AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

A Nonautochthonous U.S. Strain of *Vibrio parahaemolyticus* Isolated from Chesapeake Bay Oysters Caused the Outbreak in Maryland in 2010

Julie Haendiges,[a] Jessica Jones,[b] Robert A. Myers,[a] Clifford S. Mitchell,[a] Erin Butler,[a] Magaly Toro,[c] Narjol Gonzalez-Escalona[d]

Department of Health and Mental Hygiene, Baltimore, Maryland, USA[a]; Division of Seafood Science and Technology, Gulf Coast Seafood Laboratory, Food and Drug Administration, Dauphin Island, Alabama, USA[b]; Instituto de Nutricion y Tecnologia de los Alimentos, Universidad de Chile, Santiago, Chile[c]; Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA[d]

ABSTRACT

In the summer of 2010, *Vibrio parahaemolyticus* caused an outbreak in Maryland linked to the consumption of oysters. Strains isolated from both stool and oyster samples were indistinguishable by pulsed-field gel electrophoresis (PFGE). However, the oysters contained other potentially pathogenic *V. parahaemolyticus* strains exhibiting different PFGE patterns. In order to assess the identity, genetic makeup, relatedness, and potential pathogenicity of the *V. parahaemolyticus* strains, we sequenced 11 such strains (2 clinical strains and 9 oyster strains). We analyzed these genomes by *in silico* multilocus sequence typing (MLST) and determined their phylogeny using a whole-genome MLST (wgMLST) analysis. Our *in silico* MLST analysis identified six different sequence types (STs) (ST8, ST676, ST810, ST811, ST34, and ST768), with both of the clinical and four of the oyster strains being identified as belonging to ST8. Using wgMLST, we showed that the ST8 strains from clinical and oyster samples were nearly indistinguishable and belonged to the same outbreak, confirming that local oysters were the source of the infections. The remaining oyster strains were genetically diverse, differing in >3,000 loci from the Maryland ST8 strains. eBURST analysis comparing these strains with strains of other STs available at the *V. parahaemolyticus* MLST website showed that the Maryland ST8 strains belonged to a clonal complex endemic to Asia. This indicates that the ST8 isolates from clinical and oyster sources were likely not endemic to Maryland. Finally, this study demonstrates the utility of whole-genome sequencing (WGS) and associated analyses for source-tracking investigations.

IMPORTANCE

***Vibrio parahaemolyticus* is an important foodborne pathogen and the leading cause of bacterial infections in the United States associated with the consumption of seafood. In the summer of 2010, *Vibrio parahaemolyticus* caused an outbreak in Maryland linked to oyster consumption. Strains isolated from stool and oyster samples were indistinguishable by pulsed-field gel electrophoresis (PFGE). The oysters also contained other potentially pathogenic *V. parahaemolyticus* strains with different PFGE patterns. Since their identity, genetic makeup, relatedness, and potential pathogenicity were unknown, their genomes were determined by using next-generation sequencing. Whole-genome sequencing (WGS) analysis by whole-genome multilocus sequence typing (wgMLST) allowed (i) identification of clinical and oyster strains with matching PFGE profiles as belonging to ST8, (ii) determination of oyster strain diversity, and (iii) identification of the clinical strains as belonging to a clonal complex (CC) described only in Asia. Finally, WGS and associated analyses demonstrated their utility for trace-back investigations.**

Vibrio parahaemolyticus is an important foodborne pathogen and the leading cause of bacterial infections in United States associated with the consumption of seafood (1). *Vibrio parahaemolyticus* strains are considered pathogenic when they carry genes encoding thermostable direct hemolysin (*tdh*) and/or thermostable direct hemolysin-related hemolysin (*trh*) (2), although these potentially pathogenic strains usually represent a small fraction of all environmental strains (3). In addition to the *tdh* and *trh* genes, pathogenic *V. parahaemolyticus* strains carry other pathogenicity-related genes, such as type III secretion effectors, which are needed for producing infections at the intestinal level and are usually located in pathogenicity islands (4–7).

In the summer of 2010, two individuals became ill after eating raw oysters in two different restaurants in Baltimore, MD. In both cases, *V. parahaemolyticus* strains were isolated from their stools, and preliminary analyses by pulsed-field gel electrophoresis (PFGE) using two enzymes (SfiI and NotI) found that the two strains differed by only two bands (Table 1). On this basis, they were considered part of the same outbreak. Neither patient had traveled outside his/her home state 7 days prior to illness onset, nor did they have preexisting high-risk conditions for *Vibrio* infection. The two cases were initially linked based on the similar PFGE patterns produced by the strains

Received 29 January 2016 **Accepted** 15 March 2016

Accepted manuscript posted online 18 March 2016

Citation Haendiges J, Jones J, Myers RA, Mitchell CS, Butler E, Toro M, Gonzalez-Escalona N. 2016. A nonautochthonous U.S. strain of *Vibrio parahaemolyticus* isolated from Chesapeake Bay oysters caused the outbreak in Maryland in 2010. Appl Environ Microbiol 82:3208–3216. doi:10.1128/AEM.00096-16.

Editor: H. L. Drake, University of Bayreuth

Address correspondence to Narjol Gonzalez-Escalona, narjol.gonzalez-escalona@fda.hhs.gov.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00096-16.

TABLE 1 Characteristics of the *V. parahaemolyticus* strains used in this study[b]

Isolate	CFSAN identification no.	ST	Source	Virulence gene		PFGE profile[a]	
				tdh	*trh*	NotI	SfiI
MDVP19	CFSAN007452	8	Stool	None	2	K16S12.0156	**K16N11.0146**
MDVP20	CFSAN007453	8	Stool	None	2	**K16S12.0145**	**K16N11.0146**
MDVP21	CFSAN012491	8	Oyster	None	2	**K16S12.0145**	**K16N11.0146**
MDVP23	CFSAN012492	8	Oyster	None	2	K16S12.0153	**K16N11.0146**
MDVP24	CFSAN012493	8	Oyster	None	2	**K16S12.0145**	**K16N11.0146**
MDVP29	CFSAN012494	8	Oyster	None	2	**K16S12.0145**	**K16N11.0146**
MDVP22	CFSAN007454	676	Oyster	2	1	K16S12.0152	K16N11.0156
MDVP25	CFSAN007456	810	Oyster	None	2	K16S12.0155	K16N11.0155
MDVP26	CFSAN007457	811	Oyster	None	2	K16S12.0154	K16N11.0154
MDVP27	CFSAN007458	34	Oyster	None	2	K16S12.0150	K16S11.0005
MDVP28	CFSAN007459	768	Oyster	2	1	K16S12.0151	K16N11.0153

[a] PFGE profiles assigned by the CDC. Boldface indicates identical PFGE profiles in the same column.
[b] All strains were isolated in Maryland in 2010.

isolated from these cases and later by the oyster tags, which are collected for every positive *Vibrio* exposure.

The *V. parahaemolyticus* outbreak in Maryland in 2010 provided both an unusual opportunity and several puzzles, as causative strains are rarely isolated from food sources (5, 8–10). *V. parahaemolyticus* outbreak strains are typically available only from clinical samples. In this case, oysters containing the outbreak strain were identified. However, these oysters also contained *V. parahaemolyticus* strains that were not related to the outbreak. Furthermore, the PFGE pattern of the 2010 outbreak strains has not been detected in any subsequent *V. parahaemolyticus* cases in Maryland. Key questions include the following. What happened to the strain causing the 2010 outbreak? How related are these outbreak strains to previously archived strains? How genetically related are these outbreak strains to the other potentially pathogenic *V. parahaemolyticus* strains isolated from the same oysters? Can the genetic identity and phylogenetic relationship among these and other *V. parahaemolyticus* strains help us identify the possible origin of the outbreak strains?

Recently, scientists have been using next-generation sequencing techniques to reanalyze historical collections of pathogens and outbreak strains, in efforts to provide new insights for outbreak investigations. Whole-genome sequencing (WGS) together with single nucleotide polymorphism (SNP) (11–16) or whole-genome multilocus sequence typing (MLST) (wgMLST) (17–20) data analyses allow us to better understand both population dynamics and the mechanisms that contribute to increased virulence among foodborne bacterial pathogens.

To address the key questions about the identity, genetic makeup, and phylogenetic relationships among the *V. parahaemolyticus* strains collected during the 2010 Maryland outbreak, we sequenced the genomes of 11 of these strains: 2 from clinical samples and 9 from outbreak-implicated oysters. By comparing these genomes to each other and to the other *V. parahaemolyticus* genomes archived in GenBank, we are able to propose a possible origin for these outbreak strains as well as demonstrate the utility of WGS and associated analyses for such investigations.

MATERIALS AND METHODS

Identification of oyster isolates. A trace-back investigation was conducted, and tags from suspect oysters were used to identify the implicated growing area. Additional oysters were collected from the growing area and analyzed for *V. parahaemolyticus* by using a most-probable-number (MPN)–real-time PCR method described previously (21). Thiosulfate-citrate-bile salts-sucrose (TCBS) agar plates were streaked from PCR-positive MPN tubes for isolation of *V. parahaemolyticus*. Typical isolated colonies were confirmed by using the same real-time PCR as the one used for screening. All tdh^+ and/or trh^+ isolates were analyzed by PFGE using the standard PulseNet protocol.

Bacterial strains and media. The *V. parahaemolyticus* strains sequenced in this study, along with their assigned CFSAN identification numbers, are listed in Table 1. Isolates sequenced in our CFSAN/FDA facility are given a unique CFSAN identification number for future track-

TABLE 2 Summary report of the *de novo* assembly of the strains from this study

Isolate	GenBank accession no.	No. of contigs	Size (bp)	GC content (%)	N_{50}	Minimum contig size (bp)	Maximum contig size (bp)	Avg contig size (bp)	Avg coverage (×)
MDVP19	JNTJ00000000	37	5,127,281	45.1	527,441	504	1,521,871	138,575	99
MDVP20	JNTK00000000	39	5,126,628	45.2	567,671	511	1,521,355	131,452	105
MDVP21	JNUG00000000	35	5,126,325	45.1	526,001	511	995,851	146,466	97
MDVP23	JNUH00000000	34	5,128,956	45.2	485,136	510	1,192,043	150,852	113
MDVP24	JNUI00000000	47	5,123,442	45.2	329,201	511	635,564	109,009	79
MDVP29	JNUJ00000000	34	5,127,159	45.2	527,658	511	1,521,578	150,799	95
MDVP22	JNUO00000000	32	5,017,786	45.3	445,027	704	799,796	139,383	77
MDVP25	JNUK00000000	56	5,206,921	45.2	480,822	501	878,464	92,981	95
MDVP26	JNUL00000000	33	5,188,815	45.1	540,584	691	1,242,084	157,237	113
MDVP27	JNUM00000000	52	5,061,948	45.2	321,455	520	729,103	97,345	66
MDVP28	JNUN00000000	45	5,205,568	45.1	425,650	554	1,317,831	115,679	87

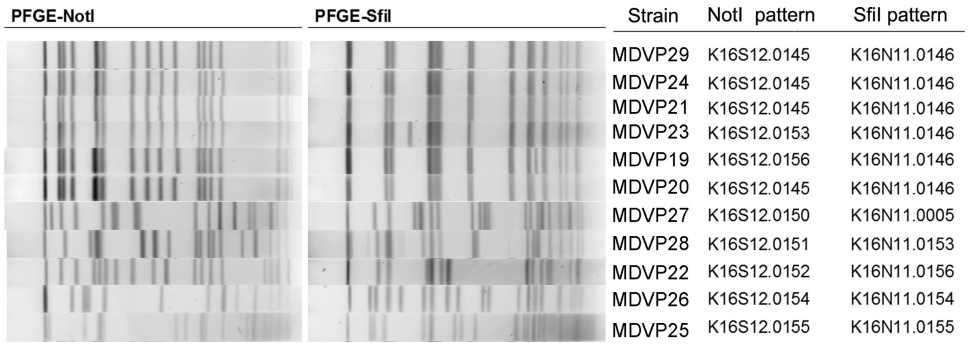

FIG 1 PFGE profiles of the *V. parahaemolyticus* strains used in this study, using two restriction enzymes (NotI and SfiI).

ing. All isolates were retrieved from storage (−80°C freezer), transferred to Luria-Bertani (LB) medium with 3% NaCl, and incubated at 37°C with shaking at 250 rpm.

DNA extraction and quantification. Genomic DNA from each strain was isolated from cultures grown overnight by using the DNeasy blood and tissue kit (Qiagen, Valencia, CA). The quality of the DNA was checked by using a NanoDrop 1000 instrument (Thermo Scientific, Rockford, IL), and the concentration was determined by using a Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit and a Qubit 2.0 fluorometer (Thermo Scientific), according to each manufacturer's instructions.

Whole-genome sequencing, contig assembly, and annotation. The genomes of the strains were sequenced by using 250-bp paired-end libraries, with a MiSeq reagent kit (v2), using a MiSeq sequencer (Illumina, San

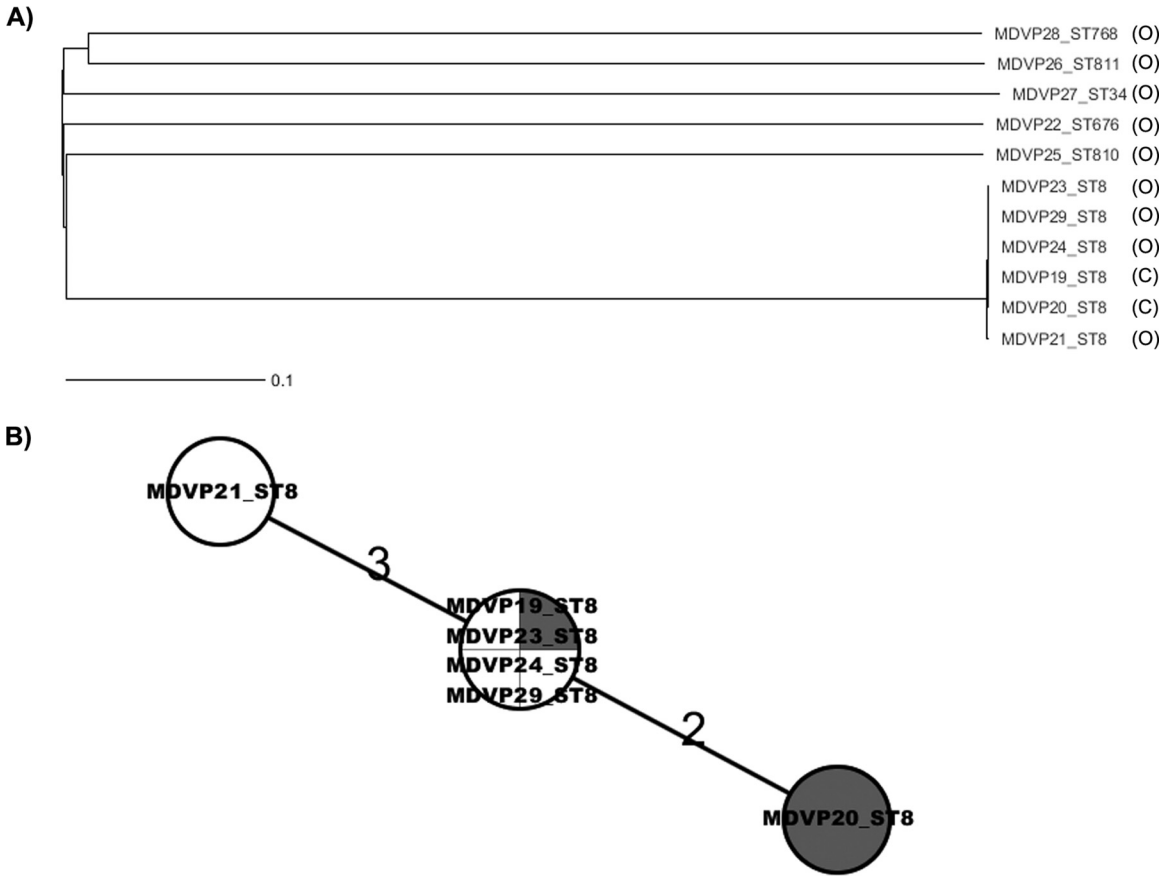

FIG 2 Phylogeny of the *V. parahaemolyticus* strains isolated during the outbreak of 2010 in Maryland assessed by wgMLST analysis. Ridom SeqSphere$^+$ identified 3,896 and 952 loci as core and accessory genes, respectively, for both chromosomes in *V. parahaemolyticus*. (A) NJ tree showing the high level of diversity of *V. parahaemolyticus* strains isolated from oysters and their relationship to the clinical samples (C, clinical; O, oysters) (3,955 loci were shared among the strains analyzed). (B) Minimum spanning tree showing the locus differences among ST8 strains from oysters (no shading) and from clinical samples (shaded). Of 4,349 loci shared by all ST8 strains, there were overall 5 loci differing among the strains, showing clonality of the strains. Also evident is that the ST8 oyster strains were indistinguishable from ST8 clinical strains. This result, combined with the epidemiological data, confirmed that the tested oysters were the source of the outbreak cases. The numbers above the connected lines are locus differences. The lines are not drawn to scale.

TABLE 3 SNP differences among the Maryland ST8 strains, their positions, and locus locations

Target	Position	Abs. position[a]	Product	Nucleotide					
				MDVP20	MDVP23	MDVP21	MDVP29	MDVP24	MDVP19
VP1807	373	1915623	Hypothetical protein	G	A	A	A	A	A
VPA0532	60	536660	Hypothetical protein	G	G	T	G	G	G
VPA1445	461	1538990	Secreted calcium-binding protein	T	T	C	T	T	T
VPA1674	2	1795177	Ribulokinase	A	T	T	T	T	T
VP1647	649	1766093	Methylcitrate synthase	G	G	A	G	G	G

[a] Base position in the genome of RIMD2210633.

Diego, CA), according to the manufacturer's instructions, at ~80× average coverage. The genome libraries were constructed by using a Nextera XT DNA sample prep kit (Illumina). Genomic sequence contigs were *de novo* assembled by using default settings within CLC Genomics Workbench v7.6.1 (Qiagen), with a minimum contig size threshold of 500 bp. The draft genomes were annotated by using the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html) (22).

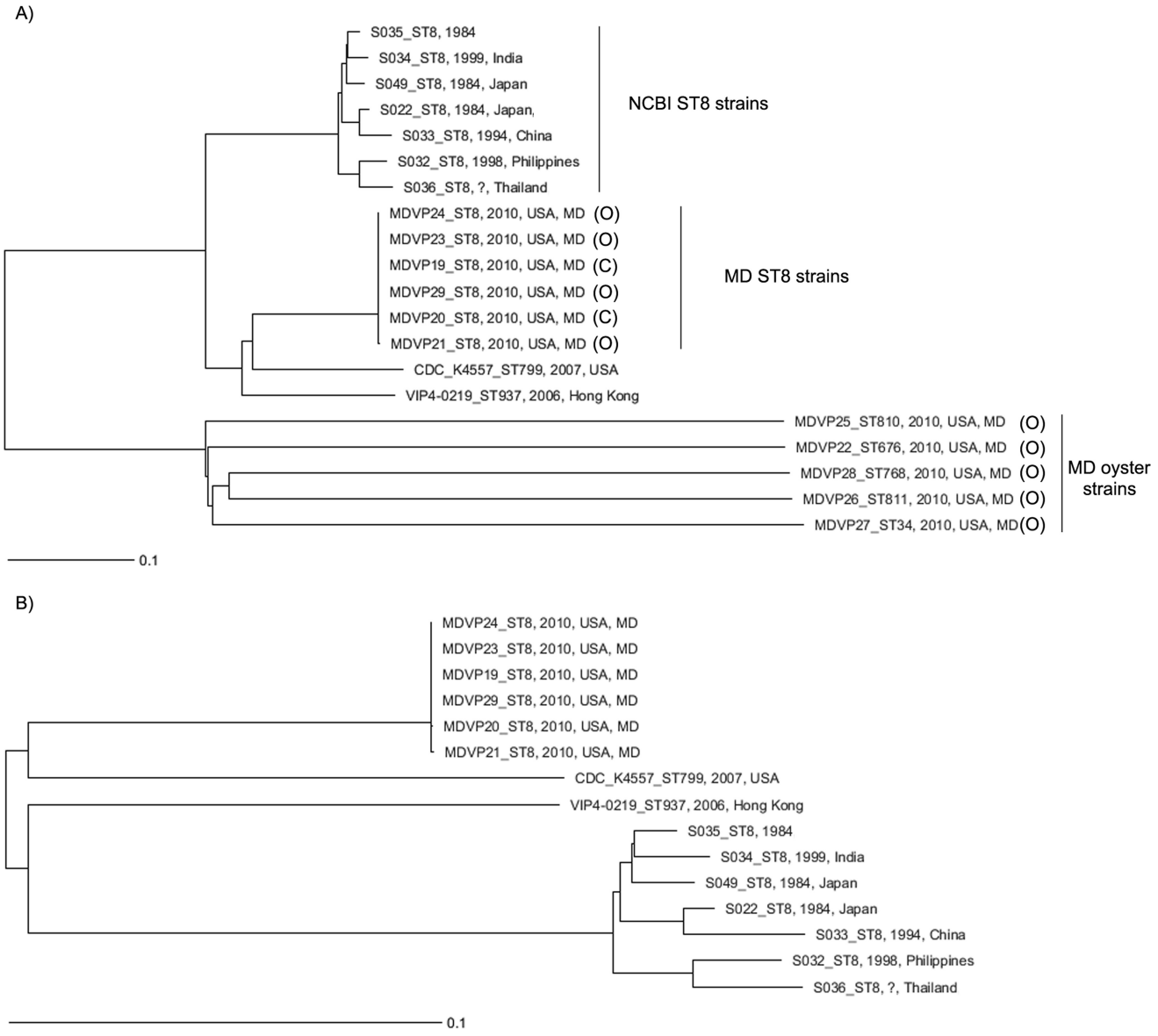

FIG 3 Genome comparison of the 2010 *V. parahaemolyticus* ST8 strains from Maryland with other members of the genome group in the NCBI database by wgMLST analysis. (A) Comparison of all Maryland 2010 strains (clinical and oyster sources) with strains belonging to the ST8 genome group in the NCBI database (using 3,376 total loci). (B) Comparison of genomes of only ST8 Maryland 2010 strains with genomes of the strains belonging to the ST8 genome group in the NCBI database (http://www.ncbi.nlm.nih.gov/genome/genomes/691?genome_assembly_id=group167998).

TABLE 4 List of genomes in GenBank that belong to the same genome group according to genomic BLAST analysis, excluding the Maryland ST8 isolates[b]

Strain	Region	Yr of isolation	Source[a]	Serotype	GenBank accession no.	ST
CDC_K4557	USA	2007	C	O1:K33	NC_021848.1, NC_021822.1	799
S049	Japan	1984	C	O4:K68	AWLM01000000	8
S036	Thailand	?	C	?	AWLZ01000000	8
S035	Japan	1984	C	O4:K53	AWMA01000000	8
S034	India	1999	C	?	AWMB01000000	8
S033	China	1994	C	O5:K60	AWMC01000000	8
S032	Philippines	1998	C	O1:K56	AWMD01000000	8
S022	Japan	1984	C	O5:K15	AWMN01000000	8
VIP4-0219	Hong Kong	2006	E	?	NZ_AXNQ01000000	937

[a] C, clinical; E, environmental.
[b] ?, unknown.

***In silico* MLST phylogenetic analysis.** The initial analysis and identification of the strains were performed by using an *in silico V. parahaemolyticus* MLST approach, based on information available at the *V. parahaemolyticus* MLST Database (http://pubmlst.org/vparahaemolyticus/), and by using Ridom SeqSphere$^+$ software v2.3 (Ridom, Münster, Germany). Seven loci (*dnaE*, *gyrB*, *recA*, *dtdS*, *pntA*, *pyrC*, and *tnaA*) described previously for *V. parahaemolyticus* (23) were used for MLST analysis. The same *V. parahaemolyticus* MLST database was also used to assign numbers for alleles and sequence types (STs).

Assignment to clonal complexes. eBURST v3 (http://eburst.mlst.net/) was used to assign STs to clonal complexes (CCs) using 1,000 bootstrap resamplings (23). In order to be included in a particular CC, isolates needed to share at least 5 of the 7 alleles with members of that CC. Single-locus variants (SLVs) shared at least 6 of the 7 alleles. Double-locus variants (DLVs) were defined as those STs which shared 5 of 7 alleles.

Phylogenomic analysis and targeted SNP analysis. The phylogenetic relationship among these isolates was assessed by wgMLST using Ridom SeqSphere$^+$ software v2.4.0 (20, 24, 25). The genome of strain RIMD2210633 was used as a reference. Core genes were defined as those that were shared among this strain and *V. parahaemolyticus* strain 10329 (GenBank accession number AFBW01000000), and accessory genes were defined as those that were present only in RIMD2210633. Ridom SeqSphere$^+$ identified 2,526 loci as core genes and 587 loci as accessory genes for chromosome I. For chromosome II, 1,370 and 365 loci were identified as core and accessory genes, respectively.

The DNA distance method described previously by Nei et al. (26) was used for calculating the matrix of genetic distance, taking only the number of same/different alleles in the core genes into consideration. After eliminating any loci that were missing from the genomes of any strains used in our analyses, we performed wgMLST analysis. The total number of core genes employed for each analysis varied depending on which strains were being analyzed. In some cases, values are missing for certain loci because that gene either was missing or became truncated due to its position at either end of the *de novo*-assembled contigs. Therefore, the total number of relevant loci in each wgMLST figure was clarified, and a neighbor-joining (NJ) tree using the appropriate genetic distances for each analysis was then constructed.

Nucleotide sequence accession numbers. The draft genome sequences for all 11 *V. parahaemolyticus* strains used in our analyses are available in GenBank under the accession numbers listed in Table 2.

RESULTS

Identification of oyster isolates. Oysters from the growing area implicated in the outbreak were screened for total and pathogenic *V. parahaemolyticus* bacteria. Of the 479 *V. parahaemolyticus* strains isolated, 11 of these isolates were determined to be potentially pathogenic. PFGE two-enzyme analysis found four of the nine *tdh*-negative *trh*$^+$ isolates to be indistinguishable from one of the human isolates (Table 1).

Draft genome assemblies. The draft genomes of the 11 *V. parahaemolyticus* strains isolated during the 2010 Maryland outbreak were generated by whole-genome sequencing using MiSeq. The estimated average coverage for these strains was between 60× and 110×. Genome assembly statistics for each strain are summarized in Table 2. The estimated genome sizes varied between 5.02 and 5.2 Mb. The average G+C content was between 45.1 and 45.3%; these values are within the range reported previously for other *V. parahaemolyticus* strains (27–29).

***In silico* MLST.** *In silico* MLST identified 7 different STs among the 11 sequenced strains. The two clinical strains were of ST8, as were the four oyster strains that matched by PFGE (Table 1 and Fig. 1). The other oyster strains appear to belong to previously undetected STs (ST810, ST811, and ST676), and one was of ST34, which is an ST commonly found in coastal areas of the United States (5, 23), mainly in the Gulf of Mexico (http://pubmlst.org/vparahaemolyticus).

Phylogenetic analysis. The phylogenetic relationships among the *V. parahaemolyticus* 2010 Maryland strains analyzed in this study were determined by using wgMLST analysis (Fig. 2). The *V. parahaemolyticus* strains isolated from outbreak-associated oysters were genetically diverse and belonged to different populations, as documented by the high number of locus differences that defined each branch (minimum spanning tree [MST]) (see Fig. S1 in the supplemental material). Interestingly, the Maryland ST8 strains differed from the other oyster strains by >3,000 loci. Nonetheless, strains with the same ST clustered together. Further wgMLST analysis using only Maryland ST8 strains showed that among the 4,349 loci used for the comparison, 4,344 of them were identical across all these strains, and two strains, MDVP20 and MDVP21, differed by only 2 and 3 loci, respectively (Fig. 2B). The differences in those loci were caused by SNPs (Table 3).

Comparison of ST8 Maryland outbreak strains with closely related *V. parahaemolyticus* genomes in GenBank. Additional wgMLST comparisons of Maryland ST8 strains with their closest relatives available in GenBank (Fig. 3 and Table 4) confirmed that they were highly related and different from the other Maryland oyster strains (Fig. 3A). However, Maryland ST8 strains had important differences from ST8 strains collected from Asia between 1984 and 1999 (Fig. 2A). wgMLST of only Maryland ST8 strains and their genome group members showed that they

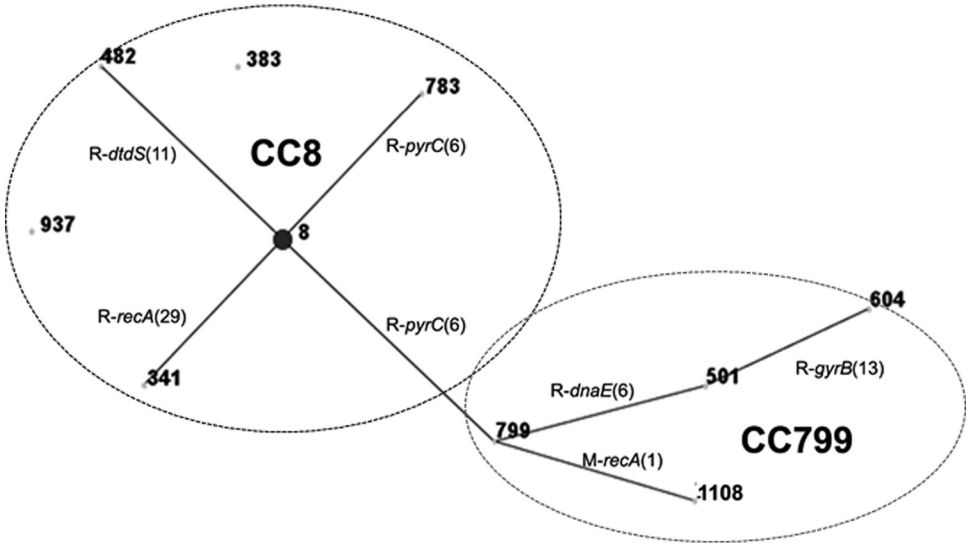

FIG 4 *V. parahaemolyticus* population "snapshot" of CC8 and SLV CC799 obtained by using eBURST v3, using data available in the MLST database. ST8 and ST799 were identified as the predicted clonal ancestors of CC8 and CC799, respectively. STs that are SLVs of each other are shown connected by black lines. Recombination events (R), loci, and the number of SNPs between the connecting STs are shown. Numbers of SNP differences are in parentheses.

were genetically more closely related to strains of other STs belonging to the same genome group that were isolated more recently (VIP4-0219 [ST937], isolated in 2006 in Hong Kong from salmon sashimi, and CDC_K4557 [ST799], isolated in 2006 from stool samples) than to other ST8 strains isolated >10 years ago (Fig. 3B).

Origin of the Maryland ST8 strains. eBURST analysis using other strains from the *V. parahaemolyticus* MLST database showed that the strains within ST8 belong to a CC that encompasses both CC8 and CC799 (Fig. 4 and Tables 4 and 5). ST8 and ST799 are SLVs, and both STs are the predicted ancestral ST of their own respective clonal complexes (Table 4). As observed by whole-genome analysis, the main sources of changes identified by MLST analysis (seven housekeeping genes) are recombination events (Fig. 4; see also Fig. S3 in the supplemental material). As shown in Table 4, 20 strains (representing 11 different STs) already found in the MLST database are members of these two CCs. Interestingly, most strains (95%) belonging to CC8 and CC799 were isolated from Asian sources. The only exception is CDC_K4557 (ST799), which was isolated from a clinical sample in the United States in 2006 (30).

Distribution of the type III secretion system and other genomic regions in *V. parahaemolyticus* strains from this study. While analyzing the genetic makeup of the studied strains, we found that potentially pathogenic *V. parahaemolyticus* strains lacked the full set of pathogenicity islands (PIs) found in pandemic strain RIMD2210633 (VPaI-1 to -7) (7) (Table 6). However, they all carried type III secretion system 2 beta (T3SS-2 beta) (region containing the *trh* gene) (6), hypothetical proteins of unknown function (NK), T3SS-1, osmotolerance (chromosome I), gametolysin and osmotolerance (chromosome II), capsule polysaccharide (CPS), type I secre-

TABLE 5 Sequence types in each of the two CCs (*V. parahaemolyticus* MLST database) identified as being related to the Maryland strains causing the outbreak in 2010 (ST8) by eBURST analysis[a]

CC	ST	Frequency (no. of strains)[b]	Variant	Country(ies) (no. of strains)	Yr of isolation	Source(s) (no. of strains)
8	8	10[c]	Ancestral type	China (5), Japan (2), Philippines (1), India (1), Thailand (1)	1984–2008	C (9), E(1)
	341	1	SLV	China	2010	C
	783	1	SLV	China	2008	C
	482	1	SLV	China	2010	C
	1016	1	SLV	NA	NA	NA
	383	1	DLV	China	2005	E
	937	1	DLV	China	2006	E
799	799	1	Ancestral type	USA	2006	C
	1108	1	SLV	China	2006	E
	501	1	SLV	China	2008	E
	604	1	DLV	China	NA	NA

[a] C, clinical; E, environmental; NA, information not available.
[b] Number of strains in the *V. parahaemolyticus* MLST database.
[c] Excluding the six ST8 strains from the Maryland 2010 outbreak.

TABLE 6 *In silico* screen for pathogenicity markers tested for *V. parahaemolyticus* strains sequenced in this study

Location and pathogenicity marker	Region or source	Presence or type of marker[a]					
		MDVP20	MDVP22	MDVP25	MDVP26	MDVP27	MDVP28
Chromosome I							
NK	VP0081–VP0092	+	+	+	+	+	+
LPS	VP0218–VP0234	−	−	−	−	−	−
VPaI-1	VP0380–VP0403	−	−	−	−	−	−
VPaI-2	VP0634–VP0643	−	+ (dif)	−	−	−	−
VPaI-3	VP1071–VP1095	−	−	−	−	−	−
T6SS	VP1386–VP1420	+	+	−	+	+	+
Phage f237	VP1549–VP1590	+ (dif)	+ (dif)	+ (dif)	+ (dif)	−	+ (dif)
T3SS-1	VP1658–VP1702	+	+	+	+	+	+
Osmotolerance	VP1719–VP1728	+	+	+	+	+	+
Integron class 1	VP1787–VP1865	+ (dif)	+ (dif)	+ (dif)	+ (dif)	+ (dif)	+ (dif)
VPaI-4	VP2131–VP2144	−	−	−	−	−	−
VPaI-5	VP2900–VP2910	−	−	−	−	−	−
Chromosome II							
Degradative	VPA0434–VPA0458	−	+	−	−	−	−
Phage f237-like	VPA0887–VPA0914	+ (dif)	+ (dif)	+ (dif)	+ (dif)	+ (dif)	−
Biofilm	VPA0950–VPA0962	−	−	−	−	−	−
Gametolysin	VPA0989–VPA0999	+	+	+	+	+	+
Osmotolerance	VPA1102–VPA1115	+	+	+	+	+	+
VPaI-6	VPA1253–VPA1270	−	−	−	−	−	−
VPaI-7 (T3SS-2 alpha)	VPA1312–VPA1395	−	−	−	−	−	−
CPS	VPA1403–VPA1412	+	+	+	+	+	+
Type I secretion	VPA1440–VPA1444	+	+	+	+	+	+
Type I pilus	VPA1503–VPA1521	+	+	+	+	+	+
Multidrug efflux	VPA1559–VPA1583	+	+	+	+	+	+
Ferric uptake	VPA1652–VPA1679	+	+	+	+	+	+
T3SS-2 beta	TH3996 *trh1* region[b]	+	+	+	+	+	+
tdh type		−	2	−	−	−	−
trh type		2	1	2	2	2	2
ST	MLST *V. parahaemolyticus* website	8	676	810	811	34	768

[a] +, present; −, absent; + (dif), present but with a different sequence or element.
[b] See GenBank accession number AB455531.

tion, type I pilus, multidrug efflux, and ferric uptake genes. Also, most of these strains carried a T6SS that differed from that of the pandemic strain.

DISCUSSION

In this study, we found that *V. parahaemolyticus* strains causing an outbreak in Maryland in 2010 and isolated from stool samples belonged to ST8. *Vibrio parahaemolyticus* strains isolated from oysters implicated in this outbreak had a PFGE pattern indistinguishable from that of the clinical strains, belonged to ST8, and were nearly indistinguishable by wgMLST (differing in 2 or 3 loci of 4,349 loci tested), with these differences being caused by SNPs. These results confirm that Chesapeake Bay oysters were the source of the vibriosis cases in 2010, since the genomic identity at the nucleotide level between oyster and clinical strains was >99.999%, which is the level of similarity that can be found between colonies of the same strain (11).

However, our research during the 2010 outbreak investigations revealed that the oysters analyzed also carried other potentially pathogenic *V. parahaemolyticus* strains that belong to different populations and differed in >3,000 loci from the outbreak strains (see Fig. S3 and Table S1 in the supplemental material). There were 115,875 SNPs within these 3,000 loci (see Table S2 in the supplemental material). One oyster strain, of ST34 and carrying *trh2*, was linked previously to clinical cases (31) (http://pubmlst.org/vparahaemolyticus). Such strains, regardless of their pathogenic potential, were not observed among any other clinical cases from 2010. This could have been because their concentration was lower in oysters and not high enough to cause illnesses but high enough to be detectable by the MPN method. Sporadic illnesses and/or small outbreaks often go unnoticed, and therefore, the amount of illness caused by any strain of *V. parahaemolyticus* is probably underreported (1).

Genomic analysis of all potentially pathogenic *V. parahaemolyticus* strains isolated from the oysters implicated in the outbreak showed that they carried several pathogenicity-related genes (i.e., type III secretion effectors) besides the *tdh* and/or *trh* genes. However, they lacked all pathogenicity islands described for pandemic strain RIMD2210633 (7). We believe that these strains probably

have their own PIs that have not yet been described, but this speculation must be interpreted with caution since a more detailed investigation using closed genomes could reveal additional elements that cannot currently be assessed.

When we compared Maryland ST8 strains with their most closely related genomic sequences available in GenBank (Table 3; see also Fig. S1 in the supplemental material), we found that the Maryland ST8 strains were more closely related to strains of other STs in their genome group than to the other ST8 strains (Fig. 3A). The significant genomic differences between both groups of ST8 strains could be explained by the time lapse between both samplings (>20 years separate the two ST8 groups), and habitat conditions/selective pressures in Asia could be different from those in the Chesapeake Bay, all of which could result in significant genomic differences.

In *V. parahaemolyticus*, the typical mechanism of evolution is believed to be recombination instead of mutation, with recombination/mutation ratios estimated to be 2.5:1 and 8.8:1 by allele and site, respectively (23). An example from the current set of isolates is MDVP23, which differs from S035 by 806 loci, but analysis at the SNP level shows that these strains differ by 7,562 SNPs within these loci, with most of the loci containing possible recombination signatures of between 5 and 6 SNPs per locus, for an estimated recombination ratio of ~9:1 by locus (see Fig. S3 in the supplemental material). Our empirical calculations suggest that strains of *V. parahaemolyticus* of ST8 are experiencing strong evolutionary pressures favoring multiple recombination events.

How nonautochthonous *V. parahaemolyticus* strains from Asia can be present in oysters harvested from the eastern coast of the United States, specifically in the Chesapeake Bay, remains a matter of speculation. This bay has a lot of maritime traffic, and ballast water from ships coming from Asia, ocean currents, or other events such as the introduction of nonnative oysters or exotic fish may have introduced the nonautochthonous strains into this area. Since 2010, ST8 strains have not been linked to any additional illnesses, and most of the outbreaks in Maryland (2012 to 2013) have been linked to autochthonous East Coast U.S. strains (e.g., ST631) (5) or pandemic strains (10). The disappearance of the ST8 strains from Maryland could have happened due to natural replacement by new or autochthonous *V. parahaemolyticus* strains and/or by the action of bacteriophages (usually present in higher levels in seawater [32]) after the probable unknown source was removed from the Chesapeake environment. The reduction or elimination of *V. parahaemolyticus* by phages has been described for pandemic strains in the south of Chile (33), and other authors have employed bacteriophage therapy to reduce *Vibrio* numbers in oysters (34–36).

Taken together, our findings demonstrate that whole-genome sequencing allowed a detailed retrospective study of outbreak and nonoutbreak strains of *V. parahaemolyticus*, revealing their phylogenetic relationships and confirming their local vector and their likely path from Asia to the Chesapeake Bay. The wgMLST method employed was easy, robust, and scalable to multiple strains to be used in future *V. parahaemolyticus* outbreak investigations. Furthermore, we demonstrate the potential consequences of nonautochthonous *V. parahaemolyticus* strains introduced into a new habitat.

ACKNOWLEDGMENTS

This project was supported by FDA Foods Program intramural funds.

We thank Lili Fox Vélez for her scientific writing assistance on the manuscript.

REFERENCES

1. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States—major pathogens. Emerg Infect Dis **17:**7–15. http://dx.doi.org/10.3201/eid1701.P11101.
2. **Turner JW, Paranjpye RN, Landis ED, Biryukov SV, Gonzalez-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:**e55726. http://dx.doi.org/10.1371/journal.pone.0055726.
3. **DePaola A, Kaysner CA, Bowers J, Cook DW.** 2000. Environmental investigations of *Vibrio parahaemolyticus* in oysters after outbreaks in Washington, Texas, and New York (1997 and 1998). Appl Environ Microbiol **66:**4649–4654. http://dx.doi.org/10.1128/AEM.66.11.4649-4654.2000.
4. **Noriea NF, III, Johnson CN, Griffitt KJ, Grimes DJ.** 2010. Distribution of type III secretion systems in *Vibrio parahaemolyticus* from the northern Gulf of Mexico. J Appl Microbiol **109:**953–962. http://dx.doi.org/10.1111/j.1365-2672.2010.04722.x.
5. **Haendiges J, Timme R, Allard MW, Myers RA, Brown EW, Gonzalez-Escalona N.** 2015. Characterization of *Vibrio parahaemolyticus* clinical strains from Maryland (2012-2013) and comparisons to a locally and globally diverse *V. parahaemolyticus* strains by whole-genome sequence analysis. Front Microbiol **6:**125. http://dx.doi.org/10.3389/fmicb.2015.00125.
6. **Park KS, Iida T, Yamaichi Y, Oyagi T, Yamamoto K, Honda T.** 2000. Genetic characterization of DNA region containing the *trh* and *ure* genes of *Vibrio parahaemolyticus*. Infect Immun **68:**5742–5748. http://dx.doi.org/10.1128/IAI.68.10.5742-5748.2000.
7. **Boyd EF, Cohen AL, Naughton LM, Ussery DW, Binnewies TT, Stine OC, Parent MA.** 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. BMC Microbiol **8:**110. http://dx.doi.org/10.1186/1471-2180-8-110.
8. **Martinez-Urtaza J, Baker-Austin C, Jones JL, Newton AE, Gonzalez-Aviles GD, DePaola A.** 2013. Spread of Pacific Northwest *Vibrio parahaemolyticus* strain. N Engl J Med **369:**1573–1574. http://dx.doi.org/10.1056/NEJMc1305535.
9. **Newton AE, Garrett N, Stroika SG, Halpin JL, Turnsek M, Mody RK.** 2014. Notes from the field: increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic Coast shellfish—2013. MMWR Morb Mortal Wkly Rep **63:**335–336.
10. **Haendiges J, Rock M, Myers RA, Brown EW, Evans P, Gonzalez-Escalona N.** 2014. Pandemic *Vibrio parahaemolyticus*, Maryland, USA, 2012. Emerg Infect Dis **20:**718–720. http://dx.doi.org/10.3201/eid2004.130818.
11. **Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.** 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13:**32. http://dx.doi.org/10.1186/1471-2164-13-32.
12. **Bakker HC, Switt AI, Cummings CA, Hoelzer K, Degoricija L, Rodriguez-Rivera LD, Wright EM, Fang R, Davis M, Root T, Schoonmaker-Bopp D, Musser KA, Villamil E, Waechter H, Kornstein L, Furtado MR, Wiedmann M.** 2011. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type. Appl Environ Microbiol **77:**8648–8655. http://dx.doi.org/10.1128/AEM.06538-11.
13. **Chin C-S, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK.** 2011. The origin of the Haitian cholera outbreak strain. N Engl J Med **364:**33–42. http://dx.doi.org/10.1056/NEJMoa1012928.
14. **Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Moller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK.** 2011. Origins of the *E. coli* strain

causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med **365:**709–717. http://dx.doi.org/10.1056/NEJMoa1106920.

15. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One **8:**e55254. http://dx.doi.org/10.1371/journal.pone.0055254.
16. **Gonzalez-Escalona N, Timme R, Raphael BH, Zink D, Sharma SK.** 2014. Whole-genome single-nucleotide-polymorphism analysis for discrimination of *Clostridium botulinum* group I strains. Appl Environ Microbiol **80:**2125–2132. http://dx.doi.org/10.1128/AEM.03934-13.
17. **Kovanen SM, Kivisto RI, Rossi M, Schott T, Karkkainen UM, Tuuminen T, Uksila J, Rautelin H, Hanninen ML.** 2014. Multilocus sequence typing (MLST) and whole-genome MLST of *Campylobacter jejuni* isolates from human infections in three districts during a seasonal peak in Finland. J Clin Microbiol **52:**4147–4154. http://dx.doi.org/10.1128/JCM.01959-14.
18. **Jolley KA, Maiden MC.** 2014. Using MLST to study bacterial variation: prospects in the genomic era. Future Microbiol **9:**623–630. http://dx.doi.org/10.2217/fmb.14.24.
19. **Schmid D, Allerberger F, Huhulescu S, Pietzka A, Amar C, Kleta S, Prager R, Preussel K, Aichinger E, Mellmann A.** 2014. Whole genome sequencing as a tool to investigate a cluster of seven cases of listeriosis in Austria and Germany, 2011–2013. Clin Microbiol Infect **20:**431–436. http://dx.doi.org/10.1111/1469-0691.12638.
20. **Kohl TA, Diel R, Harmsen D, Rothganger J, Walter KM, Merker M, Weniger T, Niemann S.** 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. J Clin Microbiol **52:**2479–2486. http://dx.doi.org/10.1128/JCM.00567-14.
21. **Nordstrom JL, Vickery MC, Blackstone GM, Murray SL, DePaola A.** 2007. 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 **73:**5840–5847. http://dx.doi.org/10.1128/AEM.00460-07.
22. **Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufo S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S, Schafer S, Tolstoy I, Tatusova T.** 2009. The National Center for Biotechnology Information's Protein Clusters Database. Nucleic Acids Res **37:**D216–D223. http://dx.doi.org/10.1093/nar/gkn734.
23. **Gonzalez-Escalona N, Martinez-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:**2831–2840. http://dx.doi.org/10.1128/JB.01808-07.
24. **Bletz S, Bielaszewska M, Leopold SR, Kock R, Witten A, Schuldes J, Zhang W, Karch H, Mellmann A.** 2013. Evolution of enterohemorrhagic *Escherichia coli* O26 based on single-nucleotide polymorphisms. Genome Biol Evol **5:**1807–1816. http://dx.doi.org/10.1093/gbe/evt136.
25. **Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D, Mellmann A.** 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. J Clin Microbiol **53:**2869–2876. http://dx.doi.org/10.1128/JCM.01193-15.
26. **Nei M, Tajima F, Tateno Y.** 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol **19:**153–170.
27. **Gonzalez-Escalona N, Gavilan RG, Brown EW, Martinez-Urtaza J.** 2015. Transoceanic spreading of pathogenic strains of *Vibrio parahaemolyticus* with distinctive genetic signatures in the *recA* gene. PLoS One **10:**e0117485. http://dx.doi.org/10.1371/journal.pone.0117485.
28. **Banerjee S, Petronella N, Chew LC, Farber J.** 2015. Draft genome sequences of four *Vibrio parahaemolyticus* isolates from clinical cases in Canada. Genome Announc **3**(1):e01482-14. http://dx.doi.org/10.1128/genomeA.01482-14.
29. **Kalburge SS, Polson SW, Boyd Crotty K, Katz L, Turnsek M, Tarr CL, Martinez-Urtaza J, Boyd EF.** 2014. Complete genome sequence of *Vibrio parahaemolyticus* environmental strain UCM-V493. Genome Announc **2**(2):e00159-14. http://dx.doi.org/10.1128/genomeA.00159-14.
30. **Ludeke CH, Kong N, Weimer BC, Fischer M, Jones JL.** 2015. Complete genome sequences of a clinical isolate and an environmental isolate of *Vibrio parahaemolyticus*. Genome Announc **3**(2):e00216-15. http://dx.doi.org/10.1128/genomeA.00216-15.
31. **Ellingsen AB, Olsen JS, Granum PE, Rorvik LM, Gonzalez-Escalona N.** 2013. Genetic characterization of *trh* positive *Vibrio* spp. isolated from Norway. Front Cell Infect Microbiol **3:**107. http://dx.doi.org/10.3389/fcimb.2013.00107.
32. **Jiang SC, Paul JH.** 1998. Gene transfer by transduction in the marine environment. Appl Environ Microbiol **64:**2780–2787.
33. **Garcia K, Bastias R, Higuera G, Torres R, Mellado A, Uribe P, Espejo RT.** 2013. Rise and fall of pandemic *Vibrio parahaemolyticus* serotype O3:K6 in southern Chile. Environ Microbiol **15:**527–534. http://dx.doi.org/10.1111/j.1462-2920.2012.02883.x.
34. **Jun JW, Kim HJ, Yun SK, Chai JY, Park SC.** 2014. Eating oysters without risk of vibriosis: application of a bacteriophage against *Vibrio parahaemolyticus* in oysters. Int J Food Microbiol **188:**31–35. http://dx.doi.org/10.1016/j.ijfoodmicro.2014.07.007.
35. **Tskhvediani A, Khukhunashvili T, Eliashvili T, Tsertsvadze G, Gachechiladze N, Tediashvili M.** 2014. The possible use of *V. parahaemolyticus*-specific bacteriophages for prevention and therapy of infections caused by *V. parahaemolyticus*. Georgian Med News **2014**(231):82–88.
36. **Pereira C, Silva YJ, Santos AL, Cunha A, Gomes NC, Almeida A.** 2011. Bacteriophages with potential for inactivation of fish pathogenic bacteria: survival, host specificity and effect on bacterial community structure. Mar Drugs **9:**2236–2255. http://dx.doi.org/10.3390/md9112236.