



Population structure and genetic diversity of *Vibrio parahaemolyticus* from a coastal area of China based on a multi-locus sequence typing (MLST) scheme

Yu Li · Hong-Qiu Yin · Jun Xia · Hong Luo · Ming-Yi Wang

Received: 1 January 2019 / Accepted: 26 February 2019 / Published online: 15 March 2019
© Springer Nature Switzerland AG 2019

Abstract The Gram-negative marine bacterium *Vibrio parahaemolyticus* has been identified as a major cause of bacterial food poisoning in China. Here, the population structure and genetic diversity of *V. parahaemolyticus* from Weihai, a coastal city in China, was studied by the multi-locus sequence typing (MLST) method. In this survey, we isolated 40 strains including environmental and clinical samples of patients with acute gastroenteritis or diarrhea; isolates from other countries were also included for comparison. DnaSP Version5, START V2, eBURST version3 and Mega 6 were used to analyse the data. We found that ST3 and ST332 were the most prevalent clones

and that they were closely associated with acute diarrhoeal diseases. These STs showed a low d_N/d_S ratio and significant linkage disequilibrium. All isolates were divided into four clonal complexes, six groups and nine singletons, showing a high degree of genetic diversity. 18 STs, mostly from environmental isolates, were recognised by the MLST analysis for the first time. In conclusion, ST3 and ST332 were the epidemic STs in the coastal area. ST332 might be a region-specific ST, which needs to be confirmed by further analysis. Thus, the long-term monitoring of *V. parahaemolyticus* plays an important role in preventing and controlling the transmission between environment and people in Weihai.

Hong Luo and Ming-Yi Wang have contributed equally to this work and should be considered co-corresponding authors.

Y. Li
Department of Clinic Laboratory Center, Yuhuangding Hospital Affiliated to Qingdao University,
Yantai 264000, Shandong, People's Republic of China

Y. Li · M.-Y. Wang (✉)
Department of Central Lab, Weihai Municipal Hospital Affiliated to Dalian Medical University,
Weihai 264200, Shandong, People's Republic of China
e-mail: wangmingyi1973@outlook.com

Y. Li · H. Luo (✉)
Department of Medical Laboratory, College of Laboratory Medicine, Dalian Medical University,
Dalian 116044, Liaoning, People's Republic of China
e-mail: dlmuluo@163.com

H.-Q. Yin
The Wendeng Osteopathic Hospital,
Weihai 264400, Shandong, People's Republic of China

J. Xia
Department of Medical Administration, Weihai Health and Family Planning Commission,
Weihai 264200, Shandong, People's Republic of China

Keywords Clonal complex · Genetic diversity · MLST · Population structure · *Vibrio parahaemolyticus*

Introduction

The Gram-negative, rod-shaped, halophilic bacterium *Vibrio parahaemolyticus* is usually found in marine and estuarine environments (Ellingsen et al. 2008). It is one of the main bacterial pathogens responsible for food-borne outbreaks and acute gastroenteritis. People are typically infected by eating raw or undercooked seafood (Pal and Das 2010). Some virulence factors may be involved in the pathogenesis, such as thermostable direct hemolysin (TDH), thermostable direct hemolysin-related hemolysin (TRH), type 3 secretion systems 1 2 (Makino et al. 2003).

Vibrio parahaemolyticus has been identified as the major cause of seafood-associated diarrhea and gastroenteritis since the pandemic O3:K6 serotype first emerged in Asia in 1996 (Okuda et al. 1997). Since then, the O3:K6 serotype and its relevant serovariants have been reported in many countries, including the United States, Mexico, Mozambique, Spain, Russia, China and elsewhere (Nair et al. 2007).

Currently, *V. parahaemolyticus* is the predominant cause of bacterial food poisoning in China (Wang et al. 2007), exerting a huge burden on national health care. Therefore, it is necessary to explore its molecular characteristics and epidemiology for effective prevention and control of *V. parahaemolyticus*. In recent years, a number of relevant studies on *V. parahaemolyticus* have been conducted in many Chinese cities, such as Beijing, Jiangsu, Zhejiang, Fujian, Guangdong and Taiwan (Li et al. 2014; Chao et al. 2009; Yu et al. 2011; Fan et al. 2013; Han et al. 2012). However Weihai, a typical coastal city with rapid development in north China, either lacks such data or has no data at all.

With the development of molecular biology, various approaches have been applied to characterise *V. parahaemolyticus* (Foxman et al. 2005), including pulsed-field gel electrophoresis, arbitrarily primed polymerase chain reaction, microarray-based comparative genome hybridization and multi-locus sequence

typing (MLST) (González-Escalona et al. 2008; Wong et al. 2007; Broberg et al. 2011).

In the present study, we selected the MLST scheme developed by González-Escalona et al. (2008). It has a better accuracy in tracking genetic differences, population genetics and phylogenetic relatedness of different strains (Banerjee et al. 2014; Ellis et al. 2012). MLST has the advantages of simple operation, time-saving, a public database and easy comparison between different laboratories. It also contributes to surveys of infection sources and transmission routes (Pérez-Losada et al. 2013).

Taking full advantage of the PubMLST database, we sought to elucidate the prevalence and genetic diversity of *V. parahaemolyticus* circulating in Weihai at the nucleotide and peptide levels. In summary, our data not only increased the richness of the database, but also identified difference from other provinces.

Materials and methods

Sampling of *V. parahaemolyticus* strains

A total of 40 *V. parahaemolyticus* isolates (including 25 clinical isolates and 15 environmental isolates) were collected for the MLST analysis. Among them, 28 strains were isolated from different regions in Weihai, China. The other 12 isolates were obtained to enrich the database's diversity, including three from Japan, one from the United States and eight from Thailand; these strains were collected and preserved by the General Hospital of the Jinan Military Region of the People's Liberation Army. All isolates were identified by VITEK2 system and 16S rRNA gene sequencing. Detailed information including isolation year, site and source are listed in Table 1. The isolates were stored in Luria–Bertani broth containing 30% glycerol at -80°C .

PCR amplification and sequencing

The PCR amplification protocols and relevant primer information are available at <http://pubmlst.org/vparahaemolyticus/info/protocol.shtml>. All isolates were PCR amplified and sequencing conducted by Thermo Fisher Scientific using M13 primers.

Table 1 Sources of the isolates analysed

Lab ref no.	Isolation year	Place of isolation	Recovered from
Cli-JP1	2007	Japan	Patient, gastro-enteritis
Cli-JP2	2008	Japan	Patient, gastro-enteritis
Cli-JP3	1951	Japan	Patient, gastro-enteritis
Cli-US1	1982	United States, Maryland	Patient, gastro-enteritis
Cli-WH1	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH2	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH3	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH4	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH5	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH6	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH7	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH8	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH9	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH10	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH11	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH12	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH13	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH14	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH15	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH16	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH17	2014	China, Weihai	Patient, gastro-enteritis
Cli-WH18	2015	China, Weihai	Patient, gastro-enteritis
Cli-WH19	2014	China, Weihai	Patient, gastro-enteritis
Cli-WH20	2014	China, Weihai	Patient, gastro-enteritis
Cli-WH21	2014	China, Weihai	Patient, gastro-enteritis
Env-WH1	2010	China, Weihai	Ocean sediments in the near-shore areas
Env-WH2	2010	China, Weihai	Ocean sediments in the near-shore areas
Env-WH3	2010	China, Weihai	Ocean sediments in the near-shore areas
Env-WH4	2011	China, Weihai	Surface water obtained from the shoreline
Env-WH5	2011	China, Weihai	Surface water obtained from the shoreline
Env-WH6	2010	China, Weihai	Shrimp
Env-WH7	2010	China, Weihai	Shrimp
Env-TH1	2009	Thailand	Shrimp
Env-TH2	2006	Thailand	Shrimp
Env-TH3	2009	Thailand	Shrimp
Env-TH4	2008	Thailand	Shrimp
Env-TH5	2008	Thailand	Shrimp
Env-TH6	2007	Thailand	Shrimp
Env-TH7	2010	Thailand	Shrimp
Env-TH8	2010	Thailand	Shrimp

The strains from Japan, United States and Thailand were preserved in the General Hospital of Jinan Military Region of PLA. The multiple clinical isolations from Weihai in the same year are from different patients and there was no link such as members of the same family

Diversity of pSTs

Seven nucleotide sequences were queried against the *V. parahaemolyticus* PubMLST database to determine the allelic profile and ST of each isolate. When a new allelic profile or ST number was discovered, the allelic nucleotide sequences were emailed to <http://www.narjol.gonzalez-escalona@fda.hhs.gov>, and then they were assigned a novel numerical identifier for allelic profile or ST.

As described by others (Urmersbach et al. 2014), each locus peptide sequence was queried against PubMLST database to be assigned to an allelic profile and the seven different allelic profiles led to a pST at the peptide level. The loci were labeled with the prefix ‘p_’, and scheme was called AA-MLST, which allows an analysis at the phenotypic level.

Isolate descriptions (such as place, source and isolation year), allelic profiles and sequencing traces (unique alleles) have been uploaded and made available in the PubMLST database.

Locus statistics

DnaSP Version5 (Librado and Rozas 2009) was implemented to calculate the number of alleles, number of polymorphic sites and nucleotide diversity for each locus. START V2 (Jolley et al. 2001) was implemented to calculate d_N/d_S through the Nei and Gojobori method. The standardised index of association (I_A^S) was also calculated by START V2.

Assignment to CCs

The CCs were assigned by eBURST version3 using STs as above-described. As reported previously (Turner et al. 2013), singletons were defined when one ST differed at two or more alleles from every other ST (Feil et al. 2004); a single locus variant (SLV) was named when two STs differed by only one single locus; and a double locus variant (DLV) was defined when two STs differed by two loci (González-Escalona et al. 2008). Likewise, pSTs were assigned to CCs by goeBURST algorithm as previously described (Urmersbach et al. 2014).

The “population snapshot” analysis was also implemented with the group definition zero of seven loci shared and to create a full-(AA)MLST, where all (p)STs were connected (Francisco et al. 2009).

Phylogenetic analysis

To estimate the genetic distance of the analysed isolates, ME trees for the in-frame concatenated sequences (*dnaE-gyrB-recA-dtdS-pntA-pyrC-tmA*) of each (p)ST were constructed with Mega 6 software using the Kimura two-parameter model, and the number of bootstrap replications was 1000 (Han et al. 2014).

Detection of virulent determinants and antimicrobial susceptibility tests

PCR for the detection of pathogenicity-related genes (TDH and TRH) was performed on all strains independently, using the primers described by Khouadja et al. (2014). Antimicrobial susceptibility was tested using the VITEK2 system.

Comparison of clinical isolates between Weihai and other regions in China

Using the *V. parahaemolyticus* PubMLST database, we searched the clinical isolates from China, and some isolates with incomplete information were excluded. Subsequently, we attempted to find the similarities and differences in clinical isolates between Weihai and other regions in China.

Results and discussion

A total of 40 isolates were collected in the present study, including 25 clinical isolates and 15 environmental isolates. Among them, 28 were isolated from different regions in Weihai, China. The other 12 isolates were isolated in other countries, including three from Japan, one from the United States and eight from Thailand (Table 1).

Diversity of peptide STs (pSTs)

Table 2 summarises data on allelic profiles and pSTs of the 40 isolates at the nucleotide and peptide levels. 23 STs were identified by MLST analysis. Among these STs, 18 STs (20 isolates) and one allele (of *tmaA*) were identified for the first time, including seven (29.2%) clinical isolates and 13 (54.2%) environmental isolates. New STs were mostly identified based on

Table 2 The allelic profiles and (p)STs of the analysed isolates

Isolates	MLST										AA-MLST										pST
	dnaE	gyrB	recA	dtdS	pntA	pyrC	tnaA	ST	p_dnaE	p_gyrB	p_recA	p_dtdS	p_pntA	p_pyrC	p_tnaA						
Cli-JP1	3	104	31	306	50	11	38	1540	2	1	1	1	1	1	1	2					
Cli-JP2	5	52	123	13	17	25	10	1547	2	1	1	1	5	1	1	125					
Cli-JP3	5	52	27	13	17	25	10	1	2	1	1	5	1	1	1	125					
Cli-US1	3	104	61	306	50	11	38	1548	2	1	1	1	1	1	1	2					
Cli-WH1	3	4	19	4	29	4	22	3	2	1	1	1	1	1	1	2					
Cli-WH2	14	30	141	78	4	37	13	332	1	1	1	1	1	2	1	3					
Cli-WH4	11	48	141	48	26	48	26	1533	2	1	1	1	1	1	1	2					
Cli-WH11	14	30	141	78	29	37	13	1532	1	1	1	1	1	2	1	3					
Cli-WH21	14	30	27	78	4	37	13	1534	1	1	1	1	1	2	1	3					
Env-WH1	34	29	256	33	25	26	17	1535	1	1	1	1	1	1	1	1					
Env-WH2	175	98	155	323	149	212	189	1536	1	10	1	14	1	1	1	237					
Env-WH3	175	329	155	113	79	58	26	1537	1	4	1	1	1	5	1	10					
Env-WH4	209	98	35	1	4	184	189	1538	1	10	1	1	1	1	1	239					
Env-WH5	175	98	321	1	17	58	26	1539	1	10	1	1	5	5	1	240					
Env-WH6	186	252	181	29	4	18	132	479	2	1	1	1	1	1	1	2					
Env-WH7	3	116	210	288	28	226	159	1541	2	1	1	1	1	47	1	177					
Env-TH1	180	452	76	323	36	150	228	1542	1	11	1	14	2	1	1	241					
Env-TH2	60	197	31	18	106	150	26	1256	1	1	1	1	2	1	1	9					
Env-TH3	51	117	35	19	26	58	189	1543	2	1	1	1	1	5	1	238					
Env-TH4	41	242	321	29	36	184	26	1545	1	1	1	1	2	1	1	9					
Env-TH6	166	155	123	126	132	3	26	1546	1	1	1	1	1	2	1	3					
Env-TH7	169	226	3	50	23	8	12	1549	1	1	1	1	1	1	1	1					
Env-TH8	42	84	31	46	63	49	246	1550	2	1	1	1	1	3	1	7					

Novel allele type and novel (p)ST in bold

Cli-WH1, Cli-WH3, Cli-WH5, Cli-WH6, Cli-WH8, Cli-WH9, Cli-WH10, Cli-WH12, Cli-WH16 and Cli-WH18 shows the same ST and pST
 Cli-WH2, Cli-WH7, Cli-WH13, Cli-WH14, Cli-WH15, Cli-WH19 and Cli-WH20 shows the same ST and pST
 Cli-WH4 and Cli-WH17 shows the same ST and pST
 Env-TH4 and Env-TH5 shows the same ST and pST

Table 3 Diversity of the seven loci in the 40 *V. parahaemolyticus* isolates

Locus	Fragment size		No of alleles		No (%) of polymorphic (variable) sites		Nucleotide diversity (π)	d_N/d_S
	ST	pST	ST	pST	ST	pST		
<i>recA</i>	729 bp	242aa	14	1	134 (18.38%)	4 (1.65%)	0.03560	0.0130
<i>gyrB</i>	592 bp	197aa	17	4	159 (26.86%)	24 (12.18%)	0.04017	0.0385
<i>dnaE</i>	557 bp	185aa	15	2	123 (22.08%)	7 (3.78%)	0.04107	0.0152
<i>dtdS</i>	458 bp	152aa	16	2	91 (19.87%)	2 (1.32%)	0.03777	0.0042
<i>pntA</i>	430 bp	143aa	14	3	92 (21.40%)	9 (6.29%)	0.02138	0.0226
<i>pyrC</i>	493 bp	163aa	15	5	83 (16.84%)	10 (6.13%)	0.01924	0.0369
<i>tmaA</i>	423 bp	141aa	12	1	77 (18.20%)	4 (2.84%)	0.01909	0.0073

the combination of alleles. Most STs were represented by a single isolate, while just four STs ranged from two to 10 isolates. The most frequently identified STs were ST3 (25%) and ST332 (17.5%) in clinical isolates, whereas no predominant ST was observed from environmental isolates.

After the in-frame nucleotide sequences were translated into peptide sequences, 13 distinct pSTs were acquired with a frequency ranging from 2.5 to 37.5%. This allows an analysis at the phenotypic level, and can be considered when non-synonymous substitutions of nucleotides lead to a different amino acid. Among these pSTs, there were five (38.5%) new pSTs, all belongs to environmental isolates. pST2 (37.5%, 15/40) and pST3 (25%, 10/40) were the two predominant pSTs. As reported previously (Han D et al. 2014), one particular pST can contain a large amount of nucleotide STs. In this study, we found that pST2 contained five different STs (ST3, ST479, ST1533, ST1540 and ST1547) and pST3 had four STs (ST332, ST1532, ST1534 and ST1546).

Locus statistics

Table 2 summarises the data on diversity of the seven loci, including fragment size, number of alleles, number of polymorphic sites, nucleotide diversity (π) and d_N/d_S ratio. The number of alleles observed for each ST ranged from 12 (*tmaA*) to 17 (*gyrB*), and the most frequently detected alleles at each locus were *recA* 19 and 141 (10), *gyrB* 4 (10), *dnaE* 3 (13), *dtdS* 4 (10), *pntA* 29 (11), *pyrC* 4 (10) and *tmaA* 22 (10), respectively. In addition, the number of alleles for each pST ranged from 1 (*tmaA/recA*) to 5 (*pyrC*). One allele was dominant for most alleles ($\sim 80\%$), with the exception of p_*dnaE* (accounting for 50.0%) and

p_*pyrC* (accounting for 62.5%). The number of polymorphic nucleotide sites ranged from 83 (*pyrC*) to 159 (*gyrB*), while the number of polymorphic peptide sites ranged from 2 (*dtdS*) to 24 (*gyrB*). The *gyrB* allele possessed the largest number of polymorphic sites at both the nucleotide and peptide levels.

The nucleotide diversity (π) ranged from 0.01909 (*tmaA*) to 0.04107 (*dnaE*). The d_N/d_S ratio was close to zero for all the seven loci (Table 3), suggesting that nonsynonymous sites were evolving more slowly than synonymous sites, and the population was mainly affected by negative selection (purifying selection) during the evolutionary process (Jolley et al. 2001; González-Escalona et al. 2008; Han et al. 2015). This finding could explain the phenomenon that the number of alleles was reduced and the loci were primarily dominated by a single allele when analysed at the peptide level (Pérez-Losada et al. 2006; Turner et al. 2013). For example, 12 different *tmaA* alleles at the nucleotide level were transformed into one p_*tmaA* at the peptide level, which was reflected by a low d_N/d_S value of 0.0073. In this study, the calculated I_A^S value was 0.762 ($P < 0.01$), indicating that these alleles were in linkage disequilibrium or not randomly distributed in the *V. parahaemolyticus* population (Lüdeke et al. 2015). To some extent, our findings supported the hypothesis that the population structure of *V. parahaemolyticus* followed the epidemic model of clonal expansion (Yu et al. 2011).

Assignment to clonal complexes (CCs)

The global ST database was referenced to query our profiles and the results showed that the 23 STs belonged to four CCs, six groups and nine singletons (ST1536, ST1537, ST1539, ST1541, ST1542,

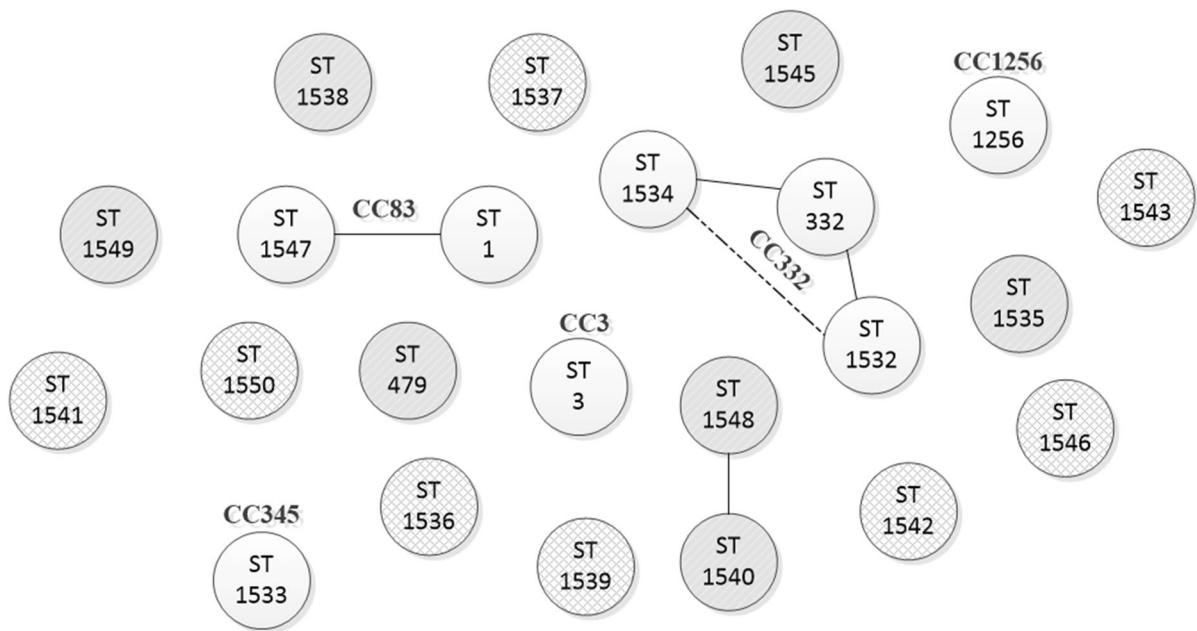


Fig. 1 The CCs assigned by eBURST version 3 using STs

ST1543, ST1545, ST1546 and ST1550) (Fig. 1). The CCs included CC3 (ST3), CC83 (ST1 and ST1547), CC332 (ST332, ST1532 and ST1534), CC345 (ST1533) and CC1256 (ST1256). In our study, CC3 was the most populated CC (10 isolates), followed by CC332. Among these CCs, this analysis revealed four SLVs (ST332 and 1532; ST332 and 1534; ST1540 and 1548; ST1 and 1547) and one DLV (ST1532 and ST1534). The nine singletons demonstrated that most of the STs identified in our study shared less than six alleles.

However, neither the relationships between the nine singletons themselves nor the relationships between the singletons could be deduced here. This suggests that goeBURST had limited application for identifying related isolates at the nucleotide level (Han et al. 2014).

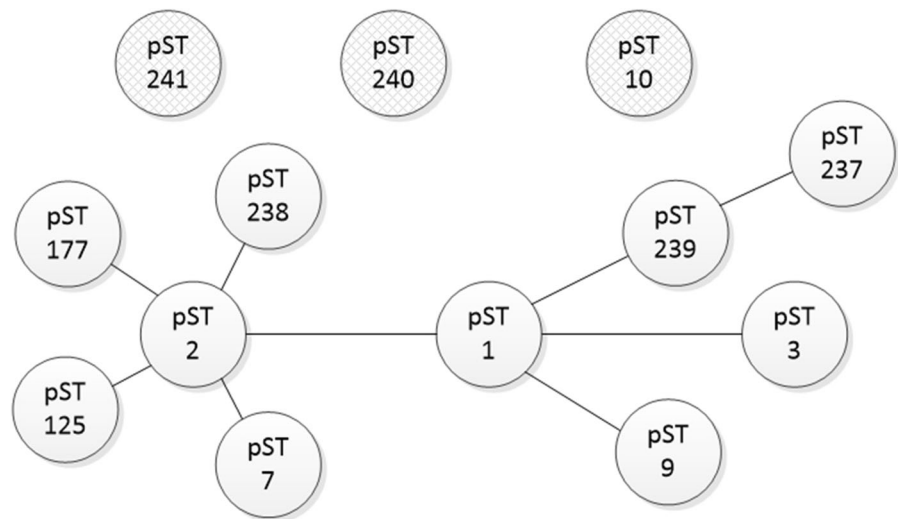
To challenge this, a “population snapshot” analysis was implemented at the peptide level (Fig. 2). The result showed that all pSTs (except for pST10, pST241 and pST240) led to a single complex founded by pST2 (founder) and pST1 (subgroup founder). Therefore, the relationship among the isolates became more reliable when analysed at the peptide level compared with the nucleotide level.

Phylogenetic analysis

A minimum evolution (ME) tree was constructed using the concatenated nucleotide sequences (Fig. 3). From the ME tree, we can see that the isolates were grouped into two clusters. Cluster II only contained four environmental isolates from Weihai. Cluster I consisted of most isolates originating from different areas and sources, showing a high degree of genetic diversity. Isolates belonging to the same CCs (CC332, CC345, CC83, CC3) or groups by eBURST were also close to each other in the ME tree. The nine singletons were spread over the branches. From the ME tree, Env-WH4 and Env-WH5 clustered together (100% bootstrap support), while the eBURST analysis identified them both as singletons. There was also a high similarity between Cli-JP1 and Cli-US1 strains, which were collected from different countries. Generally speaking, results between the eBURST and ME tree analyses were consistent. However, the ME tree provided an improved phylogenetic analysis and elucidated some genetic relationships that were not observed by eBURST.

The ME tree was also implemented based on the peptide sequences. Data showed that 13 pSTs were

Fig. 2 The CCs assigned by eBURST version 3 using pSTs



grouped into one major cluster, except for Env-WH3 (Fig. 4). The evolutionary distance was 0.017930.

Detection of virulent determinants and antimicrobial susceptibility tests

20% (8/40) *V. parahaemolyticus* strains harboured both the TDH and TRH genes, including 7 clinical isolates and 1 environmental isolate. 52.5% (21/40) strains only contained the TDH gene, while 27.5% (11/40) strains harboured neither of these genes. The antimicrobial susceptibility tests showed all strains were resistant to ampicillin and sensitive to ampicillin/sulbactam, piperacillin/tazobactam, cephalosporins and quinolone antibiotics.

Comparison of clinical isolates between Weihai and other regions in China

Clinical isolates from China were collected via the MLST database (<http://pubmlst.org/vparahaemolyticus/>). Next, we screened out them for complete sequence information. Finally, 248 strains were retained and queried against 157 STs. These isolates were collected from Taiwan, Hongkong, Zhejiang, Shanghai, Beijing and elsewhere. Among them, ST3 was predominant and found in 53 isolates (21.4%), followed by ST1464 (8, 3.2%) and ST69 (5, 2.0%). In the Weihai area, we collected a total of 21

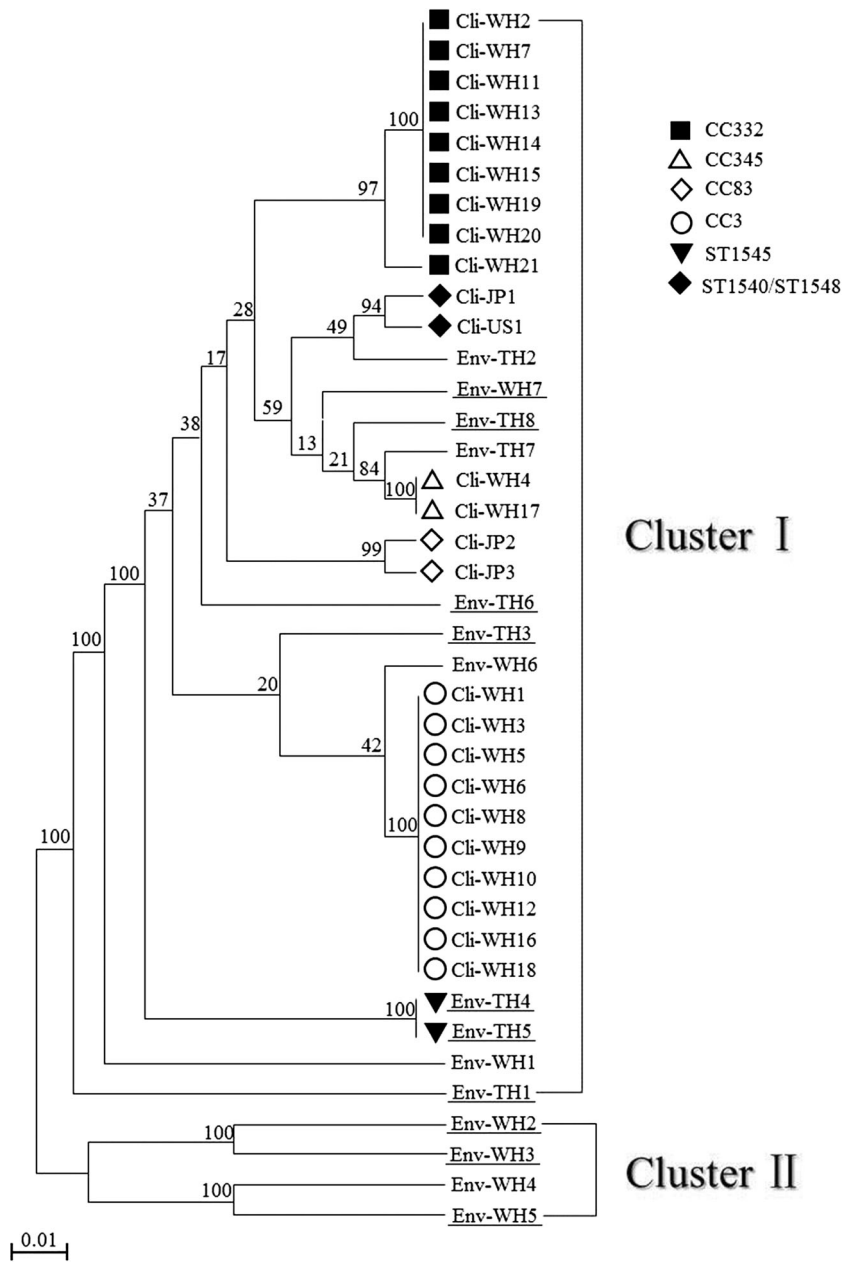
clinical isolates. Similarly, the largest number of isolates (10, 47.6%) was of ST3, followed by ST332 (7, 33.3%).

Conclusions

Weihai is a coastal city located in the east of Shandong Peninsula, facing to the Yellow Sea from three sides. Because of its geographic position and seafood-eating habits, food-borne outbreaks and acute gastroenteritis have often occurred via eating raw or undercooked seafood polluted by *V. parahaemolyticus*. Therefore, it's a necessity to understand the epidemiology and genetic diversity of *V. parahaemolyticus* strains in this region.

An MLST analysis based on a set of 10 loci, showing a high level of nucleotide diversity and a detailed analysis of recombination, has been reported (Yan et al. 2011). In the present study, we employed a protocol reported by González-Escalona et al. (2008), which has a public database and facilitates comparisons between different laboratories (Lüdeke et al. 2015). Moreover, we identified 23 STs and found 18 novel STs, which substantially contributed to the diversity and richness of the MLST database. The environmental isolates were entirely new found STs, mostly due to new combinations of alleles. Previous studies have confirmed that *V. parahaemolyticus*

Fig. 3 The ME tree constructed from analysis of concatenated nucleotide sequences

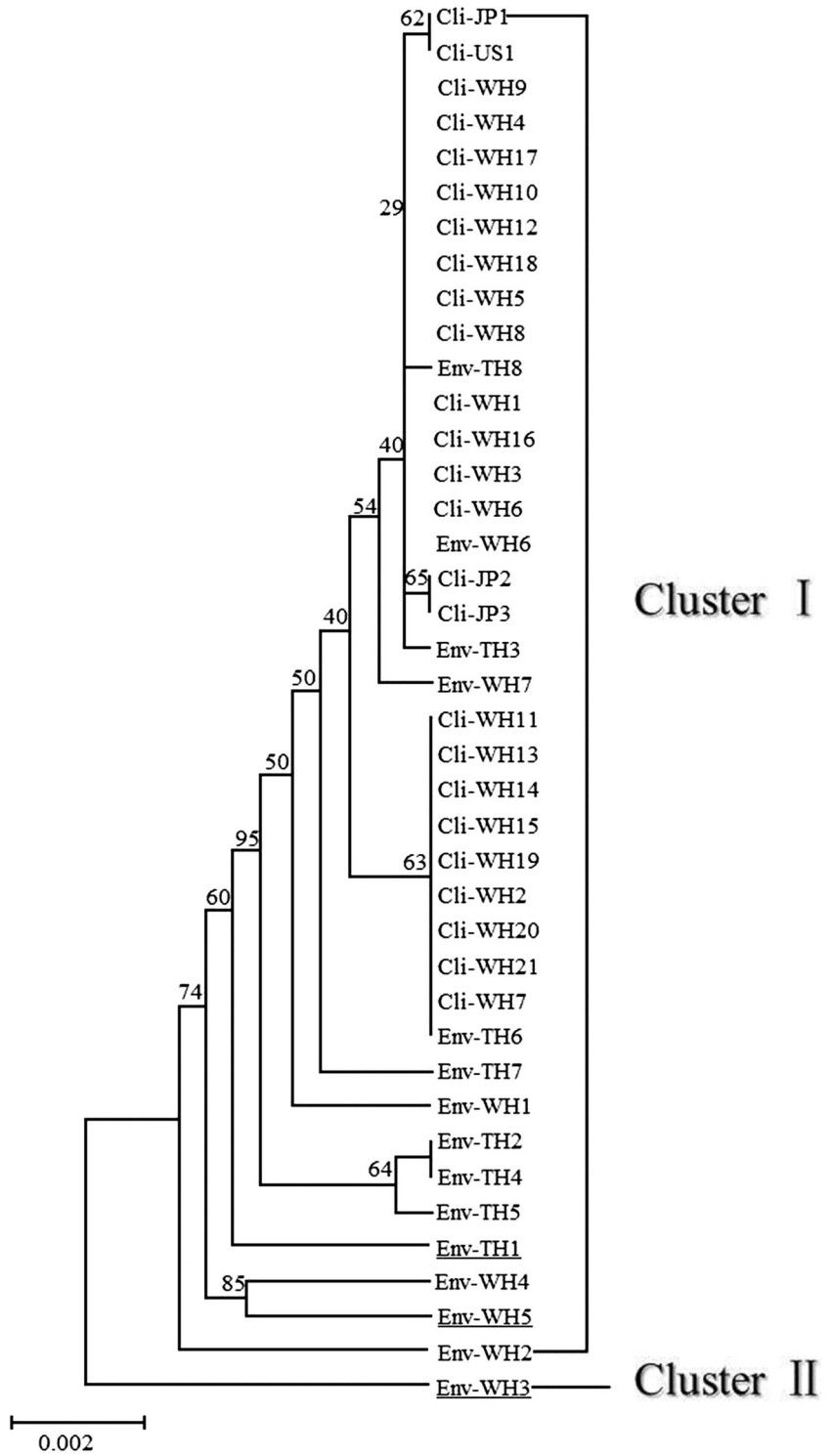


populations in a restricted locality can be as diverse as the entire worldwide pubMLST database (Johnson et al. 2009). Our results supported this observation. However, the diversity was decreased when an analysis based was carried out on the peptide sequences, as only 13 pSTs were generated. In addition, pST2 and pST3 were the most prevalent

and most were found in clinical isolates, which is in agreement with previous reports (Urmersbach et al. 2014).

The eBURST and ME tree analyses were largely consistent with each other. The isolates belonging to four CCs and groups were also clustered together in the ME tree. Some studies have reported that the

Fig. 4 The ME tree constructed from analysis of concatenated peptide sequences



isolates belonging to the same CCs may be divided into different clusters (Ansaruzzaman et al. 2005). Such discrepancy could be attributed to different analytical approaches. The goeBRUST algorithm is an allelic profile-based approach, and only SLVs are assigned to a single CC or group. Therefore, it could quickly find CCs and their ancestral STs. However, the ME tree is established based on sequences, and sequences with fewer differences are clustered together, providing an improved resolution for revealing evolutionary relationships regardless of CCs, groups or singletons. Therefore, a more reliable conclusion could be effectively acquired by combination of these two methods.

We also compared the isolates from Weihai with those from other regions of China. As reported globally, ST3 is largely responsible for *V. parahaemolyticus* outbreaks (Lüdeke et al. 2015) and is the main causative agent for the pandemic CC in most regions of China, including Weihai. 74% and 94% of *V. parahaemolyticus* infections were caused by ST3 in Beijing (Fan et al. 2013) and Zhejiang (Han et al. 2012), respectively, which are higher than that found in Weihai (47.6%), this perhaps reflects the small sample size studied here. In Weihai, the second most frequent ST was ST332. However, Han et al. have discovered that the second most frequent allelic profile includes ST305, ST8, ST8, ST301, ST670, ST432, ST91, ST1034, ST69 and ST265 in Guangdong, Jiangsu, Taiwan, Shanghai, Zhejiang, Beijing, HongKong, Fujian and Shenzhen (2014), respectively. Based on the related literature, it seems that ST332 may be restricted to the eastern coast of Shandong. But, as Han et al. (2014) reported, the *V. parahaemolyticus* database does not contain all the discovered isolates and it only contains reported from voluntary uploading of laboratory data. This may have some influence on acquisition of truly representative data. In Weihai, we, for the first time, found four isolates (ST1532, ST1533 and ST1534), which were associated with *V. parahaemolyticus* gastroenteritis. Interestingly, Turner et al. have discovered that some environmental isolates belong to ST3 by analysing a set of strains from the United States (Turner et al. 2013). Likewise, we also obtained two environmental isolates belonging to ST1 and ST 479. They were also found from clinical sources in the public database. This suggests that pathogenic strains may emerge

from the environmental condition, showing a higher potential virulence to people.

In conclusion, based on MLST analysis, ST3 and ST332 were the most epidemic clones, which accounted for a large proportion of the pathogenic strains in the coastal city of Weihai. ST332 might be a region-specific ST, which needs to be confirmed by further analysis. Some newly identified STs indicated that the genes underwent recombination frequently (Yan et al. 2011). However, the study still needs some improvement. For example, we only collected 40 isolates and this small sample size limits to some extent the conclusions that can be drawn from the findings. The MLST technique employed is also unsuited for detecting significant differences in genome inversions, transposons and plasmids. Application of pulsed-field gel electrophoresis profiling may remedy this, so further studies are needed.

Acknowledgements This work was supported by Primary Research & Development Plan of Shandong Province (No. 2016GSF121036) and the National Science and Technology Major Project of China (2013ZX10004217).

Authors' contribution Hong Luo and Ming-Yi Wang designed experiments; and Hong-Qiu Yin carried out experiments; Yu Li and Jun Xia analysed sequencing data and experimental results; Yu Li wrote the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest concerning this article.

Ethical statement This article does not contain any studies with animals performed by any of the authors.

References

- Ansaruzzaman M, Lucas M, Deen JL, Bhuiyan NA, Wang XY, Safa A, Sultana M, Chowdhury A, Nair GB, Sack DA, von Seidlein L, Puri MK, Ali M, Chaignat CL, Clemens JD, Barreto A (2005) Pandemic serovars (O3:K6 and O4:K68) of *Vibrio parahaemolyticus* associated with diarrhea in Mozambique: spread of the pandemic into the African continent. *J Clin Microbiol* 43(6):2559–2562
- Banerjee SK, Kearney AK, Nadon CA, Peterson CL, Tyler K, Bakouche L, Clark CG, Hoang L, Gilmour MW, Farber JM (2014) Phenotypic and genotypic characterization of Canadian clinical isolates of *Vibrio parahaemolyticus* collected from 2000 to 2009. *J Clin Microbiol* 52(4):1081–1088

- Broberg CA, Calder TJ, Orth K (2011) *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes Infect* 13(12–13):992–1001
- Chao G, Jiao X, Zhou X, Yang Z, Huang J, Pan Z, Zhou L, Qian X (2009) Serodiversity, pandemic O3:K6 clone, molecular typing, and antibiotic susceptibility of foodborne and clinical *Vibrio parahaemolyticus* isolates in Jiangsu, China. *Foodborne Pathog Dis* 6(8):1021–1028
- Ellingsen AB, Jorgensen H, Wagley S, Monshaugen M, Rørvik LM (2008) Genetic diversity among Norwegian *Vibrio parahaemolyticus*. *J Appl Microbiol* 105(6):2195–2202
- Ellis CN, Schuster BM, Striplin MJ, Jones SH, Whistler CA, Cooper VS (2012) Influence of seasonality on the genetic diversity of *Vibrio parahaemolyticus* in New Hampshire shellfish waters as determined by multilocus sequence analysis. *Appl Environ Microbiol* 78(10):3778–3782
- Fan YY, Zhu M, Shang XR, Wang M, Huang YF, Gu HT (2013) Virulence characteristics and multi-locus sequence type of *Vibrio parahaemolyticus* isolated from clinic (in Chinese). *Chin J Epidemiol* 36:548–552
- Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multi locus sequence typing data. *J Bacteriol* 186(5):1518–1530
- Foxman B, Zhang L, Koopman JS, Manning SD, Marrs CF (2005) Choosing an appropriate bacterial typing technique for epidemiologic studies. *Epidemiol Perspect Innov* 2(2):10
- Francisco AP, Bugalho M, Ramirez M, Carriço JA (2009) Global optimal eBURST analysis of multi locus typing data using a graphic matroid approach. *BMC Bioinform* 10:152
- González-Escalona N, Martínez-Urtaza J, Romero J, Espejo RT, Jaykus LA, DePaola A (2008) Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *J Bacteriol* 190(8):2831–2840
- Han DS, Chen X, Li ZJ, Zheng SF, Yu F, Chen Y (2012) Prevalence and multi-locus sequence typing of *Vibrio parahaemolyticus* associated with acute diarrhea in Zhejiang (in Chinese). *J Environ Health* 29:917–919
- Han D, Tang H, Lu J, Wang G, Zhou L, Min L, Han C (2014) Population structure of clinical *Vibrio parahaemolyticus* from 17 coastal countries, determined through multilocus sequence analysis. *PLoS ONE* 9(9):e107371
- Han D, Tang H, Ren C, Wang G, Zhou L, Han C (2015) Prevalence and genetic diversity of clinical *Vibrio parahaemolyticus* isolates from China, revealed by multilocus sequence typing scheme. *Front Microbiol* 9(6):291
- Johnson CN, Flowers AR, Young VC, Gonzalez-Escalona N, DePaola A, Noriega NF 3rd, Grimes DJ (2009) Genetic relatedness among *tdh* + and *trh* + *Vibrio parahaemolyticus* cultured from Gulf of Mexico oysters (*Crassostrea virginica*) and surrounding water and sediment. *Microb Ecol* 57(3):437–443
- Jolley KA, Feil EJ, Chan MS, Maiden MC (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* 17(12):1230–1231
- Khouadja S, Suffredini E, Baccouche B, Croci L, Bakhrouf A (2014) Occurrence of virulence genes among *Vibrio cholerae* and *Vibrio parahaemolyticus* strains from treated wastewaters. *Environ Monit Assess* 186(10):6935–6945
- Li Y, Xie X, Shi X, Lin Y, Qiu Y, Mou J, Chen Q, Lu Y, Zhou L, Jiang M, Sun H, Ma H, Cheng J, Hu Q (2014) *Vibrio parahaemolyticus*, Southern Coastal Region of China, 2007–2012. *Emerg Infect Dis* 20(4):685–688
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- Lüdeke CH, Gonzalez-Escalona N, Fischer M, Jones JL (2015) Examination of clinical and environmental *Vibrio parahaemolyticus* isolates by multi-locus sequence typing (MLST) and multiple-locus variable-number tandem-repeat analysis (MLVA). *Front Microbiol* 6:564
- Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima M, Nakano M, Yamashita A, Kubota Y, Kimura S, Yasunaga T, Honda T, Shinagawa H, Hattori M, Iida T (2003) Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet* 361(9359):743–749
- Nair GB, Ramamurthy T, Bhattacharya SK, Dutta B, Takeda Y, Sack DA (2007) Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin Microbiol Rev* 20(1):39–48
- Okuda J, Ishibashi M, Hayakawa E, Nishino T, Takeda Y, Mukhopadhyay AK, Garg S, Bhattacharya SK, Nair GB, Nishibuchi M (1997) Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *J Clin Microbiol* 35(12):3150–3155
- Pal D, Das N (2010) Isolation, identification and molecular characterization of *Vibrio parahaemolyticus* from fish samples in Kolkata. *Eur Rev Med Pharmacol Sci* 14(6):545–549
- Pérez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, Crandall KA (2006) Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect Genet Evol* 6(2):97–112
- Pérez-Losada M, Cabezas P, Castro-Nallar E, Crandall KA (2013) Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. *Infect Genet Evol* 16:38–53
- Turner JW, Paranjpye RN, Landis ED, Biryukov SV, González-Escalona N, Nilsson WB, Strom MS (2013) Population structure of clinical and environmental *Vibrio parahaemolyticus* from the Pacific Northwest coast of the United States. *PLoS ONE* 8(2):e55726
- Urmersbach S, Alter T, Korlagge MS, Sperling L, Gerdt G, Messelhäusser U, Huehn S (2014) Population analysis of *Vibrio parahaemolyticus* originating from different geographical regions demonstrates a high genetic diversity. *BMC Microbiol* 14:59
- Wang S, Duan H, Zhang W, Li JW (2007) Analysis of bacterial foodborne disease outbreaks in China between 1994 and 2005. *FEMS Immunol Med Microbiol* 51(1):8–13
- Wong HC, Liu SH, Chiou CS, Nishibuchi M, Lee BK, Suthienkul O, Nair GB, Kaysner CA, Taniguchi H (2007) A pulsed-field gel electrophoresis typing scheme for *Vibrio parahaemolyticus* isolates from fifteen countries. *Int J Food Microbiol* 114(3):280–287
- Yan Y, Cui Y, Han H, Xiao X, Wong HC, Tan Y, Guo Z, Liu X, Yang R, Zhou D (2011) Extended MLST-based population

genetics and phylogeny of *Vibrio parahaemolyticus* with high levels of recombination. *Int J Food Microbiol* 145:106–112

Yu Y, Hu W, Wu B, Zhang P, Chen J, Wang S, Fang W (2011) *Vibrio parahaemolyticus* isolates from southeastern Chinese coast are genetically diverse with circulation of clonal

complex 3 strains since 2002. *Foodborne Pathog Dis* 8(11):1169–1176

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