Genomic evidence of adaptive evolution in emergent *Vibrio parahaemolyticus* ecotypes

Jeffrey W. Turner[1]* • Chris T. Berthiaume[2] • Rhonda Morales[2] • E. Virginia Armbrust[2] • Mark S. Strom[3]

[1]Department of Life Sciences, Texas A&M University at Corpus Christi, Corpus Christi, Texas, United States
[2]Center for Environmental Genomics, School of Oceanography, University of Washington, Seattle, Washington, United States
[3]National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, Washington, United States

*jeffrey.turner@tamucc.edu

Abstract

The ubiquitous marine bacterium *Vibrio parahaemolyticus* is a leading cause of illness associated with seafood consumption. The emergence of two genetically distinct ecotypes (ST3 and ST36) has led to an alarming increase in the size and frequency of disease outbreaks. We conducted a genomic comparison of 30 *V. parahaemolyticus* genomes that represent a diverse collection of 15 genetically distinct ecotypes, including newly sequenced representatives of ST3 and ST36, isolated from both clinical and environmental sources. A multistep evolutionary analysis showed that genes associated with sensing and responding to environmental stimuli have evolved under positive selection, identifying examples of convergent evolution between ST3 and ST36. A comparison of predicted proteomes indicated that ST3 and ST36 ecotypes laterally acquired tens of novel genes associated with a variety of functions including dormancy, homeostasis and membrane transport. Genes identified in this study play an apparent role in environmental fitness and may confer cross protection against stressors encountered in the human host. Together, these results show the evolution of stress response is an important genetic mechanism correlated with the recent emergence of the ST3 and ST36 ecotypes.

Background

Vibrio parahaemolyticus is a Gram-negative bacterium adapted for life in the marine environment (Joseph et al., 1982). This bacterium is also a coincidental pathogen of humans and a leading cause of gastroenteritis associated with the consumption of raw or undercooked seafood (Daniels et al., 2000; Su and Liu, 2007). Historically, outbreaks of *V. parahaemolyticus* related illness were localized and attributed to a diversity of genetically distinct strains (Chowdhury et al., 2000; Matsumoto et al., 2000). Over the last two decades, the emergence of a pandemic ecotype, characterized by multilocus sequence typing (MLST) as ST3, which includes serotype O3:K6 and its many serovariants, has led to a dramatic increase in infections worldwide (Nair et al., 2007). Coincident with the global spread of ST3, larger and less localized outbreaks in North America have been attributed to additional ecotypes such as ST36 (O4:K12) (Abbott et al., 1989; Turner et al., 2013). Further, the recent transoceanic spread of ST36 to Europe in 2012 (Martinez-Urtaza et al., 2013; González-Escalona et al., 2015) has raised concern over the possible emergence of ST36 as a second pandemic ecotype.

The majority of *V. parahaemolyticus* strains are innocuous to humans (DePaola et al., 2000). Those strains that do cause illness often carry genes encoding hemolytic exotoxins such as the thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) (Shirai et al., 1990), which cause fluid loss across host cell membranes (Raimondi et al., 2000; Matsuda et al., 2010). Detection of the *tdh* and *trh* genes is a test frequently used to identify potentially virulent strains (Bej et al., 1999) although some virulent strains lack both hemolysins (Raghunath, 2014). The mechanism of pathogenesis is also associated with the secretion of effector proteins by a type III secretion system (T3SS) responsible for toxicity in various *in vitro* and *in*

vivo models (Burdette et al., 2008; Hiyoshi et al., 2010; Ham and Orth, 2012). Yet the overall mechanism of pathogenesis remains unclear and is thought to involve the concerted expression of multiple virulence-associated genes (Broberg et al., 2011; Zhang and Orth, 2013).

The predominance of ST3 and ST36 ecotypes among clinical isolates suggests these ecotypes benefit from enhanced fitness in the marine environment or the human host. Moreover, the ability of both ecotypes to cause large, non-localized outbreaks suggests parallel evolutionary strategies. Previous studies have linked the emergence of ST3 with the lateral acquisition of seven genomic islands (Hurley et al., 2006; Boyd et al., 2008). One island contains two copies of the *tdh* gene (*tdh1* and *tdh2*) and a type III secretion system (T3SS2-α) (Makino et al., 2003). Similarly, the ST36 ecotype has acquired *tdh* and *trh* gene clusters and a homologous type III secretion system (T3SS2-β) (González-Escalona et al., 2011; Paranjpye et al., 2012; Xu et al., 2015). However, the widespread occurrence of these genomic islands among diverse collections of clinical and environmental ecotypes (Paranjpye et al., 2012; Xu et al., 2015) suggests that additional forces correlate with predominance.

Given that *V. parahaemolyticus* is an environmental bacterium and the human host is an alternate and novel ecological niche, we hypothesized that the evolution of dual-purpose genes, integral to both environmental fitness and human pathogenesis, would correlate with the recent emergence of the ST3 and ST36 ecotypes. To test this hypothesis, we conducted a genomic analysis of 30 *V. parahaemolyticus* genomes to identify the evolutionary forces correlated with clinically predominant ecotypes (ST3 and ST36). We dissected two mechanisms of adaptive evolution: the mutation of genes comprising the core genome and the lateral acquisition of genes comprising the accessory genome.

Materials and methods

Bacterial genomes

We compared a total of 30 *V. parahaemolyticus* genomes (two closed, 28 draft) sequenced from clinical and environmental isolates (n = 21 and 9, respectively) collected from ten countries over a 27-year period (Table S1). Nineteen of the genomes were sequenced in this study (below) and the remaining eleven were downloaded from NCBI. The genomes were selected to represent a diverse collection of 15 genetically distinct ecotypes, and a diversity of *tdh/trh* genotypes (i.e., +/+, +/–, –/+, and –/–). The presence or absence of the virulence-associated *tdh* and *trh* genes are reported in Table S1. We consider all clinical strains as virulent; however, in lieu of animal model testing, we cannot predict the virulence or avirulence of environmental strains. The closed genome of the ST3 clinical strain *V. parahaemolyticus* RIMD2210633 (Makino et al., 2003) and the draft genome of the ST36 clinical strain *V. parahaemolyticus* 10329 (González-Escalona et al., 2011) were defined as reference genomes.

Bacterial isolates, culture conditions and isolation of genomic DNA

V. parahaemolyticus isolates (n = 19) originating from the Pacific Northwest (PNW) coast of the United States (Table S1) were selected for genome sequencing based on previous genotyping (Paranjpye et al., 2012) and multilocus sequence typing (MLST) analysis (Turner et al., 2013). In *V. parahaemolyticus*, MLST is a high-resolution genetic typing scheme based on seven conserved housekeeping genes (González-Escalona et al., 2008) that serve as the framework for a global database containing more than 2,200 isolates (http://pubmlst.org/vparahaemolyticus/). Multilocus sequence typing was chosen ahead of serotyping, as the O- and K-antigens are prone to rapid recombination and serotype conversion, and more than 20 O3:K6 serovariants have been identified (Chen et al., 2010, 2011). Late log-phase cultures were grown overnight (16–20 hours) in tryptic soy broth (TSB) at 30°C with shaking (150 rpm). Genomic DNA was isolated by the cetyltrimethyl ammonium bromide (CTAB) phenol/chloroform method (Atashpaz et al., 2010).

Sequencing, assembly and annotation

Draft genomes were generated for 19 *V. parahaemolyticus* strains by SOLiD 4 sequencing (Applied Biosystems Inc., Foster City, CA, USA). Barcoded paired-end (50 and 25 bp) libraries were constructed with a target insert size of 250 bp. Reads were error-corrected using the SOLiD Accuracy Enhancement Tool and PCR duplicated reads were removed using the fastq_nodup tool from the SEAStAR software (Iverson et al., 2012). Genomes were assembled in color-space using the CLC Assembly Cell version 4.1 (CLC Bio, Boston, MA, USA). Reads were mapped to both *V. parahaemolyticus* RIMD2210633 (Makino et al., 2003) and *V. parahaemolyticus* 10329 (González-Escalona et al., 2011) reference genomes, and the assembly recruiting the highest number of reads was retained. Unmapped reads as well as genomes mapping poorly to either reference were assembled *de novo* using a kmer length of 25. Draft genome assemblies were deposited in NCBI/GenBank (see Table S1 for accession numbers). NCBI's Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) was used to annotate the draft genomes.

Phylogenetic analysis

Relatedness between the 30 *V. parahaemolyticus* genomes was established using a whole-genome alignment (WGA) approach (Sahl et al., 2011). Briefly, genomes were aligned using Mugsy (Angiuoli and Salzberg, 2011), and conserved alignment blocks (> 500 bp in length and shared by all 30 genomes) were concatenated using the Galaxy web-based platform (Giardine et al., 2005). The concatenated alignment was trimmed using the strictplus algorithm implemented in trimAL (Capella-Gutierrez et al., 2009). A phylogenetic tree was inferred with FastTree (Price et al., 2009) using the general time-reversible (GTR) model of nucleotide evolution. Local support values (500 resampled alignments) were calculated using SEQBOOT (from the PHYLIP package, version 3.6) (Felsenstein, 1989) and the FastTree global bootstrap parameter (the intree1 option). The WGA tree was rooted to a closely related species (*Vibrio alginolyticus* ATCC17749) (Liu et al., 2015) and annotated in FigTree.

Core genes evolving under positive selection

Evidence of recent positive selection among core genes (i.e., shared by all 30 genomes) was established using TimeZone (Chattopadhyay et al., 2013). Major workflows include a modified BLASTN comparison of all genomes against a reference, extraction of orthologous genes, multiple sequence alignment, estimates of homologous recombination and phylogenetic reconstruction. The closed and well-annotated assembly of the type strain *V. parahaemolyticus* RIMD2210633 was chosen as the reference genome (Makino et al., 2003). Genes were grouped as orthologous based on a minimum sequence identify (95%) and a minimum relative length (95%). Evidence of positive selection was established by the repeated occurrence of site-specific replacement mutations (i.e. hotspot mutations). Analysis of genes evolving under positive selection was limited to non-recombinant genes containing only recent (i.e., short-term) hotspot mutations. Recombinant genes were detected using MaxChi and PhylPro as implemented in TimeZone. Genes showing evidence of positive selection were analyzed using the DAVID bioinformatic resources (Huang et al., 2009) to identify instances of enrichment and infer biological function.

Lateral acquisition of adaptive genes

Clustering of orthologous proteins among all 30 *V. parahaemolyticus* genomes was carried out in OrthoMCL (Li et al., 2003) using a modified bidirectional all-against-all BLASTP (e–10 cutoff). Results were parsed using custom Python scripts to identify the subset of orthologs unique to a given ecotype (i.e., ST3 and ST36). Functional analysis of unique orthologs was aided by NCBI's non-redundant (nr) database, the SEED Viewer (http://theSEED.org), the virulence factor database (VFDB) (Chen et al., 2012), and InterPro (Hunter et al., 2009).
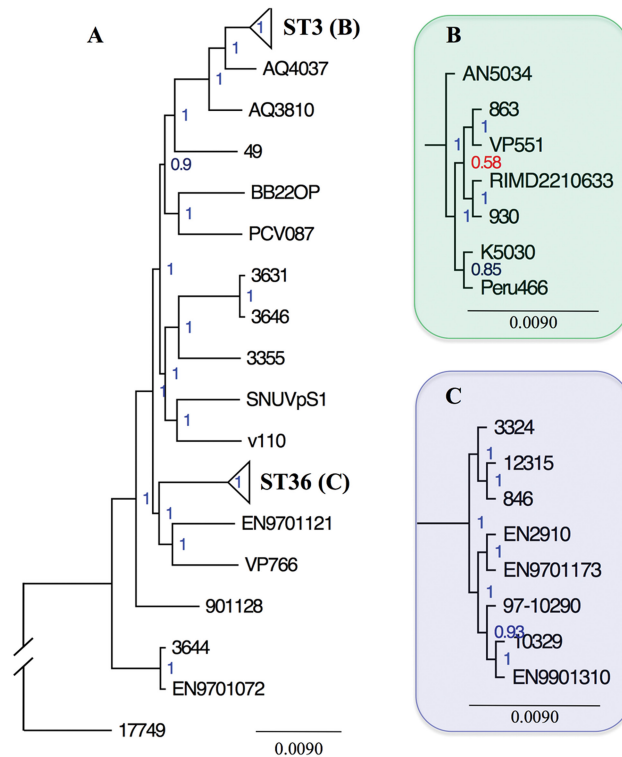
Statistics

To test whether genes located on chromosome II were more strongly influenced by positive selection (compared to chromosome I), differences in nucleotide diversity and rates of synonymous (dS) and nonsynonymous (dN) mutations between chromosomes were evaluated with a two-tailed t-test in Python using the SciPy library.

Results

We compared a total of 30 *V. parahaemolyticus* genomes that reflect a global sampling of strains (15 ecotypes from ten countries) isolated from both clinical and environmental sources over a 27-year period. The focus of our analyses was the discovery of features that distinguish the ST3 and ST36 ecotypes from the larger *V. parahaemolyticus* population.

Phylogenetic analysis

A phylogenetic tree (Figure 1A) was inferred using a whole genome alignment (WGA) approach (Sahl et al., 2011). The WGA tree was based on a gene-independent analysis of shared genomic data (~3.96 Mbp) derived from 1,343 concatenated alignment blocks (each greater than 500 bp). The analysis grouped the 30 *V. parahaemolyticus* genomes into 15 genetically distinct clades in agreement with our previous phylogenetic analysis based on a 7-loci MLST scheme (Turner et al., 2013). The WGA tree provided higher local support values and resolved the subclade diversity of ST3 (n = 7 genomes) and ST36 (n = 8 genomes) ecotypes (Figures 1B and 1C, respectively). We observed no discernible phylogenetic patterns within ST3 despite significant temporal and geographic separation between the seven isolates. In contrast, diversity within ST36 showed evidence of a temporal shift between isolates collected prior to 2001 (EN2910, EN9701173, 97-10290, 10329 and EN9901310) and those collected since 2006 (3324, 12315 and 846).

Core genes evolving under positive selection

Whole-genome comparisons with the closed reference genome (*V. parahaemolyticus* RIMD2210633) identified a core set of 1,185 orthologous genes present in all 30 genomes. Core genes showing evidence of recombination (n = 81) (see Materials and methods) were removed from additional analysis. Polymorphic genes (n = 1,104) were defined as all non-recombinant core orthologs that produced non-identical alignments consisting of at least four alleles. Those polymorphic core genes located on chromosome II displayed significantly greater nucleotide diversity (π) and significantly higher rates of synonymous (dS) and nonsynonymous (dN) mutations than those located on chromosome I (p < 0.001) (Table 1). Furthermore, recombination was detected more frequently in polymorphic core genes located on chromosome II (63.3%) than on chromosome I (38.3%).

To identify genes evolving under positive selection, we assessed the frequency of single-nucleotide polymorphisms (SNPs) resulting in repeated, site-specific amino acid replacements (i.e., hotspot mutations). Approximately one-third of polymorphic core genes (n = 344 of 1,104 genes) contained hotspot mutations (Table S2). Half encoded hypothetical proteins and half encoded putative proteins associated with a variety of functional categories (e.g., acetyltransferase activity, heterocycle biosynthesis, transcriptional regulation and nitrogen compound biosynthesis). Many of the 344 genes contained ecotype-specific hotspot mutations; 155 were found in the ST3 ecotype (n = 7 genomes), 194 were found in the ST36 ecotype (n = 8 genomes), and 105 were found in both ST3 and ST36 (n = 15 genomes) (Table S2). Functional analysis of the 105 shared hotspots clustered the proteins into 20 functional categories (e.g., transcription regulation, essential to the cell envelope, and integral to cell membrane). However, only those proteins involved in transcription regulation (n = 13 proteins) were significantly overrepresented (Table S2). These transcriptional regulators included GntR, LysR, TetR, MerR and LuxR family transcriptional regulators containing DNA-binding, helix-turn-helix domains. We also identified ten genes containing site-specific hotspot mutations shared exclusively between ST3 and ST36 ecotypes (Table 2). Of these, six resulted in identical amino acid changes. For example, analysis of TolC revealed that all ST3 and ST36 genomes (n = 15) shared a hotspot mutation resulting in an identical amino acid replacement (serine to alanine) at amino acid position 118 (Table 2).

Lateral acquisition of adaptive genes

To identify genes unique to ST3 and ST36 ecotypes but absent from all other ecotypes, we conducted an all-versus-all BLASTP comparison of the 30 predicted proteomes (Figure 2). Orthologs unique to the seven ST3 genomes (n = 42) and the eight ST36 genomes (n = 60) did not include putative toxins such as the TDH and TRH hemolysins (Honda et al., 1988; Nishibuchi et al., 1989) or the VopT and VopZ effector proteins (Kodama et al., 2007; Zhou et al., 2013). Instead, the majority of ecotype-specific orthologs (83%) were hypothetical proteins, with few defining features other than transmembrane and signal peptide domains

Table 1. Comparison of evolutionary rates among core genes located on chromosomes I and II

Parameter	Chromosome I	Chromosome II	t-test[b]
Number of core genes	746	439	-
Number of polymorphic genes[a]	686	418	-
Nucleotide diversity (π)	0.011	0.013	< 0.001
Synonymous mutations (dS)	0.039	0.044	< 0.001
Nonsynonymous mutations (dN)	0.0036	0.0051	< 0.001
Number of recombinant genes	286 (38.3%)	278 (63.3%)	-
Number of hotspot genes	184 (24.7%)	160 (36.4%)	-

[a]Genes with more than four non-identical alleles.
[b]Significance declared at P < 0.001.
doi: 10.12952/journal.elementa.000117.t001

(Tables S3 and S4, respectively). The remaining orthologs were associated with a variety of functions such as dormancy, membrane transport, capsular polysaccharide biosynthesis, and transcriptional regulation. The sole ortholog restricted to both ecotypes (locus tag VP1860 in RIMD2210633 and VP10329_14295 in 10329) was a hypothetical protein containing a domain of unknown function (DUF1706).

A large proportion of the ecotype-specific orthologs (n = 22/42 in ST3 and 17/60 in ST36) were localized to the superintegron (SI) on chromosome I. SI-localized orthologs unique to ST3 included a ReIE/ReIB toxin/antitoxin system, two acetyltransferases and numerous hypothetical proteins (Table S3). SI-localized orthologs unique to ST36 included an acetyltransferase and various hypothetical proteins (Table S4). In a different region on the same chromosome, nine orthologs unique to ST36 were localized within the capsular polysaccharide (*cps*) gene cluster (Table S4). Five of these ST36 *cps* genes encoded capsular polysaccharide proteins (glycosyl transferase, mannose-6-phosphate isomerase, nucleoside-diphosphate-sugar epimerase, and galactosyl transferase) with significant (< e–20) homology to putative virulence factors curated in the VFDB (Table 3). Comparison of ST3 and ST36 *cps* clusters revealed important differences in size, gene content, and synteny (Figure 3).

Discussion

This study describes evolutionary mechanisms correlated with two clinically important *V. parahaemolyticus* ecotypes (ST3 and ST36). These genetically distinct ecotypes are characterized by their widespread occurrence and their ability to cause large, non-localized disease outbreaks. Whole-genome comparisons across a diverse collection of 30 strains from clinical and environmental sources indicate that core genes shared between ST3 and ST36 ecotypes are converging under positive selection. In addition, the acquisition of accessory genes, novel to the ST3 and ST36 ecotypes, indicates divergent evolution. Together, these two evolutionary forces appear to be acting in concert to affect genes that play a role in sensing, responding, and adapting to extracellular stress.

Table 2. Genes containing recent site-specific hotspot mutations shared by ST3 and ST36 ecotypes

Locus tag[a]	Description	Predicted function	Position[b]	Replacement substitution[c]	
				ST3	ST36
VP0147	ComF-like protein	Nucleoside metabolism	28	A – T	A – T
VP0571	Phosphate ABC transporter	Transport	7, 15	K – R, I – T	K – R, I – T
VP1718	Cytochrome c551 peroxidase	Oxidation/reduction	169	D – N	N – D
VP1998	Outer membrane protein TolC	Transport	118	S – A	S – A
VP2639	Short-chain dehydrogenase	Oxidation/reduction	92	E – D	E – G
VPA0020	Histidine kinase-like ATPase	Signal transduction	310	M – V	M – V
VPA0651	Ferredoxin	Electron transfer	46	C – W	C – Y
VPA0835	Inosine-guanosine kinase	Purine metabolism	265	N – H	N – S
VPA1142	Transcriptional regulator MerR	Transcriptional regulation	86	V – A	V – A
VPA1298	Glutathione S-transferase	Protein binding	275	V – T	A – L

[a]Genes are identified by NCBI/GenBank locus tag.
[b]The amino acid position of the hotspot mutation.
[c]The amino acid substitution resulting from the hotspot mutation is shown for both ecotypes.
doi: 10.12952/journal.elementa.000117.t002
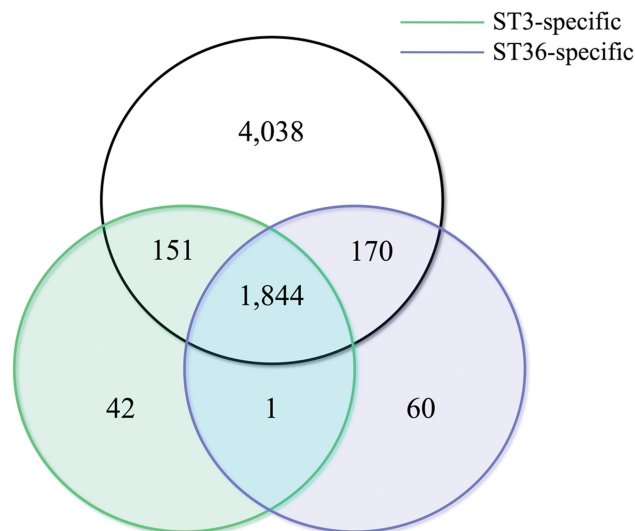
Figure 2

Venn diagram showing the distribution of orthologous proteins.

Orthologs found in all ST3 genomes (n = 7 genomes) highlighted in green while orthologs found in all ST36 genomes (n = 8 genomes) highlighted in purple. Diagram clearly illustrates the number of orthologs found exclusively in all ST3 (n = 42) and ST36 (n = 60) genomes.

doi: 10.12952/journal.elementa.000117.f002

Previous studies have inferred the relatedness of *V. parahaemolyticus* ecotypes through the phylogenetic analysis of seven housekeeping genes (González-Escalona et al., 2008). Recent studies seeking higher resolution phylogenies have turned to genome-scale analyses (Haendiges et al., 2015; Hazen et al., 2015; Loyola et al., 2015; Whistler et al., 2015). Here, we used a WGA approach to infer relatedness between major phylogenetic clades. This approach resolved the subclade diversity inherent to ST3 and ST36 ecotypes. The absence of a temporal pattern within ST3 demonstrates the strong clonality of this ecotype. Meanwhile, evidence of a temporal shift in ST36 may reflect important metabolic differences between closely related isolates. For instance, zebrafish challenged with different ST3 and ST36 strains (including strains from this study) have been shown to exhibit distinctly different survival curves (Paranjpye et al., 2013).

The diversification of bacterial ecotypes can arise from mutation and lateral gene transfer. The role played by mutation was assessed using a multistep analysis aimed at the identification of genes evolving under positive selection. A first step in this analysis was the measurement of evolutionary rates in 1,104 core genes. Significantly faster evolutionary rates (π, dS, and dN) among genes located on chromosome II (compared to chromosome I) indicate that genes located on the smaller secondary chromosome are subject to weaker purifying selection. In *Bulkolderia* and *Vibrio* species, rapidly evolving genes tend to be located on the smaller second chromosome (Cooper et al., 2010). Citing a chromosomal bias in the distribution of essential genes, it has been suggested that secondary chromosomes may act as accessory genomes for adaptation to environmental change or new ecological niches (Okada et al., 2005). In this study, the observed chromosomal bias in selection pressure supports the theory that secondary chromosomes are a reservoir for genes involved in adaptive evolution.

Table 3. Putative virulence factors unique to the ST36 ecotype

| Locus tag[a]	Description	Length[b]	BLAST hits[c]	Best BLAST hit[d]	e-value
LG43_RS11440	Glycosyltransferase	287	9	FutC alpha-1,2-fucosyltransferase	8.00e−29
				Helicobacter pylori	
LG43_RS11485	Phosphomannose isomerase	396	3	ManA mannose-6-phosphate isomerase	7.00e−50
				Haemophilus somnus	
LG43_RS11490	Glycosyltransferase	253	77	WbyL glocosyltransferase	3.00e−51
				Yersinia pseudotuberculosis	
LG43_RS11495	Epimerase	300	23	Cap8N nucleoside-diphosphate epimerase	1.00e−102
				Vibrio vulnificus	
LG43_RS11500	Galactosyltransferase	182	55	UDP-sugar phosphotransferase	4.00e−82
				Vibrio vulnificus	

[a]Genes identified by NCBI/GenBank locus tag.
[b]Number of amino acids.
[c]Number of significant hits (e−20 cutoff) in the virulence factor database (VFDB).
[d]Description of the most significant hit and the representative species.
doi: 10.12952/journal.elementa.000117.t003

A. ST3 (*V. parahaemolyticus* RIMD2210633)
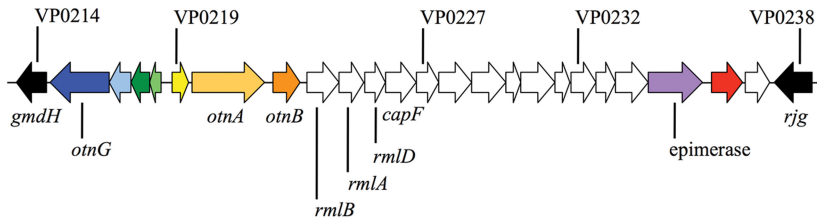
B. ST36 (*V. parahaemolyticus* 10329)
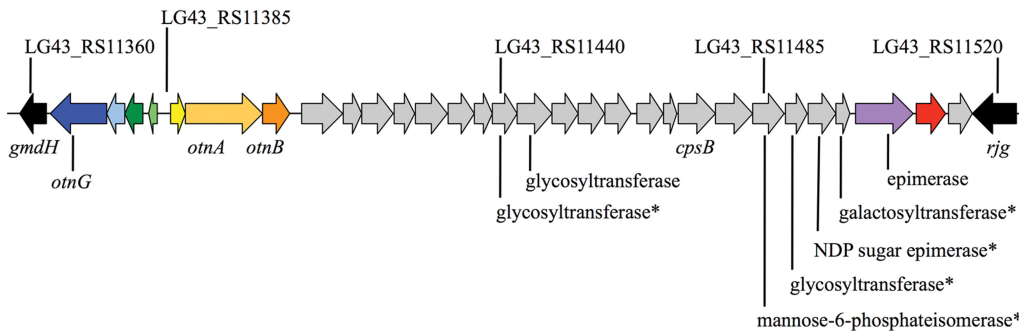
Genes containing site-specific replacement substitutions (i.e., hotspot mutations) accounted for nearly one-third of core genes. The occurrence of recent hotspot mutations is an indicator of genes evolving under positive selection (Chattopadhyay et al., 2009). Indeed, a single positively selected SNP leading to an amino acid replacement can promote adaptation to a new niche (Weissman et al., 2007). We detected an abundance of genes containing hotspot mutations and many were shared between the ST3 and ST36 ecotypes. Functional analysis of shared hotspot genes revealed a significant overrepresentation of helix-turn-helix DNA-binding transcriptional factors, which are known to regulate transcription in response to environmental stimuli like heavy metals, oxidative stress or antibiotics (Santos et al., 2009). The ability of an ecotype to respond appropriately to prevailing environmental conditions is essential to niche adaptation, as occupancy of a new niche would require the bacterium to overcome new growth conditions and extracellular stressors. For example, the ability of *V. parahaemolyticus* strains to tolerate elevated temperatures, starvation and acidity in the natural environment has been linked to cross-protection against similar stressors encountered in the human host (Koga and Takumi, 1995; Wong et al., 1998).

Since ST3 and ST36 ecotypes are both responsible for large, non-localized outbreaks, we hypothesized that these ecotypes would share adaptive SNPs in key genes associated with fitness. We identified ten site-specific hotspot mutations affecting only ST3 and ST36. The occurrence of shared hotspot mutations is a strong sign of convergent evolution (Chattopadhyay et al., 2012). Although the precise role of these genes remains unclear, four of these genes encode proteins (cytochrome c551 peroxidase, short-chain dehydrogenase, MerR family transcriptional regulator and glutathione S-transferase) predicted to mediate the detoxification of reactive oxygen species, which can accumulate as a by-product of aerobic respiration or host immune defense during phagocytosis (Cabiscol et al., 2000). Additionally, two of these exceptional hotspots occur within genes associated with membrane transport (outer membrane gene *tolC* and a phosphate ABC transporter). In *V. cholerae*, *tolC* is upregulated in response to environmental stress and is essential for *V. cholerae* bile resistance and infection (Bina and Mekalanos, 2001). Meanwhile, phosphate-binding ABC transporters are integral components of phosphate-specific transport systems connected to both homeostasis and pathogenesis (Lamarche et al., 2008).

The predominance of ST3 and ST36 ecotypes among clinical isolates has been correlated with the lateral acquisition of mobile genomic islands that carry one or more virulence-associated genes (Hurley et al., 2006; Boyd et al., 2008). We detected tens of novel genes uniquely associated with the ST3 and ST36 ecotypes. These genes were largely hypothetical and did not include putative toxins such as the TDH and TRH hemolysins. A large proportion of these genes were located within the large, hyper-variable SI. The *V. parahaemolyticus* SI was described previously (Boyd et al., 2008) and is theorized to play a role in adaptive evolution through the capture of exogenous DNA (Chen et al., 2011). In ST3, SI-bound orthologs included a toxin-antitoxin module (ReIE/ReIB). Originally described as playing a role in programmed cell death, chromosomal toxin-antitoxin modules are known to mediate persistence in response to extracellular stress (Gerdes et al., 2005;

Maisonneuve et al., 2011). Persistence and the viable but non-culturable state are similar stress-induced states of dormancy that can increase resistance to extracelluar stressors (Wong and Wang, 2004; Ayrapetyan et al., 2015). Additional SI-bound orthologs included multiple proteins containing helix and transmembrane domains, indicating these proteins may mediate unknown interactions with the environment.

Several ecotype-specific orthologs were also detected within the *cps* gene cluster. In *V. parahaemolyticus*, genes within this cluster are involved in the synthesis of capsular polysaccharides (CPS) and play a role in the formation of biofilms (Chen et al., 2010). The *cps* cluster is known to be highly variable, and previous work has shown that recombination within this locus can be driven by chitin-induced transformation (Neiman et al., 2011). Five CPS proteins, unique to ST36, showed significant homology to putative virulence factors described in other Gram-negative bacteria. In clinical *V. vulnificus* strains, CPS is a primary toxin and has been shown to protect against phagocytosis during host invasion (Jones and Oliver, 2009). Taking advantage of ecotype-specific differences in *cps* gene content, a recent study announced the development of methods to specifically detect the ST36 ecotype (Whistler et al., 2015). Similarly, detection of an open reading frame (ORF8) associated with the f237 filamentous phage (Nasu et al., 2000) was previously proposed for the specific detection of the ST3 ecotype (Iida et al., 2001).

Conclusions

This study correlated the emergence of ST3 and ST36 ecotypes with the vertical and lateral evolution of genes associated with sensing, responding, and adapting to extracellular stimuli. The identification of adaptive mutations and adaptive genes, predicted to enhance fitness in both the natural environment and the human host, is an important step in dissecting mechanisms underlying the recent spread of these clinically important ecotypes. In particular, the evolution of genes associated with transcriptional regulation, defense against oxidative stress, membrane transport, and the formation of biofilms may play key roles in both environmental fitness and human pathogenesis. A better understanding of adaptive evolution, as a basis for emergence, could prioritize the public health response and lead to the development of new therapeutics or improved methods of detection.

References

Abbott SL, Powers C, Kaysner CA, Takeda Y, Ishibashi M, et al. 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**(12): 2891–2893.

Angiuoli SV, Salzberg SL. 2011. Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics* **27**(3): 334–342. doi: 10.1093/bioinformatics/btq665.

Atashpaz S, Khani S, Barzegari A, Barar J, Vahed SZ, et al. 2010. A robust universal method for extraction of genomic DNA from bacterial species. *Microbiol* **79**(4): 538–542. doi: 10.1134/s0026261710040168.

Ayrapetyan M, Williams TC, Oliver JD. 2015. Bridging the gap between viable but non-culturable and antibiotic persistent bacteria. *Trends Microbiol* **23**(1): 7–13. doi: 10.1016/j.tim.2014.09.004.

Bej AK, Patterson DP, Brasher CW, Vickery MCL, Jones DD, et al. 1999. Detection of total and hemolysin producing *Vibrio parahaemolyticus* in shellfish using multiplex PCR amplification of *tl*, *tdh* and *trh*. *J Microbiol Methods* **36**(3): 215–225. doi: 10.1016/S0167-7012(99)00037-8.

Bina JE, Mekalanos JJ. 2001. *Vibrio cholerae tolC* is required for bile resistance and colonization. *Infect Immun* **69**(7): 4681–4685. doi: 10.1128/IAI.69.7.4681-4685.2001.

Boyd EF, Cohen AL, Naughton LM, Ussery DW, Binnewies TT, et al. 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. *BMC Microbiol* **8**: 110. doi: 10.1186/1471-2180-8-110.

Broberg CA, Calder TJ, Orth K. 2011. *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes Infect* **13**(12–13): 992–1001. doi: 10.1016/j.micinf.2011.06.013.

Burdette DL, Yarbrough ML, Orvedahl A, Gilpin CJ, Orth K. 2008. *Vibrio parahaemolyticus* orchestrates a multifaceted host cell infection by induction of autophagy, cell rounding, and then cell lysis. *P Natl Acad Sci USA* **105**(34): 12497–12502. doi: 10.1073/pnas.0802773105.

Cabiscol E, Tamarit J, Ros J. 2000. Oxidative stress in bacteria and protein damage by reactive oxygen species. *Int Microbiol* **3**(1): 3–8.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**(15): 1972–1973. doi: 10.1093/bioinformatics/btp348.

Chattopadhyay S, Paul S, Dykhuizen DE, Sokurenko EV. 2013. Tracking recent adaptive evolution in microbial species using TimeZone. *Nat Protoc* **8**(4): 652–665. doi: 10.1038/nprot.2013.031.

Chattopadhyay S, Paul S, Kisiela DI, Linardopoulou EV, Sokurenko EV. 2012. Convergent molecular evolution of genomic cores in *Salmonella enterica* and *Escherichia coli*. *J Bacteriol* **194**(18): 5002–5011. doi: 10.1128/JB.00552-12.

Chattopadhyay S, Weissman SJ, Minin VN, Russo TA, Dykhuizen DE, et al. 2009. High frequency of hotspot mutations in core genes of *Escherichia coli* due to short-term positive selection. *P Natl Acad Sci USA* **106**(30): 12412–12417. doi: 10.1073/pnas.0906217106.

Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: Toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res* **40**(Database issue): 641–645. doi: 10.1093/nar/gkr989.

Chen Y, Dai J, Morris JG Jr, Johnson JA. 2010. Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol* **10**: 274. doi: 10.1186/1471-2180-10-274.

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

Chowdhury NR, Chakraborty S, Ramamurthy T, Nishibuchi M, Yamasaki S, et al. 2000. Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. *Emerg Infect Dis* **6**(6): 631–636. doi: 10.3201/eid0606.000612.

Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. 2010. Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput Biol* **6**(4): e1000732. doi: 10.1371/journal.pcbi.1000732.

Daniels NA, MacKinnon L, Bishop R, Altekruse S, Ray B, et al. 2000. *Vibrio parahaemolyticus* infections in the United States, 1973–1998. *J Infect Dis* **181**(5): 1661–1666. doi: 10.1086/315459.

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**(11): 4649–4654. doi: 10.1128/Aem.66.11.4649-4654.2000.

Felsenstein J. 1989. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164–166.

Gerdes K, Christensen SK, Lobner-Olesen A. 2005. Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol* **3**(5): 371–382. doi: 10.1038/nrmicro1147.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. 2005. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* **15**(10): 1451–1455. doi: 10.1101/gr.4086505.

González-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**(2): e0117485. doi: 10.1371/journal.pone.0117485.

González-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus LA, et al. 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *J Bacteriol* **190**(8): 2831–2840. doi: 10.1128/JB.01808-07.

González-Escalona N, Strain EA, De Jesus AJ, Jones JL, Depaola A. 2011. Genome sequence of the clinical O4:K12 serotype *Vibrio parahaemolyticus* strain 10329. *J Bacteriol* **193**(13): 3405–3406. doi: 10.1128/JB.05044-11.

Haendiges J, Timme R, Allard MW, Myers RA, Brown EW, et al. 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. doi: 10.3389/fmicb.2015.00125.

Ham H, Orth K. 2012. The role of type III secretion system 2 in *Vibrio parahaemolyticus* pathogenicity. *J Microbiol* **50**(5): 719–725. doi: 10.1007/s12275-012-2550-2.

Hazen TH, Lafon PC, Garrett NM, Lowe TM, Silberger DJ, et al. 2015. Insights into the environmental reservoir of pathogenic *Vibrio parahaemolyticus* using comparative genomics. *Front Microbiol* **6**: 204. doi: 10.3389/fmicb.2015.00204.

Hiyoshi H, Kodama T, Iida T, Honda T. 2010. Contribution of *Vibrio parahaemolyticus* virulence factors to cytotoxicity, enterotoxicity, and lethality in mice. *Infect Immun* **78**(4): 1772–1780. doi: 10.1128/IAI.01051-09.

Honda T, Ni YX, Miwatani T. 1988. Purification and characterization of a hemolysin produced by a clinical isolate of Kanagawa phenomenon-negative *Vibrio parahaemolyticus* and related to the thermostable direct hemolysin. *Infect Immun* **56**(4): 961–965.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**(1): 44–57. doi: 10.1038/nprot.2008.211.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. 2009. InterPro: The integrative protein signature database. *Nucleic Acids Res* **37**(Database issue): 211–215. doi: 10.1093/nar/gkn785.

Hurley CC, Quirke A, Reen FJ, Boyd EF. 2006. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. *BMC Genomics* **7**: 104. doi: 10.1186/1471-2164-7-104.

Iida T, Hattori A, Tagomori K, Nasu H, Naim R, et al. 2001. Filamentous phage associated with recent pandemic strains of *Vibrio parahaemolyticus*. *Emerg Infect Dis* **7**(3): 477–478. doi: 10.3201/eid0703.010325.

Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, et al. 2012. Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. *Science* **335**(6068): 587–590. doi: 10.1126/science.1212665.

Jones MK, Oliver JD. 2009. *Vibrio vulnificus*: Disease and pathogenesis. *Infect Immun* **77**(5): 1723–1733. doi: 10.1128/IAI.01046-08.

Joseph SW, Colwell RR, Kaper JB. 1982. *Vibrio parahaemolyticus* and related halophilic Vibrios. *Crit Rev Microbiol* **10**(1): 77–124. doi: 10.3109/10408418209113506.

Kodama T, Rokuda M, Park KS, Cantarelli VV, Matsuda S, et al. 2007. Identification and characterization of VopT, a novel ADP-ribosyltransferase effector protein secreted via the *Vibrio parahaemolyticus* type III secretion system 2. *Cell Microbiol* **9**(11): 2598–2609. doi: 10.1111/j.1462-5822.2007.00980.x.

Koga T, Takumi K. 1995. Nutrient starvation induces cross protection against heat, osmotic, or H_2O_2 challenge in *Vibrio parahaemolyticus*. *Microbiol Immunol* **39**(3): 213–215.

Lamarche MG, Wanner BL, Crepin S, Harel J. 2008. The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *FEMS Microbiol Rev* **32**(3): 461–473. doi: 10.1111/j.1574-6976.2008.00101.x.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**(9): 2178–2189. doi: 10.1101/gr.1224503.

Liu XF, Cao Y, Zhang HL, Chen YJ, Hu CJ. 2015. Complete genome sequence of *Vibrio alginolyticus* ATCC 17749T. *Genome Announc* **3**(1). doi: 10.1128/genomeA.01500-14.

Loyola DE, Navarro C, Uribe P, Garcia K, Mella C, et al. 2015. Genome diversification within a clonal population of pandemic *Vibrio parahaemolyticus* seems to depend on the life circumstances of each individual bacteria. *BMC Genomics* **16**: 176. doi: 10.1186/s12864-015-1385-8.

Maisonneuve E, Shakespeare LJ, Jorgensen MG, Gerdes K. 2011. Bacterial persistence by RNA endonucleases. *P Natl Acad Sci USA* **108**(32): 13206–13211. doi: 10.1073/pnas.1100186108.

Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, et al. 2003. Genome sequence of *Vibrio parahaemolyticus*: A pathogenic mechanism distinct from that of *V. cholerae*. *The Lancet* **361**(9359): 743–749. doi: 10.1016/s0140-6736(03)12659-1.

Martinez-Urtaza J, Baker-Austin C, Jones JL, Newton AE, González-Aviles GD, et al. 2013. Spread of Pacific Northwest *Vibrio parahaemolyticus* strain. *N Engl J Med* **369**(16): 1573–1574. doi: 10.1056/NEJMc1305535.

Matsuda S, Kodama T, Okada N, Okayama K, Honda T, et al. 2010. Association of *Vibrio parahaemolyticus* thermostable direct hemolysin with lipid rafts is essential for cytotoxicity but not hemolytic activity. *Infect Immun* **78**(2): 603–610. doi: 10.1128/IAI.00946-09.

Matsumoto C, Okuda J, Ishibashi M, Iwanaga M, Garg P, et al. 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**(2): 578–585.

Nair GB, Ramamurthy T, Bhattacharya SK, Dutta B, Takeda Y, et al. 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin Microbiol Rev* **20**(1): 39–48. doi: 10.1128/CMR.00025-06.

Nasu H, Iida T, Sugahara T, Yamaichi Y, Park KS, et al. 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. *J Clin Microbiol* **38**(6): 2156–2161.

Neiman J, Guo Y, Rowe-Magnus DA. 2011. Chitin-induced carbotype conversion in *Vibrio vulnificus*. *Infect Immun* **79**(8): 3195–3203. doi: 10.1128/IAI.00158-11.

Nishibuchi M, Taniguchi T, Misawa T, Khaeomanee-Iam V, Honda T, et al. 1989. Cloning and nucleotide sequence of the gene (*trh*) encoding the hemolysin related to the thermostable direct hemolysin of *Vibrio parahaemolyticus*. *Infect Immun* **57**(9): 2691–2697.

Okada K, Iida T, Kita-Tsukamoto K, Honda T. 2005. Vibrios commonly possess two chromosomes. *J Bacteriol* **187**(2): 752–757. doi: 10.1128/JB.187.2.752-757.2005.

Paranjpye R, Hamel OS, Stojanovski A, Liermann M. 2012. Genetic diversity of clinical and environmental *Vibrio parahaemolyticus* strains from the Pacific Northwest. *Appl Environ Microbiol* **78**(24): 8631–8638. doi: 10.1128/AEM.01531-12.

Paranjpye RN, Myers MS, Yount EC, Thompson JL. 2013. Zebrafish as a model for *Vibrio parahaemolyticus* virulence. *Microbiol* **159**(12): 2605–2615. doi: 10.1099/mic.0.067637-0.

Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**(7): 1641–1650. doi: 10.1093/molbev/msp077.

Raghunath P. 2014. Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in *Vibrio parahaemolyticus*. *Front Microbiol* **5**: 805. doi: 10.3389/fmicb.2014.00805.

Raimondi F, Kao JP, Fiorentini C, Fabbri A, Donelli G, et al. 2000. Enterotoxicity and cytotoxicity of *Vibrio parahaemolyticus* thermostable direct hemolysin in *in vitro* systems. *Infect Immun* **68**(6): 3180–3185.

Sahl JW, Steinsland H, Redman JC, Angiuoli SV, Nataro JP, et al. 2011. A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect Immun* **79**(2): 950–960. doi: 10.1128/IAI.00932-10.

Santos CL, Tavares F, Thioulouse J, Normand P. 2009. A phylogenomic analysis of bacterial helix-turn-helix transcription factors. *FEMS Microbiol Rev* **33**(2): 411–429. doi: 10.1111/j.1574-6976.2008.00154.x.

Shirai H, Ito H, Hirayama T, Nakamoto Y, Nakabayashi N, et al. 1990. Molecular epidemiologic evidence for association of thermostable direct hemolysin (TDH) and TDH-related hemolysin of *Vibrio parahaemolyticus* with gastroenteritis. *Infect Immun* **58**(11): 3568–3573.

Su YC, Liu C. 2007. *Vibrio parahaemolyticus*: a concern of seafood safety. *Food Microbiol* **24**(6): 549–558. doi: 10.1016/j.fm.2007.01.005.

Turner JW, Paranjpye RN, Landis ED, Biryukov SV, González-Escalona N, et al. 2013. Population structure of clinical and environmental *Vibrio parahaemolyticus* from the Pacific Northwest coast of the United States. *PLoS One* **8**(2): e55726. doi: 10.1371/journal.pone.0055726.

Weissman SJ, Beskhlebnaya V, Chesnokova V, Chattopadhyay S, Stamm WE, et al. 2007. Differential stability and trade-off effects of pathoadaptive mutations in the *Escherichia coli* FimH adhesin. *Infect Immun* **75**(7): 3548–3555. doi: 10.1128/IAI.01963-06.

Whistler CA, Hall JA, Xu F, Ilyas S, Siwakoti P, et al. 2015. Use of whole-genome phylogeny and comparisons for development of a multiplex PCR assay to identify sequence type 36 *Vibrio parahaemolyticus*. *J Clin Microbiol* **53**(6): 1864–1872. doi: 10.1128/JCM.00034-15.

Wong HC, Peng PY, Han JM, Chang CY, Lan SL. 1998. Effect of mild acid treatment on the survival, enteropathogenicity, and protein production in *Vibrio parahaemolyticus*. *Infect Immun* **66**(7): 3066–3071.

Wong HC, Wang P. 2004. Induction of viable but nonculturable state in *Vibrio parahaemolyticus* and its susceptibility to environmental stresses. *J Appl Microbiol* **96**(2): 359–366. doi: 10.1046/j.1365-2672.2004.02166.x.

Xu F, Ilyas S, Hall JA, Jones SH, Cooper VS, et al. 2015. Genetic characterization of clinical and environmental *Vibrio parahaemolyticus* from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. *Front Microbiol* **6**: 272. doi: 10.3389/fmicb.2015.00272.

Zhang L, Orth K. 2013. Virulence determinants for *Vibrio parahaemolyticus* infection. *Curr Opin Microbiol* **16**(1): 70–77. doi: 10.1016/j.mib.2013.02.002.

Zhou X, Gewurz BE, Ritchie JM, Takasaki K, Greenfeld H, et al. 2013. A *Vibrio parahaemolyticus* T3SS effector mediates pathogenesis by independently enabling intestinal colonization and inhibiting TAK1 activation. *Cell Rep* **3**(5): 1690–1702. doi: 10.1016/j.celrep.2013.03.039.

Contributions

- Contributed to conception and design: JWT, CTB, RM, EVA, MSS
- Contributed to acquisition of data: JWT, RM
- Contributed to analysis and interpretation of data: JWT, CTB, RM
- Drafted and/or revised the article: JWT, EVA, MSS
- Approved the submitted version for publication: JWT, EVA, MSS

Competing interests

The authors declare no competing interests.

Supplemental material

- **Table S1. List of the 30 *V. parahaemolyticus* genomes examined in this study. (DOC)**
  doi: 10.12952/journal.elementa.000117.s001
- **Table S2. Distribution of 344 core genes containing hotspot mutations between ST3 and ST36 ecotypes. (DOC)**
  doi: 10.12952/journal.elementa.000117.s002
- **Table S3. List of 42 orthologs unique to the ST3 ecotype. (DOC)**
  doi: 10.12952/journal.elementa.000117.s003
- **Table S4. List of 60 orthologs unique to the ST36 ecotype. (DOC)**
  doi: 10.12952/journal.elementa.000117.s004

Data accessibility statement

The data sets supporting the results of this article are available in the NCBI/GenBank repository (http://www.ncbi.nlm.nih.gov/). Accession numbers for the 30 genomes analyzed in this study are listed in Table S1.