Molecular evolution of Vibrio pathogenicity island-2 (VPI-2): mosaic structure among Vibrio cholerae and Vibrio mimicus natural isolates

William S. Jermyn and E. Fidelma Boyd

Department of Microbiology, UCC, National University of Ireland-Cork, Cork, Ireland

Vibrio cholerae is a Gram-negative rod that inhabits the aquatic environment and is the aetiological agent of cholera, a disease that is endemic in much of Southern Asia. The 57.3 kb Vibrio pathogenicity island-2 (VPI-2) is confined predominantly to toxigenic V. cholerae O1 and O139 serogroup isolates and encodes 52 ORFs (VC1758 to VC1809), which include homologues of an integrase (VC1758), a restriction modification system, a sialic acid metabolism gene cluster (VC1773–VC1783), a neuraminidase (VC1784) and a gene cluster that shows homology to Mu phage. In this study, a 14.1 kb region of VPI-2 comprising ORFs VC1773 to VC1787 was identified by PCR and Southern blot analyses in all 17 Vibrio mimicus isolates examined. The VPI-2 region in V. mimicus was inserted adjacent to a serine tRNA similar to VPI-2 in V. cholerae. In 11 of the 17 V. mimicus isolates examined, an additional 5.3 kb region encoding VC1758 and VC1804 to VC1809 was present adjacent to VC1787. The evolutionary history of VPI-2 was reconstructed by comparative analysis of the nanH (VC1784) gene tree with the species gene tree, deduced from the housekeeping gene malate dehydrogenase (mdh), among V. cholerae and V. mimicus isolates. Both gene trees showed an overall congruence; on both gene trees V. cholerae O1 and O139 serogroup isolates clustered together, whereas non-O1/non-O139 serogroup isolates formed separate divergent branches with similar clustering of strains within the branches. One exception was noted: on the mdh gene tree, V. mimicus sequences formed a distinct divergent lineage from V. cholerae sequences; however, on the nanH gene tree, V. mimicus clustered with V. cholerae non-O1/non-O139 isolates, suggesting horizontal transfer of this region between these species.

INTRODUCTION

Vibrio cholerae, the causative agent of cholera, is a major cause of morbidity and mortality in many parts of the world where clean water is not available; in some areas of Southern Asia, cholera is endemic. V. cholerae is a natural inhabitant of the aquatic ecosystem, and most V. cholerae isolates do not possess the ability to cause cholera. Among V. cholerae natural isolates, only O1 and O139 serogroups are associated with epidemic and pandemic cholera. The V. cholerae O139 serogroup first emerged in 1992 as a cause of epidemic cholera (Albert et al., 1993; Cholera Working Group, 1993). It is now believed that O139 serogroup isolates arose from an O1 strain that acquired the O139 antigen (Berche et al., 1994; Bik et al., 1995; Stroehrer et al., 1997; Waldor & Mekalanos, 1994). All V. cholerae O1 and O139 serogroup isolates encode two main virulence factors: toxin co-regulated pilus (TCP) and cholera toxin (CT). TCP is encoded on a pathogenicity island, the Vibrio pathogenicity island (named VPI-1 in this paper), and CT is encoded by a filamentous phage, CTXφ (Karaolis et al., 1998; Kovach et al., 1996; Waldor & Mekalanos, 1996). Acquisition of TCP and CT by V. cholerae is sequential: first V. cholerae isolates acquired TCP, an essential intestinal colonization factor and the receptor for CTXφ (Taylor et al., 1987; Waldor & Mekalanos, 1996), and then they acquired CT via CTXφ (Waldor & Mekalanos, 1996). Several additional genomic regions have also been identified, predominantly among epidemic O1 and O139 serogroup isolates; these include RS1φ, Vibrio seventh pandemic island-I (VSP-I), VSP-II and VPI-2 (Faruque et al., 2002; Davis et al., 2002; Dziejman et al., 2002; Jermyn & Boyd, 2002; O’Shea et al., 2004a). RS1φ is associated with theCTX prophage in V. cholerae El Tor isolates and is required for the production of CTXφ (Davis & Waldor, 2000). VSP-I and VSP-II are genomic islands identified by microarray analysis among V. cholerae El Tor isolates (Dziejman et al., 2002). It was suggested by Dziejman et al. (2002) that the genes encoded on the VSP-I and VSP-II islands are likely to be responsible for the unique characteristics of the seventh pandemic (El Tor) strains.

Abbreviations: CT, cholera toxin; TCP, toxin co-regulated pilus; VPI-2, Vibrio pathogenicity island-2.
VPI-2 is a 57-kb chromosomal region encoding ORFs VC1758 to VC1809 which has all the characteristics of a pathogenicity island (Jermyn & Boyd, 2002). All toxigenic *V. cholerae* O1 and O139 serogroup isolates contained VPI-2, whereas non-O1/nonO139 non-toxigenic isolates lacked the region. VPI-2 encodes several gene clusters: a type-1 restriction modification system, which may protect the bacteria from viral infection (Bickle & Kruger, 1993); a nan-nag gene cluster homologous to genes involved in sialic acid metabolism, which may act as a carbon and nitrogen source (Vinmr et al., 2004); neuraminidase, which acts on higher-order gangliosides in the intestine converting them to GM1 gangliosides, with the subsequent release of sialic acids (Galen et al., 1992); and a region with homology to Mu phage. Among most *V. cholerae* O139 serogroup isolates, only a 20 kb region of VPI-2 was present, as ORFs VC1761 to VC1788 had been deleted, indicating the instability of the region (Jermyn & Boyd, 2002).

The species *V. mimicus* is closely related to *V. cholerae*; however, *V. mimicus* is phenotypically and genotypically distinct from *V. cholerae* and can be readily differentiated from *V. cholerae* (Boyd et al., 2000a; Byun et al., 1999; Davis et al., 1981; O’Shea et al., 2004a; Reen & Boyd, 2004). *V. mimicus*, unlike *V. cholerae*, is negative in sucrose, Voges–Proskauer, corn oil and Jordan tartrate reactions (Davis et al., 1981). Comparative sequence analysis of the housekeeping gene malate dehydrogenase (*mdh*) from *V. mimicus* isolates showed the mean pairwise divergence between *V. cholerae* and *V. mimicus* was approximately 10%, which is equivalent to the divergence between *Salmonella enterica* LT2 and *E. coli* K-12 (Boyd et al., 1994, 2000a; Byun et al., 1999; O’Shea et al., 2004a). In addition, analysis of groE-I and groEL-II on chromosomes 1 and 2, respectively, of *V. cholerae* and *V. mimicus* also demonstrated the divergence of both species from one another (Reen & Boyd, 2004). The natural habitat of *V. mimicus*, similar to *V. cholerae*, is the aquatic ecosystem, where it has been found both as a free-living bacterium and in association with phytoplankton and crustaceans (Acuna et al., 1999; Campos et al., 1996). Consumption of *V. mimicus*-contaminated shellfish has been linked to the development of gastroenteritis (Acuna et al., 1999). The virulence determinants of *V. mimicus* have not been well characterized. Recently, it was demonstrated that some *V. mimicus* isolates also harbour VPI-1 (TCP) and CTXφ (CT) (Boyd et al., 2000b). Several *V. mimicus* strains isolated from clinical and environmental sources were shown to produce multiple toxins, including a haemolysin, zonula occludens toxin, a heat-stable enterotoxin, as well as a CT-like toxin (Campos et al., 1996; Chowdhury et al., 1994; Ramamurthy et al., 1994; Shi et al., 1998; Spira & Fedorka-Cray, 1983, 1984).

In this paper, we examined 17 *V. mimicus* clinical and environmental isolates of various serogroups for the presence of the 57-kb VPI-2 region. ORFs VC1773 to VC1787 were found to be present in all 17 *V. mimicus* stains examined. In addition, 11 *V. mimicus* isolates contained ORFs VC1758 (integrate) and VC1804 to VC1809. We reconstructed the phylogenetic relationship of the nanH (VC1784) gene among a range of *V. cholerae* and *V. mimicus* isolates and compared it to the species tree, as deduced from the housekeeping gene *mdh*. On the *mdh* gene tree, *V. mimicus* isolates formed a separate divergent lineage from *V. cholerae*. On the nanH gene tree, however, *V. mimicus* isolates clustered with *V. cholerae* non-O1/non-O139 isolates, indicating that the *nanH* gene from these species is closely related. Analysis of additional genes spanning VPI-2 among a subset of *V. cholerae* and *V. mimicus* were in agreement with the nanH data, suggesting that VPI-2 was horizontally transferred between these species.

**METHODS**

**Bacterial strains.** A total of 22 *V. cholerae* and 17 *V. mimicus* isolates were used in this study. These isolates were derived from both clinical and environmental sources with a wide geographic distribution (Table 1). *V. cholerae* and *V. mimicus* strains were grown in Luria–Bertani (LB) broth. All strains were stored at −70°C in broth containing 20% (v/v) glycerol.

**DNA isolation.** Chromosomal DNA was extracted from each isolate using the G-nome DNA isolation kit (BIO 101). A single colony was inoculated into 3 ml LB broth and incubated overnight, with shaking, at 37°C. The cells were pelleted at 3000 r.p.m. for 5 min and resuspended in 1·85 ml suspension solution. The cells were lysed and treated with RNase and protease solutions. Following incubation at 37°C for 2 h, any remaining proteins were precipitated out and the sample was centrifuged. Supernatants were transferred to a clean Eppendorf tube (1·5 ml) and treated with both 100% ethanol and TE buffer. The DNA was spooled out and air-dried to remove excess alcohol. Following drying, the DNA was dissolved in TE buffer, pH 8, and stored at −20°C.

**Molecular analysis.** To assay for the presence of VPI-2 among *V. mimicus* isolates, 17 primer pairs that span the entire 57-kb VPI-2 region and two primer pairs that flank the island were used for PCR analysis (Jermyn & Boyd, 2002). PCR was performed in volumes of 20 μl containing 100 ng genomic DNA, 200 μM primer, 1·25 mM dNTPs, 50 mM MgCl2, and 1 μl Taq DNA polymerase (Bioline). The amplification conditions were pre-incubation at 96°C for 60 s, followed by 30 cycles of 94°C for 30 s, 48–55°C (depending on the primer pair) for 30 s and 72°C for a time chosen based on the size of the expected fragment (60 s kb−1). All PCRs were carried out in a PTC-200 Peltier Thermal Cycler (MJ Research). PCR products were resolved by electrophoresis on a 0·8% agarose gel made with 1× TAE (consisting of 40 mM Tris/acetate and 1 mM EDTA) and containing ethidium bromide. PCR amplification products were visualized by UV transillumination.

Southern hybridization was carried out with DNA probes generated from PCR products from the reference *V. cholerae* strain N16961 as a template (Heidelberg et al., 2000). PCR products were purified using the ConceRT-PCR purification kit (Gibco, BRL). For Southern hybridization analysis, DNA from each strain was digested with the restriction enzyme EcoRI (Roche Molecular Biochemicals), and the fragments were separated by electrophoresis in 0–6% TAE agarose. The gel was depurinated with 0·25 M HCl, denatured with a solution of 5 M NaCl and 10 N NaOH, and neutralized with 1 M Tris and 20× SSC. The fragments were transferred to nitrocellulose membranes by a posilbotter (Stratagene). DNA was fixed to the membrane by UV cross-linking. Approximately 100 ng probe DNA was conjugated to horseradish peroxidase using the ECL direct
nucleic acid labelling system (Amersham Pharmacia Biotech), and after hybridization overnight, was detected by the ECL chemiluminescent substrate. Hybridization membranes were washed according to manufacturer’s instructions.

**Pulsed-field gel electrophoresis.** To determine whether the ORF VC1758 encoding an integrase was located in the proximity of VC1773–VC1787, we carried out pulsed-field gel electrophoresis (PFGE). Genomic DNA was prepared from the *V. mimicus* strains PT5 and 523-80 as follows. The bacteria were grown in LB broth overnight at 37°C. The cells were centrifuged and resuspended in SE buffer (consisting of 75 mM NaCl and 25 mM EDTA) to an OD<sub>620</sub> of 1. An equal volume of the cell suspension and 2% agarose was mixed and cast in a mould. The individual agarose inserts were treated with PEN buffer (0.5 M EDTA, 1%, w/v, N-lauroylsarcosine, 1 mg Pronase ml<sup>-1</sup>) and washed with TE buffer (100 mM Tris/Cl, 10 mM EDTA). After processing, the inserts were incubated for at least 6 h with SfeI at 50°C. Electrophoresis was performed with a CHEF-DR II electrophoresis system (Bio-Rad) in 1× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA) at 13°C. Pulse ramps were 3 to 80 s for 21 h at 160 V in 1×0% agarose. After electrophoresis, the gel was stained with ethidium bromide, washed and treated for Southern hybridization analysis.

**DNA sequencing.** The primer pairs used for PCR amplification and subsequent sequencing are listed in Table 2. The nucleotide sequence of an internal fragment of six genes spanning the VPI-2 region was obtained: VC1764, encoding a methyl accepting subunit;
nonE (VC1781), encoding N-acetylmannosamine-6-phosphate 2-epimerase; nagC (VC1783), encoding N-acetylglucosamine-6-phosphate deacetylase; nanH (VC1784), encoding neuraminidase; VC1799, encoding an integrase; and VC1806, encoding a replicase. For the nanH gene, fragments were amplified from the chromosomal DNA of 22 V. cholerae strains and three V. mimicus strains, whereas for the remaining five genes, five V. cholerae strains were examined, which included a classical strain O395, an O139 serogroup strain MO2, and three non-O1/non-O139 strains V46, V47, and V52, and a V. mimicus strain PT5. PCR was performed in a 20 μl reaction mixture using conditions as described earlier. PCR amplicons were purified using the ConceRT-PCR purification kit (Gibco-BRL) in accordance with manufacturer’s instructions. Following purification, an aliquot of 10 μl was used as a sequencing template. All sequencing was carried out by MWG-Biotech.

Bioinformatic and phylogenetic analysis. The VC1764, nonE, nagC, nanH, VC1799 and VC1806 nucleotide sequences were aligned with the CLUSTAL W multiple sequence alignment program (http://www.ebi.ac.uk/clustalw/) (Higgins et al., 1996). The housekeeping locus mdh was also included in the analysis for the same set of strains (O’Shea et al., 2004a). Phylogenetic analysis was performed on the nanH and the mdh sequences with the Molecular Evolutionary Genetics Analysis (MEGA) version 2.1 (Kumar et al., 2001). Phylogenetic trees were constructed using the neighbour-joining method based on synonymous nucleotide sites, with corrections for multiple substitutions by the Jukes–Cantor method (Jukes & Cantor, 1969; Saitou & Nei, 1987; Kumar et al., 2001). Rates (per sites) of synonymous (kS) and non-synonymous (kA) substitutions were calculated by the method of Nei and Gojobori (1986). We also calculated the kS/kA ratio, which is a measure of the selective constraints on a gene. The VPI-2 region was analysed using the Basic Local Alignment Search Tool (BLAST) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/COG/).

RESULTS

Distribution of VPI-2 (ORFs VC1758–VC1809) among V. mimicus

Previously, we identified the 57.3 kb VPI-2 region among toxigenic V. cholerae isolates (Fig. 1). To determine whether the VPI-2 region was present in V. mimicus, PCR assays were carried out on 12 clinical and five environmental V. mimicus isolates using 17 primer pairs spanning the VPI-2 region (Tables 1 and 2). In all 17 V. mimicus isolates examined, positive PCR products were obtained with four of the 17 primer pairs: 148F/146B (encompassing VC1773–VC1776), 147F/146A (encompassing VC1777–VC1779), nagC1/nagA2 (encompassing VC1782–VC1783) and nanH3/nanH4 (encompassing VC1784). Thus, with primer pair 148F/146B, a 4·2 kb PCR band was obtained, and similarly with primer pair 147F/146A, a 2·3 kb PCR product was obtained from all strains examined (Fig. 2). The primer pair nagC1/nagA2 gave a 2·1 kb PCR product band and primer pair nanH3/nanH4 gave a 0·7 kb PCR product from all V. mimicus strains examined (Fig. 2). In addition, 10 clinical V. mimicus isolates and one environmental V. mimicus isolate gave positive PCR products with three additional primer pairs: nint3/nint4 (encompassing VC1758), VC3A/VC4A (encompassing VC1804–VC1807) and VC3B/VC4B (encompassing VC1806–VC1809). With the primer pair nint3/nint4, a 1·0 kb product was obtained; with the primer pairs VC3A/VC4A and VC3B/VC4B, 2·5 kb and 2·3 kb products were obtained, respectively (Fig. 2). To verify the PCR results, we carried out Southern hybridization analysis (Fig. 3). Using probes 148F, 147F, nagC and nanH, positive hybridization fragments were obtained from all 17 isolates, as expected, which span the region of interest. The probe RadC, spanning VC1786–VC1789, however, also gave a positive hybridization fragment in all isolates, even though PCR analysis failed to amplify this region. Using DNA probe nint spanning VC1758, and DNA probes VC3A and VC3B spanning VC1804–VC1809, Southern hybridization confirmed the presence of these regions in 11 V. mimicus isolates only. No hybridization fragments were obtained using probes specific for regions VC1759–VC1772 and VC1789–VC1803, demonstrating their absence in all V. mimicus isolates examined. In

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
<th>Product size</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF2</td>
<td>GATGGTCAATTGTTGAAACGGGA</td>
<td>2·9 kb</td>
<td>This study</td>
</tr>
<tr>
<td>1806R</td>
<td>TAGGATCCGCGGTGTGTTTCCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mdh1</td>
<td>ATGAATACTCCTGTTATT</td>
<td>0·9 kb</td>
<td>Boyd et al. (2000a)</td>
</tr>
<tr>
<td>mdh2</td>
<td>GTATCTAACTGCAATCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1764F</td>
<td>GCTCTAATCCTGATTTTCA</td>
<td>1·0 kb</td>
<td>This study</td>
</tr>
<tr>
<td>1764R</td>
<td>GGCCTTTGCTTCAATAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1781F</td>
<td>CCAAATCTGCTGTTCAC</td>
<td>0·6 kb</td>
<td>This study</td>
</tr>
<tr>
<td>1781R</td>
<td>GGTTGCGGTATTTGACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1783F</td>
<td>ATGGGCGCAATGTTTGG</td>
<td>0·6 kb</td>
<td>This study</td>
</tr>
<tr>
<td>1783R</td>
<td>CGGTATGCAATGGTTTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nanH3</td>
<td>TGACAGTCGCCACACAGG</td>
<td>0·7 kb</td>
<td>This study</td>
</tr>
<tr>
<td>nanH4</td>
<td>CAATTATTACGCCCATCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1799F</td>
<td>GACTTGTAGCTTGTGCGTTC</td>
<td>0·9 kb</td>
<td>This study</td>
</tr>
<tr>
<td>1799R</td>
<td>ATATCGCTCAACCCAGGGTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1806F</td>
<td>CATTCTGCCAATCCTGTAG</td>
<td>0·9 kb</td>
<td>This study</td>
</tr>
<tr>
<td>1806R</td>
<td>TAGGATCTCCTGCGTGTTC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Fig. 1.** Schematic representation and genetic organization of VPI-2 in *V. cholerae* O1 and O139 serogroups and *V. mimicus* isolates. ORFs are represented by open arrows indicating the direction of transcription. The spotted filled arrows indicate ORFs VC1773–VC1787. The dashed arrows indicate VC1758 (*int* gene) and VC1804–VC1809. Unfilled arrows represent the remaining VPI-2 ORFs. A black arrow represents the tRNA serine.

**Fig. 2.** PCR analysis of the VPI-2 region among 16 *V. mimicus* strains. Lanes: 1 and 20, 1 kb molecular mass marker; lane 2, *V. cholerae* strain N16961; lanes 3–18 *V. mimicus* isolates; lane 19, *V. cholerae* strain O395. (a) PCR amplification using the primer pair 148F and 146B; (b) PCR amplification using the primer pair nanH3 and nanH4; (c) PCR amplification using the primer pair VC3B and VC4B.
summary, examination of the results of the PCR and Southern hybridization analysis indicate that the VPI-2 in 11 *V. mimicus* isolates was 19·4 kb in size and consisted of 22 ORFs (VC1758, VC1773–VC1787 and VC1804–VC1809), whereas in six *V. mimicus* isolates the VPI-2 region consisted of a 14·1 kb region encompassing 15 ORFs VC1773–VC1804. We used PCR analysis and DNA sequencing to establish whether the region VC1773–VC1787 was contiguous with VC1804–VC1809. Using the PCR primer pair NF2 (located within the ORF VC1784) and 1806R (located within the ORF VC1806), we obtained a PCR product of 2·9 kb from the 11 isolates containing the 19·4 kb VPI-2 region. Nucleotide sequencing of the 2·9 kb PCR product from *V. mimicus* strain PT5 showed the presence of the ORFs VC1785, VC1786, VC1787, VC1804 and VC1805. Analysis of the region VC1785–VC1805 in *V. cholerae* strain N16961 revealed a 7 bp sequence AACTATT in the 3′ end of the ORF VC1787 and an identical 7 bp sequence in the 5′ end of the ORF VC1804. However, examination of the sequence in *V. mimicus* strain PT5 revealed that the last 5 bp of VC1787 and the first 12 bp of VC1804 were deleted, so only one copy of the 7 bp sequence AACTATT remained intact.

### Genomic mapping of the VPI-2 in *V. mimicus*

To determine whether the 19·4 kb VPI-2 in *V. mimicus* was inserted in the same insertion site as that in *V. cholerae*, we carried out PCR analysis using primers designed from the insertion site of the VPI-2 in *V. cholerae*. Using the primer pair tRNAF (located within the serine tRNA) and int4 (located within ORF VC1784), a PCR product of 1·3 kb was identified in 11 isolates containing the 19·4 kb VPI-2 region. Nucleotide sequencing of the 1·3 kb PCR product from *V. mimicus* strain PT5 showed that VPI-2 was located adjacent to a serine tRNA. Sequencing revealed that there

### Table 3. Sequence variation in the 611 bp nanH gene fragment of VPI-2 and in the 648 bp fragment of the chromosomally encoded mdh gene derived from *V. cholerae* and *V. mimicus* strains

<table>
<thead>
<tr>
<th>Gene and species</th>
<th>Number of strains</th>
<th>Number of sites</th>
<th>Number of sites</th>
<th>k_s</th>
<th>k_N</th>
<th>k_N/k_s</th>
</tr>
</thead>
<tbody>
<tr>
<td>nanH V. cholerae</td>
<td>22</td>
<td>49</td>
<td>41</td>
<td>8</td>
<td>0·086</td>
<td>0·004</td>
</tr>
<tr>
<td>V. mimicus</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>All</td>
<td>25</td>
<td>52</td>
<td>43</td>
<td>9</td>
<td>0·092</td>
<td>0·004</td>
</tr>
<tr>
<td>mdh V. cholerae</td>
<td>22</td>
<td>24</td>
<td>23</td>
<td>1</td>
<td>0·035</td>
<td>0·001</td>
</tr>
<tr>
<td>V. mimicus</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>All</td>
<td>21</td>
<td>82</td>
<td>79</td>
<td>3</td>
<td>0·159</td>
<td>0·003</td>
</tr>
</tbody>
</table>

Abbreviations: Total poly., total polymorphic sites; Syn., synonymous sites; Non-syn., non-synonymous sites; k_s, rate per site of synonymous substitution; k_N, rate per site of non-synonymous substitution.
was an intergenic region of 313 bp between the serine tRNA and VC1758, the first ORF of VPI-2 in *V. mimicus*.

PFGE was used to determine if VC1758 was located in close proximity to the ORFs VC1773–VC1787, since PCR analysis failed to amplify any product using a primer pair designed between these regions. Using the probes nint3/ nint4 (encompassing VC1758) and 148F/146B (encompassing VC1773–VC1776), an identical hybridization fragment of approximately 200 kb was obtained from *V. mimicus* strain PT5. Similarly for *V. mimicus* strain 523-80, an identical hybridization fragment of approximately 120 kb was obtained from each probe. This indicated that the ORFs VC1758 and VC1773–VC1776 were located in proximity to each other, on the same hybridization fragment, on the *V. mimicus* chromosome.

**Genetic variation at the nanH locus**

To determine the evolutionary history of the VPI-2 region among *V. cholerae* and *V. mimicus*, we sequenced a 611 bp (nucleotides 177–788) fragment of the nanH gene from 22 *V. cholerae* strains and three *V. mimicus* isolates. Sequencing revealed variation at 49 nucleotide sites among the 22 *V. cholerae* nanH sequences examined, 41 of which were synonymous substitutions (silent sites) and eight of which were non-synonymous substitutions (amino acid replacement sites) (Table 3, Fig. 4). No nucleotide differences were observed in *V. mimicus* nanH sequences. No allelic diversity was observed in *V. mimicus* nanH sequences.

![Fig. 4.](http://mic.sgmjournals.org) (a) Distribution of the polymorphic bases within the nanH gene of *V. cholerae* and *V. mimicus*. The nanH sequence of *V. cholerae* strain N16961 was used as the consensus sequence and only polymorphic sites in the nanH gene are shown. Nucleotides identical to those of N16961 are represented by dots. The position of each nucleotide in the gene is indicated by the value at the top of the alignment. The 1, 2 or 3 at the bottom of the alignment indicates the position of each base within a codon. The asterisks indicate non-synonymous substitutions resulting in amino acid changes. Alleles numbered 1 to 8 indicate grouping of identical sequences. (b) Distribution of the polymorphic amino acid residues within the NanH protein sequence of *V. cholerae* and *V. mimicus*. **
found between the epidemic *V. cholerae* O1 and O139 isolates examined. Of the 49 polymorphic sites identified, there were 14 singleton sites found among *V. cholerae* non-O1/non-O139 serogroup isolates. For example, *V. cholerae* O141 serogroup strain V46 contained seven singleton sites, O54 serogroup strain SG14 contained three singleton sites and O37 serogroup strain 151 contained two singleton sites. The maximum nanH sequence difference between *V. cholerae* strains was 5-23% between strains 151 and V46.

All three *V. mimicus* strains exhibited identical nanH nucleotide sequence to one another and contained three unique polymorphic sites compared to the *V. cholerae* nanH sequence (nucleotides 314, 549 and 609) (Table 3, Fig. 4). The mean pairwise percentage difference among the *V. cholerae* and *V. mimicus* nanH sequences was 2-00%.

Of the 203 NanH amino acid sites examined, nine polymorphic amino acid sites were found among the *V. cholerae* and *V. mimicus* strains (Fig. 4). Eight amino acid polymorphisms were found among *V. cholerae* isolates and three among *V. mimicus* isolates, with one amino acid polymorphic site unique to *V. mimicus*. Six of the nine amino acid replacement polymorphisms were conservative and involved substitution of amino acids with similar properties. The remaining three amino acid polymorphisms all involved substitution of a hydrophobic residue with a hydrophilic one.

Sequence analysis of the 22 *V. cholerae* mdh sequences revealed variation at 24 nucleotide sites, of which 23 were synonymous substitutions and one was a non-synonymous substitution (Table 3). All three *V. mimicus* strains exhibited identical mdh sequences to one another, and contained 58 unique polymorphic sites compared to *V. cholerae* mdh sequences, two of which resulted in amino acid changes (Table 3). The mean pairwise percentage difference among the *V. cholerae* and *V. mimicus* mdh sequences was 3-73% and the maximum difference was 12-03%.

The variability observed at the nanH locus was greater than that observed at the mdh locus for the same set of isolates. To determine whether the selective constraints at these loci are different, that is, whether nanH is under positive selection or not, we examined the synonymous (\(k_s\)) and non-synonymous (\(k_a\)) substitution rates and calculated the \(k_a/k_s\) ratio for each gene (Table 3). Amino acid replacements in a protein that incur an advantage to the cell are fixed at a higher rate than silent site substitutions resulting in a \(k_a/k_s\) ratio greater than 1. Amino acid replacements in a protein that are deleterious to the cell are fixed at a lower rate by purifying selection, with a \(k_a/k_s\) ratio less than 1. Substitutions in a protein that are neither selectively advantageous nor deleterious are neutral, with a \(k_a/k_s\) ratio equal to 1. The nanH gene \(k_a/k_s\) ratio was not high, indicating that this gene is not under positive selection (Table 3).

Evolutionary genetic relationships of VPI-2 among *V. cholerae* and *V. mimicus*

To reconstruct the evolutionary history of the VPI-2 region from *V. cholerae* and *V. mimicus*, a phylogenetic tree based on the nanH gene was constructed. The nanH sequences from 25 isolates were analysed using the Jukes–Cantor distance method and a neighbour-joining gene tree was constructed based on synonymous polymorphic sites only (Fig. 5). From Fig. 5 it can be seen that all *V. cholerae* O1 and O139 serogroup strains clustered together on the nanH gene tree. This cluster also included two *V. cholerae* non-O1/non-O139 serogroup strains: V52, a toxigenic O37 serogroup strain; and V45, a rough isolate, which gave an
identical \( \text{nanH} \) sequence to the O1 and O139 serogroup isolates. The other \( V. \text{cholerae} \) non-O1/non-O139 serogroup strains were grouped into two divergent clusters on the \( \text{nanH} \) gene tree. The first cluster contained \( V. \text{cholerae} \) strains V46, V47, V51 and SG14. The second cluster contained the strains V54, 9581, 9582, 151 and E714 in addition to \( V. \text{mimicus} \) strains PT5, PT48 and 9583.

A comparison of the \( \text{nanH} \) gene tree with the species tree, based on the chromosomal housekeeping gene \( \text{mdh} \), revealed a similar clustering of strains in each tree, with the exception of the \( V. \text{mimicus} \) isolates (Fig. 5). The \( V. \text{cholerae} \) O1 and O139 serogroup isolates formed a cluster, which included strains V45, V52 and V54 on the \( \text{mdh} \) gene tree. Similarly, \( V. \text{cholerae} \) non-O1/non-O139 strains V46, V47, V52 and SG14 clustered together, as did strains 151, 9581, 9582 and E714, on separate branches of the \( \text{mdh} \) tree. However, a notable feature of the gene trees is the placement of \( V. \text{mimicus} \) isolates: unlike the \( \text{nanH} \) gene tree, where \( V. \text{mimicus} \) isolates clustered with \( V. \text{cholerae} \) isolates, on the \( \text{mdh} \) gene tree \( V. \text{mimicus} \) isolates formed a separate distinct divergent lineage. The clustering of the \( \text{nanH} \) sequence from \( V. \text{mimicus} \) with \( V. \text{cholerae} \) indicates that horizontal transfer of \( \text{nanH} \) between \( V. \text{cholerae} \) and \( V. \text{mimicus} \) has occurred. Interestingly, the branch lengths of the \( \text{mdh} \) tree among \( V. \text{cholerae} \) isolates are short, which suggests that \( V. \text{cholerae} \) isolates are closely related and diverged from one another only very recently. The branch lengths on the \( \text{nanH} \) tree are longer than the \( \text{mdh} \) gene tree, which suggests that this region is evolving faster (Fig. 5).

To further determine the evolutionary history of the entire VPI-2 region among \( V. \text{cholerae} \) and \( V. \text{mimicus} \) isolates, we sequenced gene fragments from five additional ORFs, VC1764, VC1781, VC1783, VC1799 and VC1806, which spanned VPI-2. Nucleotide analysis of the five ORFs, for six \( V. \text{cholerae} \) strains, N16961, O395, MO2, V46, V47 and V52, provided similar findings to the \( \text{nanH} \) (VC1784) gene sequence analysis (Table 4). Strains N16961, O395, MO2 and V52 exhibited identical nucleotide sequences for the five ORFs VC1764, VC1781, VC1783, VC1799 and VC1806. Four ORFs, VC1781, VC1783, VC1799 and VC1806, of strains V46 and V47 exhibited different nucleotide sequences when compared to the consensus sequence in strain N16961. In contrast, the nucleotide sequence of the ORF VC1764 was identical between V46, V47 and N16961. Comparison of the nucleotide sequence of three ORFs, VC1781, VC1783 and VC1806, between \( V. \text{mimicus} \) strain PT5 and \( V. \text{cholerae} \) strain N16961 revealed similar findings to the \( \text{nanH} \) (VC1784) nucleotide comparison.

### DISCUSSION

In this study, we identified regions homologous to VPI-2 from \( V. \text{cholerae} \) in \( V. \text{mimicus} \) isolates. Of the 17 \( V. \text{mimicus} \) clinical and environmental isolates examined, all isolates contained a 14-kb region of VPI-2, which encompassed ORFs VC1773–1778, encoding homologues of neuraminidase and genes involved in sialic acid metabolism (Fig. 1). This region in \( V. \text{mimicus} \) was inserted adjacent to a serine tRNA similar to the VPI-2 region in \( V. \text{cholerae} \). In addition, 11 of the 17 \( V. \text{mimicus} \) isolates examined also contained VPI-2 ORFs VC1758 (encoding an integrase) and VC1804–1809. Further genomic mapping of these regions in \( V. \text{mimicus} \) indicated that ORFs VC1773–1778 were located upstream of VC1804–1809; however, PCR analysis failed to amplify any region between VC1758 and VC1773–1787 or VC1804–1809. Nonetheless, by PFGE analysis VC1758 was found located on the same genomic fragment as VC1773–1787 and VC1804–1809, indicating that additional DNA may be present between VC1758 and VC1773 which was outside the range of PCR amplification. Previously, we showed that, among \( V. \text{cholerae} \) O139 serogroup isolates, VPI-2 was truncated, with only a 20-kb region (VC1758–1760 and VC1789–1809) present, which included a number of Mu phage-like genes (Jermyn & Boyd, 2002). These data and the data presented in this study suggest that the VPI-2 is highly unstable. The size differences and instability of VPI-2 may be associated with the presence of a Mu-like phage, since

### Table 4. Number of variable sites in selected genes of VPI-2 between the consensus sequence of \( V. \text{cholerae} \) O1 strain N16961 and the sequences of \( V. \text{cholerae} \) and \( V. \text{mimicus} \) strains

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Fragment size</th>
<th>( V. \text{cholerae} )</th>
<th>( V. \text{mimicus} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>O395</td>
<td>MO2</td>
</tr>
<tr>
<td>VPI-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC1764</td>
<td>Methyl accepting subunit</td>
<td>818 bp</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VC1781</td>
<td>N-acetylmannosamine-6-phosphate 2-epimerase</td>
<td>559 bp</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VC1783</td>
<td>N-acetylglucosamine-6-phosphate deacetylase</td>
<td>481 bp</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VC1784</td>
<td>Neuraminidase</td>
<td>611 bp</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VC1799</td>
<td>Integrase</td>
<td>643 bp</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VC1806</td>
<td>Replicase</td>
<td>657 bp</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chromosomal</td>
<td>Malate dehydrogenase</td>
<td>648 bp</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Mu phage insertion and replication occurs by transposition and thus can induce chromosomal rearrangements such as deletion and insertion events.

The deletion of the type 1 restriction modification gene cluster and the nan–nag gene cluster from the *V. cholerae* O139 serogroup isolates may be the reason why this serogroup is no longer a dominant cause of epidemic cholera on the Indian subcontinent. The retention of VPI-2 in toxigenic *V. cholerae* O1 serogroup isolates suggests that this region may provide a competitive advantage in vivo, whereas the presence of the nan–nag region in all *V. mimicus* isolates examined suggests that it may be an essential region for survival in the environment. Interestingly, the nan–nag gene cluster is also present in the genomes of *Vibrio vulniificus* strains CMCP6 and YJ016 as well as *Vibrio fischeri* strain ES114, but is absent from the genome of *Vibrio parahaemolyticus* strain RIMD2210633 (W. S. Jermyn, unpublished data). Among *Vibrio* species sequenced to date, only *V. cholerae* and *V. mimicus* have been shown to encode neuraminidase (nanH), which cleaves sialic acid from sialoglycoconjugates (Corfield, 1990). *V. vulni ficus*, *V. parahaemolyticus* and *V. fischeri*, however, encode N-acetyleneuraminic acid synthase, which can synthesize sialic acid de novo, whereas *V. cholerae* and *V. mimicus* must acquire sialic acid from an external source.

There were 49 polymorphic substitutions among the nanH sequences examined in *V. cholerae*, which was approximately twice as many polymorphic sites compared to mdh in the same set of strains (Table 3). This could indicate that nanH is under some selective pressure for change or that there is relax pressure at this locus. Analysis of the nanH sequence demonstrated that most of the polymorphic sites were silent site substitutions that do not result in amino acid changes, and the $k_d/k_s$ ratio at this locus was much lower than 1, indicating that mutations in this gene are probably deleterious to the host (Table 3). In fact, there was a limited number of replacement polymorphisms at this locus, indicating functional constraints on this protein. Analysis of the sequenced nanH fragment revealed that a number of key features and motifs, which are essential for enzyme function, were conserved. For example, in the active site of NanH, two arginines that stabilize the carboxylic acid group of sialic acid were conserved. Also, the FRIP region, which is involved in substrate affinity, was conserved. This indicated that the enzyme function was maintained between *Vibrio* isolates.

In order to elucidate the evolutionary history of VPI-2 among *V. cholerae* and *V. mimicus*, comparative sequence analysis of the nanH gene tree with the species tree, deduced from the housekeeping gene mdh, was carried out. In a number of studies, the mdh locus has been shown to be an excellent indicator of the overall genetic diversity of a species, since it is not under selective pressure and is present in all isolates of a species (Boyd et al., 1994, 1996; Byun et al., 1999; O'Shea et al., 2004a; Reen & Boyd, 2004). Clustering of *V. cholerae* isolates on the nanH gene tree was similar to the clustering of isolates on the mdh gene tree; however, the congruence between the trees did not extend to nanH from *V. mimicus*. On the nanH gene tree, *V. mimicus* did not form a separate divergent branch from *V. cholerae* as it does on the mdh tree, but instead clustered with *V. cholerae* non-O1/non-O139 isolates (Fig. 5). We examined five additional loci on the VPI-2 among six *V. cholerae* isolates and one *V. mimicus* isolate (Table 4). These five loci, which span the nan–nag gene cluster and regions in the 5' and 3' of VPI-2, overall gave a similar pattern of divergence and clustering as nanH. This indicates that nanH, and by extension the nan–nag genes, were horizontally transferred between *V. mimicus* and *V. cholerae*. We favour an evolutionary scenario in which VPI-2 was present in the most recent common ancestor of *V. mimicus* and horizontally transferred to *V. cholerae* soon after these species diverged. Horizontal transfer between *V. cholerae* and *V. mimicus* has previously been documented (Boyd et al., 2000b; Faruque et al., 1999). The presence of both CTXφ and VPI in *V. mimicus* natural isolates has been demonstrated (Boyd et al., 2000b). Nucleotide sequence identity of genes from both CTXφ and VPI derived from *V. mimicus* and *V. cholerae* has been reported, which strongly suggests that contemporary horizontal transfer of DNA between these species has occurred (Boyd et al., 2000b). Both *V. cholerae* and *V. mimicus* occupy similar environmental niches and it is possible that *V. mimicus* acts as an environmental reservoir of novel DNA for *V. cholerae*. Horizontal gene transfer between *Vibrio* species appears to be an emerging theme in the evolution of this genus, since transfer of a 8.5 kb region encoded on the *Vibrio* seventh pandemic island-II (VSP-II) between *V. cholerae* and *V. vulni ficus* has also recently been described (O’Shea et al., 2004b).

**ACKNOWLEDGEMENTS**

Research in E. F. B.’s laboratory is funded by a PRTLI-3 grant from the Higher Education Authority Ireland and an Enterprise Ireland basic research grant.

**REFERENCES**


