Comparative Genomic Analyses of the Vibrio Pathogenicity Island and Cholera Toxin Prophage Regions in Nonepidemic Serogroup Strains of *Vibrio cholerae*

Manrong Li, Mamuka Kotetishvili, Yuansha Chen and Shanmuga Sozhamannan

Updated information and services can be found at: [http://aem.asm.org/content/69/3/1728](http://aem.asm.org/content/69/3/1728)

**REFERENCES**

This article cites 44 articles, 27 of which can be accessed free at: [http://aem.asm.org/content/69/3/1728#ref-list-1](http://aem.asm.org/content/69/3/1728#ref-list-1)

**CONTENT ALERTS**

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), [more »](http://aem.asm.org/cgi/alerts)

Information about commercial reprint orders: [http://journals.asm.org/site/misc/reprints.xhtml](http://journals.asm.org/site/misc/reprints.xhtml)
To subscribe to another ASM Journal go to: [http://journals.asm.org/site/subscriptions/](http://journals.asm.org/site/subscriptions/)
Comparative Genomic Analyses of the Vibrio Pathogenicity Island and Cholera Toxin Prophage Regions in Nonepidemic Serogroup Strains of *Vibrio cholerae*

Manrong Li,1,2† Mamuka Kotetishvili,1 Yuansha Chen,1 and Shanmuga Sozhamannan1,2*

Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine,1 and VA Maryland Health Care System,2 Baltimore, Maryland 21201

Received 25 June 2002/Accepted 13 November 2002

*Vibrio cholerae* is a serologically diverse, environmental, gram-negative bacterial species. Although *V. cholerae* comprises more than 200 O-antigen-based serogroups (35), only the O1 and O139 serogroup strains are known to cause epidemics of cholera, a severe diarrheal disease. Ever since cholera epidemics caused by *V. cholerae* O139 Bengal surfaced in 1992 in the Bay of Bengal region (1, 31), there has been renewed interest in the pathogenesis and evolutionary mechanisms of non-O1 vibrios. Several studies have reported (8, 9, 11, 15, 28–30, 33, 36) the presence of the two virulence genes found in epidemic strains, tcpA and ctxAB, in environmental and clinical strains of serogroups other than O1 and O139. However, the mechanisms of origin of these strains are not completely understood at the present time.

Cholera toxin encoded by ctxAB is responsible for the severe diarrheal symptoms elicited by the bacterium (20), and toxin-coregulated pilus (TCP) encoded by tcpA is responsible for efficient colonization of the human intestinal tract by the bacterium (42, 43). In addition, the O antigen is the bacterium’s major protective antigen, and therefore, changes in the O antigen of a preexisting epidemic strain may result in a new pathogen capable of causing disease in populations immune to the original epidemic strain (1, 31). For example, the *V. cholerae* O139 Bengal strain emerged from an O1 epidemic strain by genetic exchange of O-antigen biosynthesis regions (3, 26, 40), and O139 strains cause disease in persons immune to O1 strains (4, 27).

The ctxAB genes are carried in the genome of a filamentous, single-stranded DNA phage designated CTXφ (44), and their dissemination to nonpathogenic strains may, therefore, occur via phage-mediated horizontal gene transfer. Since the discovery of CTXφ, intra- and interspecies transfers of ctxAB genes in *V. cholerae* and *Vibrio mimicus*, via CTXφ-mediated transduction, have been demonstrated under both laboratory and natural conditions (6, 7, 13, 14).

The TCP structural gene, tcpA, has been mapped (21) to a gene cluster designated the vibrio pathogenicity island (VPI), and the VPI has recently been proposed (22) to also be a filamentous phage, VPIφ. Unlike for the CTXφ, convincing data are lacking for the existence or the horizontal transfer of VPIφ. At the present time, none of the 29 genes in the VPI (other than tcpA) has been assigned, based on experimental data, any function in the proposed phage’s life cycle or phage transduction. However, TCP has clearly been demonstrated (32, 43, 44) to be the receptor for CTXφ and to be the bacterium’s colonization factor.

*V. cholerae* non-O1 and non-O139 strains have been occasionally isolated from humans with gastroenteritis, extraintestinal infections, and cholera-like illness (20, 34). Strains of serogroup O37 have been implicated in localized outbreaks of cholera in the past (2, 19). In addition, Mukhopadhyay et al. (28) reported the identification and genetic characterization of potentially pathogenic non-O1 and non-O139 *V. cholerae* strains from environmental samples in the Calcutta region.
However, how frequently potentially epidemic non-O1 and non-O139 strains arise, and how widely they are distributed in nature, are not known at the present time.

Thus, the aims of our studies were to identify potentially pathogenic non-O1 and non-O139 strains and to determine the identified strains' evolutionary history. Here we report the identification and comparative genetic analyses of diverse non-O1 and non-O139 strains which possess VPI only or VPI and pre-CTXφ (a precursor of CTXφ which lacks ctxAB genes) (7) or CTXφ prophage and have distinct lineages compared to the epidemic strains, thus indicating independent horizontal transfer and acquisition of the virulence regions. While this manuscript was under preparation, Boyd and Waldor (8) also reported the evolutionary and functional analyses of several novel TcpA alleles carried by non-O1 and non-O139 strains.

### Materials and Methods

**Bacterial strains and growth conditions.** The *V. cholerae* strains used in the preliminary screening for the presence of tcpA and ctxAB included 194 serogroup strains from the Shimada type culture collection (35), 36 clinical strains, and 70 clinical and environmental strains from the Smith collection (37). N16961 and

### Table 1. Characterization of non-O1 and non-O139 *V. cholerae* strains with pathogenic potential

<table>
<thead>
<tr>
<th>Strain</th>
<th>History of isolation (country/year/source)</th>
<th>Serogroup</th>
<th>Presence of VPI Cluster&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ctxA</th>
<th>ctxB</th>
<th>ctx core&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>305</td>
<td>India/1966/diarrhea</td>
<td>O1 Cla</td>
<td>+</td>
<td>Cla</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N16961</td>
<td>Bangladesh/1975/diarrhea</td>
<td>O1 ET</td>
<td>+</td>
<td>ET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>153–94</td>
<td>Unknown/1994/CDC</td>
<td>O8</td>
<td>▲▲</td>
<td>O8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. 63</td>
<td>Japan/1991/diarrhea from travel in Thailand</td>
<td>O26</td>
<td>▼</td>
<td>O26</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>506–94</td>
<td>Thailand/1994/diarrhea</td>
<td>O44</td>
<td>▲</td>
<td>O44</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>AQ1875</td>
<td>Japan/1998/tortoise imported from Taiwan</td>
<td>O48</td>
<td>▼</td>
<td>O48</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>507–94</td>
<td>Thailand/1994/diarrhea</td>
<td>O49</td>
<td>▲▲</td>
<td>O49</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>8–76</td>
<td>India/1976/diarrhea</td>
<td>O77</td>
<td>▼</td>
<td>O77</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>1421–77</td>
<td>India/1977/diarrhea</td>
<td>O80</td>
<td>▼</td>
<td>O77</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>571–88</td>
<td>China/1988/diarrhea</td>
<td>O105</td>
<td>▲▲</td>
<td>O27</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>523–80</td>
<td>United States/1980/diarrhea</td>
<td>O115</td>
<td>▲</td>
<td>O115</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>203–93</td>
<td>India/1993/diarrhea</td>
<td>O141</td>
<td>▲▲</td>
<td>O53G</td>
<td>▼</td>
<td>▲▲</td>
</tr>
<tr>
<td>366–96</td>
<td>Japan/1996/prawn imported from Thailand</td>
<td>O191</td>
<td>▲▲</td>
<td>O191</td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>AM2</td>
<td>India/1995/diarrhea</td>
<td>O9</td>
<td></td>
<td></td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>AM107</td>
<td>India/1996/diarrhea</td>
<td>O144</td>
<td></td>
<td></td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>NRT36-S</td>
<td>Japan/1990/diarrhea</td>
<td>O31</td>
<td></td>
<td></td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>117–94</td>
<td>Korea/1994/river water</td>
<td>O35</td>
<td></td>
<td></td>
<td></td>
<td>▲▲</td>
</tr>
<tr>
<td>984–81</td>
<td>India/1981/diarrhea</td>
<td>O89</td>
<td></td>
<td></td>
<td></td>
<td>▲▲</td>
</tr>
</tbody>
</table>

<sup>a</sup> +, presence of the entire VPI cluster based on restriction mapping and hybridization.

<sup>b</sup> rstR<sup>4</sup> is different from rstR<sup>4</sup> at the following three positions: rstR<sup>4</sup>-H2, rstR<sup>4</sup>-G5D, and rstR<sup>4</sup>-R7H8.

<sup>c</sup> rstR<sup>4</sup> is SceC233 (28).

<sup>d</sup> Diffs from tcpA-env allele of reference 28 at one position, O27阴道.

<sup>e</sup> A single amino acid substitution: **M**46N**C**.573**O**141.

<sup>f</sup> O53G is SceC5 from reference 28: rstR<sup>4</sup> is a novel allele, provisionally designated rstR6.

<sup>g</sup> Abbreviations: Cla, classical; ET, EI Tor; WT, wild type; NT, novel type; ▲, insertion; ▲▲, deletion; CDC, Centers for Disease Control and Prevention.

<sup>h</sup> Presence of CTXφ genes other than ctxAB.

### Table 2. List of primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Gene, probe</th>
<th>Sequence (5'-3')</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulak 80</td>
<td>VPI-1, for-a*</td>
<td>GCA ACA GGA TGA GTA ATC GAG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 81</td>
<td>VPI-1, rev-a*</td>
<td>CTA TTG CAT AAC GTA GCT CAC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 84</td>
<td>VPI-3, for-c*</td>
<td>TGA GCC TGA AAT AAC CAG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 85</td>
<td>VPI-3, rev-c*</td>
<td>GAT GAT GAA GTG TAT ATC TAC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 86</td>
<td>VPI-4, for-c**</td>
<td>TGG GGA AGA CTT TGG CTG AAG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 87</td>
<td>VPI-4, rev-c**</td>
<td>ATC TCT TGA ATG GGC TTT ACC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 121</td>
<td>VPI-5, for-b*</td>
<td>AGC GTT AGC TCT TCC ATC GAC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 122</td>
<td>VPI-5, rev-b*</td>
<td>CAG AGC GTC TCA TCA AGA TCC AC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 123</td>
<td>VPI-6, for-b**</td>
<td>GTG AAT CTT GAT GAG AGC CTC TG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 124</td>
<td>VPI-6, rev-b**</td>
<td>GGT GAG CCA GGC TTA TTT GGG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 76</td>
<td>pgm, for</td>
<td>AAA GAT ACT CAY GCS CGT TC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 77</td>
<td>pgm, rev</td>
<td>AAC CAG CGG TTT ACC GAC GGC AAC A</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 78</td>
<td>gyrB, for</td>
<td>GAA GGB GGT ATT CAA GC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 79</td>
<td>gyrB, rev</td>
<td>GAG TCA CCC TCC ACW ATG TA</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Materials and Methods**

**Bacterial strains and growth conditions.** The *V. cholerae* strains used in the preliminary screening for the presence of tcpA and ctxAB included 194 serogroup strains from the Shimada type culture collection (35), 36 clinical strains, and 70 clinical and environmental strains from the Smith collection (37). N16961 and

**Table 2. List of primers**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Gene, probe</th>
<th>Sequence (5'-3')</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulak 80</td>
<td>VPI-1, for-a*</td>
<td>GCA ACA GGA TGA GTA ATC GAG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 81</td>
<td>VPI-1, rev-a*</td>
<td>CTA TTG CAT AAC GTA GCT CAC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 84</td>
<td>VPI-3, for-c*</td>
<td>TGA GCC TGA AAT AAC CAG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 85</td>
<td>VPI-3, rev-c*</td>
<td>GAT GAT GAA GTG TAT ATC TAC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 86</td>
<td>VPI-4, for-c**</td>
<td>TGG GGA AGA CTT TGG CTG AAG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 87</td>
<td>VPI-4, rev-c**</td>
<td>ATC TCT TGA ATG GGC TTT ACC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 121</td>
<td>VPI-5, for-b*</td>
<td>AGC GTT AGC TCT TCC ATC GAC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 122</td>
<td>VPI-5, rev-b*</td>
<td>CAG AGC GTC TCA TCA AGA TCC AC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 123</td>
<td>VPI-6, for-b**</td>
<td>GTG AAT CTT GAT GAG AGC CTC TG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 124</td>
<td>VPI-6, rev-b**</td>
<td>GGT GAG CCA GGC TTA TTT GGG</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 76</td>
<td>pgm, for</td>
<td>AAA GAT ACT CAY GCS CGT TC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 77</td>
<td>pgm, rev</td>
<td>AAC CAG CGG TTT ACC GAC GGC AAC A</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 78</td>
<td>gyrB, for</td>
<td>GAA GGB GGT ATT CAA GC</td>
<td>This study</td>
</tr>
<tr>
<td>Sulak 79</td>
<td>gyrB, rev</td>
<td>GAG TCA CCC TCC ACW ATG TA</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Materials and Methods**

**Bacterial strains and growth conditions.** The *V. cholerae* strains used in the preliminary screening for the presence of tcpA and ctxAB included 194 serogroup strains from the Shimada type culture collection (35), 36 clinical strains, and 70 clinical and environmental strains from the Smith collection (37). N16961 and
395 represented the O1 El Tor and O1 classical strains, respectively. Other strains characterized in detail are described in Table 1. Bacterial strains were grown in Luria-Bertani medium, and other methods pertaining to bacterial growth and storage conditions have been described previously (38).

PCR. PCR amplification of various segments of the VPI and CTX regions to be used as probes in dot blot and restriction fragment length polymorphism (RFLP) analyses and PCR amplification of the tcpA, rstR, pgm, recA, and gyrB genes for sequencing were performed by using Taq DNA polymerase (Promega

FIG. 1. Neighbor-joining tree constructed by the maximum-parsimony method with the nucleotide sequences of gyrB gene fragments of V. cholerae strains. Labels at the branch tips represent strain designation, serogroup, and the virulence gene (T for tcpA and C for ctxAB), if present. The horizontal length represents the genetic distance, and the vertical lengths are not meaningful. The numbers above the branch lines represent the number of changes, and the numbers in parentheses below the branch lines at the branch node are the bootstrap values. Strains for which O-antigen switching is the proposed mechanism are underlined, and strains in which O-antigen switching is supported by sequence evidence are indicated by asterisks. Strains which appear to have acquired VPI and CTXφ or pre-CTXφ by independent horizontal gene transfer are highlighted.
RESULTS

Identification of non-O1 and non-O139 strains containing VPI and CTXΦ. To determine the prevalence of tcpA and ctxAB in non-O1 and non-O139 *V. cholerae*, 300 strains were screened by DNA dot blot analysis for the presence of the genes. Fifteen non-O1 and non-O139 strains carried the tcpA gene, 9 of the 15 strains carried the ctxAB genes, and none of the 300 strains carried ctxAB alone. The 15 tcpA+ strains were further analyzed, by Southern hybridization, for the presence of other genes in the VPI and CTX regions, and the results are summarized in Table 3. All of the tcpA+ strains also carried three other genes, adaT, toxT, and int, from the left, middle, and right ends of the VPI cluster, respectively, which indicated that the entire VPI region might be present in these strains. Thirteen of the 15 tcpA+ strains carried rstR and rstA of the CTXΦ genome, and two other strains (serogroups O77 and O80) did not carry any of the CTXΦ genes. Nine of the 13 rstA+rstR+ strains also carried ctxAB, one strain (serogroup O115) carried only rstA and rstR, and three strains (serogroups O48, O53, and O65) carried rstR, rstA, and the genes of the core region, except ctxAB. Li et al. (25) previously reported the genetic characterization of four of the 15 strains (serogroups O27, O37, O53, and O65) carrying the entire VPI and a CTXΦ or a pre-CTXΦ (7), i.e., CTXΦ lacking the ctxAB genes and an epidemic genetic backbone. In the present study, we examined the 11 remaining strains which have genetic backbones different from those of the epidemic strains, and we found that seven of the strains carried the CTXΦ, one carried the pre-CTXΦ, and one only had the repeat sequence element.

Genetic relatedness of the non-O1 and non-O139 strains to the epidemic strains. To determine the genetic relatedness of the 11 tcpA+ nonepidemic strains to the tcpA+ epidemic strains, MLST was performed. A 900-bp fragment of *pgm* (phosphoglucomutase), a 650-bp fragment of *gyrB*, and *recA* for multilocus sequence typing (MLST) are listed in Table 2. For most PCRs, the amplification conditions were as follows: 92°C for 5 min, followed by 35 amplification cycles, each consisting of sequential incubation at 92°C (30 s), 55°C (1 min), and 72°C (1 to 2 min), and a final extension at 72°C (5 min). In some PCRs, optimal annealing temperatures were determined by gradient PCR analysis with a Robocycler Gradient 96 (Stratagene, Inc.) thermocycler.

MLST. A 900-bp fragment of *pgm*, a 650-bp fragment of *gyrB*, a 785-bp fragment of *recA*, a 1.8-kb fragment encompassing the tcpA gene, and a 1.5-kb fragment of the ctxAB genes were PCR amplified, and the amplified fragments were sequenced (in both directions) by using the BigDye terminator cycle sequencing kit (Applied Biosystems, Inc., Foster City, Calif.) and either an ABI 377 Prism automated sequencer or an ABI 3700 DNA analyzer (Applied Biosystems, Inc.). Sequence alignments and dendrograms were generated as described previously (24) by using the Clustal-X (18) and PAUP (D. Swoford, Sinauer Associates, Sunderland, Mass.) programs, respectively. The *rstR* gene (750 bp) was PCR amplified with primers M407 (*rstR1*) and M408 (*rstR2*) (25), subcloned in the pCR2.1 vector (Invitrogen Life Technologies, Carlsbad, Calif.), and sequenced. When two different-sized fragments were obtained in a PCR of a single strain, the fragments were gel purified, cloned separately, and sequenced.

DNA dot blot and Southern analyses. Screening for the presence of tcpA and ctxAB was performed by DNA dot blot analysis as described previously (38). The presence of the adaT, toxT, int, *rstR*, and *rstA* genes was determined by Southern hybridization of the genomic DNAs with PCR fragments of these genes amplified from an El Tor strain, N16961. The *rstA* gene was PCR amplified with primers M573 (*rstA1*) and M574 (*rstA2*) (Table 2), and all of the other primers were reported earlier (25).

RFLP analysis. Two ~50-kb genomic segments encompassing the VPI and CTXΦ prophage regions were analyzed by RFLP. Genomic DNAs were restriction digested with *XmnI* (for analysis of the VPI region) and with *EcoRI* or *SphI* (for analysis of the ctx region), electrophoresed in an agarose gel, transferred onto a Zetaprobe membrane (Bio-Rad Laboratories, Hercules, Calif.), and simultaneously hybridized with multiple nonradioactive probes prepared with an enhanced chemiluminescence kit (Amersham Biosciences, Pittsburgh, N.J.). Hybridization conditions, the various probes, the lengths and order of the restriction fragments probed, and the exact endpoints of the two regions analyzed by RFLP were reported previously (25). The identity of the individual bands was confirmed by hybridizing the same blot with individual probes.

Nucleotide sequence accession number. The nucleotide sequences of the *rstR* (accession numbers AF452585 to AF452586), tcpA (accession numbers AF452570 to AF452580), ctxA (accession numbers AF452584, AF46340, and AF463401), ctxB (accession numbers AF463402 and AF452581 to AF452583), and *gyrB* alleles (accession numbers AF501888 to AF501913) of the non-O1 and non-O139 serogroup strains have been deposited in the GenBank database.
and nonepidemic strains were topologically separated into two major branches of the phylogenetic tree. The phylogram (Fig. 1) generated for one of the genes (gyrB) revealed a tight clustering of the strains of epidemic serogroups (O1 and O139) and strains of four other serogroups (O27, O37, O53, and O65), thus indicating the clonal nature of these strains. The branch node of the epidemic cluster was supported by a value of 99% in a 1,000-replicate bootstrap analysis. Similar clustering of epidemic strains was seen with the pgm and recA gene trees (data not shown), where the number of nucleotide changes per site among the strains was greater than with gyrB. The branch carrying 10 tcpA+ non-O1 and non-O139 strains and several other non-O1 and non-O139 strains had a low bootstrap value (47%) (Fig. 1), which suggests that the strains in this group are nonclonal in nature and, accordingly, mostly formed independent secondary branches. Although a few strains within this group formed clusters with high bootstrap values (serogroups O77 to O80, O8 to O44, and O89 to O144) (Fig. 1), this clustering was not consistent with the pgm- and recA-based trees. The remaining one tcpA+ non-O1 and non-O139 strain (serogroup O115) was a V. mimicus strain and was an outlier in the tree. Apparently, this strain has acquired the VPI region via interspecies horizontal transfer. Based on these data, we conclude that the 11 tcpA+ non-O1 and non-O139 strains have distinct lineages and did not originate directly from epidemic strains by O-antigen switching.

Genetic organization of VPIs. The genetic organization of the VPI region in epidemic and nonepidemic strains was compared by RFLP analysis of genomic DNAs digested with XmnI and hybridized with various probes derived from the VPI region (Fig. 2). The VPIs of the O1 classical and O1 El Tor strains had distinct RFLP patterns (Fig. 2, lanes O1 cla and O1 ET). Ten of the 11 non-O1 and non-O139 strains had distinct RFLP patterns that were different from those of the epidemic strains, with additional XmnI sites and/or genetic rearrangements (insertions and deletions) at the left end of the VPI (Fig. 2).
strains lacked pTLC, an element carried by all of the epidemic
strains. Also, the non-O1 and non-O139 strains was found to be quite heterogeneous and distinct from
organization of that region in the non-O1 and non-O139
cctxrgn
O141 (Fig. 6). Five other previously described (25) probes,
markers O8, O26, O44, O49, O105, and O191 and two copies in
their
collections. VPIs of a majority of the non-O1 and non-O139 strains differ
fi
tti
ci
ded in this study. Taken together, these data suggest that the
ti
tti
tti
tti

c tcpA

c ctxAB
/ H9278
vealed a single copy of the CTX

Eco
deletion at the VPI
regions in the 10 strains are shown in Fig. 3. The insertions and
deletions at the VPI's left end in the non-O1 and non-O139
strains were confirmed by hybridization of SalI-digested genomic
DNAs (Fig. 4). A SalI-generated fragment of the expected
size of 4,299 bp in length was seen in the wild-type VPI (Fig.
4b, lanes O1 cla and O1 ET). However, in the non-O1 and
non-O139 strains, a fragment of >10 to 12 kb was observed
(Fig. 4b, lanes O8 to O80, O191, O105, and O141), thus indicating

c ctx

Sal
mic DNAs (Fig. 4). A

att
site and

Sal

and O141 carried a deletion and an insertion of an approximately
7-kb fragment (O8) or an approximately 5-kb fragment (O105
and O141) at the left end of the VPI. Third, strains of sero-
groups O49, O115, and O191 carried deletions of various
lengths at the left end of the VPI.

To elucidate further the origin of the VPIs found in the
above-described non-O1 and non-O139 strains, the tcpA genes
located within the VPIs were sequenced. As observed earlier in
other studies (8, 9, 15, 28–30), extensive divergence in the tcpA
gene sequences of these strains was observed. The results are
summarized in Table 1, and a phylogenetic tree based on the
amino acid sequences (12 tcpA sequences from this study and
12 previously published tcpA sequences) is shown in Fig. 5.
Four strains (serogroups O44, O53, O65, and O139) carried an
El Tor allele, and strain 203-93 (serogroup O141) carried an
allele very similar to the one found previously in strain 10259
(serogroup O53), except at three positions (O44G⁰O141,
O53⁰O141, and O65G⁰O141) (15). The tcpA alleles in sero-
groups O27 and O105 contained an insertion of an
approximately 6-kb fragment between the VPI left
end of the VPI. Second, strains of serogroups O8, O105, and
O141 carried a deletion and an insertion of an approximately
7-kb fragment (O8) or an approximately 5-kb fragment (O105
and O141) at the left end of the VPI. Third, strains of sero-
groups O49, O115, and O191 carried deletions of various
lengths at the left end of the VPI.

To elucidate further the origin of the VPIs found in the
above-described non-O1 and non-O139 strains, the tcpA genes
located within the VPIs were sequenced. As observed earlier in
other studies (8, 9, 15, 28–30), extensive divergence in the tcpA
gene sequences of these strains was observed. The results are
summarized in Table 1, and a phylogenetic tree based on the
amino acid sequences (12 tcpA sequences from this study and
12 previously published tcpA sequences) is shown in Fig. 5.
Four strains (serogroups O44, O53, O65, and O139) carried an
El Tor allele, and strain 203-93 (serogroup O141) carried an
allele very similar to the one found previously in strain 10259
(serogroup O53), except at three positions (O44G⁰O141,
O53⁰O141, and O65G⁰O141) (15). The tcpA alleles in sero-
groups O27 and O105 contained an insertion of an
approximately 6-kb fragment between the VPI left
end of the VPI. Second, strains of serogroups O8, O105, and
O141 carried a deletion and an insertion of an approximately
7-kb fragment (O8) or an approximately 5-kb fragment (O105
and O141) at the left end of the VPI. Third, strains of sero-
groups O49, O115, and O191 carried deletions of various
lengths at the left end of the VPI.

FIG. 4. RFLP analysis of the left end of the VPI region in V.
cholerae strains. The schematic diagram of the left end of the VPI
region and the expected SalI- and XmnI-generated fragments 1 to 3 as
shown in Fig. 3 are indicated at the top (A). SalI-digested genomic
DNAs were hybridized with the b probe (B). Strains of serogroups
O49 and O115 have a deletion in the region of the probe. ET, El Tor;
cla, classical.

Genetic organization of the ctx region. The organization of the
ctx region in the seven non-O1 and non-O139 strains that were
called tcpA⁰ (serogroups O8, O26, O44, O49, O105, O141,
and O191) was determined by RFLP analysis. EcoRI does not
cut within the CTXα genome; thus, digestion of the genomic
DNAs with EcoRI and hybridization with ctxAB probes revealed
a single copy of the CTXα genome in strains of sero-
groups O8, O26, O44, O49, O105, and O191 and two copies in
O141 (Fig. 6). Five other previously described (25) probes,
ctxrgn 1 to 5, were used to scan the ctx region, and the genetic
organization of that region in the non-O1 and non-O139
strains was found to be quite heterogeneous and distinct from
that of the epidemic strains. Also, the non-O1 and non-O139
strains lacked pTLC, an element carried by all of the epidemic
strains, and they possessed an El Tor-like RTX cassette (data
not shown).

To understand further the origin of the CTXαs present in the
non-O1 and non-O139 strains, the nucleotide sequences of the
rstR and ctxAB genes were determined (the allelic types
based on their sequences are listed in Table 1). The rstR se-
quences were diverse among the strains examined in this study.
Four strains (serogroups O8, O44, O48, and O49) had two
alleles (double lysogens), one of which was an El Tor type in
the O8, O44, and O49 serogroups, three strains carried the
rstRα allele (serogroups O26, O141, and O191), two strains
carried only an rstRβ allele (serogroups O105 and O115) and
two new alleles were found, rstRα⁰ (serogroups O8, O44, and
O49), rstRα⁰ (serogroups O8, O44, and O49). rstR⁰ is a variant of the rstRα⁰ type (28), and
rstR⁰ is a novel type, which we provisionally designate
rstR6, and is similar to a sequence already available in the
GenBank database (accession number AF302794). A Clustal-
X alignment of the nucleotide sequences of the rstR type alleles
and two variants is shown in Fig. 7. Twenty-four rstR sequences
available in the GenBank database were aligned. They fall into 5 \(rst\)R types: El Tor type (5 sequences), classical type (4 sequences), Calcutta type (4 sequences), \(rstR4\) type (9 sequences), and \(rstR6\) (2 sequences). The sequences upstream and downstream of \(rstR\) (Fig 1 and \(rstA\)) are highly conserved in all of the alleles while the \(rstR\) region showed extensive variations, as do their respective \(RstR\) proteins.

The \(CtxA\) protein was conserved in all seven strains, except at one position in two strains (\(\text{wt}^S46N\text{O105, O141}\)). The \(CtxB\) protein was of the El Tor type in serogroups O8 and O49 and was of the classical type in serogroups O141 and O191. The classical and El Tor \(CtxB\) alleles differed from each other at two positions (\(\text{ET}Y39\text{H}^{\text{cla}}\) and \(\text{ET}I68\text{T}^{\text{cla}}\)). In the O26, O44, and O105 serogroups, novel alleles with four amino acid substitutions (\(\text{wt}^T36A, \text{wt}^F46L, \text{wt}^K55N, \text{and} \text{ET}I68\text{T}^{\text{cla, O26, O44, O105}}\)) were found.
Southern blot analysis of Eco al. (28) reported the identification and characterization of sim-
VPI or the VPI and CTX prophage (regions originally thought
sidered nonpathogenic and nonepidemic and which carry the
mental strains of V. cholerae
genetic backgrounds and virulence regions compared to the
O49, O77, O80, O105, O115, O141, and O191 have distinct
mechanism. Second, strains of serogroups O8, O26, O44, O48,
region in an O37 serogroup strain (25) support that emergence
that regard, previous studies (5, 26) and the results we ob-
transfer and exchange of O-antigen biosynthesis regions. In
epidemic strains, and they most likely emerged by horizontal
O65 have very similar genetic backgrounds and cluster with the
V. cholerae
intermediates in the evolution of the CTX
Waldor (8) reported several additional isolates (serogroups
strains. (7). Also, as previously reported (7),
elements of the inserts and the deletion end-points need to be determined in order to
understand whether translocation of the same region has oc-
curred in the non-O1 and non-O139 strains described in this
paper.
Extensive site-directed mutagenesis analysis of the tcpA
gene (10, 23, 41) has yielded useful insights into the structure-
function relationship of the TCP. Based on functional analysis
of mutant TcpA pilins, the TcpA protein has been proposed to have
different domains, N-terminal, C-terminal structural,
and C-terminal interaction domains (10, 23, 41). As observed
earlier (8), the various tcpA alleles identified in our study have
variations in the amino acid residues previously demonstrated
(23, 41) to have functional significance, i.e., the majority of the
changes are found in the C-terminal structural and interaction
domains (Fig. 8). The novel finding, however, is further delin-
cation of the region between amino acid residues 50 and 120,
which carries two mutational hotspots, residues 53 to 75 and 90
to 105, where multiple variations are seen (Fig. 8). The role of
this region in the structure-function relationship of TcpA pilin
remains to be determined. Since there are multiple variations
within any given tcpA allele, assessing the effect of these
changes may be difficult, especially if the changes in one do-
main are the result of intragenic suppression of mutations
elsewhere in the gene. Site-directed mutagenesis studies of this
region (residues 50 to 120) in a wild-type tcpA may reveal the
functional significance of this domain.
Horizontal gene transfer plays an important role in the evo-
lution of pathogenic bacteria, including V. cholerae, which has
become a paradigmatic organism for studying horizontal gene
transfer mechanisms. All three of the major known V. cholerae
virulence markers (CTXφ, VPI, and O-antigen biosynthesis
regions) are believed to have been acquired by horizontal gene
transfer. Two of these markers (CTXφ and VPI) have been
introduced into the V. cholerae chromosome via phage-medi-
ated transduction (22, 44); the mechanisms responsible for
acquiring the O-antigen biosynthesis regions are unknown at
the present time. However, acquisition of VPI and CTXφ
appears to be more frequent than exchange of O antigens in
epidemic serogroups, which may reflect mechanistic differ-
ences in their horizontal transfer. Acquisition of VPI and
CTXφ involves single-stranded phages and a site-specific re-
combination process, whereas acquisition of the O-antigen
region probably involves generalized transducing phages and a
homologous recombination mechanism. Hence, it is tempting
to speculate that horizontal transfer of O-antigen regions
might be subjected to DNA restriction and recombination bar-
riers.

The tcpA, rstR, and ctxAB diversity observed in this study is
similar to that reported in previous publications. For example,
the two new rstR alleles identified in the present work (rstR^O8
and rstR^O44) are very similar to the rstR^4 allele (28) and to
another sequence deposited in the GenBank database (acces-
sion number AF302794). In addition, an insertion and deletion
observed at the left end of the VPIs in some non-O1 and
non-O139 strains has been reported by Mukhopadhyay et al.
(28). However, the sizes of the inserted fragments observed by
us are much larger than the 1.6-kb segment of chromosome II
reported, in the above-referenced study (28), to have been
translocated to the left end of the VPI on chromosome I in
place of a 300-bp segment. The genetic contents of the inserts
and the deletion end-points need to be determined in order to
understand whether translocation of the same region has oc-
curred in the non-O1 and non-O139 strains described in this
paper.

DISCUSSION

In this study we utilized a panel of 300 clinical and environ-
mental strains of V. cholerae to identify and characterize 15
novel pathogens which belong to serogroups traditionally con-
sidered nonpathogenic and nonepidemic and which carry the
VPI or the VPI and CTX prophage (regions originally thought
to be exclusive to O1 and O139 serogroups). Mukhopadhyay et
al. (28) reported the identification and characterization of sim-
ilarly unusual V. cholerae pathogens (serogroups O8, O10,
O11, O27, O35, O42, and O69) isolated from environmental
samples from the Calcutta region, and very recently, Boyd and
Waldor (8) reported several additional isolates (serogroups
O8, O37 and O141). Thus, our study further expands the rep-
ertoire of novel V. cholerae serogroups (O26, O37, O44, O48,
O49, O53, O65, O77, O80, O105, O115, O141, and O191)
found to carry virulence genes.

The 15 non-O1 and non-O139 strains carrying the VPI clus-
ter could be divided into two groups, based on comparative
genomic analyses (MLST and RFLP) (Fig. 1 and 2 to 4, re-
respectively). First, strains of serogroups O27, O37, O53, and
O65 have very similar genetic backgrounds and cluster with the
epidemic strains, and they most likely emerged by horizontal
transfer and exchange of O-antigen biosynthesis regions. In
that regard, previous studies (5, 26) and the results we ob-
tained during our sequencing of the O-antigen biosynthesis
region in an O37 serogroup strain (25) support that emergence
mechanism. Second, strains of serogroups O8, O26, O44, O48,
O49, O77, O80, O105, O115, O141, and O191 have distinct
 genetic backgrounds and virulence regions compared to the
epidemic strains, and they most likely arose by independent
acquisition of the VPI. Also, several strains in the second
group subsequently acquired a CTXφ or a pre-CTXφ. None of
the 300 strains screened in our study had the CTXφ alone,
which supports the two-step sequential model for the acquisi-
tion of VPI and CTXφ (7). Also, as previously reported (7),
intermediates in the evolution of the CTXφ have also been
found, i.e., a pre-CTXφ lacking ctxAB has been found in three
different non-O1 serogroups (O48, O53, and O65).

FIG. 6. Analysis of the CTXφ prophage in V. cholerae strains.
Southern blot analysis of EcoRI-digested genomic DNAs probed with
the ctxAB gene probes is shown; each band represents one copy of the
CTXφ prophage genome. ET, El Tor; Cla, classical.
FIG. 7. Clustal-X alignment of rstR sequences. A total of 24 rstR sequences available in the GenBank database, which included alleles described in this study, were used for the analysis. Only the rstR type sequences that encode different RstR proteins (5 type sequences, i.e., El Tor [ET], classical [Cla], Calcutta [Cal], rstR4**, and rstR6) and two variants are shown in the figure. One sequence (accession number AF133308, designated rstR5 from strain SCE264) of the Calcutta type shows extensive variations, its RstR protein exhibited the least similarity to all of the other alleles (28), and hence, it was not included in the alignment. The strain designations of the published sequences are as follows (the serogroup [if known] and the GenBank accession numbers are in parentheses): El Tor (ET) type, E7946 (O1 ET, U83795), E7946 (O1 ET, U83796), SC511 (O1 ET, AF511000), N16961 (O1 ET, AE004224), and JS9803 (O139, AY101180); classical (Cla) type, O39 (O1 Cla, AF262318), SC9773 (O1 ET, AF510999), 569B (O1 Cla, AF05890), and 86015 (O1 ET, AF220606); Calcutta (Cal) type, AS207 (O139, AF110029), SCE188 (O44, AF133310), FJ98352 (O139, AF511001), and SCE264 (O42, AF133308); rstR4** type, SCE225 (O27, AF133807), 365-96 (O27, AF390570), JX4484 (O139, AF511001), VCE22 (O36, AY145124), VCE228 (O27, AY145125), VCE22 (O4, AY145129), VCE223 (O27, AF452885), VCE223 (O4, AY145127), 153-94 (O8, AF452885), VCE223 (O10, AF133334); Novel type, 506-94 (O44, AF452886) and 9803 (O139, AF302794). Nucleotides CT and AG are in green and red, respectively. ig-1, intergenic region 1; rstA, the 5' end of the rstA gene. Nucleotides that are identical in all sequences are indicated by asterisks.
The present study provides further support to the growing body of evidence that the classical *V. cholerae* virulence markers, *ctxAB* and *tcpA*, are not unique to epidemic strains and that they are found in at least some nonepidemic, non-O1 and non-O139 *V. cholerae* serogroups. However, the epidemic potential of these strains is not clear at the present time, especially since—and in clear contrast to O1 and O139 isolates—none of the non-O1 and non-O139 strains has been associated with past epidemics and/or pandemics of cholera. However, these observations raise the intriguing possibility that the presence of *ctxAB* and *tcpA* in *V. cholerae*, while critical, is not sufficient for the full-blown epidemics of cholera. In this context, Dziejman et al. (12) recently reported comparative genome analyses of 11 *V. cholerae* strains of epidemic serogroups, and they identified two clusters of genes differentiating pandemic strains from other strains. Similar comparative genomic analysis of epidemic strains with the nonepidemic, non-O1 and non-O139 strains identified in this study is likely to provide invaluable information regarding the mechanisms responsible for the emergence of the epidemic *V. cholerae* serogroups and strains.

**FIG. 8.** Domain structure of TcpA. The primary amino acid sequence of the El Tor TcpA is shown. The residues shown to have functional significance, based on site-directed mutagenesis (10, 23, 41), are indicated in boldface type, the two cysteine residues predicted to be essential for the structure and function of TcpA pilin are indicated by boxes, and the proposed domains are indicated by bars (10, 23, 41). The variations in the amino acid residues found in various tcpA alleles described in other studies (9, 15, 28–30) and in this work (highlighted) are indicated above the primary sequences. The variations seen in the classical allele are underlined. The changes unique to a single strain are indicated in italics. The O37 serogroup has a classical TcpA with a single change (K184 to E) indicated by a shaded box.
ACKNOWLEDGMENTS

We thank Arnold Kreger, Alexander Sulakvelidze, and Colin Stine for helpful comments and editorial assistance, Judy Johnson for providing access to the Smith strain collection, and Glenn Morris for support and encouragement during the course of this study.

M.K. was supported by an International Training and Research in Emerging Infectious Diseases grant from the Fogarty International Center at the National Institutes of Health. Funding for our studies was provided by a University of Maryland intramural grant (to S.S.) and BRFER intramural support from the Department of Veterans Affairs (to S.S.).

REFERENCES


