Genomic stability among O3:K6 *V. parahaemolyticus* pandemic strains isolated between 1996 to 2012 in American countries

Abraham Guerrero[1,2], Bruno Gomez-Gil[2] and Marcial Leonardo Lizarraga-Partida[3*]

Abstract

Background: The *V. parahaemolyticus* pandemic clone, results in the development of gastrointestinal illness in humans. Toxigenic strains of this species are frequently isolated from aquatic habitats and organisms such as mollusks and crustaceans. Reports on the isolation of the pandemic clone started in 1996, when a new O3:K6 clone was identified in Asia, that rapidly spread worldwide, becoming the predominant clone isolated from clinical cases. In this study whole genome sequencing was accomplished with an Illumina MiniSeq platform, upon six novel *V. parahaemolyticus* strains, that have been isolated in Mexico since 1998 and three representative genomes of strains that were isolated from reported outbreaks in other American countries, and were deposited in the GenBank. These nine genomes were compared against the reference sequence of the O3:K6 pandemic strain (RIMD 2210633), which was isolated in 1996, to determine sequence differences within American isolates and between years of isolation.

Results: The results indicated that strains that were isolated at different times and from different countries, were highly genetically similar, among them as well as to the reference strain RIMD 2210633, indicating a high level of genetic stability among the strains from American countries between 1996 to 2012, without significant genetic changes relative to the reference strain RIMD 2210633, which was isolated in 1996 and was considered to be representative of a novel O3:K6 pandemic strain.

Conclusions: The genomes of *V. parahaemolyticus* strains isolated from clinical and environmental sources in Mexico and other American countries, presented common characteristics that have been reported for RIMD 2210633 O3:K6 pandemic strain. The major variations that were registered in this study corresponded to genes non associated to virulence factors, which could be the result of adaptations to different environmental conditions. Nevertheless, results do not show a clear pattern with the year or locality where the strains were isolated, which is an indication of a genomic stability of the studied strains.

Keywords: *V. parahaemolyticus* outbreaks, *V. parahaemolyticus* in American countries, *V. parahaemolyticus* O3:K6 pandemic clone

* Correspondence: lizarra@cicese.mx
[3]Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE), Ensenada, Baja California, Mexico
Full list of author information is available at the end of the article

Guerrero *et al. BMC Genomic Data*　　　(2021) 22:38

Page 2 of 9

Background

V. parahaemolyticus is a Gram-negative bacterium that is commonly distributed in marine environments. This bacterium has been associated with foodborne infections, causing 3 major syndromes: gastroenteritis, wound infections, and septicemia [1]. Most *V. parahaemolyticus* strains capable of causing infection express genes that encode thermostable direct hemolysin (*tdh*), *tdh*-related hemolysin (*trh*), or both. These genes have been identified as the primary virulence factors [2, 3] and are commonly used to identify pathogenic strains [4, 5].

Infections that are caused by *V. parahaemolyticus* in the US and Mexico have historically been associated with the O4:K12 serotype [6]; however, after the first outbreak of a novel O3:K6 clone (O3:K6/*tdh*$^+$/*trh*$^-$), reported in India in 1996, the O3:K6 clone spread throughout Southeast Asia and various geographical zones, including countries in the American continent, resulting in a pandemic event [7] and becoming the most common serotype that was associated with infection, globally [1].

The detection of this pandemic strain in American countries was first reported in Peru, in 1996; however, the first outbreak was registered in 1997–1998 [8, 9], and since then, outbreaks have been reported in other countries in the American continent [10]. The US recorded 416 infection cases in 1998, most of which were linked to the consumption of raw seafood (oysters) from the Gulf of Mexico [11]. In Chile, greater than 16,000 infection cases have been reported since 1998, representing the highest recorded number of infections [10, 12, 13]. Brazil presented the lowest incidence of infections, with only 18 clinical cases reported from 2001 to 2002 [14]. Guerrero et al. [15], found that the O3:K6 clinical strain had been isolated in Mexico by INDRE (Mexican Epidemiological Institute) between 1998 to 2009, in different states, and Revilla-Castellanos et al. [16] identified the strain in 2012 among hull biofouling samples from a ship with Japanese provenance. O3:K6 strains were also isolated in 2004, from the only recorded outbreak in Mexico, which resulted in more than 1200 clinical cases [17].

This study represents the first attempt to explore genomic variations among the pandemic *V. parahaemolyticus* O3:K6 strains isolated from 1998 to 2012 in Mexico and their genetic similarities to strains that have been isolated from other outbreaks in American and Asian countries (Table 1).

Methods

Collection of strains

V. parahaemolyticus strains, collected during various years and varying locations in Mexico, were selected for this study. The reference genomes of strains from outbreaks that were registered in Latin America and the US and the reference genome of the strain RIMD 2210633, isolated in Japan were also included as representative genomes of the O3:K6 pandemic clone. The list of strains is presented in Table 1.

Sequencing

Six O3:K6 *V. parahaemolyticus* strains (Table 1), described as belonging to sequence type 3 (ST3) by multilocus sequence typing (MLST) [15], were sequenced for the present study. Genomic DNA from a single colony of each strain was extracted using the cetrimonium bromide (CTAB) method. The obtained DNA was sequenced using the Illumina MiniSeq platform (2 × 150-bp paired-end reads) at the Mazatlán Unit for Aquaculture and Environmental Management A.C. (CIAD-Mazátlan).

Genome assembly

The reads for each strain were used to obtain the coverage depth, using SAMtools V1.2 [18] and BWA-MEM V0.7.12 [19], implemented with coverage.sh script developed by our team (https://github.com/cabraham03/coverage/blob/main/coverage.sh), and the metrics were visualized in Qualimap V2.2 [20]. Each genome was assembled using SPAdes V3.8 [21], the de novo assembly was performed with careful mode and -k 21,33,55,77,99,127 parameters. The resulting contigs were submitted to the Rapid Annotation Subsystem Technology (RAST [22];).

The pan- and core-genome plots were constructed based on the shared gene families of the 6 genomes described here and the 3 genomes of the O3:K6 pandemic strains isolated from outbreaks reported in South America, ATC210 (Chile, GenBank: LFUN00000000), Peru-466 (Peru, GenBank: ACFM00000000), North America CDC_K5058 (USA, GenBank: MITP00000000), and the genome from the reference strain RIMD 2210633 (GenBank: BA000031 and BA000032), which was isolated in Asia. The analyses were implemented as follows. Contigs of each genome were used to generate gene predictions, using Prokka V1.14.6 [23], and then Roary V3.12.0 [24], was implemented to obtain the pan-genome. Gene presence/absence matrices were visualized using roary_plots.py V1.01 (https://github.com/sanger-pathogens/Roary/tree/master/contrib/roary_plots). The predicted amino acid were generated with Prodigal V2.6.3 [25], and CMG-Biotools [26] was implemented to obtain a pairwise comparison with blastmatrix program. If a BLAST hit showed a 50% identity match in the alignment, and the length of the alignme nt was at least 50% of the longest gene (50/50 cut-off), it was considered to be a protein family.

Table 1 List of studied O3:K6 strains isolated in American countries

Strain	Year	State	GenBank	Characteristics
CICESE-170*	1998	Hgo	JAABPG000000000	R72H, *tlh, tdh, orf8*, O3:K6 (+), *trh* (−)
CICESE-186*	1999	Hgo	JAABPH000000000	R72H, *tlh, tdh, orf8*, O3:K6 (+), *trh* (−)
CICESE-187*	2000	Tams	JAAIFK000000000	R72H, *tlh, tdh, orf8*, O3:K6 (+), *trh* (−)
CAIM 1400*	2004	Sin	JAAIFJ000000000	R72H, *tlh, tdh, orf8*, O3:K6 (+), *trh* (−)
CICESE-188*	2009	NL	JAAHBO000000000	R72H, *tlh, tdh, orf8*, O3:K6 (+), *trh* (−)
CICESE-273*	2012	BC	JAABPI000000000	R72H, *tlh, tdh, orf8*, O3:K6 (+), *trh* (−)
RIMD 2210633	1996	Japan	BA000031/BA000032	O3:K6, *tdh* (+), *trh* (−)
Peru-466	1996	Peru	ACFM00000000	O3:K6, *tdh* (+), *trh* (−)
CDC_K5058	2007	USA	MITP00000000	O3:K6, *tdh* (+), *trh* (−)
ATC210	1998	Chile	LFUN00000000	O3:K6, *orf 8, tdh* (+), *trh* (−)

The novel (*) strains isolated in Mexico at the states of Hidalgo (Hgo), Tamaulipas (Tams), Sinaloa (Sin), Nuevo Leon (NL) and Baja California (BC).

Comparative analysis

To visualize the similarities and differences between strains, contigs from each genome were assembled in 2 supercontigs for each chromosome (–I and -II), using the MeDuSa web server (http://combo.dbe.unifi.it/medusa; [27]), and aligned with blast ring image generator (BRIG, V0.95 [28];). The comparative analyses were focused on the primary mobile genetic elements that were associated with the pandemic strain, such as the 7 VPaIs, phage f237, super integron (SI), and secretion systems (TSS).

Results

The number of reads for the obtained sequences ranged from 619,896 for CICESE-188 to 6,256,738 for CAIM 1400. The resulting reads presented average lengths of between 71 bp for CICESE-170 to 145 bp for CAIM 1400. The coverage depth showed differences between strains, wherein the lowest registered coverage was 14.2×, for CICESE-188, and the highest was registered for CAIM 1400, at 155.9×. Genomes were assembled from between 143 contigs (N50 = 194,688 and L50 = 8) for CICESE-170 to 1590 contigs (N50 = 5908 and L50 = 244) for CICESE-188 (Table 2). Among the Mexican genomes, CAIM 1400 presented the highest number of detected genes, with 4901 genes, including 283 new genes and 280 new gene families; CICESE-170 had the lowest number of genes, with 4643 detected genes, including 87 new genes and 69 new gene families (Table 2).

The pan- and core-genomes, new genes and new gene families for each strain, are show in Fig. 1. The strains Peru-466, isolated in 1996, CAIM 1400, isolated in 2004, and CICESE-188, isolated in 2009, presented the most new genes and new gene families relative to the reference genome of the strain RIMD 2210633, which was isolated in 1996, without a clear pattern associated with the year of isolation. The pan-genome presented an increase in the number of genes, from 4654 genes in the RIMD 2210633 strain to 5825 genes that were registered for CICESE-273, an environmental strain isolated in 2012; in contrast, the core-genome decreased, from 4654 genes in RIMD 2210633 to 4013 in CICESE-188, which was isolated in 2009.

The alignment of chromosomes I and II for the O3:K6 strains isolated in American countries, relative to the reference strain RIMD 2210633, which was isolated in Japan, are presented in Fig. 2. For both chromosomes, the figure indicates the most common genetic mobile elements that were associated with the pandemic strains, which were identified as the 7 VPaIs, the phage f237, the secretion system (TSS), and the type I pilus, among others. The positions of each element corresponded with those in the genome of the reference strain RIMD 2210633, as described in Hurley et al. [29], Boyd et al. [30], and Chen et al. [31].

From the RAST web server analysis, the genes commonly associated with pathogenic strains were detected, in categories such as resistance to antibiotics and toxic compounds, phages, prophages, iron acquisition systems, stress responses, toxins, the regulation of virulence genes, secretion systems, flagellar motility, capsular and extracellular polysaccharides, siderophores, colonization, and biofilm formation. Based on the RAST results, using a function-based comparison tool, most of the differences relative to the reference genome RIMD 2210633 were identified in genes that were associated with categories other than pathogenicity.

Mobile genetic elements that were associated with O3:K6 pandemic strains were detected in both chromosomes of the studied O3:K6 strains (Fig. 2). Five of the VPaIs (VPaI-1 to – 5) were detected in chromosome I, whereas the other 2 (VPaI-6 and -7) were detected in chromosome II. The results indicated that the VPaIs that were detected in the Mexican strains contained most of the genes that have been described for the reference

Table 2 Metrics obtained for the Mexican O3:K6 strains

Genome	No reads	bp	Coverage	GC %	Contigs	N50	L50	Total genes	New genes/families	Year
CAIM 1400	6,256,738	145	155.9X	45.2	180	204,016	7	4901	283/280	2004
CICESE-170	3,940,460	71	52.9X	45.3	143	194,688	8	4643	87/69	1998
CICESE-186	1,045,832	134	26.5X	45.4	464	25,261	58	4699	102/96	1999
CICESE-187	1,202,227	121	27.5X	45.5	760	16,608	94	4732	121/121	2000
CICESE-188	619,896	120	14.2X	46.0	1590	5908	244	4778	313/310	2009
CICESE-273	1,890,618	117	33.7X	45.6	934	10,218	147	4759	109/108	2012

bp = Average base pair. Year = year of isolation. N50 = minimum conting needed to cover 50% of the genome. L50 = number of contigs whose length sum makes up 50% of the genome size.

genome RIMD 2210633, with high similarity (> 96.4%). The few variations that were identified were associated with non-coding bases.

The VPaIs 1, 2, 3, 4, and 6 were present in each genome of the Mexican strains, containing the same number of genes that were previously reported for RIMD 2210633. VPaI-5 was not detected in CICESE-188; however, this strain presented most of the mobile genetic elements, including transposase, hypothetical protein, lipopolysaccharides (LPS), phage f237, 2 secretion systems (T3SS-1 and – 2), an osmotolerance gene cluster, integron class I, type I pilus, a multidrug efflux gene cluster, ferric uptake, gametolysin, biofilm, degradative genes, and *tdh* gene, which are characteristic of pandemic strains. In VPaI-7, most of the Mexican genomes presented differences with the reference genome, in the genes *VPA1312*, *VPA1313*, *VPA1314*, *VPA1316*, *VPA1318*, and *VPA1357*. In addition, in most of the Mexican strains, the phage f237-like was not detected; a gap registered in chromosome II, associated with phage f237-like (6 o'clock), may be due to a close association with phage f237, which was registered in chromosome I.

Based on the BLAST matrix comparison, the homology between and within predicted amino acids is presented in Fig. 3. The analyzed genomes from strains isolated in various American countries registered a maximum of 4901 proteins within 4696 families, for CAIM 1400, whereas the reference strain RIMD 2210633 contained a total of 4832 proteins within 4654 families. The similarities in shared proteins between the American strains and RIMD 2210633 ranged from 79.6% (CICESE-188) to 88.8% (ATC210). The American strains
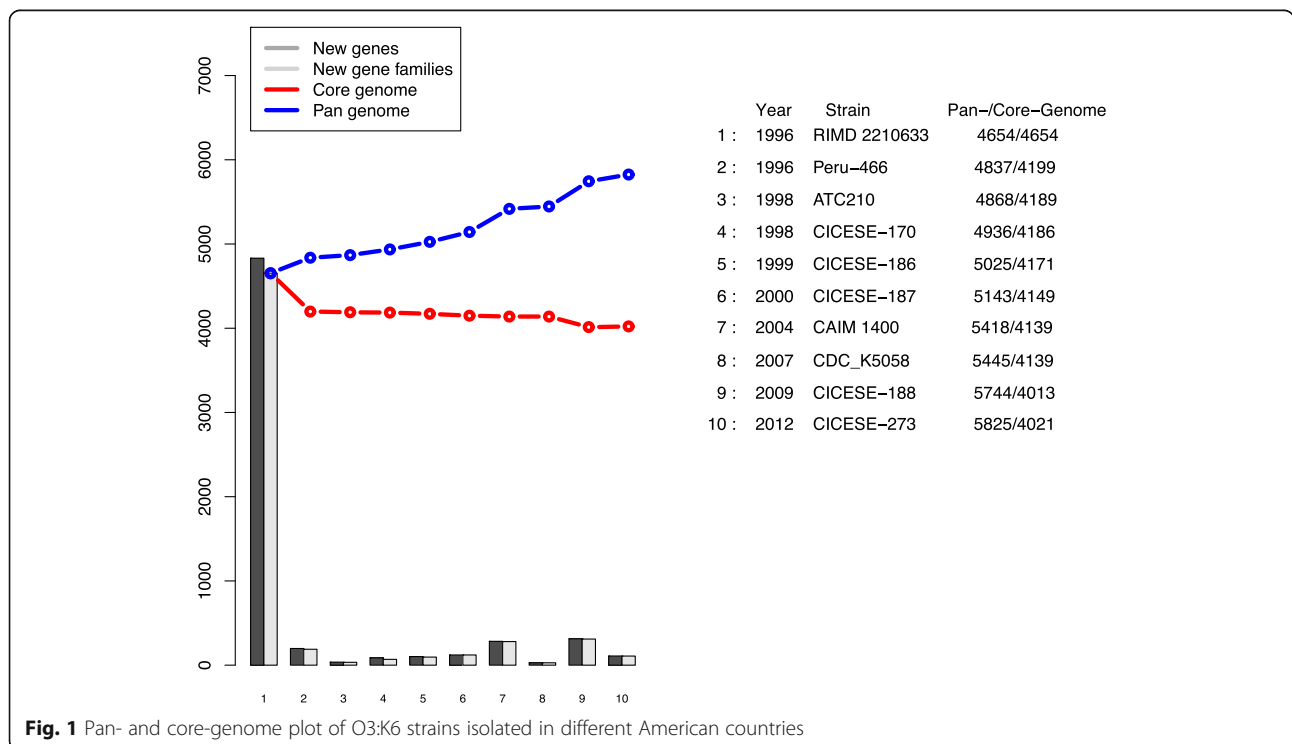

	Year	Strain	Pan-/Core-Genome
1 :	1996	RIMD 2210633	4654/4654
2 :	1996	Peru–466	4837/4199
3 :	1998	ATC210	4868/4189
4 :	1998	CICESE–170	4936/4186
5 :	1999	CICESE–186	5025/4171
6 :	2000	CICESE–187	5143/4149
7 :	2004	CAIM 1400	5418/4139
8 :	2007	CDC_K5058	5445/4139
9 :	2009	CICESE–188	5744/4013
10 :	2012	CICESE–273	5825/4021

Fig. 1 Pan- and core-genome plot of O3:K6 strains isolated in different American countries
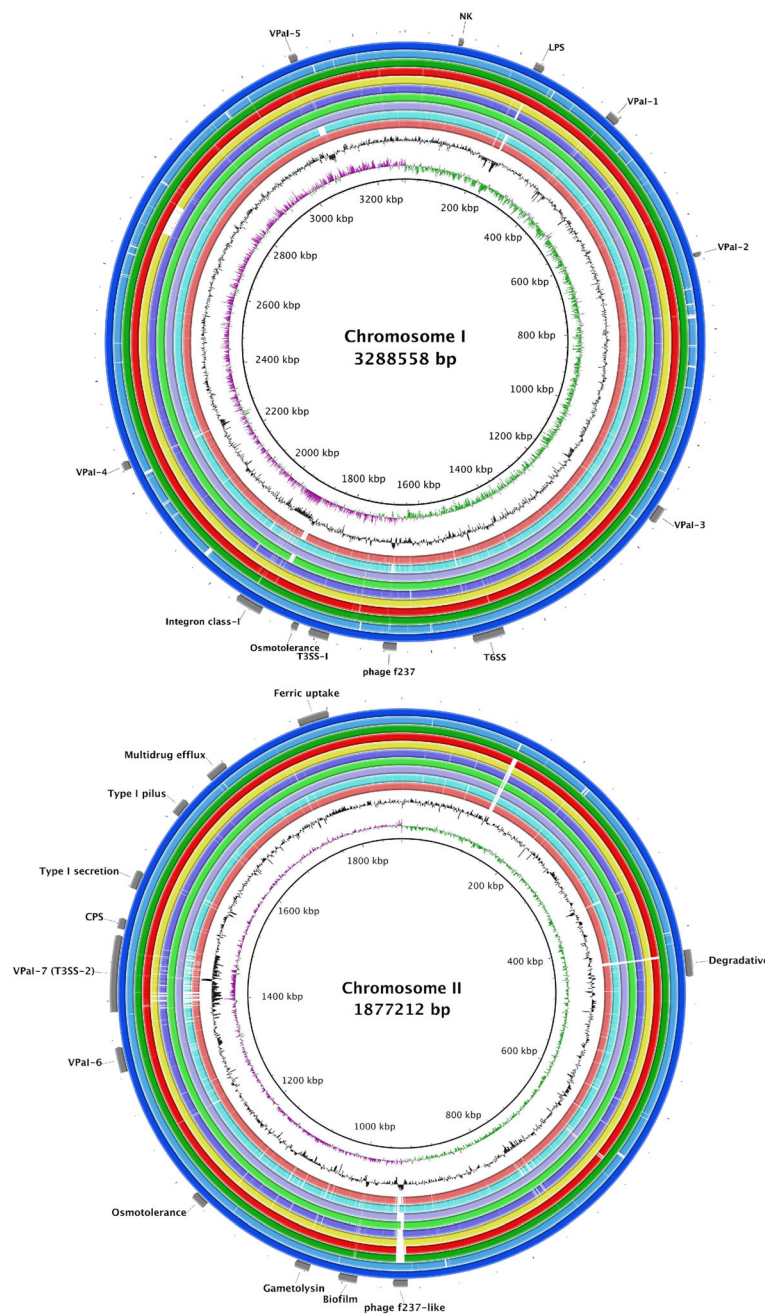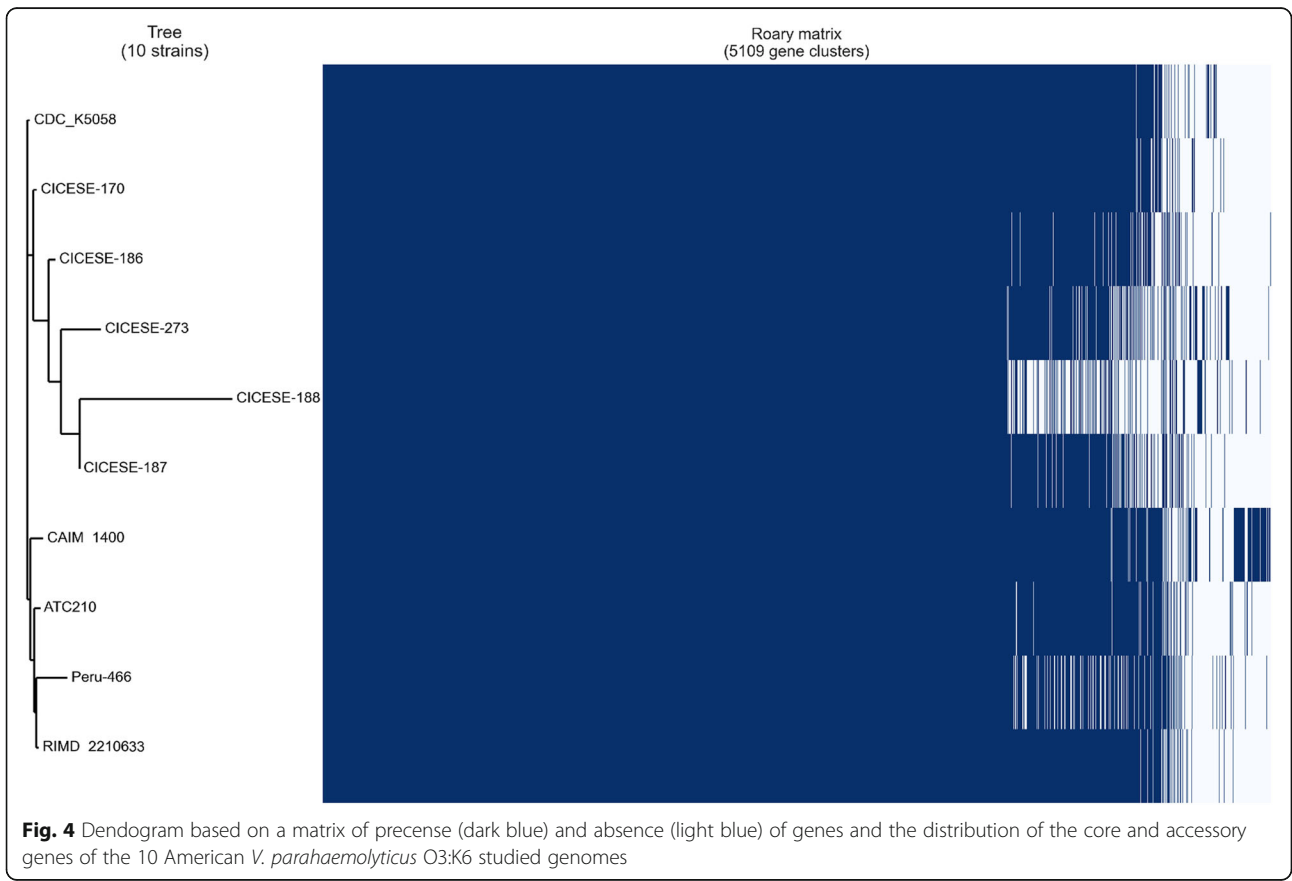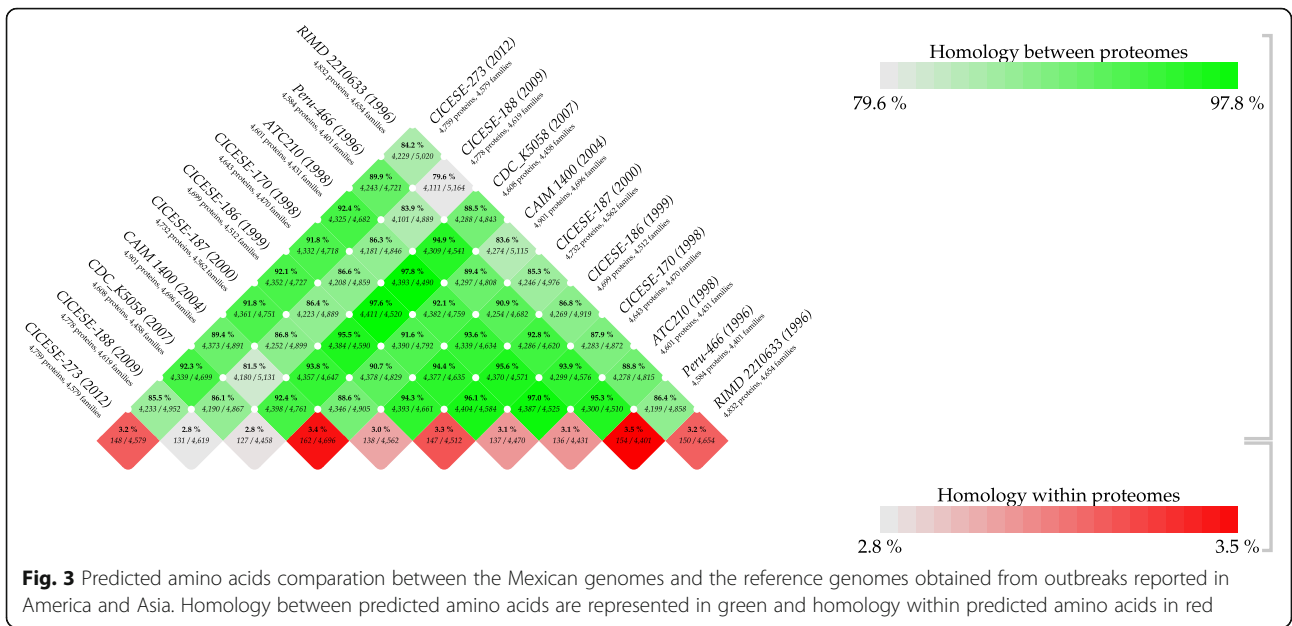
Fig. 2 Alignment of the reference and novel genomes. From the outer ring to inner, RIMD 2210633 (■, Japan), Peru-466 (■, Peru), ATC210 (■, Chile), CICESE-170 (■, Mexico), CICESE-186 (■, Mexico), CICESE-187 (■, Mexico), CAIM 1400 (■, Mexico), CDC_K5058 (■, USA), CICESE-188 (■, Mexico), CICESE-273 (■, Mexico), GC Content (■). Most common elements associated to O3:K6 strains are indicated in the figure

presented higher percentages of similarities with each other, with CICESE-170 sharing 97.6% (4411) of proteins with CDC_K5058 and ATC210 sharing 97.8% (4393) of proteins with CDC_K5058. CICESE-188, isolated in 2009 in northeastern Mexico, presented the lowest percentage of shared proteins with all studied genomes. The homology among predicted amino acids ranged

from 2.8 to 3.5%. Although no clear pattern was associated with isolation year, high percentages of shared proteins were registered between the Peru-466 strain, isolated in 1996, and the strains CICESE-170, isolated in Mexico in 1998, and ATC210, isolated in Chile in 1998, and these countries were the first American countries in which the pandemic clone was isolated.

Guerrero et al. BMC Genomic Data   (2021) 22:38

Page 6 of 9

Fig. 3 Predicted amino acids comparation between the Mexican genomes and the reference genomes obtained from outbreaks reported in America and Asia. Homology between predicted amino acids are represented in green and homology within predicted amino acids in red

Fig. 4 Dendogram based on a matrix of precense (dark blue) and absence (light blue) of genes and the distribution of the core and accessory genes of the 10 American *V. parahaemolyticus* O3:K6 studied genomes

The results presented in the pangenome of Fig. 4, indicated high similarity among the American O3:K6 strains and the reference strain RIMD 2210633. Two clusters could be distinguish in the dendrogram of this figure; one that groupe strains isolated in the USA in 2007 (CDC_K5058) and those isolated in Mexico in 1998 (CICESE-170), 1999 (CICESE-186), 2000 (CICESE-187) and 2009 (CICESE-188) and 2012 (CICESE-273). The strain CICESE-188, which was isolated in a northeastern state of Mexico, showed the lowest degree of similarity with the other strains in this group, and also displayed the lowest percentage of shared proteins. A second cluster was formed by those strains isolated in Peru (Peru-466) in 1996, in Chile (ATC 210) in 1998, CAIM 1400 isolated in 2004 in México and the reference strain RIMD 2210633 isolated in Japan in 1996.

Discussion

The pandemic O3:K6 *V. parahaemolyticus* strains are closely related genetically and commonly clustered into a single group [32, 33]. These strains are genetically different from non-pandemic strains, including the old O3: K6 (*trh*+/*tdh*-) clone [7]. Molecular approaches have demonstrated that O3:K6 pandemic strains presented a clonal origin [34], in which most of the strains presented almost identical molecular patterns [35, 36] or the same sequences for *toxRS* [7]. ST3 was the most common MLST found among the pandemic strains [32, 37]. Based on these studies, the clonal term has been commonly used to describe O3:K6 pandemic strains. However, whole-genome analyses have detected differences between the closely genetically related strains, which could not be discovered by sequencing only a few genes (*toxRS* or MLST).

The present study included 6 Mexican strains, isolated from 1998 to 2012, most of which were isolated from clinical cases [15], and 3 strains that were isolated in American countries from 1996 to 2007. Differences were found between the numbers of new genes and new gene families, shared proteins, and phylogenetic relatedness (Figs. 1–4); however, the homology among predicted amino acids ranged from 79.6 to 97.8% (Fig. 3), which indicated that these strains are closely related genetically. Based on the genome analysis (Fig. 1), American O3:K6 strains presented new genes and new families of genes, which were not detected in the reference genome of the strain RIMD 2210633; most of these differences were associated with genes that have been classified as hypothetical proteins, with unknown function or with genes that are not necessary for infection.

These *V. parahaemolyticus* O3:K6 pandemic strains presented mobile genetic elements, such as the filamentous phage f237, which has been widely associated with this serotype [38]. Some of these mobile genetic

elements (ORF8) have been used as molecular markers for pandemic strains [39]. However, the most common elements that were associated with the O3:K6 pandemic strains are the VPaIs [30, 33], which are mobile elements that harbor multiple putative virulent genes, encoding hydrolases, colicins, M proteins, cytotoxin integrase, methyltransferase (MTase), and the TDH toxin, as well as genes that are associated with the T3SS2 gene cluster and phage-like proteins [2, 29, 31, 40–42]. The presence of these elements has been correlated with pathogenicity [43]. Hurley et al. [29] hypothesized that these elements increase the fitness or the infectivity of O3:K6 pandemic strains. VPaI-1, VPaI-4, VPaI-5, and VPaI-6 have been reported to be found exclusively in O3:K6 pandemic strains [29, 30], in addition to the mobile elements in VPaI-7, in chromosome II. These elements encode several virulent factors, such as *tdh*, which encodes the TDH toxin, and the T3SS2, which has been implicated in host cell invasion [44]. These elements were detected in all of the American strains, with only a few genetic variations relative to the reference strain RIMD 2210633 (Fig. 2).

Conclusions

These results showed that the genomes of *V. parahaemolyticus* strains isolated from clinical and environmental sources in Mexico and other American countries presented common characteristics that have been reported for the O3:K6 RIMD 2210633 pandemic strain, and the major variations that were registered in this study corresponded with genes that have been categorized as non-pathogenic, which could be the result of adaptations that were necessary for the different environmental conditions of the localities from which they were isolated [45]. However, the studied genomes did not present any clear patterns according to the year or region of isolation, based on pan/core genomes, shared proteins, and phylogenetic relatedness, which is concerning. Therefore, the results from these *V. parahaemolyticus* pandemic strains indicated a high level of genetic stability among the strains from American countries between 1996 to 2012, without significant genetic changes relative to the reference strain RIMD 2210633, which was isolated in 1996 and was considered to be representative of a novel O3:K6 pandemic strain. The pangenome presented in Fig. 4, show the association of strains isolated in the North of the continent as well as the association of those strains isolated in the South of the continent, with the exception of CAIM 1400 that was isolated in 2012, during the only outbreak registered in México. Nevertheless, results do not show a clear pattern with the year or locality where the strains were isolated, which is an indication of a genomic stability of the studied strains.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have not competing interest,

Author details
[1]Cátedras CONACyT-CIAD, Food Research and Development Center A.C. Mazatlán Unit (Centro de Investigación en Alimentación y Desarrollo, A.C. Unidad Mazatlán), Mazatlán, Sinaloa, Mexico. [2]CIAD, Food Research and Development Center A.C. Mazatlán Unit for Aquaculture, A.P. 711, Mazatlán, Sinaloa, Mexico 82100. [3]Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE), Ensenada, Baja California, Mexico.

References
1. Nair GB, Ramamurthy T, Bhattacharya SK, Dutta B, Takeda Y, Sack DA. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin Microbiol Rev. 2007;20(1):39–48. https://doi.org/10.1128/CMR.00025-06.
2. Ceccarelli D, Hasan NA, Huq A, Colwell RR. Distribution and dynamics of epidemic and pandemic *Vibrio parahaemolyticus* virulence factors. Front Cell Infect Microbiol. 2013;3:97.
3. Nishibuchi M, Fasano A, Russell RG, Kaper JB. Enterotoxigenicity of *Vibrio parahaemolyticus* with and without genes encoding thermostable direct hemolysin. Infect Immun. 1992;60(9):3539–45. https://doi.org/10.1128/iai.60.9.3539-3545.1992.
4. Bej AK, Patterson DP, Brasher CW, Vickery MC, Jones DD, Kaysner CA. Detection of total and hemolysin-producing *Vibrio parahaemolyticus* in shellfish using multiplex PCR amplification of *tl*, *tdh* and *trh*. J Microbiol Methods. 1999;36(3):215–25. https://doi.org/10.1016/S0167-7012(99)00037-8.
5. Nordstrom JL, Vickery MCL, Blackstone GM, Murray SL, DePaola A. Development of a multiplex real-time PCR assay with an internal amplification control for the detection of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters. Appl Environ Microbiol. 2007;73(18):5840–7. https://doi.org/10.1128/AEM.00460-07.
6. Abbott S, Powers C, Kaysner CA, Takeda Y, Ishibashi M, Joseph SW, et al. Emergence of a restricted bioserovar of *Vibrio parahaemolyticus* as the predominant cause of *Vibrio*-associated gastroenteritis on the west coast of the United States and Mexico. J Clin Microbiol. 1989;27(12):2891–3. https://doi.org/10.1128/jcm.27.12.2891-2893.1989.
7. Matsumoto C, Okuda J, Ishibashi M, Iwanaga M, Garg P, Rammamurthy T, et al. Pandemic spread of an O3:K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and *toxRS* sequence analyses. J Clin Microbiol. 2000;38(2):578–85. https://doi.org/10.1128/JCM.38.2.578-585.2000.
8. Gavilan RG, Zamudio ML, Martinez-Urtaza J. Molecular epidemiology and genetic variation of pathogenic *Vibrio parahaemolyticus* in Peru. PLoS Negl Trop Dis. 2013;7(5):e2210. https://doi.org/10.1371/journal.pntd.0002210.
9. Gil AI, Miranda H, Lanata CF, Prada A, Hall ER, Barreno CM, et al. O3:K6 serotype of *Vibrio parahaemolyticus* identical to the global pandemic clone associated with diarrhea in Peru. Int J Infect Dis. 2006;11:324–8.
10. Velazquez-Roman J, León-Sicairos N, Hernández-Díaz LJ, Canizalez-Roman A. Pandemic *Vibrio parahaemolyticus* O3:K6 on the American continent. Front Cell Infect Microbiol. 2014;3:110.
11. Daniels NA, Ray B, Easton A, Marano N, Kahn E, McShan AL, et al. Emergence of a new *Vibrio parahaemolyticus* serotype in raw oysters: a prevention quandary. JAMA. 2000;284(12):1541–5. https://doi.org/10.1001/jama.284.12.1541.
12. García K, Torres R, Uribe P, Hernandez C, Rioseco ML, Romero J, et al. Dynamics of clinical and environmental *Vibrio parahaemolyticus* strains during seafood-related summer diarrhea outbreaks in southern Chile. Appl Environ Microbiol. 2009;75(23):7482–7. https://doi.org/10.1128/AEM.01662-09.
13. González-Escalona N, Cachicas V, Acevedo C, Rioseco ML, Vergara JA, Cabello F, et al. *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. Emerg Infect Dis. 2005;11(1):129–31. https://doi.org/10.3201/eid1101.040762.
14. Leal NC, Da Silva SC, Cavalcanti VO, Figueiroa AC, Nunes VV, Miralles IS, et al. *Vibrio parahaemolyticus* serovar O3:K6 gastroenteritis in Northeast Brazil. J Appl Microbiol. 2008;105(3):691–7. https://doi.org/10.1111/j.1365-2672.2008.03782.x.
15. Guerrero A, Lizárraga-Partida ML, Gómez-Gil Rodríguez B, Licea-Navarro AF, Revilla-Castellanos VJ, Wong-Chang I, et al. Genetic Analysis of *Vibrio parahaemolyticus* O3:K6 strains that have been isolated in Mexico since 1998. PLoS One. 2017;12:e0169722.
16. Revilla-Castellanos VJ, Guerrero A, Gomez-Gil B, Navarro-Barrón E, Lizárraga-Partida ML. Pathogenic *Vibrio parahaemolyticus* isolated from biofouling on commercial vessels and harbor structures. Biofouling. 2015;31(3):275–82. https://doi.org/10.1080/08927014.2015.1038526.
17. Cabanillas-Beltrán H, LLausás-Magaña E, Romero R, Espinoza A, Garcia-Gasca A, Nishibuchi M, et al. Outbreak of gastroenteritis caused by the pandemic *Vibrio parahaemolyticus* O3:K6 in Mexico. FEMS Microbiol Lett. 2006;265(1):76–80. https://doi.org/10.1111/j.1574-6968.2006.00475.x.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. et alThe sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352.
19. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
20. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 2012;28(20):2678–9. https://doi.org/10.1093/bioinformatics/bts503.
21. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77. https://doi.org/10.1089/cmb.2012.0021.
22. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9(1):75. https://doi.org/10.1186/1471-2164-9-75.
23. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9. https://doi.org/10.1093/bioinformatics/btu153.

24. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3. https://doi.org/10.1093/bioinformatics/btv421.

25. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11(1):119. https://doi.org/10.1186/1471-2105-11-119.

26. Vesth T, Lagesen K, Acar Ö, Ussery D. CMG-biotools, a free workbench for basic comparative microbial genomics. PLoS One. 2013;8(4):e60120. https://doi.org/10.1371/journal.pone.0060120.

27. Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lió P, et al. MEDUSA: a multi-draft based scaffolder. Bioinformatics. 2015;31(15):2443–51. https://doi.org/10.1093/bioinformatics/btv171.

28. Alikhan NF, Petty NK, Zakour NLB. Beatson SABLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12(1):402. https://doi.org/10.1186/1471-2164-12-402.

29. Hurley CC, Quirke AM, Reen FJ, Boyd EF. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. BMC Genomics. 2006;7(1):104. https://doi.org/10.1186/1471-2164-7-104.

30. Boyd EF, Cohen ALV, Naughton LM, Ussery DW, Binnewies TT, Stine OC, et al. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. BMC Microbiol. 2008;8(1):110. https://doi.org/10.1186/1471-2180-8-110.

31. Chen Y, Stine OC, Badger JH, Gil AI, Nair GB, Nishibuchi M, et al. Comparative genomic analysis of *Vibrio parahaemolyticus*: serotype conversion and virulence. BMC Genomics. 2011;12(1):294. https://doi.org/10.1186/1471-2164-12-294.

32. González-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.

33. Han H, Wong HC, Kan B, Zhaobiao G, Zeng X, Yin S, et al. Genome plasticity of *Vibrio parahaemolyticus*: microevolution of the 'pandemic group'. BMC Genomics. 2008;9(1):570. https://doi.org/10.1186/1471-2164-9-570.

34. Okuda J, Ishibashi M, Hayakawa E, Nishino T, Takeda Y, Mukhopadhyay AK, et al. 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. 1997;35(12):3150–5. https://doi.org/10.1128/jcm.35.12.3150-3155.1997.

35. Chowdhury NR, Chakraborty S, Eampokalap B, Chaicumpa W, Chongsa-Nguan M, Moolasart P, et al. Clonal dissemination of *Vibrio parahaemolyticus* displaying similar DNA fingerprint but belonging to two different serovars (O3:K6 and O4:K68) in Thailand and India. Epidemiol Infect. 2000a;125(1):17–25. https://doi.org/10.1017/S0950268899004070.

36. Chowdhury NR, Chakraborty S, Ramamurthy T, Nishibuchi M, Yamasaki S, Takeda Y, et al. Molecular evidence of clonal pandemic strains Emerg Infect Dis. 2000b;6:631–6.

37. Turner JW, Paranjpye RN, Landis ED, Biryukov SV, González-Escalona N, Nilsson WB, et al. Population structure of clinical and environmental *Vibrio parahaemolyticus* from the Pacific Northwest coast of the United States. PLoSE One. 2013;8:–e55726.

38. Nasu H, Iida T, Sugahara T, Yamaichi Y, Park KS, Yokoyama K, et al. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. J Clin Microbiol. 2000;38(6):2156–61. https://doi.org/10.1128/JCM.38.6.2156-2161.2000.

39. Myers ML, Panicker G, Bej AK. PCR detection of a newly emerged pandemic *Vibrio parahaemolyticus* O3:K6 pathogen in pure cultures and seeded waters from the Gulf of Mexico. Appl Environ Microbiol. 2003;69(4):2194–200. https://doi.org/10.1128/AEM.69.4.2194-2200.2003.

40. Ham H, Orth K. The role of type III secretion system 2 in *Vibrio parahaemolyticus* pathogen. J Microbiol. 2012;50(5):719–25. https://doi.org/10.1007/s12275-012-2550-2.

41. 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. https://doi.org/10.1016/S0140-6736(03)12659-1.

42. Wang H, Wong MM, O'Toole D, Mak MM, Wu RS, Kong RY. Identification of a DNA methyltransferase gene carried on a pathogenicity island-like element (VPAI) in *Vibrio parahaemolyticus* and its prevalence among clinical and environmental isolates. Appl Environ Microbiol. 2006;72(6):4455–60. https://doi.org/10.1128/AEM.02095-05.

43. Izutsu K, Kurokawa K, Tashiro K, Kuhara S, Hayashi T, Honda T, et al. Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in Kanagawa phenomenon-positive *Vibrio parahaemolyticus* strains. Infect Immun. 2008;76(3):1016–23. https://doi.org/10.1128/IAI.01535-07.

44. Zhang L, Krachler AM, Broberg CA, Li Y, Hamid Mirzael H, Gilpin CJ, et al. Type III effector VopC mediates invasion for Vibrio species. Cell Rep. 2012;1(5):453–60. https://doi.org/10.1016/j.celrep.2012.04.004.

45. Li L, Wong HC, Nong W, Cheung MK, Law PT, Kam KM, et al. Comparative genomic analysis of clinical and environmental strains provides insight into the pathogenicity and evolution of *Vibrio parahaemolyticus*. BMC Genomics. 2014;15(1):1135. https://doi.org/10.1186/1471-2164-15-1135.

Publisher's Note

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