## GENES AND GENOMES

# Sequence Analyses of Type IV Pili from Vibrio cholerae, Vibrio parahaemolyticus, and Vibrio vulnificus

Alisha M. Aagesen · Claudia C. Häse

Received: 8 July 2011 / Accepted: 3 February 2012 / Published online: 2 March 2012 © Springer Science+Business Media, LLC 2012

Abstract Bacterial surface structures called pili have been studied extensively for their role as possible colonization factors. Most sequenced Vibrio genomes predict a variety of pili genes in these organisms, including several types of type IV pili. In particular, the mannose-sensitive hemagglutinin (MSHA) and the PilA pili, also known as the chitin-regulated pilus (ChiRP), are type IVa pili commonly found in Vibrio genomes and have been shown to play a role in the colonization of Vibrio species in the environment and/or host tissue. Here, we report sequence comparisons of two type IVa pilin subunit genes, mshA and pilA, and their corresponding amino acid sequences, for several strains from the three main human pathogenic Vibrio species, V. cholerae, V. parahaemolyticus, and V. vulnificus. We identified specific groupings of these two genes in V. cholerae, whereas V. parahaemolyticus and V. vulnificus strains had no apparent allelic clusters, and these genes were strikingly divergent. These results were compared with other genes from the MSHA and PilA operons as well as another Vibrio pili from the type IVb group, the toxin co-regulated pilus (TCP) from V. cholerae. Our data suggest that a selective pressure exists to cause these strains to vary their MSHA and PilA pilin subunits. Interestingly, V. cholerae strains possessing TCP have the same allele for both mshA and pilA. In contrast, V. cholerae isolates without TCP have polymorphisms in their mshA and pilA sequences similar to what was observed for both V. parahaemolyticus and V. vulnificus. This data suggests a possible linkage between host interactions and maintaining a highly conserved type IV pili sequence in V.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00248-012-0021-2) contains supplementary material, which is available to authorized users.

 A. M. Aagesen · C. C. Häse (⊠)
Department of Biomedical Sciences, Oregon State University, Corvallis, OR 97331, USA
e-mail: hasec@science.oregonstate.edu *cholerae.* Although the mechanism underlying this intriguing diversity has yet to be elucidated, our analyses are an important first step towards gaining insights into the various aspects of *Vibrio* ecology.

## Introduction

*Vibrio* species are marine bacteria that naturally inhabit aquatic environments worldwide and are commonly associated with marine organisms. Some *Vibrio* species are pathogenic bacteria capable of producing life-threatening infections in humans typically following consumption of contaminated food, including seafood. Although the specific factors that contribute to the pathogenicity of vibrios in humans are well studied, little is known about the bacterial factors involved in the association of the bacteria with environmental organisms.

Bacteria display a variety of mechanisms that enable them to specifically interact with target cells. Many bacteria produce hair-like surface structures, called pili or fimbriae, which are often important for survival [1-3]. These adhesins have been clustered into groups based on amino acid sequence similarities among their pilin subunits [4]. One type of pili, the type IV group, are known to be involved in adhesion, immune escape, microcolony formation, transformation, and phage transduction [4] and are commonly found in Gramnegative bacteria, including numerous pathogens [4, 5]. Type IV pili are known to assist many bacterial species in survival in various environments, ranging from attachment to a variety of surfaces for biofilm formation [6–9] to colonizing the host [10–17]. These pili begin as prepilins possessing a hydrophilic leader peptide and are processed by a unique peptidase that cleaves the leader sequence to form a mature pilin protein [18]. After processing, mature pilin subunits assemble together to form pili through interactions between the conserved N-termini in the pilin cores, leaving the variable C-terminal regions to

interact with the environment [4]. Type IV pili are divided into two subclasses based on differences in amino acid sequence and length. Type IVa pili have both a shorter leader peptide and mature protein sequence, while type IVb pili have considerably longer leader sequences and overall length [4, 5, 18]. In addition to similarities in their amino acid composition, all type IV pili appear to have analogous architecture [4].

When examining the genomes of Gram-negative bacteria possessing type IV pili, type IVa pili biogenesis genes are scattered throughout the genome, but the genes or gene clusters are almost always flanked by the same genes, typically housekeeping genes. In addition, homologous gene sets for type IVa pili are found in virtually identical locations throughout more than 150 sequenced genomes. Considering these genes have not been found on any identifiable pathogenicity island, it suggests that these pili are ancient to many of the bacterial phyla possessing these genes [18]. In contrast, type IVb pili genes are fewer in number than type IVa genes and are typically found clustered within the genome. Moreover, the gene sequence order does not appear to be conserved amongst different organisms possessing the type IVb pili except for the universally conserved core proteins. In addition, when comparing N-terminal sequence homology, type IVa pilin subunits are more similar among themselves than to type IVb pilins or within the type IVb pili group. Furthermore, type IVa pili occur in bacteria with a broad host range, while type IVb pili have only been identified in colonizers of the human intestinal Vibrio species possess many type IV pili from both tract [4]. type IVa and b groups, but only a select few have been studied for their role in environmental and/or host survival. One thoroughly studied pili from the type IVb group is the toxin coregulated pilus (TCP) from Vibrio cholerae, and it is known for its key role in virulence [19-21]. It is expressed by V. cholerae classical and El Tor biotypes from the O1 and O139 serogroups [22]. TCP is composed of TcpA subunits and appears as thick bundles on the electron microscope [4]. TcpA is processed by a TCP-specific signal peptidase, TcpJ, to form mature pilin subunits for assembly [22, 23]. The structure of TCP consists of the conserved N-terminal  $\alpha$ -helices of TcpA buried in the core of the pilus, maximizing contact between subunits to provide overall strength. The structurally variable regions of the pilins interact to hold the core units together and coat the surface where interactions take place with the environment, i.e., the intestines [4]. In addition to colonization, TCP is the receptor for the CTX $\Phi$  phage [24, 25].

An additional well-studied *V. cholerae* type IV pilus is the mannose-sensitive hemagglutinin (MSHA), which belongs to the type IVa group. When examining operon composition, MSHA in *V. cholerae* consists of two operons where one operon encodes five prepilin subunits, including the major pilus subunit MshA, and the other contains genes involved in assembly and secretion [26]. In *V. cholerae*, the PilD peptidase has been shown to process the MshA subunits for

assembly of the mature pilus structure [27, 28]. The MSHA pilus hemagglutinates red blood cells [29, 30] and is a receptor for filamentous phage [31-33]. It has been studied extensively in V. cholerae to identify any involvement in host colonization [19, 21, 34]. In V. cholerae, only the El Tor biotypes produce functional MSHA pili [29, 30], and during human colonization studies, the protein was repressed [35]. Expression of the MSHA pilus was tightly regulated so that when TCP was expressed, the MSHA protein was repressed; therefore, the MSHA pilus is considered an anticolonization factor in human disease [36]. When the MSHA pilus was constitutively expressed during colonization, it resulted in immune system recognition [35]. Thus, the MSHA pilus does not appear to be a virulence factor for V. cholerae, suggesting that expression of the gene product is for utilization in the environment. Studies have shown that the MSHA pilus is used to adhere to zooplankton exoskeletons as a survival strategy in the aquatic environment [37, 38], presumably by forming biofilms. V. cholerae and Vibrio parahaemolyticus are known to use the MSHA pilus to form biofilms on various surfaces [6, 8, 38], including chitin [39], which provides some supporting evidence for the role of the MSHA pilus in environmental survival.

Another pilus found in Vibrio spp. is the type IVa PilA pilus, also known as the chitin-regulated pilus (ChiRP). The PilA operon is composed of five open reading frames that constitute a single operon, consistent with other type IVa pili [28]. A mature PilA pilus is composed of PilA subunits that were processed by the PilD peptidase [28], the same peptidase that processes the MshA pilin subunits [27, 28]. The PilD peptidase is the fourth open reading frame in the PilA operon [28]. The PilA type IVa pilus is an integral player in the V. cholerae chitin utilization program [39]. Expression of the PilA protein has been shown to be induced by chitin in both V. cholerae [39] and V. parahaemolyticus [6]. PilA is involved in biofilm formation [6, 10], adherence to human epithelial cells [10], and colonization of oysters [11]. It has been implicated as a virulence factor for V. vulnificus [10], although direct evidence of its role in virulence has not been clearly described in other human pathogenic vibrios.

Taken together, the studies of the type IVa pili MSHA and PilA in various *Vibrio* spp. suggest that these proteins might be utilized by vibrios for environmental survival by attaching to chitinous substrates such as zooplankton. In contrast, the type IVb pilus, TCP, from *V. cholerae*, is critical for host colonization and has not be implicated in environmental survival, pointing out the possibility of two very distinct roles for the different subclasses of type IV pili.

During our efforts to investigate the roles of MSHA and PilA in *V. parahaemolyticus* colonization of the Pacific oyster, *Crassostrea gigas*, we noted sequence heterogeneities in these genes. This led us to examine these genes in other human pathogenic *Vibrio* species, such as *V. cholerae* and *V. vulnificus*. Here, we present a comparative sequence analysis of the *mshA* and *pilA* pilin genes from several strains of *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*. These sequence analyses suggest that a selective environmental pressure has been applied to these genes, resulting in the observed sequence heterogeneities for all three *Vibrio* species examined.

## **Materials and Methods**

# **Bacterial Strains**

Thirteen of the *V. parahaemolyticus* bacterial strains sequenced were kindly provided by Dr. Yi-Cheng Su, Oregon State University Seafood Laboratory, Astoria, OR, USA. Genomic DNA for five tdh/trh negative strains of *V. parahaemolyticus* was obtained from Dr. Narjol-Gonzalez-Escalona, FDA, College Park, MD, USA. Genomic DNA for ten of the *V. vulnificus* strains sequenced were provided by Dr. Paul Gulig, University of Florida, Gainesville, FL, USA. Five of the *V. vulnificus* strains sequenced were provided by Dr. Kathy O'Reilly, Oregon State University, Corvallis, OR, USA. Bacterial strains were grown on Luria–Bertani agar supplemented with sodium chloride to a final concentration of 2%. All strains used in this study are listed in Table 1.

#### Sequencing

Genomic DNA from *V. parahaemolyticus* and *V. vulnificus* strains were isolated using the Qiagen DNeasy blood and tissue kit, following the protocol for DNA isolation included in the kit. Primers for sequencing each gene were designed for the region approximately 100 base pairs upstream from the start codon and 100 bp downstream of the stop codon for the gene of interest (Table 2). Polymerase chain reaction (PCR) was conducted using Invitrogen Platinum HiFi Supermix, following their standard protocol for PCR. PCR samples were quantified using the NanoDrop Spectrophotometer ND-1000. Sanger sequencing reactions for *V. parahaemolyticus* and *V. vulnificus* PCR products were performed at the Center for Genomic Research and Bioinformatics (CGRB), Oregon State University, Corvallis, OR, USA.

## In Silico Analyses

The *in silico* sequence data for all the *V. cholerae* strains and additional *V. parahaemolyticus* and *V. vulnificus* strains were obtained from the Department of Energy Joint Genome Institute website: http://img.jgi.doe.gov/cgi-bin/pub/main. cgi. The *V. parahaemolyticus* and *V. vulnificus* sequenced DNA was translated into their predicted amino acid sequences using SeqTool and sequence alignments were created in ClustalW at the bioinformatics website for the CGRB: http://

bioinfo.cgrb.oregonstate.edu/. Maximum likelihood phylogenetic trees were constructed using the MEGA 5 program: http://www.megasoftware.net/ using the Tamura–Nei model with nucleotide substitutions. Bootstrap values were calculated with 500 replicates. For the analysis of synonymous and nonsynonymous substitutions, calculations were made using the Synonymous Non-synonymous Analysis Program (SNAP): www.hiv.lanl.gov [40]. The program is based on the Nei and Gojobori [41] method for calculating synonymous and nonsynonymous rates of substitution with the incorporation of Ota and Nei [42] statistics. The package is described by Ganeshan et al. [43].

## Results

# Sequence Alignments

Overall, the sequence alignments for the DNA encoding the mshA and pilA genes from different strains of V. cholerae, V. parahaemolvticus, and V. vulnificus showed considerable sequence heterogeneity (Supplemental Figs. 1 and 2). Although the immediate 5' regions are highly conserved in both genes, most of the gene sequences varied depending on the strain. Interestingly, V. cholerae exhibited distinct groupings for both genes, separating most clinical isolates from environmental isolates. In contrast, V. parahaemolyticus and V. vulnificus strains did not appear to group based on isolate origin or any other phenotype. Sequence alignments of the predicted amino acid sequences of MSHA and PilA from V. cholerae, V. parahaemolyticus, and V. vulnificus are shown in Figs. 1 and 2. For V. parahaemolyticus and V. vulnificus, the predicted amino acids sequences for MSHA and PilA from both environmental and clinical isolates displayed notable sequence heterogeneity. With V. cholerae strains, most clinical isolates had conserved sequences for both MSHA and PilA. Most environmental isolates exhibited marked sequence heterogeneity, comparable to what was observed for the V. parahaemolyticus and V. vulnificus isolates.

## Phylogenetic Trees

Maximum likelihood (ML) phylogenetic trees were constructed from the *mshA* (Fig. 3) and *pilA* (Fig. 4) sequences for the *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus* isolates. Similar to the DNA and amino acid alignments, the *mshA* (Fig. 3a) and *pilA* (Fig. 4a) ML phylogenetic trees for *V. cholerae* clustered most clinical isolates into one branch, while environmental isolates exhibited various branching patterns. When ML phylogenetic trees were constructed for these two gene sequences from *V. parahaemolyticus* (Figs. 3b and 4b) and *V. vulnificus* (Figs. 3c and 4c), no discernable

Table 1 Strains used in this study

Vibrio cholerae			
Strain	Serogroup	Biotype	Relevant genotype
MAK757	O1, Ogawa	El Tor	TCP+, CTxA+
NCTC 8457	O1, Inaba	El Tor	TCP+, CT-
B33	O1, Ogawa	Hybrid Classical/	TCP+, CTxA+
O395	O1, Ogawa	El Tor Classical	TCP+, CT+
2470-80	O1, Inaba	El Tor	TCP+, CT-
N16961	O1, Inaba	El Tor	TCP+, CT+
V51	O141		TCP+, CT-
V52	O37		TCP+, CT+
MO10	0120		TODIOTI
MOIO	0139		ICP+, C1+
MZO-2	014		TCP- CT-
1587	012		ТСР–, СТ–
D. (2005	0.105		
RC385	0135		TCP-, C1-
623-39	non-01/-0139		TCP- CT-
AM-19226	039		TCP-, CI-
MJ-1236	O1, Inaba	El Tor "Matlab	TCP+, CT+
		variant	
VL426	non-O1/-O139	Albensis	TCP-, CT-
TM 11079-80	O1, Ogawa	El Tor	TCP-, CT-
RC9	O1, Ogawa	El Tor	TCP+, CT+
TMA21	non-O1/-O139		TCP-, CT-
BX 330286	O1, Inaba	El Tor	TCP+, CT+
CIRS 101	O1. Inaba	El Tor.	TCP+, CTxA+
M(( )	01	classical ctxB	TODIOT
W100-2	01		ICr⊤, C1−
CT 5369-93			TCP-, CT-
RC27	01	Classical	TCP+, CT-
INDRE 91/1	O1, Inaba	El Tor	TCP+, CTxA+
Khuio nau-Lau-	utiona		
strain	Serotype	Relevant	Strain information
Stam	Serviype	genotype	Saun mormation
RIMD 2210633	O3:K6	tdh+/trh-	Clinical strain from

tdh+/trh+

tdh+/trh+

tdh+/trh+

tdh+/trh+

O4:K68

O1:K2

Strain information	GenBank (PilA,MshA)
Pre-7th pandemic patient	ZP_01954444, ZP_01953747
isolate, mildly toxigenic,	
from Celebes Islands in 1937	ZB 01071084 ZB 01070012
non-pandemic	21_019/1084, 21_019/0915
Clinical isolates Beira,	ZP_04401766, ZP_01974939
Mozambique 2004	
Clinical isolate, strain of the 6th	YP_001217923, YP_001218677
isolated in 1965 from India	
Nontoxigenic environmental water	ZP 01677376, ZP 01677345
isolate from the Gulf Coast, 1980,	
clonal with TCP+ CT+ isolates	
Clinical isolate from Bangladesh, 19/1	NP_232053, NP_230063
Clinical Isolate from the	ZP_00748678, ZP_00749816
Clinical isolate from Sudan	ZP 00746513 ZP 00747249
limited epidemic, not endemic	21_00710010, 21_00717217
Clinical isolate from India and	ZP_00758906, ZP_00758992
Bangladesh outbreak 1992,	
clinical isolate from Bangladesh	ZP 01979911 ZP 01978309
patients with diarrhea in 2001	21_01070011, 21_01070500
Clinical isolate from Lima, Peru,	ZP_01949212, ZP_01949969
1994, limited epidemic, not	
Persistent and luminescent	ZP 00751854 ZP 00753463
environmental plankton isolate,	21_00751054, 21_00755405
Chesapeake Bay, 1998	
Environmental water isolate	ZP_01983325, ZP_01981981
from Bangladesh, 2002 Clinical isolate from	ZP 04962566 ZP 04962108
Bangladesh, 2001	21_04902300, 21_04902108
Clinical isolate from patients	YP_002877671, YP_002876957
with acute diarrhea, Matlab,	
Bangladesh 1994 Disaasad fish from Elba Piyar	7P 04413813 7P 04414508
Germany	21_04413813, 21_04414308
Environmental sewage isolate	ZP_04410672, ZP_04409450
from Brazil, 1980	
Clinical isolate from Kenya 1985	ZP_04408866, ZP_04409227
Environmental seawater isolate	ZP_04403511, ZP_04402128
Irom Brazil, 1982 Water isolate from	ZP 04397022 ZP 04396555
Australia, 1986	<u></u>
Clinical isolate from Dhaka,	ZP_05420851, ZP_05417751
Bangladesh 2002	
1937 outbreak Indonesia, pre-7th pandemic isolate	YP_002811095, YP_002809171
Sewage, Brazil 1993	ZP 06048448, ZP 06049688
Indonesia 1991	ZP 06036099, ZP 06035331
Mexico 1991, first	ZP_06030661, ZP_06028721
case of 7th pandemic	
in Mexico	
GenBank (PilA,MshA)	
NP 798902, NP 799077	
ZP_05904882, ZP_05905780	

Bangladesh 1998ZP\_05891366, ZP\_05889900Isolated from sediment<br/>at Goose Point oyster farm,<br/>Willapa Bay, Washington,<br/>October 2002JF923890, JF923914

ZP\_05776528, ZP\_05778018

Osaka, Japan 1996

Peru-466

AN-5034

SFL1009

K5030

Table 1 (continued)

513	

Table 1 (continu	ucu)				
SFL1027	05	tdh+/trh+	Isolated from sediment at Oregon Oyster farm Yaquina Bay, Oregon, December 2002	JF923885, JF923903	
SFL1050	O1:K7	tdh+/trh+	Isolated from sediment at Goose Point oyster farm, Willapa Bay, Washington, July 2003	JF923892, JF923904	
SFL1079	O1:K4	tdh+/trh+	Isolated from seawater at Goose Point oyster farm, Willapa Bay, Washington, July 2003	JF923894, JF923905	
SFL1080	O5:K1	tdh+/trh-	Isolated from sediment at Goose Point oyster farm, Willapa Bay, Washington, July 2003	JF923891, JF923906	
10290	O4:K12	tdh+/trh+	1997 Washington outbreak strain	JF923888, JF923916	
10292	O6:K18	tdh+/trh+	1997 Washington outbreak strain	JF923893, JF923915	
BE98-2029	O3:K6	tdh+/trh-	1998 Texas outbreak strain	JF923886, JF923901	
O27-1C1	O5:K15	tdh+/trh+	1997 Oregon outbreak strain	JF923887, JF923908	
M25-0B	O4	tdh-/trh-	Environmental isolate from Washington, 1993	JF923900, JF923913	
UCM-V586	O8:K22	tdh-/trh-	Environmental isolate from Spain, 2003	JF923899, JF923912	
UCM-V441	O4	tdh-/trh-	Environmental isolate from Spain, 2002	JF923898, JF923911	
049-2A3	O4:K29	tdh-/trh-	Environmental isolate from Oregon, 1997	JF923896, JF923909	
357-99	O3	tdh-/trh-	Clinical isolate from Peru, 1999	JF923897, JF923910	
ATCC 17802		trh+	Shirasu food poisoning, Japan 1965	JF923889, JF923907	
vibrio vuinificus	G 1		<b>T</b> 7* 1		
Strain	Capsule	vcg	viruience group	Strain information	GenBank (PIIA, MsnA)
98-783 DP-A1	1	E	2	Environmental isolate	JF923941, JF923921
99-520 DP-B8	2	E	3	Environmental isolate	JF923932, JF923920
99-581 DP-C/	2	E	3	Environmental isolate	JF923930, JF923917
99-584 DP-B12	2	E	2	Environmental isolate	JF923931, JF923922
99-736 DP-C7	2	E	2	Environmental isolate	JF923934, JF923924
99-738 DP-B5	2	E	5	Environmental isolate	JF923942, JF923923
S1-13	1	E	4	Environmental isolate	JF923933, JF923919
ATL-9580	1	С	5	Clinical isolate	JF923935, JF923918
CMCP6	1	С	4	Clinical isolate	NP_760518,NP_760356
YJ016	2	С	5	Clinical isolate, Taiwan	NP_935571, NP_935733
CP Mussel 10 PT					JF923939, JF923929
95-10-15 PT				isolated 10/18/05	JF923938, JF923928
960926 -1/4c PT				isolated 10/19/05	JF923937, JF923927
ATCC 27562				clinical isolate, Florida	JF923936, JF923926
OLOL-1				Katrina, isolated 10/18/05	JF923940, JF923925

grouping patterns appeared for either species, unlike the *V. cholerae* phylogenetic trees.

# **Substitution Analyses**

We analyzed *mshA* and *pilA* for the rate of synonymous (silent)  $(d_S)$  and nonsynonymous (structural)  $(d_N)$  changes

for the *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus* isolates. For *mshA* from *V. cholerae*, the rate of synonymous  $(d_S)$  was 0.759, while the rate of nonsynonymous  $(d_N)$  was 0.471, with a  $d_N/d_S$  ratio of 0.621 (Table 3). The rate of synonymous changes for *V. parahaemolyticus* was 0.746 and that for *V. vulnificus* was 0.662. The rate of nonsynonymous changes for *V. parahaemolyticus* and *V. vulnificus* was 0.431 and 0.384, respectively. This resulted in a  $d_N/d_S$ 

Table 2 Primers used in this study

Locus	Gene	Primer name	Primer sequence
pilA	VP2523	Shorter5' VP2523 SpeI P1	5'-GATAATTGGGGGGCATATCAACCTCTATAGTTTG-3'
		New3'VP2523 NotI P4	5'-ACCATGGGTGCATTCGTTGCAACCATCTGGATT-3'
		VP2523 sequence	5'-GCAACTTTCTACCAAAGAGTTTTACCTCACTCG-3'
	VV2278	VV pilA seq primer	5'-GTAAGTAACCAGATGTAAATAAAG-3'
		3'VV pilA seq P2	5'-GCCAAAAATCGCGCTTAGCTG-3'
mshA	VP2698	New5'VP2698 SpeI P1	5'-CGTAAACGCATTAAAGCCGCGATGCGCTATCCG-3'
		New3'VP2698 NotI P4	5'-CCATTAAGGTGAAACCACGAGTTTTCATTCAGT-3'
		5'VP2698 seq2	5'-TCGTCATTCTGCTCAAGCGGTAGA-3'
	VV2940	VV mshA seq primer	5'-CAAATGCTAAATGTACTTATATTC-3'
		VV 3' mshA seq P2	5'-CTGCCAGTGCCAATATAGCGACTG-3'

of 0.577 for *V. parahaemolyticus* and 0.580 for *V. vulnificus* (Table 3). For *pilA*, the rate of synonymous changes was 1.109 for *V. cholerae*, 1.691 for *V. parahaemolyticus*, and 1.186 for *V. vulnificus*. The rate of nonsynonymous changes was 0.629 for *V. cholerae*, 0.642 for *V. parahaemolyticus* and 0.503 for *V. vulnificus*. This resulted in a  $d_N/d_S$  of 0.567, 0.380, and 0.424 for *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*, respectively (Table 3).

## **Region Analyses**

To compare the diversity of *mshA* and *pilA*, we examined neighboring genes from their respective operons, mshC and pilB, as well as the type IV pilin peptidase, pilD. The rate of synonymous and nonsynonymous changes for mshC was 0.135 and 0.039 for V. cholerae, 0.229 and .017 for V. parahaemolyticus, and 0.042 and 0.015 for V. vulnificus. This resulted in a  $d_N/d_S$  ratio of 0.290 for V. cholerae, 0.072 for V. parahaemolyticus, and 0.356 and V. vulnificus (Table 3). For *pilB*, the rates of synonymous and nonsynonymous for V. cholerae, V. parahaemolyticus, and V. vulnificus was 0.176 and 0.008, 0.288 and 0.037, and 0.208 and 0.016 respectively. The  $d_N/d_S$  ratio for *pilB* was 0.047 for V. cholerae, 0.127 for V. parahaemolyticus, and 0.074 for V. vulnificus. For pilD, the synonymous and nonsynonymous rates calculated for V. cholerae were 0.122 and 0.005 with a  $d_{\rm N}/d_{\rm S}$  of 0.039. The V. parahaemolyticus strains used to calculate the synonymous and nonsynonymous rates of substitution for *pilD* had identical sequences; thus, the synonymous and nonsynonymous rates of substitution were zero, and the  $d_N/d_S$  ratio cannot be calculated. These rates are comparable with data from Chattopadhyay et al. [46], which calculated the rates of synonymous and nonsynonymous substitutions for pilD from V. vulnificus as 0.092 and 0.007 with a  $d_{\rm N}/d_{\rm S}$  ratio of 0.076.

#### TcpA and TcpJ

To compare the findings for *mshA* and *pilA* with another type IV pilin and its corresponding peptidase, we calculated the rates of synonymous and nonsynonymous substitutions for the toxin co-regulated pilus pilin subunit *tcpA* from *V. cholerae* and its processing leader peptidase *tcpJ* (Table 3). Only 13 *V. cholerae* strains out of the available 25 possess *tcpA* and *tcpJ*. The  $d_S$  and  $d_N$  for *tcpA* was 0.486 and 0.052 with a  $d_N/d_S$  ratio of 0.106. For *tcpJ*, the  $d_S$  and  $d_N$  was 0.003 and 0.000 with a  $d_N/d_S$  ratio of 0.000.

## Discussion

The results from our sequence analyses of the mshA and pilA genes from several strains of three human pathogenic Vibrio species, V. cholerae, V. parahaemolyticus, and V. vulnificus, suggested that the various alleles observed were the result of selective pressure. When examining the V. cholerae predicted amino acid alignment (Fig. 1a) and phylogenetic tree (Fig. 3a) for the mshA gene, one distinct grouping emerged with highly conserved sequences for the MSHA pilin subunit. In fact, the isolates in this group, identifiable as one branch of the phylogenetic tree (Fig. 3a), were primarily from the O1 serogroup (13 out of 15) and clinical isolates (11 out of 15). This differs considerably from the remaining V. cholerae isolates examined, which were predominately environmental, non-O1/O139 strains (9 out of 10) with no apparent grouping pattern in the phylogenetic tree (Fig. 3a). When comparing the predicted amino acid alignments and phylogenetic trees for the V. parahaemolyticus (Figs. 1b and 3b) and V. vulnificus (Fig. 1c and 3c) strains sequenced, no grouping could be established based on either isolation source or phenotype, in contrast to what was observed for V. cholerae.

а

037\_clinical\_V52 01\_clinical\_CIRS\_101 0139\_clinical\_M010 01\_RC27 01\_clinical\_INDRE\_91/1 01\_clinical\_N16961 01\_environmental\_BX\_330286 Ol\_environmental\_BX Ol\_clinical\_RC9 Ol\_clinical\_MJ-1236 Ol\_clinical\_M66-2 Ol\_clinical\_B33 Ol\_clinical\_NCTC\_84 Ol\_clinical\_NCTC\_84 Ol\_clinical\_0395 01 environmental\_2740-80 non-01/0139\_environmental VL42 environmental\_CT\_5369-93 012\_clinical\_1587 RC385 0135\_environmental 0141\_clinical\_V51 014\_clinical\_MZO-2 non-01/0139\_environmental\_TMA 039\_clinical\_AM-19226 non-01/0139\_environmental 623-Ol\_environmental\_TM\_11079-80 037\_clinical\_V52 01\_clinical\_CIRS\_101 0139\_clinical\_MO10 01\_RC27 O1\_clinical\_INDRE\_91/1 O1\_clinical\_N16961 Ol\_environmental\_BX\_330286 Ol\_environmental\_BX Ol\_clinical\_RC9 Ol\_clinical\_MJ-1236 Ol\_clinical\_M66-2 Ol\_clinical\_B33 Ol\_clinical\_NCTC\_84 Ol\_clinical\_NAK757 Ol\_clinical\_0395 8457 01\_environmental\_2740-80 non-01/0139\_environmental VL42 environmental\_CT\_5369-93 012\_clinical\_1587 \_RC385 0135\_environmental 0141\_clinical\_V51 014\_clinical\_MZO-2 non-O1/O139\_environmental\_TMA\_ 039\_clinical\_AM-19226 non-01/0139\_environmental 623-Ol\_environmental\_TM\_11079-80 037\_clinical\_V52 01\_clinical\_CIRS\_101 0139\_clinical\_MO10 01\_RC27 O1\_clinical\_INDRE\_91/1 O1\_clinical\_N16961 O1\_environmental\_BX\_330286 01 \_clinical\_RC9 O1\_Clinical\_MC9 O1\_clinical\_MJ-1236 O1\_clinical\_M66-2 O1\_clinical\_B33 O1\_clinical\_NCTC\_84 O1\_clinical\_MAK757 O1\_clinical\_0395



KRQGGFTLIELVVVIVILGILAVTAAPRFLNLQGDAR

Figure 1 Amino acid sequence alignment of MshA from Vibrio cholerae (a), Vibrio parahaemolyticus (b), and Vibrio vulnificus (c). The predicted amino acid sequence alignments of MshA for V. cholerae (a), V. parahaemolyticus (b), and V. vulnificus (c) were constructed using the

8457

Ol environmental 2740-80

0141\_clinical\_V51 014\_clinical\_Mz0-2

039\_clinical\_AM-19226

non-01/0139\_environmental environmental\_CT\_5369-93 012\_clinical\_1587 0135\_environmental\_RC385

ClustalW program. White indicates normal residues. Green are similar residues. Pink are identical residues. Black indicates globally conserved residues

ASL GL

~	
tdh+/trhenviron_SFL1080	MKRQGGFTLIELVVVIVILGIL <mark>A</mark> VTAAPRFLNLQ <mark>S</mark> DARESALQGLKGAIDGASGI <mark>VF</mark> GKA
tdh+/trh+_environ_SFL1050	MKRQGGFTLIELVVVIVILGILAVTAAPRFLNLQSDARESALQGLKGAIDGASGIVFGKA
tdh+/trh+_environ_SFL1027	MKRQGGFTLIELVVVIVILGIL <mark>A</mark> VTAAPRFLNLQ <mark>S</mark> DARESALQGLKGAIDGASGIVFGKA
tdh+/trh+_clinical_027-1C1	MKRQGGFTLIELVVVIVILGILAVTAAPRFLNLQSDARESALQGLKGAIDGASGIVFGKA
tdh+/trh+ clinical 10290	MKRQGGFTLIELVVVIVILGIL VTAAPRFLNLQSDARESALQGLKGAIDGASGIVFGKA
tdh-/trh- environ M25-0B	MKRQGGFTLIELVVVIVILGILAVTAAPRFLNLQNDARESALQGLKGA <mark>L</mark> DGASGIVFGKA
tdh+//trh-clinical RIMD 221063	MKRQGGFTLIELVVVIVILGIL <mark>A</mark> VTAAPRFLNLQSDARESALQGLKGAIDGASGIVFGKS
tdh+/trh+ Peru-466	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLOSDARESALOGLKGAIDGASGIVFGKS
tdh+/trh+ AN-5034	MKROGGFTLIELVVVIVILGIL <mark>A</mark> VTAAPRFLNLOSDARESALOGLKGAIDGASGIVFGKS
tdh+/trh+ K5030	MKROGGFTLIELVVVIVILGIL <mark>A</mark> VTAAPRFLNLO <mark>sdaresalo</mark> glkgaidgasgivfgks
tdh+/trh- clinical BE98-2029	MKROGGFTLIELVVVIVILGIL <mark>A</mark> VTAAPRFLNLOSDARESALOGLKGAIDGASGIVFGKS
trh+ clinical ATCC 17802	MKRQGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNSALQGLKGALDGAAGIVIGKA
tdh-/trh- environ 357-99	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLONDARASSLOGLKGATEGAGIVFGKA
tdh-/trh- environ UCM-V441	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLOSDARASSLDGLKGAMOGASGIVFGKA
tdh-/trh- environ UCM-V586	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNAVLEGLSGATOGASGIVTGKA
tdh+/trh+ environ SFL1009	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNAVLEGLSGATOGASGIVTGKA
tdh-/trh- environ 049-2A3	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNATLEGLAGAINGASGITTGKA
tdh+/trh+ clinical 10292	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNSALEGLKGATAGAAGITTGKA
tdh+/trh+ environ SFL1079	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARDS7LBGLKGAT7GASGIVTGKS
+dh+/+rh- environ SEL1080	
tdh+/trh+ environ SFL1050	AT NELL TVSSCK
tdh+/trh+ environ SFL1027	
tdh+/trh+ clinical 027-101	
tdh+/trh+ clinical 10290	
tdh-/trh- environ W25-0B	
tdh+/trh-slinical BIND 221063	
tdh+/trh+ Doru_A66	
+dh+/+rh+ 3N_5034	
tdh+/trh+_KN-SUS4	
tdh+/trh- alinical BE98 2020	
trb+ clinical ATCC 17802	
tdh_/trh_ onviron 357-99	
tdh_/trh_ environ UCW_VAA1	
tdh_/trhenviron_UCM_V596	
tdh+/trh+ environ SEL1009	TO CHORE TO UNDI CHNUN CEVERITUTI TARE AVUTO TO A DED VOIOTCO
tdh-/trh- environ 049-213	
tdh+/trh+ clinical 10292	
tdh+/trh+ environ SFL1079	
tdh+/trh-environ_artions	
tdn+/trnenviron_SFL1080	TSVVAT VSGDLKAGYGNTLKTITDCKCYVVNBASASTAGNEASTSIEDSC
tdh+/trh+_environ_SFL1050	TSVAT VSGDLKAGIGNTLKTITDCKCIVVMNLASASTAGNEASTSIEDSGC
tdn+/trn+_environ_SFL102/	TSVVAT VSGDLKAGIGNTLKTITDCKCIVVMNLASASTAGNEASTSIEDSGC
tdn+/trn+_clinical_02/-1Cl	TSVAT VSGDLKAGIGNTLKTITDCKUVVVNLASASTAGNEASTSIEDSC
tdh+/trh+_clinical_10290	ASVANV SEDI AGYGNALKTITDEKCIVVMNA SASTAGNEASASIEDSEC
tdn-/frn- environ_M25-08	SSVAN VRDGLSKCONSG-RESTITGTSCHTOMNOANSSAAGFEAKVATEKGC
tdn+//clinical_kimD_221063	TPNKLVAT I SOULKKGLDAGNAAKVKACNCI VITT SATSAAVSTAAVTDT CC
tdn+/trn+_Peru-466	TPNKE VAT I SOULKKOLDAGN AAKVKACNCI VITIT JAASAAVSTAAVTOT CC
tdn+/trn+_AN-5034	TPNKEVATTI SGDLKKGLDAGNTAAKVKACNCI VITTI SATSAAVSTAAVIDTGC
tdh+/trh+_K5030	TPNKL VAT I SODIKKGLDAGNAAKVKACNCYVTYTSATSAAVSTAAVTDTGC
tdh+/trhclinical_BE98-2029	PNKL WAY I SOUTHKELDAGN AAKVKACNO AVTYT SANSAAVS TAAVTOT GC
the dinical_ATCC_17802	NAITFCIEGNDASNOSUOUSMLPTAQEPIITVNOCON
tan-/trnenviron_35/-99	ADTSTTIGFKGAKTTNEGIUTHSQUAGTAAASAAKATVTLGSACK-
tdn-/trnenviron_UCM-V441	AGSGAAPVITSADADC
tun-/trnenviron_UCM-V586	
Additional and an and a second	
tdh+/trh+_environ_SFL1009	EADNDQWVTF IANYTNOC VKYTM ADANTPAVVEVTTTTSE TAP
tdh+/trh+_environ_SFL1009 tdh-/trhenviron_049-2A3	EADNDQWVTF IANYTNQC KKTMAADANTPAVVEVTTTTSE TAP YSPLNGGPAW-SAAAV IAGYTKAC VVYGAANENTAGKAELTET TH KQ-
tdh+/trh+_environ_SFL1009 tdh-/trhenviron_049-2A3 tdh+/trh+_clinical_10292	EADNDQWTTFGIANYTNQC VKYTMAADANTPAVVEVTTTTSECTAP YSPLNGGPAV-SAAAVGIAGYTKACVVYYGAANENTAGKAELTETOTHCKQ- NAVTFCCDGYTDKCVTYTQATSSAAATVKTVTTC

Figure 1 (continued)

Reviewing the sequence data for the PilA pilin subunit from *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*, the *pilA* sequences exhibited a trend similar to what was observed for the MSHA pilin subunit. For *V. cholerae* strains, a group of highly conserved PilA sequences emerged and were primarily from the O1 serogroup (13 out of 14) and of clinical origin (11 out of 14). The remaining isolates were predominately non-O1/O139 (10 out of 11) and from an environmental source (6 out of 11). They did not have any clear pattern to their alignment (Fig. 2a) or tree branching (Fig. 4a). Consistent with the *mshA* findings, no apparent grouping pattern was observed for either the amino acid alignment or branching on the phylogenetic tree for any of the *V. parahaemolyticus* (Figs. 2b and 4b) and *V. vulnificus* (Figs. 2c and 4c) *pilA* genes sequenced. Taken together, our hypothesis is that a selective pressure has caused the differences observed in these two type IVa pili in *V. cholerae*, *V. parahemolyticus*, and *V. vulnificus*.

To test for selective pressure, the synonymous and nonsynonymous nucleotide substitution rates were calculated to determine a  $d_N/d_S$  ratio [41]. In protein-coding sequences, synonymous substitutions ( $d_S$ ) are structurally silent, while nonsynonymous substitutions ( $d_N$ ) result in a change to the amino acid sequence. When a  $d_N/d_S$  ratio is calculated,

С		
environmental 99-581	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARSNAAEGLKGALAGAA	GIVYGKS
clinical ATCC 27562	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARSAAAEGLKGA <mark>I</mark> AGAA	GIVYGKS
CP Mussel 10 PT	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARSAAAEGLKGA <mark>I</mark> AGAA	GIVYGKS
95-10-15 PT	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARSAAAEGLKGATAGAA	GIVYGKS
environmental 99-584	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLOCDARASALOGLKGANDGAA	GIVYGKS
clinical CMCP6	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARSSALOGLKGAN GAA	GIVYGR
environmental 99-520	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARASL GLKGAMEGAA	GIVYGR
clinical ¥J016	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARAASLOGLKGAMBGAA	GIVYGRA
environmental 98-783	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARAASLOGLKGAMBGAA	GIVYGRA
environmental 51-13	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDAR <mark>AASLO</mark> GLKGAM <mark>B</mark> GAA	GIVYGRA
960926 1/4c PT	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARESALOGLAGAMOGAA	SIVYGES
clinical ATL-9580	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNATLEGLRGALOGGA	GIVYGKS
environmental 99-736	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARNATLEGLAGATOGAA	GIVYGRA
OLOL-1	MKROGGFTLIELVVVIVILGILAVTAAPRFLNLODDARVATLEGLRGATAGAS	GIVYGK
environmental_99-581	AINGTENAQTSSVAVDGTVIATARCOPAQTSAALSAVVSGL	ATDWTLV
clinical_ATCC_27562	AINGTENAQTSSVTVDGTVIATTFGYPAQTSAALSAVVSGL	ATDWTLV
CP_Mussel_10_PT	AINGTENAQSSSVTVNGTPIATTFGYPAQTSAALSAVVSGL	ATDWTLV
95-10-15_PT	AINGTENAQSSSVAVNGTPLATTCOTPAQTSAALSAVVSGL	ATDWTLV
environmental_99-584	ATEGEETVSKCAATAPAAEGVTEHFGYPTAETTGIQVAVNGL	PEDNDTA
clinical_CMCP6	ATEGLEAISSCRSISGANGNNINLVFGYPEASANGIGRAVAGI	DEDWSVI
environmental_99-520	ALDCKEGSATCQAVSVGSDTVKLTWGYCEASTDGIGKAVSGL	STDWAVV
clinical_YJ016	AIDGKEGSATCQAVSVGSDTVKLTWGYPEASTDGIGKAVSGL	STDWAVV
environmental_98-783	AIDGKEGNPTCQAVSVGSDTVKLTWGYPEASTDGIGKAVSGL	STDWAVV
environmental_S1-13	AIDGKEGDATCQAVSVGSDTVKLEWGYPEASTDGIGNAVSGL	STDWAVV
960926_1/4c_PT	ALECEEAAVSCAVTVGNATINTAFGYTAAVDGIGNSLDIDCF/	A-ALTNT
clinical_ATL-9580	ATEGEBAVSSATVTVASGNVDTAFGYPKATSTALSNTIQGLG	-DPFVFI
environmental_99-736	ATEGIETTSGAWEASCOTINTERCENATEADISNIVQGICA	TEPFRFI
OLOL-1	AVAGLEATERDATATPPVAPVVENGN GN DLET GYPT STTGVMLAIDVSN	EFAVI
environmental 99-581	SGATNGIGYGTASNADCHUTOAPASSAGEPT	LVNCN
clinical ATCC 27562	SGAUNGIGIGTASNADCHUNYOAPASSAGEPT	LVNCN
CP Mussel 10 PT	SGATNGIGYGTASNADCHUTYOAPASSAGEPTIS	LVNCR
95-10-15 PT	SGATNGIGYGTASNADCHUTYQAPASSAGEPTIS	LVNCR
environmental 99-584	VSGSCAAGAQYLAT SSDPALTAAQIVASNCFVSTTEA DANN PVVTVT	STGC
clinical CMCP6	SSCTCTITYG TENN TACFUTAINATAATAQEP ALT	VTECN
environmental 99-520	AS-TETTSISYSPOSGVTGATAGKCSV/VALNTSPSNAOPTIT	VKACE
clinical ¥J016	AS-TGTTSISYSFOSGVTGATADKCSV/HALNTSPSNAOPTIT	VKACR
environmental 98-783	AS-TGTTSISYSFKSGVTGATADKCSVTYVLNTSPSNAOPTIT	VKDCN
environmental S1-13	AS-TGTTSISYSFKSGVTGATANKCSVTVALNTSPSNAOP	VNDCN
960926 1/4c PT	TOATDSIDYGPENYTTNCVRUTOATSOAOATARVIAGIS	GTACGAO
clinical ATL-9580	KAISCTPNSATICIPNYTTNCVONTEA DANTPAVARVVSAG	SNSICSN
environmental 99-736	QVVSCASPAAAAFCIPNYTTQCVTTQAQDANTPATVTVGTA-	-NCVAP-
OLOL-1	DSANDDNTDNTNAWIDIGITGYTNNCURUTAATNATTPATAQVV8G-	-AGFCGQ
		_

Figure 1 (continued)

typically the value suggests whether the substitutions are largely neutral ( $d_N/d_S=1$ ), under a negative selection ( $d_N/d_S<1$ ), or a positive selection  $(d_N/d_S \ge 1)$  [44]. Table 3 shows the calculations for  $d_{\rm S}$ ,  $d_{\rm N}$ , and  $d_{\rm N}/d_{\rm S}$  for the *mshA* and *pilA* genes from the different Vibrio strains analyzed, and the data suggest that a selective pressure has been applied to these two genes for all three Vibrio species. To further analyze the selective pressure applied to the type IV pili examined, we compared *mshA* and pilA with another gene in their corresponding operon, the neighboring genes mshC gene and pilB, respectively, to determine if a selective pressure has been applied strictly to the gene encoding the pilin subunit or to the entire operon. When comparing the  $d_{\rm N}/d_{\rm S}$  value for *mshA* with *mshC* and *pilA* with *pilB* for all three vibrios, the  $d_N/d_S$  values for the pilin subunits (*mshA* and *pilA*) are considerably larger than the neighboring gene in the operon (mshC and pilB) (Table 3). These results suggest that the neighboring genes (mshC and pilB) in both the MSHA and PilA

operons are more conserved than their corresponding pilin subunits (*mshA* and *pilA*). Thus, it is possible that the pilin subunits are not under the same selective pressure as their neighboring genes.

Both *mshA* and *pilA* encoded pilins are processed by the same type IV prepilin peptidase, *pilD* [27, 28, 45]. When examining the  $d_N/d_S$  value for *pilD*, it was evident that the *pilD* gene maintained a highly conserved sequence. We calculated the *pilD*  $d_N/d_S$  for *V. cholerae* (0.039) but were unable to calculate it for *V. parahaemolyticus* because the sequences were identical for  $d_S$  (0.000) and  $d_N$  (0.000) so the  $d_N/d_S$  for *V. parahaemolyticus*, the results for *V. cholerae pilD* (0.039) were congruent with what was found for *V. vulnificus* (0.076) by Chattopadhyay et al. [46]. This suggests that a strong purifying selection has maintained the highly conserved *pilD* sequence in contrast to the general observation

O141_clinical_V51	IKAYKNKQQKGF7JJIBJMHVVAICQVJAVKDVVIC
014 clinical MZO-2	MKAYKNKQQQGFTLTELMIVVAIIGVLAAVAIPAYKDYVKKSEAA
037 clinical V52	MEAVENETOOGPTICTEIMIVANTGULAATAUPOYONYVEKSEA
012 clinical 1597	MEAVENEDO OPPLITELY VUATIOUTAAUAT PAVEVOU PATERA
012_CIINICAL_ISO/	ANALAS CONTRACTOR AND ANALAS
0135_environmental_RC385	GVIAATAVPOYOKY VAKSEAA
O39_clinical_AM-19226	GVLAAIAVPQYQKYVAKSEAA
non-01/0139 environmental 623-	MKAYKNKOOGFTLIELMIVVAIIGVLAAIAVPOYOKYVAKSEAA
environmental CT 5369-93	MNAYKNKOOGETTLT RIMTYVAVIGULAATAUPOYOKYVAKSRAA
non 01/0120 environmental WIA2	NT BAT WEATHINGTON AND AND AND AND AND AND AND AND AND AN
non-oi/oisy_environmental_vL42	ALTALATOTTTEGT ATTAINDOOD TTTEGT ATTAIN OF DAALANT OT ATTAIN
O1_Clinical_INDRE_91/1	MKAYKNKQQKGFTLIELMIVVAVIGVLAAIAIPQYQNYVKKSAIG
01_RC27	MKAYKNKQQKGFTLIELMIVVAVIGVLAAIAIPQYQNYVKKSAIGV
O1 clinical 0395	MKAYKNKOOKGFTLIELMIVVAVIGVLAAIAIPOYONYVKKSAIGV
Ol clinical MAK757	MEAYENKOOKGETTITIMIVVAVIGVLAATAIPOVONYVKSAIGV
Ol clinical CIPS 101	WEAVENUE OF OPPOTTEL VIEWAUTOW AATA TO OVON WERE TO
01_environmental_2/40-80	
Ol_clinical_NCTC_8457	MKAYKNKQQKGPTLIELMIVVAVIGVLAAIAIPQYQNYVKKSAIGV
0139 clinical MO10	MLTALNTOTPTEEIMKAYKNKOOKGFTLIELMIVVAVIGVLAAIAIPOYONYVKKSAIGV
Ol clinical MJ-1236	MLTALNTOTPTER IMEANENEORGETTITEIMIVVAVIGVIAAIAIPOVONVVESAIGV
01 clinical W66-2	MI TAL NTOT DEPET INVAVIATION OF OPPT TELEVIANT OUT AATA TROVONY UPPERATOR
OI_CIINICAL_NI6961	METAENTOTPTBEIMKAYKNKOOKGPTEIEEMIVVAVIGVEAAIAIPDYONYVKKSAIG
Ol_environmental_BX_330286	MLTALNTQTPTEEIMKAYKNKQQKGPTLIELMIVVAVIGVLAAIAIPQYQNYVKKSAIGV
O1 clinical B33	MLTALNTOTPTEEIMKAYKNKOOKGFTLIELMIVVAVIGVLAAIAIPOYONYVKKSAIGV
Ol clinical RC9	MLTALNTOTPTERIMEAXENKOOKGETTIRIMIVVAVIGVLAAIAIPOVONYVESAIGV
Ol onwironmontal TW 11070 90	NI TAL NEOTRE CITY AND YOU COOPERT TO LATE OT A APRILE AND ANO DE THE AND A STATE OF
OI_environmental_IM_II0/9-80	ALTALATOTFTEGTAATAAAQOOOFTOTEDAATVAATOTDAATAVFATODITAAATAS
non-01/0139_environmental_TMA_	MLTALNTQTPTEGI <u>MAAVKNKOQOGYTJIEJMIVVAL GDJAA</u> FAVPAYONYTK <mark>K</mark> AHASE
0141 clinical V51	FITHKSVI PARLEFOENCKIS SASAVDVLGVSSGANALGSLT
014 clinical MZO-2	ATATERATIC PARTE VORK - ITAAANI STDIGSI S GANNIGI TTS
027 aliniant WE2	
US7_CIINICAL_VS2	AVATLESS FILL STRADTON FIDETALCANSNANREGALTEGANEVELV
012_clinical_1587	AATTVRG LTNIDM QQEVGSFPTDITKVCGTTTMNAFCSIALASTV
O135_environmental_RC385	ALASITCHRINVETIVLENCEPTTAQLATPSASIGTIQYGTAQ
039 clinical AM-19226	ALASITCHRONVET VLENGSPPTSSAVPVPS
non-01/0139 environmental 623-	AT A ST WARD THUR TO UVEN SERVING SCALETERS
non-oi/oiss_environmental_ois-	
environmental_CT_5369-93	AGASI TGARARY SSEVLENGTEPSASLAVPTSPAGAISIVSA
non-01/0139_environmental_VL42	ALAT C KTNABAY TVETGSFPTDTQQAQLGTPSSAMCSIAYANSGN
Ol_clinical_INDRE_91/1	GLANITALKTNIEDYIATEGSFPAT <mark></mark> TAGTAAGFT <mark></mark> RLGTVEDMGDGKIVIA
01 RC27	GLANITALKTNIEDYIATEGSFPAT-TAGTAAGFTRLGTVEDMGDGKIVIA
Ol clinical 0395	GT ANT WALL WANT PROT A WEGE PAGE TAGE A GET A COMPANY AND CHILDREN CONCERNING
Ol alinian WARTER	
OI_CIINICAI_MAK/S/	GLANITALKTNIEDIIATEGSFPAT <mark></mark> TAGTAAGFT <mark></mark>
01_clinical_CIRS_101	GLANITALKTNIEDYIATEGSFPAT <mark>TAGTAAGFT</mark> RLGTVEDMGDGKIVIA
Ol_environmental_2740-80	GLANITALKTNIEDYIATEGSFPAT <mark>T</mark> AGTAAGFT <mark></mark> RLGTVEDMGDGKIVIA
Ol clinical NCTC 8457	GLANITALKTNIRDYIATEGSEPAT - TAGTAAGET
0139 clinical MOIO	OT ANTEST VENTERVISEROSPERATE BACESSOR
OI_CIINICAL_MJ-1236	GLANITALKTNIEDYIATEGSFPAT <mark>TAGTAAGFT</mark>
Ol_clinical_M66-2	GLANITALKTNIEDYIATEGSFPAT <mark></mark> TAGTAAGFT <mark>RLGTVEDMGDGKIVI</mark>
O1 clinical N16961	GLANITALKTNIEDYIATEGSFPAT <mark></mark> TAGTAAGFT <mark>RLGTVEDMGDGKIVIA</mark>
Ol environmental BX 330286	GLANITALKTNIRDYIATEGSFPAT-TAGTAAGFTRLGTVRDMGDGKIVIA
Ol clinical B33	
OI_CIINICAI_RC9	GLANITALKTNIEDYIATEGSFPATTAGTAAGFT
Ol_environmental_TM_11079-80	FPKALA_A_LAV_VCAHENASDETSFISNCVSGSNGVPASFTLN-NIDIQVKTAANAGVS
non-01/0139 environmental TMA	IVSAAA F GICLLDGOADCTATKG-GVPGEOTFKKNDNDDFKITSSV OTTI
Oldi aliniani WEL	
0141_CIINICAL_V51	-IFSDNQIQFAINSGAAAGARFVISKDISGHSCIITQPAAVSAAAAAQCAQ
014_Clinical_MZO-Z	ELVSSKPTLKFEFGANSSMTSNDTLTFTREDTGWKCAKTTNVPAIDGCQ
O37_clinical_V52	SNTDS FAS DEVKMAKDAT GLWTCTVPTGVTLKGC A GSGSGSGSGSGSGSGSGSG
O12 clinical 1587	SGGTATFTFEDG-ALKTEGTGNTAAKEIYTKDNATGNTCTHTINDE-SVVESCC
0135 environmental RC385	ASGAGS TVF AFGT SGASP SVVSKN VTGARD GNGAWKCT TT DTNI APKGCAAP
039 clinical AM-19226	ASCARSTIFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
035_01/0120	
non-01/0139_environmental_623-	QSAAGTIEFKFKTSGVSPDVVGKTVUEENTSCNECKTIAADIKPKGCTVTP
environmental_CT_5369-93	ATSSCGIQFEFEQLGVSPDVVSKKVTLSRTAACAMSCTTDISGDLKPKGC
non-01/0139 environmental VL42	SGRACTITFSFTGT-VSPSLKSSAIRCKRDDSCNMSCESKTGLDSGVIPKGCNSGTF
Ol clinical INDRE 91/1	PTASGALGGTIKYTFDAGVVSSSKIOLARDANGLWTCSTTVTSEIAPKGCTAGATI
01 8027	
UI_CIINICAL_U395	PTASGALGGTIKITFDAGVVSSSKIQLARDANGLWTCSTTVTSEIAPKGCTAGATI
Ol_clinical_MAK757	PTASGALGGTIKYTFDAGVVSSSK-IQLARDANGLWTCS-TTVTSBIAPKGCTAGAT
01_clinical_CIRS_101	PTASGALGGTIKYTFDAGVVSSSKIQLARDANGLWTCSTTVTSEIAPKGCTAGATI
01 environmental 2740-80	PTASGALGGTIKYTFDAGVVSSSKIÖLARDANGLWTCSTTVTSEIAPKGCTAGATI
Ol clinical NCTC 8457	PERSON COT INTERDACIVESSE TO APPRINCE WESS TRUESET APPRCEMENT
0120 aliniant Main	
UISS_CIINICAL_MOID	PTASGALGGTT KTTT DAGVVSSSKPULARDANGLATUSTTVTSETAPKGCTAGAT
Ol_clinical_MJ-1236	PTASGALGGTIKITFDAGVVSSSK <mark></mark> IQLARDANGLWTCS <mark></mark> TTVTSBIAPKGCTAGATI
O1_clinical_M66-2	PTASGALGGTIKYTFDAGVVSSSKIQLARDANGLWTCSTTVTSBIAPKGCTAGATI
Ol clinical N16961	PTASGALGGTIKITFDAGVVSSSKIÖLARDANGLWTCSTTVTSEIAPKGCTAGATI
Ol environmental By 330286	PERSON COT INTERNATIVESSES TO APPRINCE WECK TRUTSET ABROCEMONE
Ol clinical B33	TROOKDOOTTKTTFDAOTTSSSATQDAKDARGERTCSTTFTSSTAPKGCTAGAT
UI_CIINICAL_B33	
	PTASGALGGTIKYTFDAGVVSSSK <mark></mark> IQLARDANGLWTCS <mark></mark> TTVTSBIAPK <u>GCTAGATI</u>
O1_Clinical_RC9	PTASGALGGTIKYTFDAGVVSSSK <mark>-I</mark> QLARDANGLWTCS <mark>-T</mark> TVTSEIAPKGCTAGATI PTASGALGGTIKYTFDAGVVSSSK <mark>-I</mark> QLARDANGLWTCS <mark>-</mark> TTVTSEIAPKGCTAGATI
Ol_clinical_RC9 Ol_environmental TM 11079-80	PTASGALGGTIKYTFDAGVVSSSKIQLARDANGLWTCSTTVTSEIAPKGCTAGATI PTASGALGGTIKYTFDAGVVSSSKIQLARDANGLWTCSTTVTSEIAPKGCTAGATI GAVAVEAKAKTAKGPIAKDESFIMAAAYKPEGLEWKSSCKDAAGKAOTSYCPD
Ol_clinical_RC9 Ol_environmental_TM_11079-80 non-Ol/O139 environmental TMA	PTASGALGGTIKTFPDAGVVSSSKIQLARDANGLWTCSTTVTSEIAPKGCTAGATI PTASGALGGTIKTFPDAGVVSSSKIQLARDANGLWTCSTTVTSEIAPKGCTAGATI GAVAVEAKAKTAKGPIAKDESFIMAAAYKPEGLEWKSSCKDAAGKAQTSYCPD GTVDADTAITATV-KKGALPANSTVVLTPKLTASCVTWAVTCTGDGKEDWCPVK

Figure 2 Amino acid sequence alignment of PilA from *Vibrio cholerae* (a), *Vibrio parahaemolyticus* (b), and *Vibrio vulnificus* (c). The predicted amino acid sequences of PilA for *V. cholerae* (a), *V. parahaemolyticus*  (b), and *V. vulnificus* (c) were aligned using the ClustalW program. *White* indicates normal residues. *Green* are similar residues. *Pink* are identical residues. *Black* indicates globally conserved residues

b

+dh+/+rh+ Barn_466	NERSYARDAGEST TELNT WAT TOUT A SALE AVAILY UNDER WORKS AND TRACK AN UNAT TODA ST
tdh+//tab clinical BIND 221062	
tdh+//trh-clinical_kimb_221005	
tdn+/trn+_AN-5034	MKESKOKKOOGFTLISLMIVVAIIGVLAAAAIPAYONIVTRSEVISGLATVKALITPASL
tdh+/trh+_K5030	MKHSKOKKOOGFTLIELMI VVAI IGVLAAAAIPAYONYVTRSEVTSGLATVKALI TPAEL
tdh+/trhclinical_BE98-2029	MKESKOKKOOGFTLIELMIVVAIIGVLAA <u>A</u> ATPAYONYVTESEVTSCLATVKALITPAEL
tdh-/trhenviron_357-99	MKHSKQKKQQGFTLIELMIVVAVIGVLAAIAIPQYQNYVAKSELGAGLATITSVRTNVED
tdh-/trhenviron_UCM-V586	MKHSKQKKQQGFTLIELMIVVAVIGVLAAIAIPQYQHYVAKSELGAGLATITSVRTNVED
tdh+/trh+ clinical 10292	MKQSKQKKQQGFTLIELMIVVAVIGVLAS <mark>I</mark> AIPQYQKYVAKSEVA <mark>SA</mark> LATLTGVKTNVEA
tdh+/trh+ environ SFL1050	MKESKQKKQQGFTLIELMIVVAIIGVLAAVAIPAYONYVQKTEVASASATVRGLLTNIDM
tdh-/trh- environ UCM-V441	MKESKOKKOOGFTLIELMI VVAIIGVLAAVAIPAYONYVOKTEVASASATVRGLUTNIDM
tdh+/trh+ environ SFL1079	MKHSKOKKOOGFTULELMI VVAIIGVLSAVAIPAYKSYVAKSEAATAAATVRGLLTNIDM
trh+ clinical ATCC 17802	MKHSKOKKOOGP TATEIMI WYCIIGIISALAWPAYKSYVIETEANTAYCYPBALLANYDI
tdh+/trh- environ SEL1080	NEWSKOKKOOGP TATELATUVCI TOTISE LAWPS VKSVVI TEANTS VCVPRATISAVUD
tdh+/trh+ clinical 10290	NY SKOW COOP TITELATING TOTTS TAUDA VY SVUT TEAM TO VOUD DIT AN UDI
tdh+/trh+ onviron EFL1009	
tdh+/trh+_environ_SFL1009	
tdh+/trn+_environ_SFL102/	MARSAOKAOOGFTI I SIMI VVAI I GI JAAFAVFAI ORITIMRAHASEMI SASSASKTAASI
tdn+/trn+_clinical_027-1C1	RKHSKOKKOOGFFULTELMI VVAI IGI LAAFAVPAYONYTMRAHASEMU SASSAFKTAASI
tdh-/trhenviron_M25-0B	NKESKOKKOOGFTIJELMIVVAIIGILAA PAVPAYONYTMRAHASEMISASAAMKTAVGI
tdh-/trh- environ 049-2A3	<u>NIXOSKOKKOOGFUUIDIMIGUANIGUAAAFAUPAYONY</u> TK <b>R</b> AHASEM <u>M</u> HASNAFK <u>B</u> AVGU
tdh+/trh+ Peru-466	HYQENG-IAAAA-TLAQZGVDVAANDLGAIDSALAGGSP-TLAFTEDADSS
tdh+//trh- clinical RIMD 221063	HYOENG-IAAAA-TLAO GVDVAANDLGATDSALAGGSP-TLTTTFDADSS
tdh+/trh+ AN-5034	HYDENG-IAABA-TLAOUGVDVBANDLGAIDSALACGSP-TLTTTPDADSS
tdh+/trh+ \$5030	HYDENG-TAABA-TLAOLGUDVAANDLGATDSALAGGSP-TLEFTEDADSS
tdh+/trh- clinical BE98-2029	HYORNG IAAAA TIAOI GUDUAANDI GATDSALAGGSP_TUTETTDADSS
tdh /trb onviron 357-00	TUTUT CEPDER TALOT TELL CULORY NO. TEPPANE - TITERACE
tdh /trh onviron UCV VERC	TUTIOFFDAT TAQUITDEVILLAN AND ALTOPANS ATTACAS
tdh-/trhenviron_ocm-v500	
tdh+/trh+_clinical_10292	FAVENCKFPDGS-STRETEADLGVPTTIPSG-SITFTAASSSAGTIAVEDDSGVSNL
tdh+/trh+_environ_SFL1050	IQUENGGT-FPNNANLVGGTSTMNAIGTITLAPVETSGENANNALVEGSLKG
tdh-/trhenviron_UCM-V441	IODENGGT-FPNNANLVGGTSTMNAIGTITLAPVETSGENA FAFVEGSLKG
tdh+/trh+_environ_SFL1079	IQDEVES-FPIDIKKVGGTSTMNAFELIALAST-TSGCAACTTBTDGSLKTGG
trh+_clinical_ATCC_17802	FVORKERYPNST-QTAD AAIGAAIDMSAM TL TPDAD SEYCDIE IGSNAS
tdh+/trhenviron_SFL1080	FVDEKCKYPNSN-QTADCAAIGAAIDMSAMCTLAITPDADCSEYCDIEFTIGSNAS
tdh+/trh+_clinical_10290	FV2EKCKYPDST-QTADCAAIGAAIDMSAMCTLAITPDADCPEYCDIECTIGSNAS
tdh+/trh+_environ_SFL1009	KLOODELSGVVADEGGSASHALG <u>NITT</u> SGANISAALLETTNT-GS
tdh+/trh+_environ_SFL1027	CLLDGEVNCTSGNGGVPTNQQFAK-NATDDFTVSSTIEQAALCGAGAVTATVAA-
tdh+/trh+_clinical_027-1C1	CLLDGEVNCTSGNGGVPTNQQFAK-NATDDFTVSSTIEQAALCGAGOVTATVAA-
tdh-/trh- environ M25-0B	CPMDGNADCLSATANNGVPTAQNLGDFTVASLAGNVVTATVNAN
tdh-/trh- environ 049-2A3	CLLSSETDCTDGTGGVPSTOSFDKGTGDDSFSTSSVLQTTLGTVDTDTFISATIDAT
+dh+/+rh+ Peru-466	
tdh+/itrh- clinical RIND 221063	
tdh+/trb+ av 5034	
tdh+/trn+_KR-JUJ4	
tdh+/trhclinical_BE98-2029	APAIAVOG S
tdn-/trnenviron_35/-99	LDAKLEPKSSKDANL-
tdh-/trhenviron_UCM-V586	LDAKLEPKSSKDANL-
tdh+/trh+_clinical_10292	VTSKNFELVASDGTMT QGSSASPVTDDLLPKNOR
tdh+/trh+_environ_SFL1050	VPTDSRPNSTATY
tdh-/trhenviron_UCM-V441	VPTDSRPNSTATY
tdh+/trh+_environ_SFL1079	AGNNAATIIIYTKDNST CIISCTHAIQDTSVVPSG
trh+_clinical_ATCC_17802	QDLKGATTARSTN-CIKCTHDTGQDLKGATTPATP
tdh+/trhenviron_SFL1080	QDLKGATTPARSTN-GIKGTHDTGQDLKGATTPATP
tdh+/trh+ clinical 10290	ODLKGATTPATP
tdh+/trh+ environ SFL1009	TVLONON
tdh+/trh+ environ SFL1027	KGALAPNSTVVLTPTITG GVTMAVTC GLGNA W-PER
tdh+/trh+ clinical 027-1C1	KGALAPNSTVVLTPTITG GVT AVTC GLGNA WAPER
tdh-/trh- environ M25-0B	-GKGSLSATDTIEITPTLGA GVTUAVACUSTNANVDSGUWAPTP
+dh_/trh_ environ 040_213	
cun-/crnenviron_049-2A5	Dergentret FALTERDag VV DVICADIGGARA F GFIN

Figure 2 (continued)

for the *mshA* and *pilA* sequences. When examining the predicted amino acid sequences for both *mshA* (Fig. 1) and *pilA* (Fig. 2) for all three vibrios, it was clear that the N-termini remain highly conserved while the C-termini varied considerably. The N-termini region is recognized by the PilD peptidase for processing the protein into a mature pilin subunit [4]. If the N-terminal region of the type IVa pili proteins MSHA and PilA varied, it is possible that PilD would no longer process these proteins into mature subunits, while variations in the Ctermini should still result in a mature pilin subunit. Thus, it appears that PilD has maintained a highly conserved sequence unlike the MSHA and PilA proteins it processes.

To further understand the variations observed in the MSHA and PilA pilins, the *V. cholerae mshA* and *pilA* sequences were compared to the type IVb pilin TCP from *V. cholerae*. The *tcpA* gene encodes the major pilin subunit of TCP and is processed by its own type IV pili peptidase TcpJ, encoded by *tcpJ* [23]. Contrary to *tcpA* that exhibit some variability in its sequences with mostly synonymous substitutions ( $d_S$  of 0.486) and few nonsynonymous

519

С

environmental_99-581	MMKKLEKTKKQQGFTLIELHIVVAIIGILSAVAVPATKNTVAKSEAATALGSIRALVTPA
clinical_¥J016	MMKKL#KTKKQQGFTLIELMIVVGIIGILSALAVPAYKSYVIKTBANTAVGIPRTILSXI
environmental_98-783	MMKKLEKTKKQQGFTLIELMIVVAIIGVLS <mark>AV</mark> AIPATKSYVAKSEAA <mark>T</mark> AAAT <mark>VRGLLTEI</mark>
environmental_99-738	MMKKLEKTKKQQGFTLIELMIVVAIIGVLS <mark>AV</mark> AIPATKSYVAKSEAA <mark>T</mark> AAAT <mark>VRGLLTEI</mark>
clinical_CMCP6	MMKKLÆKTKKQQGFTLIELMIVVAIIGVLÄAVAIPATONYVQK <b>T</b> EVA <b>S</b> ASAT <mark>VRGLLTHI</mark>
environmental_99-584	MMKKLEKTKKQQGFTLIELMIVVAIIGVLAAVAVPATODYVKKSEAASALAT <mark>LKS</mark> L <mark>I</mark> TPA
environmental_S1-13	MMKKL#KTKKQQGFTLIELMIVVAIIGVLA <mark>AVAV</mark> PATODYVKKSEAA <mark>S</mark> ALAT <mark>LKSLI</mark> TPA
environmental_99-736	MMKKLNKTKKQQGFTLIELMIVVAIIGVLAA <mark>I</mark> AIPATONYVKKSEAAI <mark>G</mark> LAT <mark>AKS</mark> L <mark>I</mark> TNV
environmental_99-520	MMKKLNKTKKQQGFTLIELMIVVAIIGVLAA <mark>I</mark> AIPAYQNYVKKSEAAI <mark>G</mark> LAT <mark>AKS</mark> L <b>I</b> TNV
clinical_ATCC_27562	MMKKLEKTKKQQGFTLIELMIVVAIIGVLAA <mark>I</mark> AIPATONYVKKSEAAI <mark>G</mark> LAT <mark>AKS</mark> L <mark>I</mark> THV
clinical_ATL-9580	MMKKL#KTKKQQGFTLIELMIVVAIIGVLAA <mark>I</mark> AIPAYONYVKKSEAAI <mark>G</mark> LAT <mark>AKS</mark> L <mark>I</mark> THV
960926-174C_PT	MMKKLHKTKKQQGFTLIELMIVVA <mark>V</mark> IGVLAA <mark>IAV</mark> POXOKYVAKSEAATALA <mark>SI</mark> TGHRTNV
95_10_15_PT	MMKKLHKTKKQQGFTLIELMIVVAVIGVLAAIAVPOYOKYVAKSEAATALASITGHRTNV
CP Mussel 10 PT	MMKKL <mark>H</mark> KTKKQQGFTLIELMIVVA <mark>V</mark> IGVLAA <mark>I</mark> AIPQYQKYVAKABVASALATLIGLKTNV
OLOL-1	MMKKLNKTKKQQGFTLIELMIVVAIIG <mark>I</mark> LAAFAIPAYODYTKRATMAEFPRVAS <mark>S</mark> TKLAV
environmental 99-581	EMICEN TISG GUSAL GSASH-AC
clinical VJ016	DEVI USKOAFPETANIADVOGATOMSALOGUALT-KOGTTATNOVLI PTI SNGOA-
environmental 98-783	DAYOUSVESEPTDITKYCGTOTABAFCSIALA-ST-TSGGTATETPTOGSLKTG
environmental 99-738	DAYOUSVG-SEPTDITKYGGTOTMEAFCSTALA-ST-TSGGTAUFTUTDGSLKTG
clinical CMCP6	DAVOUENCG-TEPNNANLVCGTSTMNALCTITLL-PVGTSGGTATEAFTEGTLKG-
environmental 99-584	FLYYOENGTATAAGLEDLOST SANDLOST LASALTGSG SAVPTITEST GSHSS-
environmental S1-13	FLYYON CTATAAGLSDLCSTTSANDLCTLASAITGSCSALVPTLTCSCSHSS-
environmental 99-736	DAVIOSKONFPINDST GFTAVGATTIME KLOTIS FESISGANGTIKFLENNKSS-
environmental 99-520	DAVIOSKONFPINDST GFTAVGATTIME KLOTIS FESISGANGTIKFLENNKSS-
clinical ATCC 27562	DAVIORKONFPINDST GFTAVGATTINEKLOTIS FESTISCINGTIKELENNKSS-
clinical ATL-9580	DAVIOSKONFPINDAT GFTAVGATTIME KLOTIS FESISGAN GTIKFLENNKSS-
960926-174C PT	ET VULIN SFPSTSGALAIPIS-PTCVITYENPTS A DIK SINSSGVS
95 10 15 PT	ET VULEN SEPTANSLAUPUSPIGUSYVSAASSS GTOPLEKSTGVS
CP Mussel 10 PT	FAVAUEN CAFPDGTTSGOSVTDL CAPUTT - PSCSVT FTAGANGACSTTFFFGATGVS
01.011	FLCAN HAADGAGEKSKCYLAAN GIPAFFTINNIKUTPTTGT SCSUDITUAATADK
environmental 99-581	SLVNKKTALT DAST KTYDAGVPLDKAN
clinical ¥J016	SLNNKKITYTR GG-GWKCTHDIPTTL-TEELSSKATTA000TPSS
environmental 98-783	GAGNNAATIIYT DIST GWSCTHTIODT SVV SGC
environmental 99-738	GAGNNAATIIYT DNSTGWSCTHTIODTSVVPSG
clinical CMCP6	KTASVOYSKNNTT WISCATKNVPADSRPNSTATY
environmental 99-584	MTTTDKLTFSRSTTEGNSCARSGTVPAVDG-0
environmental \$1-13	MT TTKLTFS85 TEGUSCARSGTVPAVDGO
environmental 99-736	IDT-AFVSIARTN-SGMSCSFTTSNKLTNEEIPEAGOP
environmental 99-520	IDT-AFVSTAKTN-SGUSCSFTTSNKLTNEEIPEAGOP
clinical ATCC 27562	IDU-AFVSIANTN-SGUSCSFTTSNKLTNEEI PRAGOP
clinical ATL-9580	IDU-AFVSIAKTN-SGUSCSFTTSNKLTNEEIZKAGOP
960926-174C PT	P-DIVSKHVTLSR SAGUISCATDVATDLK CG
95 10 15 PT	P-DVVSKKVTLSR TAGA TSTDVDSDLK G
CP Mussel 10 PT	N-LINSKKFLLARDSSGT TODGTAATPVTDDLLPNNK
OLOL-1	GSIK GEKYILTA YNTTGITWEAVCKDAGGTVOTAY P

Figure 2 (continued)

substitutions ( $d_N$  of 0.052), tcpJ has relatively few substations overall ( $d_{\rm S}$  of 0.003 and  $d_{\rm N}$  of 0.000). The  $d_{\rm N}/d_{\rm S}$  for *tcpA* is 0.106 and that for *tcpJ* is 0.000, suggesting that these genes are under strong negative selection to maintain their sequences and structures. When examining the V. cholerae phylogenetic trees constructed for the *mshA* and *pilA* genes, the strains that possess TCP are all from the O1 serogroup and on a single branch (Figs. 3a and 4a). Looking at the amino acid alignment data, it was evident that the V. cholerae isolates containing all three type IV pili were highly conserved (Figs. 1a and 2a). To break it down further, the  $d_{\rm N}/d_{\rm S}$  ratio for mshA and pilA from the V. cholerae strains possessing TCP were also calculated, and the  $d_{\rm S}$  and  $d_{\rm N}$  for both genes were 0.000, resulting in an undefined  $d_{\rm N}/d_{\rm S}$  ratio (Table 3). Therefore, V. cholerae strains possessing all three type IV pili appear to be under a strong purifying selection. Even though some O1 V. cholerae isolates in this conserved branch were from environmental or unknown sources (3 out of 13), the fact that they possess TCP implies they could cause cholera. Taken together, the evidence suggests a connection between host interactions and highly conserved type IV pili in *V. cholerae*.

A previous study by Chattopadhyay et al. [46] analyzed *pilA* from 55 *V. vulnificus* strains of various origins and also determined that *pilA* is highly divergent. A total of 25 unique alleles were identified from the 55 analyzed strains, and the authors did not determine any relationship between the various alleles and pathogenicity of *V. vulnificus* [46]. They concluded that the genetic diversity of *pilA* in *V. vulnificus* was higher than neighboring genes (*pilBCD*) and thus was under strong positive, diversifying selection [46]. This conclusion was made despite the fact that the  $d_N/d_S$  ratio calculated for *pilA* was <1. The usefulness of the  $d_N/d_S$  ratio to detect positive selection is reduced when comparing gene polymorphisms within a single





Figure 3 Bootstrap maximum likelihood phylogenetic trees for mshA from Vibrio cholerae (a), Vibrio parahaemolyticus (b), and Vibrio vulnificus (c). The bootstrap maximum likelihood phylogenetic trees for mshA from V. cholerae (a), V. parahaemolyticus (b), and V.

vulnificus (c) were constructed using the gene sequences for mshA in the Molecular Evolutionary Genetics Analysis (MEGA) 5 software. All bootstrap values are listed

population compared to divergent populations [47]. Our results are consistent with their findings and also demonstrate that MSHA and PilA from V. cholerae, V. parhaemolyticus, and V. vulnificus exhibit higher genetic diversity than other genes in their corresponding operon (mshC and pilB and pilD).

Chattopadhyay et al. [46] suggested various ideas to explain their observation, including that the allelic variability in PilA for V. vulnificus could be the result of oyster innate immune system [46]. It was noted that since V. vulnificus commonly associate with shellfish in the environment and infections in humans are typically opportunistic, the selective pressure applied to this gene was probably not in response to an adaptive immune system [46]. Shellfish have an innate immune system that recognizes highly conserved motifs while lacking a well-developed adaptive immunity [48, 49]. Thus, the driving force behind the variations observed in the PilA protein could be the result of the innate immunity of shellfish, such as oysters, in part based on a previous study showing that PilA was involved in oyster colonization by V. vulnificus [11, 46]. Data from our laboratory also indicated that PilA and MSHA play a role in V. parahaemolyticus colonization of the Pacific oyster, C. gigas (Aagesen, A.M., and C.C. Häse, unpublished results), further supporting the idea that the shellfish immune system might be involved in applying pressure to

tdh+/trh+ Peru-466





**Figure 4** Bootstrap maximum likelihood phylogenetic trees for *pilA* from *Vibrio cholerae* (**a**), *Vibrio parahaemolyticus* (**b**), and *Vibrio vulnificus* (**c**). The bootstrap maximum likelihood trees for *pilA* from *V. cholerae* (**a**), *V. parahaemolyticus* (**b**), and *V. vulnificus* (**c**) were

constructed using the gene sequences for pilA in the Molecular Evolutionary Genetics Analysis (MEGA) 5 software. All bootstrap values are listed

these pili proteins, thus causing variability. Studies using different strains expressing the various alleles for MSHA and PilA from *V. cholerae*, *V. parahaemolyticus*, and *V. vulni-ficus* in shellfish interaction experiments are required to fully address this issue.

In addition to the shellfish immune system, other selective pressures in the environment could exist to cause the observed allelic diversity in MSHA and PilA, such as protozoan grazing, bacteriophages and DNA uptake [46]. Ideally, various alleles for MSHA and PilA from *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus* would need to be examined to better understand the role of bacteriophages as a selective pressure causing the variations observed for these proteins. However, future studies using various alleles for MSHA and PilA are required to support these hypotheses.

In summary, this study illustrates significant diversity of the MSHA and PilA pilin subunits from *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus*. For all three vibrios examined in this study, mshA and pilA had considerably higher  $d_N/d_S$ ratios than any of the other genes examined, suggesting these genes are under a possible positive selection while the other genes examined are not. Another interesting finding was that V. cholerae strains that possess TCP also maintain highly conserved MSHA and PilA sequences, suggesting a connection with the host. Even though a selective pressure appears to exist causing the allelic variations observed for mshA and pilA, the mechanism(s) driving this diversification have yet to be determined. Several suggestions can be made, yet evidence to support these ideas awaits further experimental analyses. In addition, our observations raise an important point about the use of these genes in detection methods for these important human pathogens. In particular, some PCR-based detection methods utilize certain pathogen-associated genes as targets, including type IV pili genes [50, 51]. Realizing that the Vibrio mshA and pilA genes can be extremely variable at the 3' ends of the genes is important to consider when designing primers Table 3Analysis of synonymousmous and nonsynonymousnucleotide substitutions forgenes involved in type IV Pilifunction from V. cholerae, V.parahaemolyticus, and V.vulnificus

Gene locus	Organism	Sequence length (bp)	Number of strains	$d_{\rm S}$	$d_{\rm N}$	$d_{\rm N}/d_{\rm S}$
mshA	V. cholerae	438–537	25	0.759	0.471	0.621
	V. parahaemolyticus	456-504	19	0.746	0.431	0.577
	V. vulnificus	447-510	15	0.662	0.384	0.580
mshC	V. cholerae	489–513	25	0.135	0.039	0.290
	V. parahaemolyticus	131 (1-131 5'end)	15	0.229	0.017	0.072
	V. vulnificus	94 (1-94 5'end)	12	0.042	0.015	0.356
pilA	V. cholerae	420-504	25	1.109	0.629	0.567
	V. parahaemolyticus	405–486	19	1.691	0.642	0.380
	V. vulnificus	402–453	15	1.186	0.503	0.424
pilB	V. cholerae	1,689	24	0.176	0.008	0.047
	V. parahaemolyticus	248 (1-248 5'end)	19	0.288	0.037	0.127
	V. vulnificus	122 (1-122 5' end)	14	0.208	0.016	0.074
pilD	V. cholerae	876	24	0.122	0.005	0.039
	V. parahaemolyticus	870	3	0.000	0.000	_
tcpA	V. cholerae	675	13	0.486	0.052	0.106
tcpJ	V. cholerae	762	13	0.003	0.000	0.000
mshA	V. cholerae with TCP		13	0.000	0.000	_
	V. cholerae without TCP		12	0.663	0.377	0.568
pilA	V. cholerae with TCP		13	0.000	0.000	-
	<i>V. cholerae</i> without TCP		12	1.270	0.615	0.484

to target these genes. Therefore, it is possible that a PCR protocol designed to amplify *mshA* and *pilA* from various *V*. *cholerae*, *V*. *parahaemolyticus*, and *V*. *vulnificus* strains may not detect these genes simply due to the variations observed in this study. This is certainly something to consider when utilizing these genes in a PCR protocol.

Acknowledgments We thank Dr. Yi-Cheng Su and Dr. Narjol-Gonzalez-Escalona for providing the *V. parahaemolyticus* strains and Dr. Paul Gulig and Dr. Kathy O'Reilly for providing the *V. vulnificus* strains. This study was funded by grants from the USDA/CREES NRI and the OSU General Research Fund.

## References

- Kline KA, Falker S, Dahlberg S, Normark S, Henriques-Normark B (2009) Bacterial adhesins in host–microbe interactions. Cell Host Microbe 5:580–592
- Proft T, Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria—structure, assembly and their role in disease. Cell Mol Life Sci 66:613–635
- Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol 16:33–40
- Craig L, Pique ME, Tainer JA (2004) Type IV pilus structure and bacterial pathogenicity. Nat Rev Microbiol 2:363–378
- Strom MS, Lory S (1993) Structure–function and biogenesis of the type IV pili. Annu Rev Microbiol 47:565–596
- Shime-Hattori A, Iida T, Arita M, Park KS, Kodama T, Honda T (2006) Two type IV pili of Vibrio parahaemolyticus play

different roles in biofilm formation. FEMS Microbiol Lett 264:89-97

- Enos-Berlage JL, Guvener ZT, Keenan CE, McCarter LL (2005) Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. Mol Microbiol 55:1160–1182
- Watnick PI, Fullner KJ, Kolter R (1999) A role for the mannosesensitive hemagglutinin in biofilm formation by *Vibrio cholerae* El Tor. J Bacteriol 181:3606–3609
- Barken KB, Pamp SJ, Yang L, Gjermansen M, Bertrand JJ, Klausen M, Givskov M, Whitchurch CB, Engel JN, Tolker-Nielsen T (2008) Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in Pseudomonas aeruginosa biofilms. Environ Microbiol 10:2331–2343
- Paranjpye RN, Strom MS (2005) A Vibrio vulnificus type IV pilin contributes to biofilm formation, adherence to epithelial cells, and virulence. Infect Immun 73:1411–1422
- Paranjpye RN, Johnson AB, Baxter AE, Strom MS (2007) Role of type IV pilins in persistence of *Vibrio vulnificus* in *Crassostrea* virginica oysters. Appl Environ Microbiol 73:5041–5044
- 12. Hang L, John M, Asaduzzaman M, Bridges EA, Vanderspurt C, Kirn TJ, Taylor RK, Hillman JD, Progulske-Fox A, Handfield M, Ryan ET, Calderwood SB (2003) Use of in vivo-induced antigen technology (IVIAT) to identify genes uniquely expressed during human infection with *Vibrio cholerae*. Proc Natl Acad Sci U S A 100:8508–8513
- Forslund AL, Salomonsson EN, Golovliov I, Kuoppa K, Michell S, Titball R, Oyston P, Noppa L, Sjostedt A, Forsberg A (2010) The type IV pilin, PilA, is required for full virulence of *Francisella tularensis* subspecies *tularensis*. BMC Microbiol 10:227
- Wu HY, Zhang XL, Pan Q, Wu J (2005) Functional selection of a type IV pili-binding peptide that specifically inhibits *Salmonella* Typhi adhesion to/invasion of human monocytic cells. Peptides 26:2057–2063

- Zhang XL, Tsui IS, Yip CM, Fung AW, Wong DK, Dai X, Yang Y, Hackett J, Morris C (2000) Salmonella enterica serovar Typhi uses type IVB pili to enter human intestinal epithelial cells. Infect Immun 68:3067–3073
- Mahmoud KK, Koval SF (2010) Characterization of type IV pili in the life cycle of the predator bacterium *Bdellovibrio*. Microbiology 156:1040–1051
- Stabb EV, Ruby EG (2003) Contribution of pilA to competitive colonization of the squid Euprymna scolopes by *Vibrio fischeri*. Appl Environ Microbiol 69:820–826
- Pelicic V (2008) Type IV pili: e pluribus unum? Mol Microbiol 68:827–837
- Attridge SR, Manning PA, Holmgren J, Jonson G (1996) Relative significance of mannose-sensitive hemagglutinin and toxin-coregulated pili in colonization of infant mice by *Vibrio cholerae* El Tor. Infect Immun 64:3369–3373
- Taylor RK, Miller VL, Furlong DB, Mekalanos JJ (1987) Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc Natl Acad Sci U S A 84:2833–2837
- Tacket CO, Taylor RK, Losonsky G, Lim Y, Nataro JP, Kaper JB, Levine MM (1998) Investigation of the roles of toxin-coregulated pili and mannose-sensitive hemagglutinin pili in the pathogenesis of *Vibrio cholerae* O139 infection. Infect Immun 66:692–695
- 22. Manning PA (1997) The tcp gene cluster of *Vibrio cholerae*. Gene 192:63–70
- 23. Kaufman MR, Seyer JM, Taylor RK (1991) Processing of TCP pilin by TcpJ typifies a common step intrinsic to a newly recognized pathway of extracellular protein secretion by gram-negative bacteria. Genes Dev 5:1834–1846
- 24. Faruque SM, Kamruzzaman M, Meraj IM, Chowdhury N, Nair GB, Sack RB, Colwell RR, Sack DA (2003) Pathogenic potential of environmental *Vibrio cholerae* strains carrying genetic variants of the toxin-coregulated pilus pathogenicity island. Infect Immun 71:1020–1025
- Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910–1914
- Marsh JW, Taylor RK (1999) Genetic and transcriptional analyses of the *Vibrio cholerae* mannose-sensitive hemagglutinin type 4 pilus gene locus. J Bacteriol 181:1110–1117
- Marsh JW, Taylor RK (1998) Identification of the *Vibrio cholerae* type 4 prepilin peptidase required for cholera toxin secretion and pilus formation. Mol Microbiol 29:1481–1492
- Fullner KJ, Mekalanos JJ (1999) Genetic characterization of a new type IV-A pilus gene cluster found in both classical and El Tor biotypes of *Vibrio cholerae*. Infect Immun 67:1393–1404
- Jonson G, Sanchez J, Svennerholm AM (1989) Expression and detection of different biotype-associated cell-bound haemagglutinins of *Vibrio cholerae* O1. J Gen Microbiol 135:111–120
- Hanne LF, Finkelstein RA (1982) Characterization and distribution of the hemagglutinins produced by *Vibrio cholerae*. Infect Immun 36:209–214
- Campos J, Martinez E, Izquierdo Y, Fando R (2010) VEJ{phi}, a novel filamentous phage of Vibrio cholerae able to transduce the cholera toxin genes. Microbiology 156:108–115
- 32. Jouravleva EA, McDonald GA, Marsh JW, Taylor RK, Boesman-Finkelstein M, Finkelstein RA (1998) The Vibrio cholerae mannosesensitive hemagglutinin is the receptor for a filamentous bacteriophage from V. cholerae O139. Infect Immun 66:2535–2539
- Faruque SM, Bin Naser I, Fujihara K, Diraphat P, Chowdhury N, Kamruzzaman M, Qadri F, Yamasaki S, Ghosh AN, Mekalanos JJ

(2005) Genomic sequence and receptor for the *Vibrio cholerae* phage KSF-1phi: evolutionary divergence among filamentous vibriophages mediating lateral gene transfer. J Bacteriol 187:4095–4103

- Thelin KH, Taylor RK (1996) Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. Infect Immun 64:2853–2856
- Hsiao A, Liu Z, Joelsson A, Zhu J (2006) Vibrio cholerae virulence regulator-coordinated evasion of host immunity. Proc Natl Acad Sci U S A 103:14542–14547
- Hsiao A, Toscano K, Zhu J (2008) Post-transcriptional cross-talk between pro- and anti-colonization pili biosynthesis systems in *Vibrio cholerae*. Mol Microbiol 67:849–860
- Chiavelli DA, Marsh JW, Taylor RK (2001) The mannosesensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. Appl Environ Microbiol 67:3220–3225
- Moorthy S, Watnick PI (2004) Genetic evidence that the *Vibrio* cholerae monolayer is a distinct stage in biofilm development. Mol Microbiol 52:573–587
- Meibom KL, Li XB, Nielsen AT, Wu CY, Roseman S, Schoolnik GK (2004) The Vibrio cholerae chitin utilization program. Proc Natl Acad Sci U S A 101:2524–2529
- Korber B (2000) HIV signature and sequence variation analysis. In: Rodrigo AG, Learn GH (eds) Computational analysis of HIV molecular sequences. Kluwer Academic, Dordrecht, pp 55–72
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426
- Ota T, Nei M (1994) Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol Biol Evol 11:613–619
- 43. Ganeshan S, Dickover RE, Korber BT, Bryson YJ, Wolinsky SM (1997) Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. J Virol 71:663–677
- Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–503
- 45. Paranjpye RN, Lara JC, Pepe JC, Pepe CM, Strom MS (1998) The type IV leader peptidase/N-methyltransferase of Vibrio vulnificus controls factors required for adherence to HEp-2 cells and virulence in iron-overloaded mice. Infect Immun 66:5659–5668
- Chattopadhyay S, Paranjpye RN, Dykhuizen DE, Sokurenko EV, Strom MS (2009) Comparative evolutionary analysis of the major structural subunit of *Vibrio vulnificus* type IV pili. Mol Biol Evol 26:2185–2196
- Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/ dS. PLoS Genet 4:e1000304
- Zampini M, Canesi L, Betti M, Ciacci C, Tarsi R, Gallo G, Pruzzo C (2003) Role for mannose-sensitive hemagglutinin in promoting interactions between *Vibrio cholerae* El Tor and mussel hemolymph. Appl Environ Microbiol 69:5711–5715
- Pruzzo C, Gallo G, Canesi L (2005) Persistence of vibrios in marine bivalves: the role of interactions with haemolymph components. Environ Microbiol 7:761–772
- Gubala AJ (2006) Multiplex real-time PCR detection of Vibrio cholerae. J Microbiol Methods 65:278–293
- Jagadeeshan S, Kumar P, Abraham WP, Thomas S (2009) Multiresistant *Vibrio cholerae* non-O1/non-O139 from waters in South India: resistance patterns and virulence-associated gene profiles. J Basic Microbiol 49:538–544