sampling assembly

We are interested in determining the minimum coverage depth required to assemble an accurate sequence. By exploring the relationships between sequencing depth, with step sizes of 2(2–10), 10(10–100) and 100(>100), and sequence accuracy, we aimed to identify the threshold at which sufficient coverage enables the reliable reconstruction of a high-quality sequence. This investigation involves analyzing sequencing data at varying depths and assessing the impact of coverage and identity on assembly quality. Through this study, we seek to elucidate the optimal coverage depth necessary for robust sequence assembly and to gain insights into the relationship between sequencing depth and sequence fidelity.

Genotyping by the *MtlA* gene

All the genome sequences of *V. parahaemolyticus* strains (9,700 sequences up to 2022) were downloaded from NCBI. Using the seven MLST genes (*dnaE*, *gyrB*, *recA*, *dtdS*, *pntA*, *pyrC*, and *tnaA*) and the *mtlA* gene from the reference genome RIMD 2210633 as reference sequences, we compared these loci against 9,700 genomes via BLASTN analysis. The result showed that only 2,504 genomes contained the full-length genes of all seven MLST genes, while 7,573 genomes contained the *mtlA* gene (Additional file 1), and the *mtlA* gene sequence

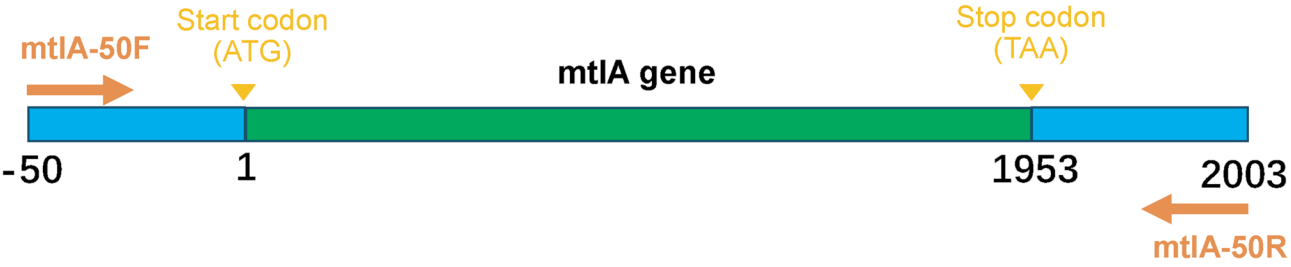

Fig. 1 The schematic diagram of *mtlA* primer

Table 1 The distribution of consensus sequence lengths and quality

Num of consensus sequence	Total consensus sequence bases	consensus sequence length N50	Q8	Q10	Q15	Q20	Q30
64,086	133,619,947	2,085	100%	100%	100%	99.4%	92.86%

was extracted from each genome. Sequence alignment of *mtlA* gene was performed using blastclust with 100% identity.

Results

Sequencing of the *MtlA* gene

To amplify the full length of *mtlA*, a pair of primers were designed with a forward primer (mtlA-50F: 5'-CGAAC CGTTTACGACAAC-3') 50 bp upstream and a reverse primer (mtlA-50R: 5'-AGGGGGTGACAGCTGACC-3') 50 bp downstream of the *mtlA* gene (Fig. 1). A 16 bp barcode was added to both forward and reverse primers to distinguish the 33 different samples. The 2,085 bp amplicons of each strain were subsequently mixed in equal amounts and sequenced with the Geneus G-seq500 platform.

A total of 199,658 long reads, with an average length of 17,306 bp and a median length of 22,424 bp, were produced. Owing to the nature of the ligation reactions employed to create templates, the resulting products are covalently closed circular molecules that contain two complementary sequences. It is possible to generate a consensus sequence from multiple reads of both the sense and antisense strands, all of which are derived from a single molecule. The polymerase reads are trimmed of adapters to yield subreads. An approximate draft consensus is called from subreads to provide a starting template for polishing, and a final consensus sequence is generated by iteratively polishing the template using subreads. Then, 0.13 Gb of consensus sequence bases were generated from the 1.81 Gb of long reads, with an average of 12 subreads for every consensus sequence.

The length of the consensus sequence is enriched in 2,085 bp, with average 1,942 consensus sequences produced for each of the 33 samples, and the quality is enriched in 35 (Table 1; Fig. 2). As the genomes of 33 *V. parahaemolyticus* strains had been sequenced by shotgun sequencing using the Illumina X-10 platform [18], we used the *mtlA* gene sequences of these genomes as reference to evaluate the accuracy of sequence result of G-seq500. The average accuracy of each consensus sequence ranged from 99.88 to 99.93% (Table 2). Then the final *mtlA* gene was generated by aligning consensus sequences from the same strain, which reached 100% accuracy when four consensus sequences were aligned together (Table 3; Fig. 2d and e).

Typing and tracing of *V. parahaemolyticus* strains with the *MtlA* gene

To facilitate the typing and tracing of *V. parahaemolyticus*, we have developed a naming rule based on *mtlA* clustering of 7,573 *mtlA* sequences from NCBI together with the 33 *mtlA* sequences of this study (Additional file 2). The largest cluster composed of 2,321 strains is designated as mtlA0001, followed by mtlA0002 containing 564 strains, and then mtlA0003,..., mtlA1371. There are 962 clusters each containing only one strain.

Nine *V. parahaemolyticus* strains isolated by Shanghai Customs were grouped into mtlA0001, represented by strain FDAARGOS_191 (isolated in India in 1996), RIMD_2210633 (isolated in Thailand in 1996), and VPD14 (isolated in China: Shanghai in 2012) (Table 4). Another three strains isolated by Shanghai Customs were grouped into mtlA0002, represented by strain BB22OP (isolated in Bangladesh in 2012).

The other 21 strains only presented 98.57-99.90% identity with known *V. parahaemolyticus* strains isolated from other regions (Table 4).

By comparing the ST typing and *mtlA* typing of 33 strains collected by Shanghai Customs, we observed a strong correlation between the two methods. All nine strains identified as ST3 were consistently typed as mtlA0001, and all three ST800 strains were typed as mtlA0229. Among the ten strains belonging to different ST types, each displayed a unique *mtlA* type. Additionally, the remaining 12 strains that could not be typed by ST (mostly due to the absence of a housekeeping gene)

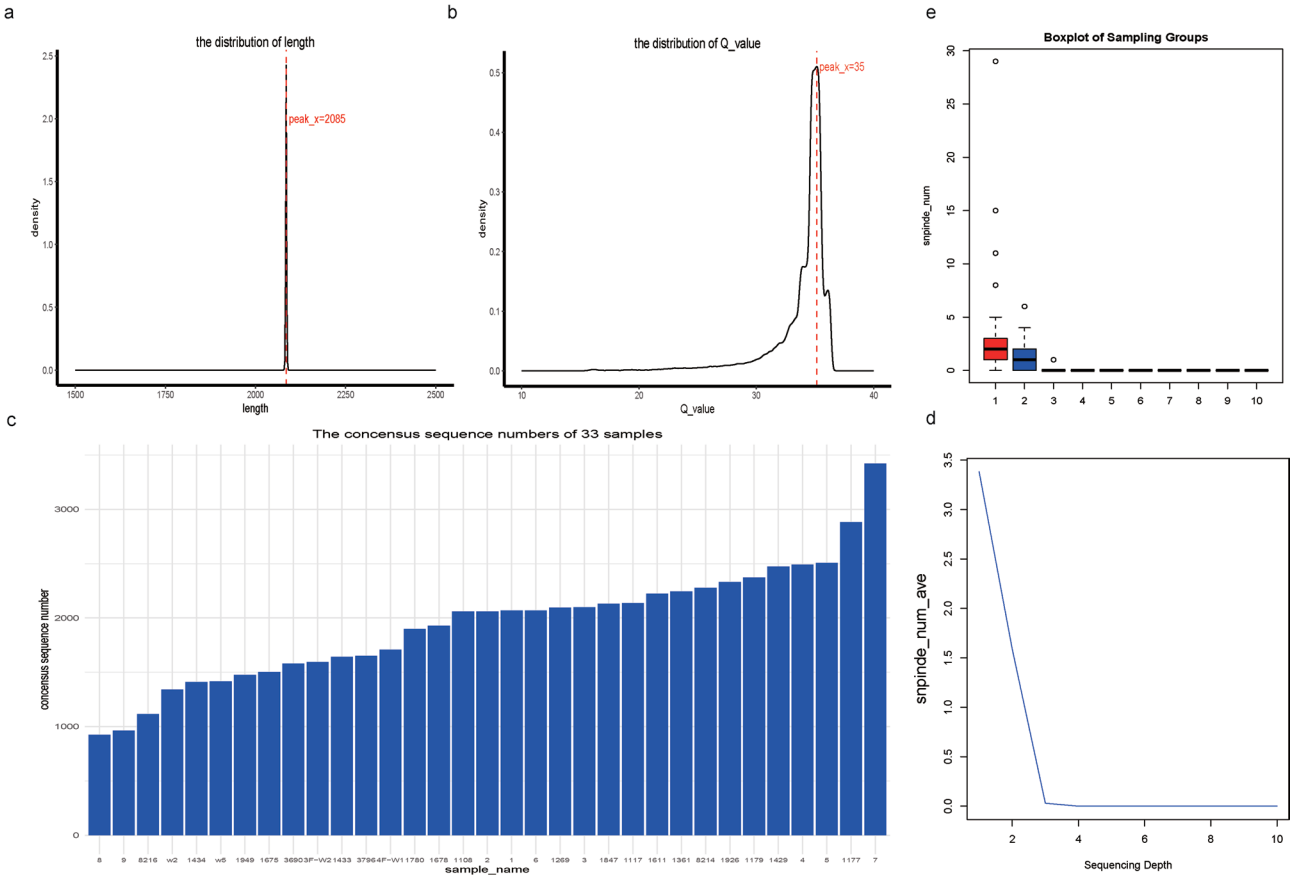

Fig. 2 The overview of the consensus sequences (**a**) The distribution of consensus sequence lengths. (**b**) The distribution of consensus sequence quality. (**c**) The distribution of consensus sequence number among 33 samples. (**d**) The relationship between average snp/indel number and sequencing depth for 33 samples. (**e**) The boxplot between snp/indel number and sequencing depth for 33 samples

still carried the *mtlA* gene and could therefore be classified using mtlA typing (Table 5).

To further evaluate the discriminatory power of the *mtlA* gene, we analyzed the number of mtlA types within each ST type, and vice versa, based on data in Additional file 1 (see Additional files 3 and 4). The most prevalent ST type, ST3, corresponded almost exclusively to mtlA0001 (99.73%). Similarly, the second most common type, ST631, matched exclusively with mtlA0004. For the remaining strains, the results from both typing methods remained highly consistent (Additional file 3). Conversely, mtlA0001, the most frequent *mtlA* type, corresponded almost exclusively to ST3 (96.58%). Notably, among the 5,066 strains lacking ST typing (66.9% of the total), mtlA typing still provided highly discriminatory results (Additional file 4).

Given that data from non-complete genomes may be of lower quality, we further focused our analysis on complete genome level data extracted from Additional file 1 (Table 6). We found that strains sharing the same *mtlA* type always belonged to the same ST type, whereas strains with the same ST type could have different *mtlA* types. This finding suggests that mtlA typing offers higher resolution than traditional ST typing. For example, three closely related strains—20140829008-1, 20140624012-1, and 20140722001-1—were all grouped as ST415 by ST typing, while *mtlA* typing was able to distinguish them into two types: mtlA0170 (two strains) and mtlA0063 (one strain).

Discussion

To verify the feasibility and convenience of directly using the full-length *mtlA* gene for typing *V. parahaemolyticus*, we sequenced the *mtlA* genes of 33 *V. parahaemolyticus* strains with known genomes [18], using the third-generation sequencing platform Geneus G-seq500.

The average sequence accuracy of the G-seq500 is 93.25%. By recognizing the adapter and barcode, on average, 12 reads were split from one long read and clustered into a consensus sequence. The accuracy of the consensus sequences range from 99.88 to 99.93%. However, when we generated the final *mtlA* gene by aligning consensus sequences from the same strain and compared it with the genome sequence, the accuracy reached 100% when at least four consensus sequences clustered together. With greater sequencing depth, the chances of errors in the

Table 2 The average identity of consensus sequence and source country for 33 samples

sample	ave_ identity	source country
SHP-1108	0.9988	The Philippines
SHP-1117	0.9989	Vietnam
SHP-1177	0.9988	The Philippines
SHP-1179	0.9989	Thailand
SHP-1269	0.9990	China
SHP-1361	0.9992	Japan
SHP-1429	0.9990	Thailand
SHP-1433	0.9991	The Philippines
SHP-1434	0.9988	Thailand
SHP-1611	0.9993	Thailand
SHP-1675	0.9990	Thailand
SHP-1678	0.9991	Thailand
SHP-1780	0.9990	Malaysia
SHP-1847	0.9993	Thailand
SHP-1926	0.9990	Thailand
SHP-1949	0.9992	Thailand
SHP-1	0.9990	China
SHP-2	0.9989	China
SHP-3690	0.9990	Thailand
SHP-3796	0.9992	Thailand
SHP-3 F-W2	0.9991	Panama
SHP-3	0.9993	China
SHP-4 F-W1	0.9988	Haryana
SHP-4	0.9990	China
SHP-5	0.9989	China
SHP-6	0.9991	China
SHP-7	0.9989	China
SHP-8214	0.9990	Malaysia
SHP-8216	0.9988	Indonesia
SHP-8	0.9992	China
SHP-9	0.9993	China
SHP-w2	0.9992	Panama
SHP-w5	0.9991	Cosmos

Table 3 The average sequencing errors at corresponding depth for 33 samples

depth	snp_inde_num_ave
1	3.382353
2	1.588235
3	0.029412
4	0
5	0
6	0
7	0
8	0
9	0
10	0

sequences decrease, leading to a more reliable and consistent outcome. This trend is attributed to the increased depth of sequencing data, which allows for better error correction and consensus calling. Since the G-seq500 platform produced 64,086 consensus sequences of *mtlA* in one sequencing chip, it can complete high-precision sequencing and typing of over 10,000 strains at one time.

In comparison, traditional MLST analysis defines sequence types (STs) for uploaded *V. parahaemolyticus* strains using internal fragment sequences of seven housekeeping genes [21]. Currently, there are 3,708 STs of *V. parahaemolyticus* in the PubMLST database (https://pubmlst.org/organisms/vibrio-parahaemolyticus). To provide higher resolution and discriminatory power, core genome MLST (cgMLST) analysis was developed in 2017 using 2,254 core genes [22]. Although these methods have advanced epidemiological surveillance, they require multiple amplification steps or whole-genome data, making them less efficient for high-throughput applications.

To simplify and streamline strain typing, we established a nomenclature system based on *mtlA* gene clustering, designating the largest cluster as mtlA0001, followed by others in descending size. Based on this naming system, we observed that among the 33 newly isolated *V.*

Table 4 MtlA cluster of 33 strains in this study

mtlA cluster	Strains	Matched strain in NCBI	identity	Isolation source	Isolation time	ST
mtlA0001	SHP-1780	VPD14	100%	China: Shanghai	2012	3
	SHP-1926	RIMD_2210633	100%	Tailand	1996	3
	SHP-1847 SHP-1678 SHP-1434 SHP-1433 SHP-1269 SHP-1117 SHP-1108	FDAARGOS_191	100%	India	1996	3
mtlA0003	SHP-8216 SHP-8214 SHP-1179	BB22OP	100%	Bangladesh	2012	88
mtlA0229	SHP-W2 SHP-4 F-W1 SHP-3 F-W2	FDAARGOS_662	99.44%	USA: KY	/	36
mtlA0913	SHP-9 SHP-8	S107-1	99.59%	Canada	/	2484
mtlA0008	SHP-3796 SHP-3690	FB-11	99.44%	China: Hangzhou	2017	165
mtlA0914	SHP-W5	S107-1	99.59%	Canada	/	2484
mtlA0052	SHP-1177	S107-1	99.59%	Canada	/	2484
mtlA0907	SHP-1361	S107-1	99.69%	Canada	/	2484
mtlA0023	SHP-1429	LH80	99.90%	China: Tangshan	2017	1800
mtlA0134	SHP-1611	DHO76	99.59%	China: Donggang	2018	1750
mtlA0126	SHP-1675	S107-1	98.98%	Canada	/	2484
mtlA0005	SHP-1949	20160303005-1	99.54%	China: Weifang	2016	150
mtlA0193	SHP-1	VP120	98.57%	China: Tianjin	2015	1
mtlA0908	SHP-2	DLM1805	99.39%	China: Dalian	1905	1802
mtlA0909	SHP-3	TJ-187	99.90%	China: Tianjin	2017	424
mtlA0910	SHP-4	TJA114	99.64%	China: Tianjin	2017	1743
mtlA0095	SHP-5	20140829008-1	99.74%	China: Jiangmen	2014	415
mtlA0911	SHP-6	S107-1	99.08%	Canada	/	2484
mtlA0912	SHP-7	FORC_023	99.69%	South Korea: Pusan	2014	2015

parahaemolyticus strains, 12 could be grouped into two major clusters (mtlA0001 and mtlA0002), indicating a worldwide prevalence. The other 21 strains had unique *mtlA* genes, with identities ranging from 98.57 to 99.90% when compared to known strains.

We further compared the typing performance of *mtlA*-based typing versus MLST using 7,573 *V. parahaemolyticus* genomes from NCBI. For strains with available ST designations, *mtlA* typing results were largely concordant. While MLST provided slightly higher resolution in some cases (e.g., among 2,312 *mtlA0001* strains, 2,233 strains belonged to ST3 type, and MLST distinguished another 38 strains into seven additional ST types) (Additional file 4), *mtlA* typing overall showed superior resolution, particularly for the 69 complete genomes and 33 isolates from Shanghai Customs (Tables 5 and 6). For strains lacking ST assignments, *mtlA* typing still demonstrated strong discriminatory power. These findings underscore the efficiency and higher resolution of *mtlA* typing, making it a compelling alternative to conventional MLST.

Beyond accuracy, *mtlA* typing offers significant advantages in cost, time, and scalability. While MLST requires amplification and sequencing of seven genes per strain, *mtlA* typing involves only one PCR step and minimal bioinformatics analysis. A single G-seq500 chip can process thousands of samples simultaneously, enabling high-throughput surveillance. Although third-generation sequencing infrastructure may be cost-prohibitive for some laboratories, *mtlA* typing becomes increasingly cost-effective at larger scales (e.g., > 100 strains), delivering 100% accuracy at high resolution.

The enhanced resolution and simplicity of *mtlA* typing have important implications for public health and food safety. It enables rapid discrimination of outbreak-related strains and detection of emerging virulent clones. For instance, distinguishing the pandemic *mtlA0001* lineage from rarer variants supports precise risk assessment and regulatory action. Integrating *mtlA* typing data with geographic metadata can further enhance source tracking and transmission modeling, aligning with the One Health

Table 5 ST and *MtlA* typing of 33 samples

sample	ST	mtlA cluster
SHP-1108	3	mtlA0001
SHP-1117	3	mtlA0001
SHP-1269	3	mtlA0001
SHP-1433	3	mtlA0001
SHP-1434	3	mtlA0001
SHP-1678	3	mtlA0001
SHP-1847	3	mtlA0001
SHP-1926	3	mtlA0001
SHP-1780	3	mtlA0001
SHP-1429	8	mtlA0023
SHP-1177	17	mtlA0052
SHP-3796	199	mtlA0008
SHP-2	533	mtlA0908
SHP-1611	635	mtlA0134
SHP-3 F-W2	800	mtlA0229
SHP-4 F-W1	800	mtlA0229
SHP-W2	800	mtlA0229
SHP-5	1820	mtlA0095
SHP-6	2094	mtlA0911
SHP-3	2723	mtlA0909
SHP-3690	2936	mtlA0008
SHP-4	2963	mtlA0910
SHP-1179	-	mtlA0003
SHP-1361	-	mtlA0907
SHP-1675	-	mtlA0126
SHP-1949	-	mtlA0005
SHP-1	-	mtlA0193
SHP-7	-	mtlA0912
SHP-8	-	mtlA0913
SHP-9	-	mtlA0913
SHP-8214	-	mtlA0003
SHP-8216	-	mtlA0003
SHP-W5	-	mtlA0914

framework by bridging human, animal, and environmental surveillance.

Future applications of *mtlA* typing include expanding its use to underrepresented regions such as Africa and South America to identify novel variants and refine global transmission maps. Investigating the role of *mtlA* in virulence and environmental adaptation may provide further insight into its evolutionary and epidemiological relevance. Additionally, adapting this approach to other *Vibrio* species (e.g., *V. cholerae*) could broaden its utility. To improve accessibility, development of user-friendly kits and cloud-based analytical platforms is recommended, enabling adoption by resource-limited laboratories. Longitudinal surveillance of *mtlA* allele shifts, especially under pressures such as climate change or aquaculture practices, could also inform proactive risk mitigation strategies.

Though our study highlights the economic and temporal advantages of *mtlA* typing over traditional MLST, several limitations warrant consideration. First, the need for advanced sequencing infrastructure may hinder adoption in low-resource settings. This challenge could be mitigated by centralized sequencing hubs and the development of cost-effective platforms. Second, existing regulatory frameworks favor conventional MLST, potentially delaying clinical implementation of *mtlA*-based methods. Finally, the harmonization of *mtlA* nomenclature with established ST and serotyping systems is urgently needed to ensure standardization and broader acceptance.

Conclusion

In summary, our study confirms the high sequencing accuracy of Geneus G-seq500, and the effectiveness of using the full-length *mtlA* gene for *V. parahaemolyticus* strain typing. The method's simplicity and efficiency surpass traditional MLST and cgMLST, offering a 100% accurate consensus when four or more consensus sequences clustered. Our *mtlA*-based naming system

Table 6 ST and *MtlA* typing of strains from NCBI with complete genome

Strain	ST	mtlA cluster
VP120	1	mtlA0066
TJ-20	1	mtlA0066
64	1	mtlA0066
RIMD_2210633	3	mtlA0001
VPD14	3	mtlA0001
FDAARGOS_191	3	mtlA0001
2013 V-1174	34	mtlA0015
10,329	36	mtlA0002
2010 V-1106	36	mtlA0002
FDAARGOS_51	36	mtlA0002
FDAARGOS_662	36	mtlA0002
2014 V-1125	36	mtlA0002
2014 V-1066	36	mtlA0002
2015AW-0174	36	mtlA0002
2013 V-1146	36	mtlA0002
BB22OP	88	mtlA0003
VP157	122	mtlA0024
PB1937	150	mtlA0207
20160303005-1	150	mtlA0431
2013 V-1244	162	mtlA0017
Vb0624	163	mtlA0020
FB-11	165	mtlA0182
FORC_022	233	mtlA0133
CHN25	395	mtlA0340
19-021-D1	413	mtlA0115
20-082A3	413	mtlA0115
20-082E4	413	mtlA0115
20140829008-1	415	mtlA0063
20140624012-1	415	mtlA0170
20140722001-1	415	mtlA0170
LH24	424	mtlA0102
TJ-187	424	mtlA0102
UCM-V493	471	mtlA0189
2013 V-1181	631	mtlA0004
2012AW-0224	679	mtlA0027
XMO116	722	mtlA0012
XMM117	722	mtlA1360
CDC_K4557	799	mtlA0101
160,807	809	mtlA0045
FORC_008	984	mtlA0029
FORC_018	984	mtlA0029
20151116002-3	1166	mtlA0088
20130629002S01	1166	mtlA0309
FORC_004	1628	mtlA0222
FORC_014	1629	mtlA0645
FORC_006	1630	mtlA0644
2013 V-1136	1716	mtlA0009
2012AW-0353	1727	mtlA0148
2012AW-0154	1727	mtlA0148
TJA114	1743	mtlA0120
DHO76	1750	mtlA0612
LVP66	1798	mtlA0318
DLM1799	1799	mtlA0084

Table 6 (continued)

Strain	ST	mtlA cluster
LH80	1800	mtlA0698
DLM1805	1805	mtlA0074
AM46865	1815	mtlA0255
MVP1	1913	mtlA0295
FORC_023	2015	mtlA0646
LVP1	2150	mtlA0194
AM43962	2290	mtlA0211
FORC_072	2482	mtlA0357
S107-1	2484	mtlA0081
D3112	2513	mtlA0267
FORC_071	2514	mtlA0647
AM51552	2517	mtlA0138
LVP2	2521	mtlA0046
MAVP-R	2556	mtlA0004
R13	-	mtlA0387
FDAARGOS_667	-	mtlA0643

simplifies strain identification and reveals both global prevalence and genetic diversity among strains. This approach enhances our ability to conduct epidemiological studies and investigate outbreaks, providing a valuable tool for public health and microbiological research.

Abbreviations

V. parahaemolyticus	Vibrio parahaemolyticus
MLST	Multi Locus Sequence Typing
ST	Sequence Type
SBS	Sequencing by synthesizing
PCR	Polymerase chain reaction
BLASR	Basic local alignment with successive refinement
cgMLST	Complete Genome Multilocus SequenceTyping

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-025-04078-1.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements
We would like to thank Xiaoyun Ma, Sunmei Nie and Leiping Zhang for invaluable help in sample collection.

Author contributions
Zilong Zhang, Huan Hu and Huajun Zheng designed the project. Rensi Zhou, Yingju Qi and Lei Zhou performed the experiment. Rensi Zhou, Yingju Qi, Danlei Liu, Yaozhong Zou, and Ting Xian participated in sample collection, bioinformatics analysis and figures illustration. Huan Hu 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.

Data availability
The consensus sequences of mtlA gene have been submitted to the GenBank Sequence Read Archive (accession number PRJNA1213999).

Declarations

Ethics approval and consent to participate
No applicable.

Consent for publication
No applicable.

Competing interests
Yingju Qi, Yaozhong Zou, and Huan Hu are affiliated with Geneus Technologies (Chengdu) Co.,Ltd, the manufacturer of the sequencing instrument (G-seq500) used in this research. However, this affiliation did not influence the design, execution, or interpretation of the study. All authors declare that they have no other competing interests that could affect the objectivity or integrity of the reported results.

Received: 31 December 2024 / Accepted: 26 May 2025
Published online: 31 May 2025

References

1. Morris JG Jr., Black RE. Cholera and other vibrioses in the united States. N Engl J Med. 1985;312(6):343–50. https://doi.org/10.1056/nejm198502073120604.
2. Balakrish G, Thandavarayan N, Sujit R, Basabjit KB, Yoshifumi D, David T. Global dissemination of Vibrio parahaemolyticus serotype O3:K6 and its serovariants. Clin Microbiol Rev. 2007;20(1). https://doi.org/10.1128/cmr.00025-06.
3. Han D, Yu F, Chen X, Zhang R, Li J. Challenges in Vibrio parahaemolyticus infections caused by the pandemic clone. Future Microbiol. 2019;14:437–50. https://doi.org/10.2217/fmb-2018-0308.
4. Zhou L, Liu D, Zhu Y, Zhang Z, Chen S, Zhao G, et al. Advance typing of Vibrio parahaemolyticus through the MtlA and Aer gene: A high-resolution, cost-effective approach. Heliyon. 2024;10(3):e25642. https://doi.org/10.1016/j.heliyon.2024.e25642.
5. He M, Lei T, Jiang F, Zhang J, Zeng H, Wang J, et al. Genetic diversity and population structure of Vibrio parahaemolyticus isolated from clinical and food sources. Front Microbiol. 2021;12:708795. https://doi.org/10.3389/fmicb.2021.708795.
6. Abanto M, Gavilan RG, Baker-Austin C, Gonzalez-Escalona N, Martinez-Urtaza J. Global expansion of Pacific Northwest Vibrio parahaemolyticus sequence type 36. Emerg Infect Dis. 2020;26(2):323–6. https://doi.org/10.3201/eid2602.190362.

7. van der Graaf-van Bloois L, Chen H, Wagenaar J, Zomer A. KaptiveDevelopment of databases for O- and K-antigen genotyping. Microb Genomics. 2023;9(5). https://doi.org/10.1099/mgen.0.001007.
8. Li J, Xue F, Yang Z, Zhang X, Zeng D, Chao G, et al. Vibrio parahaemolyticus strains of pandemic serotypes identified from clinical and environmental samples from Jiangsu, China. Front Microbiol. 2016;7:787. https://doi.org/10.3389/fmicb.2016.00787.
9. Aanensen DM, Spratt BG. The multilocus sequence typing network: mlst.net. Nucleic Acids Res. 2005;33(Web Server issue):W728-33; https://doi.org/10.1093/nar/gki415
10. Jesser KJ, Valdivia-Granda W, Jones JL, Noble RT. Clustering of Vibrio parahaemolyticus isolates using MLST and Whole-Genome phylogenetics and protein motif fingerprinting. Front Public Health. 2019;7:66. https://doi.org/10.3389/fpubh.2019.00066.
11. Zhou L, Liu D, Zhang Z, Zhu Y, Zheng H. mtlA, a novel typing gene offering enhanced resolution for pandemic Vibrio species. Microbiol Spectr. 2024;e0110524. https://doi.org/10.1128/spectrum.01105-24.
12. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998;95(6):3140–5. https://doi.org/10.1073/pnas.95.6.3140.
13. Boers SA, van der Reijden WA, Jansen R. High-throughput multilocus sequence typing: bringing molecular typing to the next level. PLoS ONE. 2012;7(7):e39630. https://doi.org/10.1371/journal.pone.0039630.
14. van Dijk E, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418–26. https://doi.org/10.1016/j.tig.2014.07.001.
15. Scarano C, Veneruso I, De Simone RR, Di Bonito G, Secondino A, D'Argenio V. The Third-Generation Sequencing Challenge: Novel Insights for the Omic Sciences. Biomolecules. 2024;14(5):568. https://doi.org/10.3390/biom14050568
16. Fuller CW, Kumar S, Porel M, Chien M, Bibillo A, Stranges PB, et al. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. Proc Natl Acad Sci U S A. 2016;113(19):5233–8. https://doi.org/10.1073/pnas.1601782113.
17. Stranges PB, Palla M, Kalachikov S, Nivala J, Dorwart M, Trans A, et al. Design and characterization of a nanopore-coupled polymerase for single-molecule DNA sequencing by synthesis on an electrode array. Proc Natl Acad Sci U S A. 2016;113(44):E6749–56. https://doi.org/10.1073/pnas.1608271113.
18. Liu D, Zhou L, Zhang Z, Zhang Y, Wang Z, Li S, et al. Epidemiological and genomic analysis of Vibrio parahaemolyticus isolated from imported travelers at the Port of Shanghai, China (2017–2019). BMC Microbiol. 2024;24(1):145. https://doi.org/10.1186/s12866-024-03303-7.
19. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3–new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115. https://doi.org/10.1093/nar/gks596.
20. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics. 2012;13:238. https://doi.org/10.1186/1471-2105-13-238.
21. Gonzalez-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus LA, DePaola A. Determination of molecular phylogenetics of Vibrio parahaemolyticus strains by multilocus sequence typing. J Bacteriol. 2008;190(8):2831–40. https://doi.org/10.1128/JB.01808-07.
22. Gonzalez-Escalona N, Jolley KA, Reed E, Martinez-Urtaza J. Defining a core genome multilocus sequence typing scheme for the global epidemiology of Vibrio parahaemolyticus. J Clin Microbiol. 2017;55(6):1682–97. https://doi.org/10.1128/JCM.00227-17.

Publisher's note

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