

# Determination of Molecular Phylogenetics of *Vibrio parahaemolyticus* Strains by Multilocus Sequence Typing<sup>∇†</sup>

Narjol González-Escalona,<sup>1,5\*</sup> Jaime Martínez-Urtaza,<sup>2</sup> Jaime Romero,<sup>3</sup> Romilio T. Espejo,<sup>3</sup> Lee-Ann Jaykus,<sup>1</sup> and Angelo DePaola<sup>4</sup>

Department of Food Science, North Carolina State University, Raleigh, North Carolina<sup>1</sup>; Instituto de Acuicultura, Universidad de Santiago de Compostela, Campus Universitario Sur, 15782 Santiago de Compostela, Spain<sup>2</sup>; Laboratorio de Biotecnología, Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Santiago, Chile<sup>3</sup>; Gulf Coast Seafood Laboratory, Food and Drug Administration, Dauphin Island, Alabama<sup>4</sup>; and Center for Food and Applied Nutrition, Food and Drug Administration, College Park, Maryland<sup>5</sup>

*Vibrio parahaemolyticus* is an important human pathogen whose transmission is associated with the consumption of contaminated seafood. There is a growing public health concern due to the emergence of a pandemic strain causing severe outbreaks worldwide. Many questions remain unanswered regarding the evolution and population structure of *V. parahaemolyticus*. In this work, we describe a multilocus sequence typing (MLST) scheme for *V. parahaemolyticus* based on the internal fragment sequences of seven housekeeping genes. This MLST scheme was applied to 100 *V. parahaemolyticus* strains isolated from geographically diverse clinical ( $n = 37$ ) and environmental ( $n = 63$ ) sources. The sequences obtained from this work were deposited and are available in a public database (<http://pubmlst.org/vparahaemolyticus>). Sixty-two unique sequence types were identified, and most (50) were represented by a single isolate, suggesting a high level of genetic diversity. Three major clonal complexes were identified by eBURST analysis. Separate clonal complexes were observed for *V. parahaemolyticus* isolates originating from the Pacific and Gulf coasts of the United States, while a third clonal complex consisted of strains belonging to the pandemic clonal complex with worldwide distribution. The data reported in this study indicate that *V. parahaemolyticus* is genetically diverse with a semiclinal population structure and an epidemic structure similar to that of *Vibrio cholerae*. Genetic diversity in *V. parahaemolyticus* appears to be driven primarily by frequent recombination rather than mutation, with recombination ratios estimated at 2.5:1 and 8.8:1 by allele and site, respectively. Application of this MLST scheme to more *V. parahaemolyticus* strains and by different laboratories will facilitate production of a global picture of the epidemiology and evolution of this pathogen.

*Vibrio parahaemolyticus* is a natural inhabitant of coastal waters and is the leading cause of seafood-borne gastroenteritis (34). Since 1996 an increasing number of *V. parahaemolyticus* infections and outbreaks caused by strains belonging to a pandemic clonal complex have been observed throughout the world (2, 8, 10, 17, 18, 20, 32, 33, 38, 39). The emergence of this clonal complex has elevated public health concerns for pandemic spread previously uncharacteristic of *V. parahaemolyticus*. Serology, the historic mainstay of *V. parahaemolyticus* surveillance, has been unreliable in tracking the spread of the epidemic clonal complex because in addition to the original O3:K6 serotype, at least 11 other serovariants have been identified (2, 7, 9). These new serovariants are highly similar to or indistinguishable from the original O3:K6 strains by a variety of molecular fingerprinting techniques, including arbitrarily primed PCR, pulsed-field gel electrophoresis, ribotyping, the intergenic spacer region between the 16S and 23S rRNA genes, and direct genome restriction enzyme analysis (8, 18, 20, 33, 38). Accordingly, the molecular subtyping techniques, which

usually demonstrate a high degree of diversity for *V. parahaemolyticus*, provide very limited information on the genetic relatedness of strains. Further, these methods are of limited value in elucidating the evolution of the clonal complex of *V. parahaemolyticus* or its relationship to other *V. parahaemolyticus* strains or marine bacteria. The existence of other clonal complexes of *V. parahaemolyticus* is highly probable. However, until now there has been no description of these clonal complexes and how they may be evolving relative to the pandemic clonal complex.

Multilocus sequence typing (MLST) is based on sequence analysis of chosen housekeeping (HK) genes and is becoming the method of choice for determining the global epidemiology of bacterial pathogens (e.g., *Neisseria meningitidis* and *Staphylococcus aureus*) (29, 30). Being sequence-based, MLST provides a definitive characterization of bacterial isolates that is consistent from one laboratory to the next. The nuclei acid sequences are typically stored in a public database that can be readily accessed via the Internet (<http://www.mlst.net> or <http://pubmlst.org/>). Previous MLST studies have led to better understanding of the genetic relatedness of strains within a species and have identified the relative evolutionary importance of mutations and lateral transfer events (13, 16, 35).

A previous MLST study was conducted to investigate the evolution of the new pandemic strains of *V. parahaemolyticus*, and the authors hypothesized their evolution from a common

\* Corresponding author. Mailing address: FDA CFSAN, 5100 Paint Branch Parkway, College Park, MD 20740. Phone: (301) 436-1937. Fax: (301) 436-2644. E-mail: narjol@gmail.com.

† Supplemental material for this article may be found at <http://jba.asm.org/>.

O3:K6 ancestor (9). That study was limited primarily to pandemic strains and examined only four genes, all located in chromosome I. Usually six to eight genes are examined to provide a more comprehensive picture of the genetic characteristics of the organism analyzed (29). Additionally, *Vibrio* spp. (including *V. parahaemolyticus*) possess two chromosomes, and it would be useful to determine whether both of these are subjected to similar evolutionary pressures. Thus, the development of a more comprehensive MLST scheme for *V. parahaemolyticus* and one which is comparable to that used for other bacteria is warranted.

We report on the first MLST scheme for *V. parahaemolyticus* using sequences of internal fragments of seven HK genes. In order to better represent their population structure, three genes from chromosome I and four from chromosome II were chosen. A well-characterized and geographically diverse set of *V. parahaemolyticus* isolates were examined. Isolates belonging to the *V. parahaemolyticus* pandemic clonal complex isolated from eight countries in four continents were also analyzed.

## MATERIALS AND METHODS

**Bacterial strains and media.** One hundred *V. parahaemolyticus* isolates were included in this analysis, 37 from clinical and 63 from environmental sources (see Table S1 in the supplemental material). These isolates were temporally (collected from 1951 to 2005) and geographically (collected from Asia, Europe, South America, and the United States) diverse. Twenty-five of them were classified previously as belonging to the *V. parahaemolyticus* pandemic clonal complex and were isolated in the United States, Chile, Spain, Peru, Korea, Japan, Thailand, and Bangladesh; most of these were of the O3:K6 serotype. Strains were grown overnight at 37°C in Luria-Bertani (LB) medium with 3% NaCl at 250 rpm.

**DNA extraction and PCR amplification.** Bacterial DNA was extracted using the DNeasy kit in accordance with the manufacturer's instructions (Qiagen, Valencia, CA). PCR amplification was carried out using primers (IDT, Coralville, IA) detailed on the *V. parahaemolyticus* MLST website (<http://pubmlst.org/vparahaemolyticus>). The seven loci analyzed by MLST were dispersed on both chromosomes. These genes were chosen based on two previously published reports of MLST data for *V. parahaemolyticus* and for *V. vulnificus* (3, 9). For chromosome I, the HK genes chosen were *recA* (RecA protein), *dnaE* (DNA polymerase III, alpha subunit), and *gyrB* (DNA gyrase, subunit B). For chromosome II, the HK genes were *dtlS* (threonine 3-dehydrogenase), *pntA* (transhydrogenase alpha subunit), *pyrC* (dihydro-oroate), and *tnaA* (tryptophanase). PCR conditions were as follows; denaturation at 96°C for 1 min, primer annealing at 58°C for 1 min, and extension at 72°C for 1 min for 30 cycles, with a final extension step at 72°C for 10 min. In a few cases the temperature of annealing was decreased to obtain PCR product. Reagent concentrations in 100 µl of PCR mixture were 1.5 mM MgCl<sub>2</sub>, 0.125 mM deoxynucleoside triphosphates, 0.5 µM each primer (forward and reverse for each locus), and 1 U of Platinum *Taq* DNA high-fidelity polymerase (Invitrogen, Carlsbad, CA). One nanogram of DNA was used as template per PCR, and PCR products (10 µl each) were analyzed on 2% agarose gels run at 75 V for 1:30 h in 1 × Tris-acetate-EDTA with a molecular weight standard (Lambda DNA/HindIII; Promega, Madison, WI) for determining amplicon concentrations. Amplification products were visualized by ethidium bromide staining. All PCR products consisted of a single band, and these were cleaned using the QIAquick PCR purification kit (Qiagen). PCR products were sequenced in both directions by Mclab (South San Francisco, CA) with primers M13F and M13R. DNA sequences were individually inspected and manually assembled. The alignments of these sequences were determined using BioEdit (21). Numbers for alleles and sequence types (STs) were assigned according to the database created for *V. parahaemolyticus* (<http://pubmlst.org/vparahaemolyticus>) (25).

**Assignment to clonal complexes.** The program eBURST v 3.0 was used to identify the different clonal complexes (<http://eburst.mlst.net>). The most restrictive group definition was used to define the clonal complexes, i.e., at least six of the seven alleles had to be identical to be included in the same group or clonal complex (15). The statistical confidences for the ancestral types were assessed using 1,000 bootstrap resamplings. Two different STs are considered single-locus

TABLE 1. Nucleotide sequence variation for each MLST locus

Chromosome and locus	Fragment size (bp)	No. of alleles	Nucleotide diversity (SE)	No. of polymorphic sites	$d_S/d_N$ ratio ( $P$ value) <sup>a</sup>
<b>I</b>					
<i>dnaE</i>	557	38	0.032 (0.004)	112	9.420 (0.000)
<i>gyrB</i>	592	36	0.031 (0.004)	94	6.517 (0.000)
<i>recA</i>	729	33	0.029 (0.003)	112	5.230 (0.000)
<b>II</b>					
<i>dtlS</i>	458	37	0.036 (0.005)	67	4.992 (0.000)
<i>pntA</i>	430	34	0.012 (0.003)	34	5.529 (0.000)
<i>pyrC</i>	493	34	0.013 (0.003)	37	2.455 (0.015)
<i>tnaA</i>	423	28	0.012 (0.003)	31	4.201 (0.000)

<sup>a</sup> The hypothesis tested was neutrality ( $d_S = d_N$ ). The  $P$  value indicates the significance of the difference between  $d_S$  and  $d_N$  and must be <0.05 for hypothesis rejection at a 5% level.

variant (SLV) when they differ from each other at a single locus. Double-locus variants (DLVs) are any two different STs differing in two loci.

**Phylogenetic analysis.** Minimum-evolution (ME) trees for each locus and for the concatenated sequences of each ST (3,682 bp) were constructed by Mega 3.1 software (27) using the Kimura two-parameter model to estimate the genetic distances. The statistical support of the nodes in the ME tree was assessed by 1,000 bootstrap resamplings. The nucleotide diversity of each locus and their respective standard errors were determined using Mega 3.1 as described elsewhere (30).

**Estimates of recombination rates.** Estimation of recombination rates was done as described previously (14, 35), where the per-allele and per-site recombination/mutation ( $r/m$ ) parameter was calculated empirically. Briefly, any SLV allele differing by one nucleotide and not observed elsewhere in the database as part of another ST was considered to have arisen by mutation. An SLV allele differing by multiple nucleotides or containing a single nucleotide change observed as part of another ST in the database was considered to have originated by recombination.

**Test for recombination.** The START version 2.0 software package (26) was used to calculate the "standardized" index of association ( $I^S_A$ ) (22). This statistical method tests for the null hypothesis of linkage equilibrium; i.e., if  $I^S_A = 0$ , then alleles are independently distributed at all loci analyzed (alleles are in linkage equilibrium) and recombination occurred frequently. The ratio between the numbers of synonymous ( $d_S$ ) and nonsynonymous ( $d_N$ ) substitutions was calculated by the method of Nei and Gojobori with the Jukes-Cantor correction implemented in Mega 3.1 (27). This measures the type of selection occurring at each locus. The hypothesis tested was for neutrality ( $d_S = d_N$ ); if  $d_S/d_N > 1$ , then nonsynonymous sites are under selective constraint or purifying pressure (negative selection);  $d_S/d_N < 1$  indicates positive selection, and  $d_S/d_N = 1$  indicates neutrality. Congruence among the seven genes was determined as described by Brown et al. (4) employing the incongruence length difference (ILD) test (11). The version of the ILD test employed here is available in PAUP\* v.4.0b (41). An exception was that for the ILD tests a heuristic model was performed instead of branch and bound searches. Both split-tree generation for individual loci and the phi test ( $\phi_w$ ) for recombination were done using the SplitsTree v 4.8 software (23).

**Nucleotide sequence accession numbers.** *recA*, *dnaE*, *gyrB*, *dtlS*, *pntA*, *pyrC*, and *tnaA* sequences were deposited in GenBank under accession numbers EU051383 to EU051622 and are also available at <http://pubmlst.org/vparahaemolyticus>.

## RESULTS

**Nucleotide diversity at each locus and STs generated by MLST.** One hundred *V. parahaemolyticus* isolates were analyzed by MLST using the sequences generated from internal fragments of seven HK genes, and these data are summarized in Table 1. These internal fragments ranged in size from 423 bp (*tnaA*) to 729 bp (*recA*). The number of alleles observed for each locus ranged from 28 (*tnaA*) to 38 (*dnaE*). The number of polymorphic sites observed varied per locus from 31 (*tnaA*) to

112 (*dnaE*). The most frequently found alleles per locus analyzed were *dnaE3* (24), *gyrB4* (26), *recA19* (27), *dtdS4* (25), *pntA29* (27), *pyrC4* (26), and *tnaA22* (27). These are the alleles present in the pandemic strain, and their abundance probably reflects merely the abundance of isolates from this clone among the collected strains. The nucleotide diversity ranged from 0.012 to 0.036, with the lowest degrees of diversity observed for *pntA* (0.012), *pyrC* (0.013), and *tnaA* (0.012). The ratio of synonymous to nonsynonymous substitutions ( $d_S/d_N$ ) was higher than 1 for each locus analyzed using a selection test for neutrality (Table 1). This indicated that nonsynonymous sites are evolving slower than synonymous sites and suggests that most amino acid substitutions seem to be deleterious. This effect is usually observed in HK genes, where purifying selection dominates (6).

**Identification of clonal complexes.** Sixty-two different STs were identified by MLST, indicating a high degree of genotypic diversity (Table 2). Fifty of the STs contained single isolates, while 12 STs included between 2 and 22 isolates. ST-3 was most frequent (22 out of 100 isolates) and was composed of strains belonging to the *V. parahaemolyticus* pandemic complex. These isolates originated from four continents and belonged to four different serotypes (see Table S1 and S2 in the supplemental material). Other STs were also represented by more than one strain: ST-36 (seven isolates); ST-30 (three isolates); and ST-12, ST-17, ST-23, ST-34, ST-35, ST-43, ST-50, ST-58, and ST-59 (two isolates each).

The 62 STs generated in this data set were separated by eBURST into three clonal complexes, six doublets, and 38 singletons (Fig. 1 and Table 2). We designated the clonal complexes CC3, CC36, and CC34. CC3 corresponded to the pandemic strain and was comprised of 25 strains with four different STs (ST-3, ST-42, ST-27, and ST-51); ST-3 was defined by eBURST as the ancestral type or founder for this clonal complex. CC36 was comprised of 10 strains, for which ST-36 was identified as the ancestral type. CC34 contained six strains, for which ST-34 was the ancestral type. Doublet 6 (D6) contained ST-21 and ST-59 and was highly related to CC36, as those STs were DLVs of the founder of that clonal complex (ST-36). Similarly, ST-32 may belong to CC34 as well, since it is a DLV of the founder of that clonal complex (ST-34). The singletons did not belong to any of the clonal complexes or other groups identified in this study. However, some of these singletons (Fig. 1) (S43-S44, S5-S23, and S6-S7) shared five of the seven loci; suggesting that they were more related to each other than to the rest of the population.

**Geographical distribution of the STs.** The geographical distributions of the different STs are shown in Table 2. ST-3, consisting of *V. parahaemolyticus* pandemic strains, was isolated in four continents. Three *V. parahaemolyticus* strains representing ST-27, ST-42, and ST-51 were SLVs of ST-3 (all at the *dnaE* locus), and together they constitute the pandemic CC3. CC3 strains were isolated primarily in Chile, India, Korea, Japan, Peru, Bangladesh, and Thailand.

Most of the nonpandemic *V. parahaemolyticus* isolates were isolated in the United States ( $n = 54$ ) and Chile ( $n = 12$ ). All Chilean isolates were environmental isolates in the region of Puerto Montt in the southern region of the country. They all belonged to different STs, although some of them showed a degree of genetic relatedness (shared one or two alleles' types

[e.g., ST-6 and ST-7, ST-10 and ST-1]) (Table 2). U.S. *V. parahaemolyticus* strains originated from the Pacific (including Alaska), Gulf, and Atlantic coasts. With the possible exception of strain NY-3483 (ST-36), which was isolated from a patient in New York who consumed oysters of unknown origin, isolates belonging to CC36 originated from the Pacific coast. Most of the isolates belonging to this clonal complex were of the O4:K12 serotype, a serotype historically associated with *V. parahaemolyticus* outbreaks corresponding to the consumption of raw oysters harvested from the U.S. Pacific coast (1). Strains belonging to CC34 originated principally from Gulf oysters. However, ST-34 isolates originating from both Louisiana and the Atlantic coast (Massachusetts) were noted. Based on allelic profiles, two groups isolated from Gulf oysters (D1 and D2) appeared to belong to the same clonal complex. These were not previously detected by eBURST as belonging to the same clonal complex because the ancestral type was missing or extinct (see "Clustering and phylogenetic analysis" below). A Connecticut oyster isolate (ST-53) was an SLV of ST-49 isolated in Chile. Other than the pandemic *V. parahaemolyticus* isolates, this was the only instance of an SLV occurring in two distant geographical locations.

**Contribution of mutation and recombination to clonal diversification. (i)  $r/m$  parameter.** Only 14 SLVs were observed in our data, and 7 belonged to the three clonal complexes. Most SLVs (10) arose from a recombination event, whereas only 4 arose by mutation (Table 3). This resulted in a per-allele  $r/m$  parameter of 2.5:1. In the case of the per-site analysis, the  $r/m$  parameter ratio was 8.8:1. These two parameters suggest that the initial steps of *V. parahaemolyticus* clonal diversification at allele or individual nucleotide sites are 2.5- and 8.8-fold more likely to occur by recombination than by point mutation, respectively.

**(ii) Index of association.** The "standardized" index of association ( $I_A^S$ ) of the entire isolate collection was 0.7626 ( $P < 0.05$ ), indicating that the alleles were in linkage disequilibrium or were nonrandomly distributed. Linkage disequilibrium can be observed when a recent, more pathogenic clone arises (40). However, even after eliminating the pandemic isolates from the data set, significant linkage disequilibrium still was detected ( $I_A^S = 0.5513$ ). A similar  $I_A^S$  (0.5911) was observed among U.S. isolates, the only geographical group for which there were enough isolates to generate meaningful information. This suggests a nonrandom distribution of alleles in the *V. parahaemolyticus* population in general, although recombination may also be occurring within different subpopulations.

**Clustering and phylogenetic analysis.** In order to validate the clustering and evolutionary model generated by eBURST, an ME tree was generated from the concatenated sequences of the seven loci of the 62 STs. For the most part, clustering, eBURST and ME tree analysis were consistent with one another, but ME tree gave better resolution and uncovered some phylogenetic relationships among groups or singletons not observed or resolved by eBURST (Fig. 2).

Based on eBURST analysis, CC3, CC36, and CC34 formed individual clusters in the ME tree (Fig. 2). The different groups found by eBURST also grouped together in the ME tree. Additionally, ST-21 and ST-59 belonging to D6 and DLVs of ST-36 (CC36 ancestral type by eBURST) were closely related to this clonal complex in the ME tree, confirming their inclu-

TABLE 2. Sequence types, allele profiles, and geographic locations of the *V. parahaemolyticus* strains analyzed

ST	Allele profile ( <i>dnaE</i> , <i>gyrB</i> , <i>recA</i> , <i>dddS</i> , <i>pntA</i> , <i>pyrC</i> , <i>tnaA</i> )	No. of strains	Clonal complex, group (G), or singleton (S)	Country (no. of strains)
3 <sup>a</sup>	3, 4, 19, 4, 29, 4, 22	22	CC1	Bangladesh (6), Chile (2), India (5), Japan (3), Peru (1), Spain (1), Thailand (2), United States (2)
51 <sup>a</sup>	29, 4, 19, 4, 29, 4, 22	1	CC1	Bangladesh
27 <sup>a</sup>	17, 4, 19, 4, 29, 4, 22	1	CC1	Korea
42 <sup>a</sup>	22, 4, 19, 4, 29, 4, 22	1	CC1	United States
36	21, 15, 1, 23, 23, 21, 16	7	CC2	United States
37	21, 15, 1, 23, 23, 20, 16	1	CC2	United States
38	21, 15, 1, 23, 22, 21, 16	1	CC2	United States
39	21, 29, 1, 23, 23, 21, 16	1	CC2	United States
59	36, 15, 1, 25, 23, 21, 16	2	G6	United States
21	17, 15, 1, 25, 23, 21, 16	1	G6	United States
35	20, 7, 15, 13, 7, 11, 5	2	CC3	United States
34	20, 25, 15, 13, 7, 11, 5	2	CC3	United States
32	20, 25, 15, 6, 7, 11, 4	1	CC3	United States
33	20, 25, 15, 16, 7, 11, 5	1	CC3	United States
24	17, 19, 13, 36, 33, 24, 26	1	G1	United States
60	36, 19, 13, 36, 33, 24, 26	1	G1	United States
25	17, 19, 12, 36, 31, 32, 26	1	G2	United States
26	17, 19, 13, 36, 31, 32, 26	1	G2	United States
49	28, 17, 21, 30, 20, 23, 24	1	G3	Chile
53	31, 17, 21, 30, 20, 23, 24	1	G3	United States
4	3, 5, 22, 12, 20, 22, 25	1	G4	United States
50	29, 5, 22, 12, 20, 22, 25	2	G4	United States
30	19, 31, 8, 31, 32, 29, 20	3	G5	United States
31	19, 31, 7, 31, 32, 29, 20	1	G5	United States
1	1, 3, 28, 3, 34, 18, 26	1	S1	Chile
2	2, 1, 11, 1, 18, 9, 23	1	S2	Chile
5	3, 16, 13, 36, 15, 30, 26	1	S5	United States
6	4, 13, 11, 38, 18, 9, 23	1	S6	Chile
7	4, 12, 11, 38, 16, 9, 23	1	S7	Chile
8	5, 27, 27, 13, 17, 25, 10	1	S8	Japan
9	6, 6, 3, 17, 11, 35, 26	1	S9	United States
10	7, 36, 4, 26, 6, 18, 26	1	S10	Chile
11	8, 24, 4, 14, 12, 14, 12	1	S11	Chile
12	9, 21, 15, 13, 4, 10, 26	2	S12	United States
13	10, 5, 22, 34, 20, 22, 15	1	S13	Chile
14	11, 4, 16, 35, 29, 15, 22	1	S14	Bangladesh
15	12, 34, 25, 32, 19, 12, 26	1	S15	United States
16	12, 13, 15, 22, 3, 1, 12	1	S16	Chile
17	13, 10, 19, 27, 28, 27, 21	2	S17	Spain
18	14, 30, 5, 11, 27, 11, 13	1	S18	United States
19	15, 11, 30, 10, 1, 3, 9	1	S19	Peru
20	16, 33, 32, 36, 13, 11, 19	1	S20	United States
22	17, 37, 6, 28, 21, 28, 7	1	S22	United States
23	17, 16, 13, 36, 15, 31, 26	2	S23	United States
28	17, 16, 13, 7, 24, 16, 20	1	S28	Chile
29	18, 18, 9, 19, 4, 28, 18	1	S29	United States
40	22, 24, 17, 13, 9, 19, 14	1	S40	United States
41	22, 28, 33, 13, 8, 19, 14	1	S41	United States
43	23, 29, 10, 7, 14, 24, 2	2	S43	United States
44	22, 4, 19, 4, 29, 4, 22	1	S44	United States
45	24, 22, 18, 37, 18, 11, 8	1	S45	Spain
46	25, 35, 4, 8, 33, 31, 6	1	S46	United States
47	26, 16, 13, 24, 31, 33, 2	1	S47	United States
48	27, 26, 20, 15, 5, 2, 28	1	S48	Chile
52	30, 32, 23, 20, 2, 13, 11	1	S52	Spain
54	31, 8, 3, 13, 30, 7, 2	1	S54	United States
55	32, 23, 26, 21, 26, 17, 23	1	S55	Chile
56	33, 22, 24, 5, 10, 5, 1	1	S56	Chile
57	34, 29, 2, 33, 25, 26, 17	1	S57	United States
58	35, 20, 31, 29, 23, 6, 27	2	S58	United States
61	37, 14, 14, 9, 14, 34, 26	1	S61	United States
62	38, 2, 29, 2, 29, 4, 22	1	S62	Chile

<sup>a</sup> ST belonging to the pandemic *V. parahaemolyticus* clonal complex.

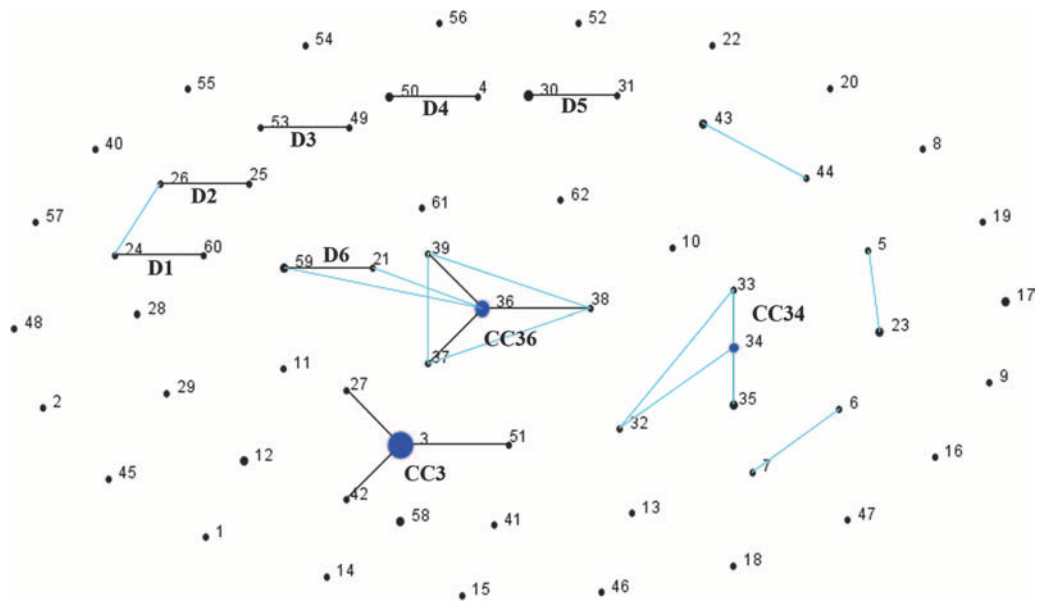


FIG. 1. *V. parahaemolyticus* “population snapshot” obtained using eBURST v3. Nine groups were defined using stringent criteria (6/7 shared alleles). Among those nine groups, three were identified as clonal complexes, and their predicted clonal ancestors are shown in blue (ST-3, ST-36, and ST-34 are the founders of each complex). STs that are SLVs of each other are shown connected by black lines. DLV STs are shown connected by aqua lines. The sizes of the circles are relative to the number of strains in the ST.

sion in the CC36 clonal complex. A similar case was observed for ST-32 (DLV of ST-34 and ST-33), which we hypothesized above belonged to CC34 and was the closest ST to that cluster in the ME tree. While D1 and D2 were loosely connected by eBURST (ST-24 and ST-26 were DLVs), the evolutionary relationship between these two groups was not detected because the putative SLV strain of both STs connecting these two groups was missing from the isolate collection analyzed. The ME tree definitively established the relationship between D1 and D2 even in the absence of the isolate linking these two groups. The ME tree showed that D3, D4, and ST-13 were closely related, a relationship also not observed by eBURST. Similarly, the ME tree showed that ST-47 and ST-46 were highly related to D5, with ST-46 more closely related to D5 than ST-47.

A test for incongruence was conducted in order to determine the impact of recombination on the *V. parahaemolyticus* populations analyzed. The ILD testing for all seven HK genes when partitioned separately led to the conclusion of incongruence ( $P = 0.001$ ). To facilitate the isolation of the gene(s) responsible for the incongruence, each of the seven genes was partitioned against a combined matrix consisting of the remaining six. All genes were highly incongruent with the combined HK matrix ( $P = 0.001$ ), suggesting that each of the seven genes contribute significantly to the overall signal of incongruence observed for the entire data set. Furthermore, when all possible pairwise ILD comparisons were performed, all of them exhibited incongruence among themselves ( $P = 0.001$ ). Taken as a whole, all of this indicates that recombination in those seven HK genes is probably frequent in *V. parahaemolyticus*.

Split trees were generated in an effort to visualize the impact of recombination in each locus and to verify the results obtained with the ILD tests (Fig. 3; see Fig. S1 in the supple-

mental material). The split trees generated for each locus, as well as for the concatenated sequences, showed reticulated structures. The majority of the strains also showed a star phylogeny radiating from the same central point, which suggests frequent recombination (23, 37). Furthermore, the phi test for recombination was significant in all loci analyzed ( $P < 0.05$ ) (see the supplemental material).

**Relationship between serotypes and ST.** Serotyping is used customarily for the characterization of *V. parahaemolyticus*. At least 35 different serotypes were included in the present study (see Table S2 in the supplemental material). Frequently, isolates belonging to genetically distant STs shared the same serotype (e.g., ST-20 and ST-40) and isolates with the same ST displayed different serotypes (e.g., ST-3). Multiple serotypes were observed in each of the three clonal complexes and three of the six groups.

## DISCUSSION

This study describes an MLST scheme that was applied to a geographically diverse panel of *V. parahaemolyticus* strains for investigation of population structure and determination of the influence of recombination/mutation on the clonal expansion of this species. A database from this study was created and is currently freely available on the Internet (<http://pubmlst.org/vparahaemolyticus>). This database can be used for unambiguous comparison of data generated from laboratories around the world. Our results unequivocally establish the clonal relationship of the pandemic complex and further confirm the first reported pandemic spread of *V. parahaemolyticus* (8, 36).

The MLST scheme described here is based on the allelic variation of seven HK genes, three from chromosome I (*recA*, *dnaE*, and *gyrB*) and four from chromosome II (*dtdS*, *pntA*, *pyrC*, and *tnaA*). The  $d_s/d_N$  ratios were higher than 1 for all the

TABLE 3. SLV allele variants found among the three clonal complexes and the different groups identified by eBURST in this study and identification of the events responsible for their evolution

Ancestral ST	ST of SLV	Variant locus	SLV allele	No. of nucleotide differences	Other STs containing SLV allele	Genetic event <sup>a</sup>
3	27	<i>dnaE</i>	17	10	21, 22, 23, 24, 25, 26, 27, 28	R
	42	<i>dnaE</i>	22	8	40, 41, 42	R
	51	<i>dnaE</i>	29	10	50, 51	R
36	37	<i>pyrC</i>	20	1	None	M
	38	<i>pntA</i>	22	1	None	M
	39	<i>gyrB</i>	29	5	39, 43, 57	R
34	33	<i>dddS</i>	16	1	None	M
	35	<i>gyrB</i>	7	10	35	R
NF <sup>b</sup>	21	<i>dnaE</i>	17	12	21, 22, 23, 24, 25, 26, 27, 28 (for SLV17); 60 (for SLV36)	R
	59	<i>dnaE</i>	36			
	30	<i>recA</i>	8	4	None	R
	31	<i>recA</i>	7			
	50	<i>dnaE</i>	29	10	51 (for SLV29); 3 (for SLV3)	R
	4	<i>dnaE</i>	3			
	53	<i>dnaE</i>	31	3	54 (for SLV31); none (for SLV28)	R
	49	<i>dnaE</i>	28			
	26	<i>recA</i>	13	1	5, 23, 24, 26, 28, 47, 60 (for SLV13); none (for SLV12)	M
	25	<i>recA</i>	12			
60	<i>dnaE</i>	36	12	59 (for SLV 36); same that above	R	
24	<i>dnaE</i>	17				

<sup>a</sup> R, recombination; M, mutation.

<sup>b</sup> NF, not found.

genes analyzed, indicating that they are under purifying pressure such that most amino acid substitutions are deleterious (6). Among the 100 *V. parahaemolyticus* strains used in this study; 62 different allelic combinations and an average of 33 alleles per locus were identified, indicating a high degree of genotypic diversity at slowly evolving loci. Three of the loci from chromosome II displayed lower nucleotide diversity (0.012 to 0.013) than the loci analyzed from chromosome I (0.029 to 0.032). It appears that the HK genes from chromosome II chosen for this study may be under different selective pressure than other regions of this chromosome.

This MLST scheme identified three major clonal complexes and six minor groups. Further sequence analysis (ME tree of concatenated sequences) shifted some groups into clonal complexes (e.g., D6 into CC36 and D1 and D2 into a potential clonal complex that were not identified by eBURST). The absence of the SLV strain linking these groups (D1 and D2) obscured this finding by eBURST. Additionally, numerous singletons were observed, demonstrating the high discriminatory capability of this scheme. The chosen genes appear to be well suited for broader population structure studies of *V. parahaemolyticus*. The observation of a high degree of genetic diversity in this study may be partially attributed to selection of particular *V. parahaemolyticus* strains in our collection. However, *V. parahaemolyticus* occurs in different habitats, such as seawater, sediment, gastrointestinal tracts of fish, chitin of zooplankton,

etc., where nutrient contents, temperatures, and other physiochemical properties (e.g., pH) differ. Shifts in these parameters require frequent bacterial adaptation. These environments are often heavily populated with phages that can facilitate gene transfer among vibrios and other marine bacteria (24). We encourage other investigators to populate the database with data generated from *V. parahaemolyticus* strains isolated from other sources and geographical locations in order to further delineate the extent of diversity within the species.

In the current study, ST-3 was the only ST with an international distribution, and this was determined to be the ancestor of CC3. The other SLVs within CC3 were also internationally distributed (Korea, Bangladesh, and the United States) and were identical to ST-3 except for differences in *dnaE*, apparently resulting from different recombination events. Similar findings were reported by Chowdhury et al. (9) using an MLST scheme based on the fragments of four genes from chromosome I. Those investigators showed that 51 of 54 *V. parahaemolyticus* pandemic isolates were indistinguishable in the four loci analyzed and that the 3 remaining isolates differed only in the *recA* locus. Variability in the *dnaE* gene in the current study and in the *recA* gene in the previous study suggests that these genes are evolving more rapidly in the pandemic clonal complex, but analysis of a much larger set of strains would be necessary to confirm this observation. The analysis of seven genes instead of four genes as done in previous studies further

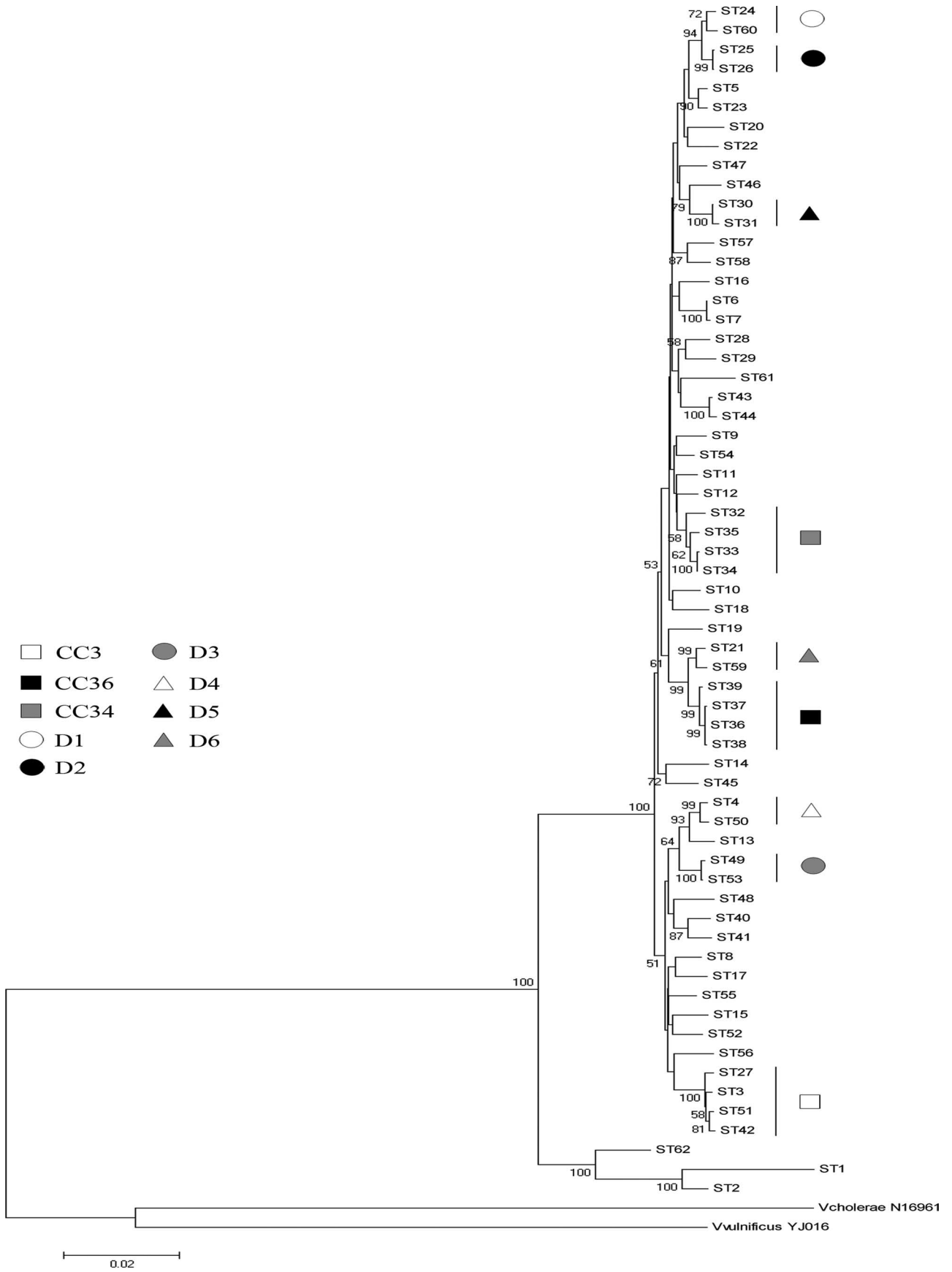


FIG. 2. An ME tree was constructed using the concatenated sequences of the seven loci of each of the 62 STs obtained in this study. Squares, circles, and triangles with different shading represent the three clonal complexes and the six doublets observed by eBURST. The scale represents the evolutionary distance, and bootstrap values over 50% are shown in the branches.

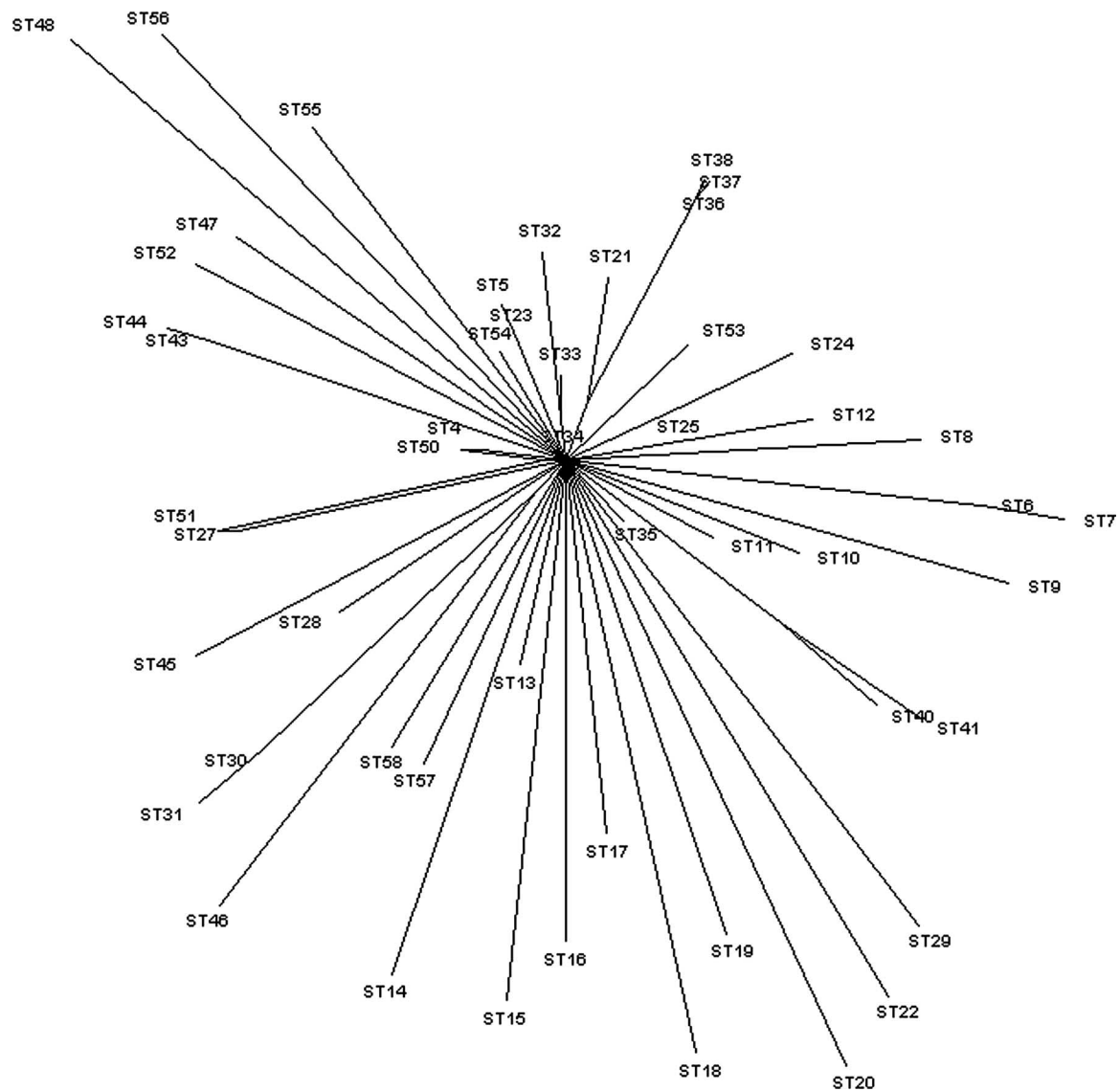


FIG. 3. Split decomposition analysis of the concatenated sequences of the seven chosen loci for the 62 STs obtained in this study.

confirms the homogenous nature of the pandemic clonal complex, independent of geographical site of isolation. While new variants are arising, they apparently have not yet become well established and are not replacing, to any significant degree, the ancestor type (ST-3) that continues to cause outbreaks in some countries (e.g., Chile in 2006) (17).

The persistence of the same ST over extended periods (e.g., ST-3 in numerous countries from 1996 to 2005 and ST-36 and ST-50 in the on U.S. Pacific coast from 1988 to 1997 and 1997 to 2004, respectively) is indicative of a clonal population structure. The clonal population structure is also supported by the existence of significant linkage disequilibrium between the MLST alleles ( $I_A^S = 0.7626$ ;  $P < 0.05$ ). The  $I_A^S$  decreased when pandemic strains were excluded or the analysis was limited to CC36. This phenomenon has also been reported by Miragaia et al. (35) for *Staphylococcus epidermidis*. However, the almost threefold per-allele and ninefold per-site higher frequencies of recombination events relative to mutations are

uncharacteristic of highly clonal bacteria. A history of frequent recombination events is further supported by the lack of any observed congruence among the genes analyzed (4). Furthermore, split trees for both concatenated sequences and individual loci resulted in star phylogeny structures radiating from the same central point, which suggest frequent recombination (37). Statistically significant evidence for recombination at each locus was also detected with the phi test (23). These apparently contradictory results for *V. parahaemolyticus* clonality parallel those reported for other bacteria, such as *V. cholerae* (19) and *S. epidermidis* (35), for which an epidemic population structure has been proposed. Therefore, we suggest that the *V. parahaemolyticus* population structure follows this “epidemic” model of clonal expansion in bacteria, where more adapted clonal complexes emerge from a background of highly recombinogenic bacteria (12). These clones then diversify predominantly by recombination rather than by point mutation. The paradox



between clonality and high diversity should be resolved as more strains are added to the MLST database.

Two other clonal complexes were identified among the analyzed strains (CC36 and CC34). This is the first definitive demonstration of *V. parahaemolyticus* clonal complexes other than the pandemic clonal complex and further supports the epidemic model. CC2 has been linked almost exclusively to outbreaks associated with the consumption of raw oysters harvested from the U.S. Pacific coast since the 1970s (1). This clonal complex also included a 1998 clinical isolate from New York that was linked to oysters with an unknown harvest location (NY-3483, ST-36); it has been speculated that those oysters originated from the state of Washington (5). While CC36 has been geographically restricted, it is similar to the pandemic clonal complex in that it displays multiple serotypes (O4:K12 and O12:K12) within the ancestor ST (ST-36) and consists of at least six STs (if D6 is included into this clonal complex). Furthermore, one DLV of ST-36 isolated in 1982 (ST-21) is consistent with its earlier emergence (CC36) relative to the pandemic CC3. Finally, ST-21 is the earliest isolate in CC36 and was most closely related to recent ST-59 isolates from Alaska (2004). CC34 consisted of six isolates with four STs (ST-32, -33, -34, and -35) and two serotypes (O4:K8 and O4:K9). Five of the isolates were from oysters collected in Alabama, Louisiana, and Massachusetts; the remaining clinical isolate was from a California patient with unknown food consumption history. This clonal complex appears to be more diverse than either CC3 or CC36, and its association with human illnesses is less certain.

Consistent with the results of Chowdhury et al. (9), there does not appear to be a linkage between serotype and ST among pandemic strains, as four different serotypes were observed among the ST-3 isolates and the three SLVs in CC3 were all O3:K6. If, as hypothesized (31), seroconversion occurs by lateral transfer of genes that participate in the synthesis of either O or K antigens, it appears that these genetic acquisitions are independent and probably arise from donors other than those which might affect the HK genes analyzed in this study. Taken together and consistent with other studies (7, 28, 31), these findings support the conclusion that serotyping for *V. parahaemolyticus* need not be considered obligatory for further characterization of isolates and that serotyping may actually be misleading and of limited epidemiological value relative to molecular-based strain typing tools.

Analysis of the ME tree of the concatenated sequences of all loci showed that *V. parahaemolyticus* forms a very homogeneous, well-supported group. However, three of the STs (ST-1, -2, and -62) formed a separate and distinct cluster outside the main group. This deviation from the main group appeared to be due to recombination with other, non-*V. parahaemolyticus* vibrios in some loci (e.g., *gyrB*, *recA* [ST-1], and *dnaE*). Analysis of individual trees for each locus suggests independent evolution of each gene and the importance of examining multiple genes for establishing phylogenetic relationships between *V. parahaemolyticus* strains (data not shown). A multiple gene sequence approach provides more detailed information on the genetic relationships among different *V. parahaemolyticus* strains because it allows for a buffering effect on the impact of lateral gene transfer such as was observed for strains that possessed ST-1, -2, and -62.

Overall, the data reported in this study indicate that *V. parahaemolyticus* is genetically diverse with a semiclonal population structure and that frequent recombination events seem to play an important role in the first steps of clonal diversification. Broader application of this MLST scheme will enhance understanding of the molecular epidemiology and evolution of this pathogen. This MLST scheme provides a universally available mechanism for timely recognition of evolutionary trends and emergence of *V. parahaemolyticus* clonal complexes, thus providing an early warning system. Prompt application of interventions (i.e., ballast water controls and harvest restrictions) could reduce public health consequences if new clonal complexes with enhanced virulence emerge or spread.

#### ACKNOWLEDGMENTS

We thank Keith Jolley, University of Oxford, for development of the database and website. We also thank Jessica L. Nordstrom for her assistance with the preparation of the manuscript and Eric Brown for his help in performing the ILD tests and his assistance in the revision of the manuscript.

This study was supported by a grant from the U.S. Department of Agriculture, Cooperative State Research, Education and Extension Service, National Research Initiative, Competitive Grants Program, Epidemiological Approaches to Food Safety, project no. 2004-35212-14882.

#### REFERENCES

1. Abbott, S. L., C. Powers, C. A. Kaysner, Y. Takeda, M. Ishibashi, S. W. Joseph, and J. M. Janda. 1989. 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.* **27**:2891–2893.
2. Ansaruzzaman, M., M. Lucas, J. L. Deen, N. A. Bhuiyan, X. Y. Wang, A. Safa, M. Sultana, A. Chowdhury, G. B. Nair, D. A. Sack, L. von Seidlein, M. K. Puri, M. Ali, C. L. Chaignat, J. D. Clemens, and A. Barreto. 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**:2559–2562.
3. Bisharat, N., D. I. Cohen, R. M. Harding, D. Falush, D. W. Crook, T. Peto, and M. C. Maiden. 2005. Hybrid *Vibrio vulnificus*. *Emerg. Infect. Dis.* **11**:30–35.
4. Brown, E. W., M. L. Kotewicz, and T. A. Cebula. 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. *Mol. Phylogenet. Evol.* **24**:102–120.
5. Centers for Disease Control and Prevention. 2006. *Vibrio parahaemolyticus* infections associated with consumption of raw shellfish—three states, 2006. *MMWR Morb. Mortal. Wkly. Rep.* **55**:854–856.
6. Charlesworth, J., and A. Eyre-Walker. 2006. The rate of adaptive evolution in enteric bacteria. *Mol. Biol. Evol.* **23**:1348–1356.
7. Chowdhury, A., M. Ishibashi, V. D. Thiem, D. T. Tuyet, T. V. Tung, B. T. Chien, L. L. Seidlein, D. G. Canh, J. Clemens, D. D. Trach, and M. Nishibuchi. 2004. Emergence and serovar transition of *Vibrio parahaemolyticus* pandemic strains isolated during a diarrhea outbreak in Vietnam between 1997 and 1999. *Microbiol. Immunol.* **48**:319–327.
8. Chowdhury, N. R., S. Chakraborty, T. Ramamurthy, M. Nishibuchi, S. Yamasaki, Y. Takeda, and G. B. Nair. 2000. Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. *Emerg. Infect. Dis.* **6**:631–636.
9. Chowdhury, N. R., O. C. Stine, J. G. Morris, and G. B. Nair. 2004. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequence typing. *J. Clin. Microbiol.* **42**:1280–1282.
10. DePaola, A., C. A. Kaysner, J. Bowers, and D. W. Cook. 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.
11. Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. *Cladistics* **10**:315–319.
12. Feil, E. J. 2004. Small change: keeping pace with microevolution. *Nat. Rev. Microbiol.* **2**:483–495.
13. Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day. 2003. How clonal is *Staphylococcus aureus*? *J. Bacteriol.* **185**:3307–3316.
14. Feil, E. J., M. C. Enright, and B. G. Spratt. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a

- comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. Res. Microbiol. **151**:465–469.
15. Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**:1518–1530.
  16. Feil, E. J., M. C. Maiden, M. Achtman, and B. G. Spratt. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol. Biol. Evol. **16**:1496–1502.
  17. Fuenzalida, L., L. Armijo, B. Zabala, C. Hernandez, M. L. Rioseco, C. Riquelme, and R. T. Espejo. 2007. *Vibrio parahaemolyticus* strains isolated during investigation of the summer 2006 seafood related diarrhea outbreaks in two regions of Chile. Int. J. Food Microbiol. **117**:270–275.
  18. Fuenzalida, L., C. Hernandez, J. Toro, M. L. Rioseco, J. Romero, and R. T. Espejo. 2006. *Vibrio parahaemolyticus* in shellfish and clinical samples during two large epidemics of diarrhoea in southern Chile. Environ. Microbiol. **8**:675–683.
  19. Garg, P., A. Aydanian, D. Smith, J. Glenn, G. B. Nair, and O. C. Stine. 2003. Molecular epidemiology of O139 *Vibrio cholerae*: mutation, lateral gene transfer, and founder flush. Emerg. Infect. Dis. **9**:810–814.
  20. Gonzalez-Escalona, N., V. Cachicas, C. Acevedo, M. L. Rioseco, J. A. Vergara, F. Cabello, J. Romero, and R. T. Espejo. 2005. *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. Emerg. Infect. Dis. **11**:129–131.
  21. Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**:95–98.
  22. Haubold, B., and R. R. Hudson. 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics **16**:847–848.
  23. Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254–267.
  24. Jiang, S. C., and J. H. Paul. 1998. Gene transfer by transduction in the marine environment. Appl. Environ. Microbiol. **64**:2780–2787.
  25. Jolley, K. A., M. S. Chan, and M. C. Maiden. 2004. mlstDbNet—distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics **5**:86.
  26. Jolley, K. A., E. J. Feil, M. S. Chan, and M. C. Maiden. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics **17**:1230–1231.
  27. Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. **5**:150–163.
  28. Laohaprertthisan, V., A. Chowdhury, U. Kongmuang, S. Kalnauwakul, M. Ishibashi, C. Matsumoto, and M. Nishibuchi. 2003. Prevalence and serodiversity of the pandemic clone among the clinical strains of *Vibrio parahaemolyticus* isolated in southern Thailand. Epidemiol. Infect. **130**:395–406.
  29. Maiden, M. C. 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. **60**:561–588.
  30. Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95**:3140–3145.
  31. Martinez-Urtaza, J., A. Lozano-Leon, A. DePaola, M. Ishibashi, K. Shimada, M. Nishibuchi, and E. Liebana. 2004. Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. J. Clin. Microbiol. **42**:4672–4678.
  32. Martinez-Urtaza, J., L. Simental, D. Velasco, A. DePaola, M. Ishibashi, Y. Nakaguchi, M. Nishibuchi, D. Carrera-Flores, C. Rey-Alvarez, and A. Pousa. 2005. Pandemic *Vibrio parahaemolyticus* O3:K6, Europe. Emerg. Infect. Dis. **11**:1319–1320.
  33. Matsumoto, C., J. Okuda, M. Ishibashi, M. Iwanaga, P. Garg, T. Ramamurthy, H. C. Wong, A. DePaola, Y. B. Kim, M. J. Albert, and M. Nishibuchi. 2000. 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. **38**:578–585.
  34. Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe. 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. **5**:607–625.
  35. Miragaia, M., J. C. Thomas, I. Couto, M. C. Enright, and L. H. de. 2007. Inferring a population structure for *Staphylococcus epidermidis* from multilocus sequence typing data. J. Bacteriol. **189**:2540–2552.
  36. Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda, and D. A. Sack. 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin. Microbiol. Rev. **20**:39–48.
  37. Octavia, S., and R. Lan. 2006. Frequent recombination and low level of clonality within *Salmonella enterica* subspecies I. Microbiology **152**:1099–1108.
  38. Okuda, J., M. Ishibashi, E. Hayakawa, T. Nishino, Y. Takeda, A. K. Mukhopadhyay, S. Garg, S. K. Bhattacharya, G. B. Nair, and M. Nishibuchi. 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**:3150–3155.
  39. Quilici, M. L., A. Robert-Pillot, J. Picart, and J. M. Fournier. 2005. Pandemic *Vibrio parahaemolyticus* O3:K6 spread, France. Emerg. Infect. Dis. **11**:1148–1149.
  40. Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90**:4384–4388.
  41. Swofford, D. L. 1999. Phylogenetic analysis using parsimony (PAUP\* V. 4.0.3b) program and documentation. The Smithsonian Institution, Washington, DC.