Microbiology (2002), **148**, 3681–3693 Printed in Great Britain

Characterization of a novel *Vibrio* pathogenicity island (VPI-2) encoding neuraminidase (*nanH*) among toxigenic *Vibrio cholerae* isolates

William S. Jermyn and E. Fidelma Boyd

Author for correspondence: E. Fidelma Boyd. Tel: +353 21 4903624. Fax: +353 21 4903101. e-mail: f.boyd@ucc.ie

Department of Microbiology, University College Cork, National University of Ireland, Cork, Ireland

Acquisition of virulence genes encoded on mobile genetic elements has played an important role in the emergence of pathogenic isolates of *Vibrio cholerae*, the causative agent of the diarrhoeal disease cholera. The genes encoding cholera toxin (*ctxAB*), the main cause of profuse secretory diarrhoea in cholera, are encoded on a filamentous bacteriophage CTXϕ. The toxin coregulated pilus (TCP), an essential intestinal colonization factor, was originally designated as part of a pathogenicity island named the *Vibrio* pathogenicity island (VPI), but this island has more recently been proposed to be the genome of a filamentous phage, VPIϕ. In this study, it is shown that *nanH*, which encodes neuraminidase, maps within a novel pathogenicity island designated VPI-2. The 57·3 kb VPI-2 has all of the characteristic features of a pathogenicity island, including the presence of a bacteriophage-like integrase (*int*), insertion in a tRNA gene (serine) and the presence of direct repeats at the chromosomal integration sites. Additionally, the G+C content of VPI-2 (42 mol%) is considerably lower than that of the entire genome (47 mol%). VPI-2 encodes several gene clusters, such as a restriction modification system (*hsdR* and *hsdM*) and genes required for the utilization of amino sugars (*nan-nag* region) as well as neuraminidase. To determine the distribution of VPI-2 among *V. cholerae*, 78 natural isolates were examined using PCR and Southern hybridization analysis for the presence of this region. All toxigenic *V. cholerae* O1 serogroup isolates examined contained VPI-2, whereas non-toxigenic isolates lacked the island. Of 14 *V. cholerae* O139 serogroup isolates examined, only one strain, MO2, contained the entire 57·3 kb island, whereas 13 O139 isolates contained only a 20·0 kb region with most of the 5′ region of VPI-2 which included *nanH* deleted in these strains.

Keywords: virulence factors, bacteriophage, restriction modification

INTRODUCTION

Vibrio cholerae, a Gram-negative bacterium, is the aetiologic agent of the severe diarrhoeal disease cholera. Cholera has a predilection for pandemic spread and of the nearly 200 recognized serogroups of *V. cholerae*, only the O1 and O139 serogroups have been associated with epidemic cholera (Kaper *et al.*, 1995). The O1 serogroup is divided into two distinct biotypes, classical and El Tor, on the basis of a number of biochemical characteristics. To date, seven pandemics of cholera have been recorded. Little is known about the first four; however, the fifth and sixth pandemic were caused by the classical biotype and the present ongoing seventh pandemic, which began in 1961, is caused by the El Tor biotype (Barua, 1992). In 1992, a new epidemic serogroup, O139 Bengal, emerged and temporarily replaced the predominant O1 serogroup (Albert *et al.*, 1993; Cholera Working Group, 1993; Ramamurthy *et al.*, 1993). The emergence of this new O139 epidemic variant

Abbreviations: CT, cholera toxin; IS, insertion sequence; TCP, toxin coregulated pilus; VPI, *Vibrio* pathogenicity island.

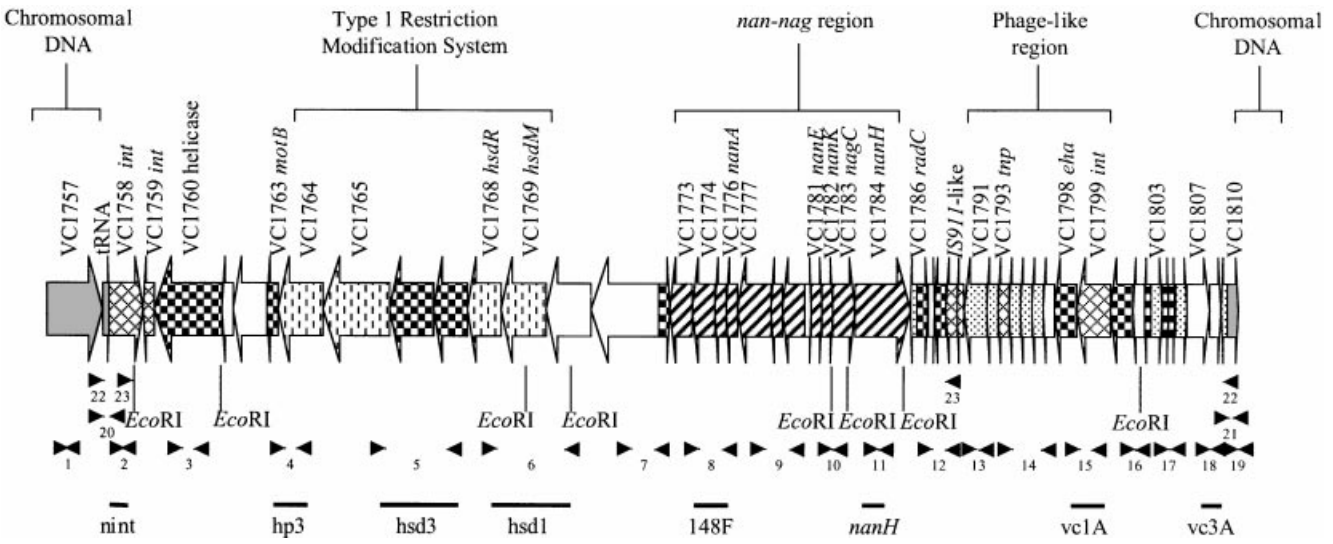

Fig. 1. Schematic representation and genetic organization of the 57·3 kb VPI-2. The position and direction of transcription of the ORFs are indicated by the direction of the arrows. The dash-patterned arrows indicate the restriction modification system, diagonal-patterned arrows indicate the *nan-nag* region, spotted arrows indicate genes of bacteriophage origin, chequered arrows indicate genes encoding hypothetical proteins, hatched arrows indicate potential mobility genes, white arrows indicate genes that are unlike any other gene in the database and grey arrows indicate chromosomal genes flanking VPI-2. The black, bold horizontal lines indicate the location of the probes used in hybridization. The small black arrowheads indicate the position of the PCR primer pairs. Vertical black lines indicate the position of restriction enzyme sites.

is thought to have resulted from the horizontal transfer of the genes encoding the LPS O-antigen to an El Tor progenitor strain via a bacteriophage (Berche *et al*., 1994; Waldor & Mekalanos, 1994; Bik *et al*., 1995, 1996; Comstock *et al*., 1995; Stroeher *et al*., 1997).

Interestingly, the main virulence factors of *V. cholerae* are also encoded on mobile genetic elements and are acquired via horizontal gene transfer. For example, the major symptoms of cholera result from the production of cholera toxin (CT) in the small intestine (Sears & Kaper, 1996) and the *ctxAB* genes encoding CT reside on a lysogenic filamentous phage, CTXϕ (Waldor & Mekalanos, 1996). The receptor for CTXϕ on the *V. cholerae* cell is a type IV pilus, the toxin-coregulated pilus (TCP), that also functions as an essential intestinal colonization factor (Taylor *et al*., 1987). The genes encoding the biosynthesis of TCP were initially shown to reside on a pathogenicity island, designated the *Vibrio* pathogenicity island (VPI) (Kovach *et al*., 1996; Karaolis *et al*., 1998). As defined by Hacker *et al*. (1997) a pathogenicity island is a large unstable chromosomal region that encodes several virulence genes; is present in pathogenic isolates and absent from non-pathogenic isolates; has a G + C content that differs from the rest of the genome; is associated with a tRNA gene; has IS and/or repeat sequences near the site of integration; and contains a bacteriophage-like integrase. The VPI fulfils all the criteria of a pathogenicity island and the identification of integrase and transposase genes at each end of the VPI suggests that they could be involved in the transfer and integration of this region (Karaolis *et al*., 1998). More recently, Karaolis and colleagues have suggested that the VPI is actually the genome of a novel filamentous bacteriophage VPIϕ; however, transfer of VPIϕ between *V. cholerae* O1 and O139 serogroup isolates, the predominant cause of epidemic cholera, was not shown (Karaolis *et al*., 1999). Instead, a recent report by O'Shea & Boyd (2002) demonstrated efficient transfer of the VPI region between *V. cholerae* O1 serogroup strains via CP-T1 generalized transduction.

In *Vibrio cholerae* neuraminidase, encoded by *nanH*, is thought to increase the sensitivity of host cells to CT (Galen *et al*., 1992). It has been suggested that *nanH* from a number of bacterial pathogens has been acquired by horizontal gene transfer (Roggentin *et al*., 1993). In this study, we examined the *nanH* gene and its flanking sequences among *V. cholerae* toxigenic (CTXϕ-positive) and non-toxigenic (CTXϕ-negative) isolates and found that the *nanH* gene is encoded within a 57·3 kb region that showed all the characteristics of a pathogenicity island which we named VPI-2 (Fig. 1).

METHODS

Bacterial strains. A total of 78 *V. cholerae* isolates from our laboratory collection were used in this study. The 78 strains examined were temporally (1947–1996) and geographically widespread (Africa, Asia, North and South America), of which 55 were toxigenic (CTXϕ-positive) and 23 were non-toxigenic (CTXϕ-negative), representing 14 serogroups, in-

Table 1. *V. cholerae* toxigenic and non-toxigenic strains used in Southern hybridization analysis

All strains were from clinical sources.

Strain	Serogroup (biotype)	*ctxAB*	Location	Year
O395	O1 (classical)	+	India	1964
C5	O1 (El Tor)	+	Indonesia	1957
569B	O1 (classical)	+	India	1947
SM115	O1 (El Tor)	+	Bahrain	1978
N16961	O1 (El Tor)	+	Bangladesh	1975
1528-79	O1	−	Louisiana	1979
MO2	O139	+	Madras	1992
MO10	O139	+	Madras	1992
MO45	O139	+	Madras	1992
AS207	O139	+	Calcutta	1996
AS209	O139	+	Calcutta	1996
AS210	O139	+	Calcutta	1996
V52	O37	+	Sudan	1968
SG3	O32	−	Calcutta	1992/1993
SG6	O45	−	Calcutta	1992/1993
SG7	O56	−	Calcutta	1992/1993
SG10	O69	−	Calcutta	1992/1993

cluding 36 O1 serogroup isolates, 14 O139 serogroup isolates and 28 isolates of the following serogroups: O8, O10, O11, O32, O35, O37, O42, O45, O56, O66, O69 and O141. *V. cholerae* isolates used for Southern hybridization analysis are listed in Table 1. Bacterial strains were stored at −70 °C in Luria–Bertani (LB) broth containing 20% glycerol. The antibiotic kanamycin was used at 40 μg ml^{-1}.

Molecular analysis. Total genomic DNA from each bacterial isolate was extracted using the G-nome DNA isolation kit (Bio101). PCR analysis was carried out using 23 primer pairs (Table 2), the location of which can be seen in Fig. 1. PCR primers were designed based on the genome sequence of *V. cholerae* O1 serogroup strain N16961 (Heidelberg *et al.*, 2000). PCR was performed in volumes of 20 μl containing 10 ng genomic DNA, 10 pmol primer and 1 U *Taq* DNA polymerase. The amplification conditions were pre-incubation at 96 °C for 1 min, 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 (1 min per kb). All PCRs were carried out in a PTC-200 Peltier Thermal Cycler (MJ Research).

Southern hybridization was carried out with eight DNA probes generated by PCR from the reference strain N16961 as template. 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 *Eco*RI (Roche Molecular Biochemicals) and the fragments were separated by electrophoresis in 0·6% TBE agarose. The gel was depurinated with 0·25 M HCl for 10 min, denatured with a solution of 5 M NaCl and 10 M NaOH for 40 min and neutralized with 1 M Tris and 20× SSC for 40 min. The fragments were transferred to nitrocellulose membrane by a Posiblotter (Stratagene). DNA was fixed to the membrane by UV cross-linking. Approximately 200 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.

Sequence analysis. A region spanning 57·3 kb from position 1896092 to 1953461 of the *V. cholerae* genome from strain N16961 (Heidelberg *et al.*, 2000) was analysed for sequence similarities using the BLAST algorithm (Altschul *et al.*, 1997). DNA sequence analysis was carried out on both DNA strands of the fk3/fk4 purified PCR product by MWG-Biotech and the resulting sequence was examined using BLAST (Altschul *et al.*, 1997).

RESULTS

Association of *nanH* with toxigenic *V. cholerae*

To determine whether there is a correlation between the presence of *nanH* and the presence of CTXϕ among *V. cholerae* isolates, the distribution of the *nanH* gene among 78 toxigenic (CXTϕ-positive) and non-toxigenic (CTXϕ-negative) isolates of *V. cholerae* was examined. Of the 30 O1 and 11 non-O1/O139 serogroup toxigenic isolates examined by PCR analysis with the primer pair nanH1 and nanH2, all gave a positive 1·9 kb PCR band. In contrast, 20 of the 23 non-toxigenic *V. cholerae* isolates examined failed to give a PCR amplification product. The three non-toxigenic strains that gave a 1·9 kb positive PCR band, 468-83, 2740-80, E714, were all O1 serogroup strains. Of the 14 toxigenic O139 serogroup isolates examined, only one strain, MO2,

Table 2. Oligonucleotide primer sequences used for PCR

All primers were designed in this study.

Primer no.	Name	Sequence (5′–3′)	Product size (kb)
1	acrB1	GGCTACATCAGCTACCTC	2·9
	acrB2	GCGGGATCACGACTAAG	
2	nint3	ATCTGATGGCGGCAATC	1·0
	nint4	GCGGCTTCAATGACATC	
3	helF	GATCAGCACCATTGTTCG	2·3
	helR	CGTTCCTGGAAAACCCTAC	
4	hp3	CGGTCTTTCTTCTGAATTGC	0·9
	hp4	TCGGCGAATGAGTTACGAG	
5	hsd3	GCTTAGTCGTCATATAGACTTC	3·9
	hsd4	TCCCTATCGCTGGAATG	
6	hsd1	TCCATGCCCACATTATG	4·5
	hsd2	CGCTGCATTAACCTTGC	
7	ukf	GTTAGCACTGGCTCTACC	3·4
	repr	CCGCAAGATAAGTATTGATGTC	
8	148F	CATTAGAGTTTTCTCCCC	4·2
	146B	AACAGGTTTGATTGCTGC	
9	147F	CCCATAACCATCAGAGC	2·3
	146A	CGCAATACTTACTTTGAGTG	
10	nagC1	CATACAGCGTTACCAGATAG	2·1
	nagA2	ACGTTGCAGTTATGTTCG	
11	nanH1	GACAGTCCAGCCAAACAG	1·9
	nanH2	CGTTAGCGTTGTTAGCCTC	
12	radC1	CACGGAATACTTACGCTG	1·8
	isB2	CCAAGAACCAAAGCATCGTTAC	
13	isA1	TCGGCTCAGCTTTACAC	1·4
	isA2	TTTCGGTAATCGACTGC	
14	is93	ATGGGGATCTCACTGCTG	3·2
	mu4	GATGTGCCGAAATGATCG	
15	vc1A	CCTCTTTGGTTGGGTTC	2·2
	vc2A	GACTACCAGAGACTTACGC	
16	vc1B	GGTCAGGAAATCCAAATCTAGG	0·9
	vc2B	TTACGGCTCGCTTTAGAG	
17	vc3A	CCTATTTTGCAGTTGGTGC	2·5
	vc4A	CGCTAATGTTTTTCCAGAGTGG	
18	vc3B	CGTACCATTTCCAATGATGCAG	2·3
	vc4B	GAAGTGCTAACTATGACTGG	
19	vc5B	GCATGATCGCTCAATCC	1·2
	vc6B	GCCTGTTTTTCGAGGTTG	
20	lr1	GGATTCGGTCGATACTGTC	1·6
	lr2	TCGTAGCCTTCCATTGC	
21	1808F	ATATGAGAGCAAGGGAAGTG	2·8
	vc6A	CGCTCTGTTGTAACGAGATG	
22	fk1	CTGGCTATGAGCTGATTTG	1·0
	fk2	AGGGATTGGCTTTGAGG	
23	fk3	TCTAGCCGTAATCCAAAGG	3·8
	fk4	AGATCACCCGCTACATTATC	

which was clinically isolated in Madras in 1992, gave a positive PCR band with the *nanH* primer pair (Table 3).

The PCR results were verified by Southern hybridization analysis using a 1·9 kb probe for the *nanH* gene generated by PCR from strain N16961 genomic DNA with the primer pair nanH1/nanH2. An identical hybridization fragment was obtained for seven toxigenic *V. cholerae* strains examined and as expected, no hybridization fragment was obtained for the five non-

Table 3. PCR analysis of VPI-2 among 78 toxigenic and non-toxigenic *V. cholerae* isolates

	acrB1/ acrB2	nint3/ nint4	helF/ helR	hp3/ hp4	hsd3/ hsd4	hsd1/ hsd2	ukf/ repr	148F/ 146B	147F/ 146A	nagC1/ nagA2	nanH1/ nanH2	radC1/ isB2	isA1/ isA2	is93/ mu4	vc1A/ vc2A	vc1B/ vc2B	vc3A/ vc4A	vc5B/ vc6B
O1 toxigenic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Non-O1/non-O139 toxigenic	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
O139*	+	+	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+
Non-toxigenic†	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+

* One O139 strain, MO2, isolated from Madras, India, in 1992 contained a full VPI-2, whereas all other O139 strains examined contained a truncated 20·0 kb VPI-2 region.
† Includes both O1 and non-O1/non-O139 serogroup isolates. All non-toxigenic isolates examined lacked VPI-2 except for three strains: 468-83, 2740-80 and E714.

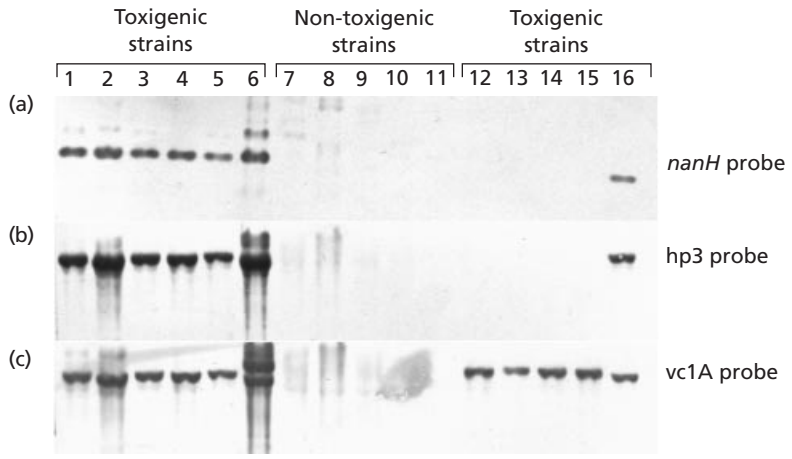

Fig. 2. Southern hybridization analysis of *Eco*RI-digested chromosomal DNA to verify the presence of VPI-2 in toxigenic *V. cholerae* isolates and absence of VPI-2 in non-toxigenic *V. cholerae* isolates. Lanes: 1, 569B; 2, SM115; 3, C5; 4, N16961; 5, MO2; 6, V52; 7, 1528-79; 8, SG3; 9, SG6; 10, SG7; 11, SG10; 12, AS207; 13, AS209; 14, MO10; 15, MO45; 16, O395. Probes used are indicated on the right. (a) Hybridization of DNA with the single-stranded *nanH* probe produced the expected 3·1 kb fragment and was obtained in all VPI-2-positive strains (lanes 1–6 and 16). No fragment was obtained in the O139 strains lacking the 5′ region of the island (lanes 12–15). (b) Hybridization of DNA with the single-stranded hp3 probe specific for a 5′ region of VPI-2 produced the expected 14·3 kb fragment in all VPI-2-positive strains (lanes 1–6 and 16). No fragment was obtained in the O139 strains lacking the 5′ region of the island (lanes 12–15). (c) Hybridization of DNA with the single-stranded vc1A probe specific for a 3′ region of VPI-2 produced the expected 11·9 kb fragment in all VPI-2-positive strains (lanes 1–6 and 16) and the O139 strains containing the 3′ region of the island (lanes 12–15). For each probe used no fragment was obtained in VPI-2-negative strains (lanes 7–11).

toxigenic strains examined (Fig. 2a and Table 1). Thus, a 3·1 kb hybridization fragment was obtained for O1 serogroup strains O395 (classical biotype), 569B (classical biotype), C5 (El Tor biotype), SM115 (El Tor biotype) and N16961 (El Tor biotype), one O139 serogroup strain MO2 and a non O1/O139 serogroup strain V52, which was associated with the cholera outbreak in Sudan in 1968. Therefore, it appears that the *nanH* gene is present in toxigenic *V. cholerae* and is absent from non-toxigenic isolates.

Association of *nanH* with a pathogenicity island

The association of *nanH* with toxigenic *V. cholerae* isolates prompted us to examine the region flanking *nanH* in the genome sequence of strain N16961 (Heidelberg *et al.*, 2000). A putative bacteriophage-like integrase gene (*int*), which is located adjacent to a serine tRNA gene, was identified 35·9 kb upstream of *nanH*. To determine whether there was an association between the presence of *int* and *nanH*, we carried out PCR analysis with the PCR primers for the *int* gene, nint3 and nint4 (Table 2), to examine the distribution of *int* among *nanH*-positive and *nanH*-negative *V. cholerae* isolates. Of the 45 *nanH*-positive strains examined, all gave an expected 1·0 kb PCR band, indicating the presence of *int* in these strains. In contrast, PCR analysis of the 20 *nanH*-negative *V. cholerae* non-toxigenic isolates failed to amplify a PCR band with nint3 and nint4 primers, indicating the absence of *int* in these strains. Unexpectedly, PCR analysis of the 13 *nanH*-negative O139 serogroup strains gave a 1·0 kb PCR band, indicating the presence of *int* in these strains (Table 3). Southern hybridization analysis of chromosomal DNA from *nanH*-positive and *nanH*-negative isolates with an *int* probe verified the absence of *int* in *nanH*-negative *V. cholerae* isolates.

We used PCR analysis to investigate whether the DNA sequence between the *int* and the *nanH* genes is present among *V. cholerae* isolates. Eight primer pairs (3–10) were designed to encompass the 35·9 kb region of interest (Fig. 1 and Table 2). Positive PCR bands were obtained with all eight primer pairs for the 45 *nanH*- and *int*-positive *V. cholerae* strains. No PCR products were obtained for *nanH*- and *int*-negative *V. cholerae* strains. Similarly, no PCR products were obtained for the 13 *nanH*-negative O139 serogroup strains that tested positive for *int* (Table 3). The absence of 35·9 kb between *int* and *nanH* among non-toxigenic *V. cholerae* was verified by Southern hybridization using six DNA probes (nint, hp3, hsd3, hsd1, 148F and *nanH*) derived from PCR fragments generated using primer pairs from Table 2, which span the regions of interest (Fig. 1). No hybridization fragments were obtained with the DNA probes (hp3, hsd1, hsd3 and 148F) for the nine *nanH*-negative strains tested (1528-79, SG3, SG6, SG7, SG10, AS207, AS209, MO10 and MO45) and, as expected, positive hybridization bands were obtained for all *nanH*-positive *V. cholerae* strains (Fig. 2b and data not shown).

PCR analysis was also used to investigate whether the region immediately upstream of serine tRNA and *int* is present among *V. cholerae* isolates. The primer pair acrB1 and acrB2 was designed to PCR-amplify the gene

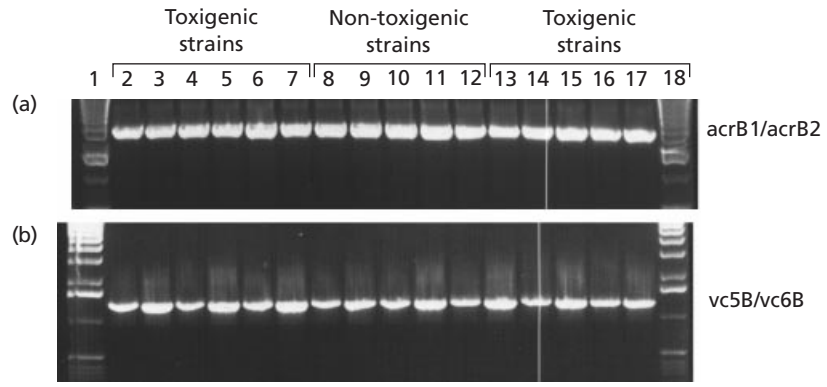

Fig. 3. PCR analysis of the chromosomal junction of VPI-2. Lanes: 1, 18, 1 kb molecular mass marker; 2, 569B; 3, SM115; 4, C5; 5, N16961; 6, MO2; 7, V52; 8, 1528-79; 9, SG3; 10, SG6; 11, SG7; 12, SG10; 13, AS207; 14, AS209; 15, MO10; 16, MO45; 17, O395. (a) An identical 2·9 kb product was obtained in all strains, both VPI-2-positive and -negative (lanes 2–17), using the primer pair acrB1 and acrB2. This PCR identified the chromosomal region flanking VPI-2 at the 5′ end. (b) An identical 1·2 kb product was obtained in all strains, both VPI-2-positive and -negative (lanes 2–17), using the primer pair vc5B and vc6B. This PCR identified the chromosomal region flanking VPI-2 at the 3′ end.

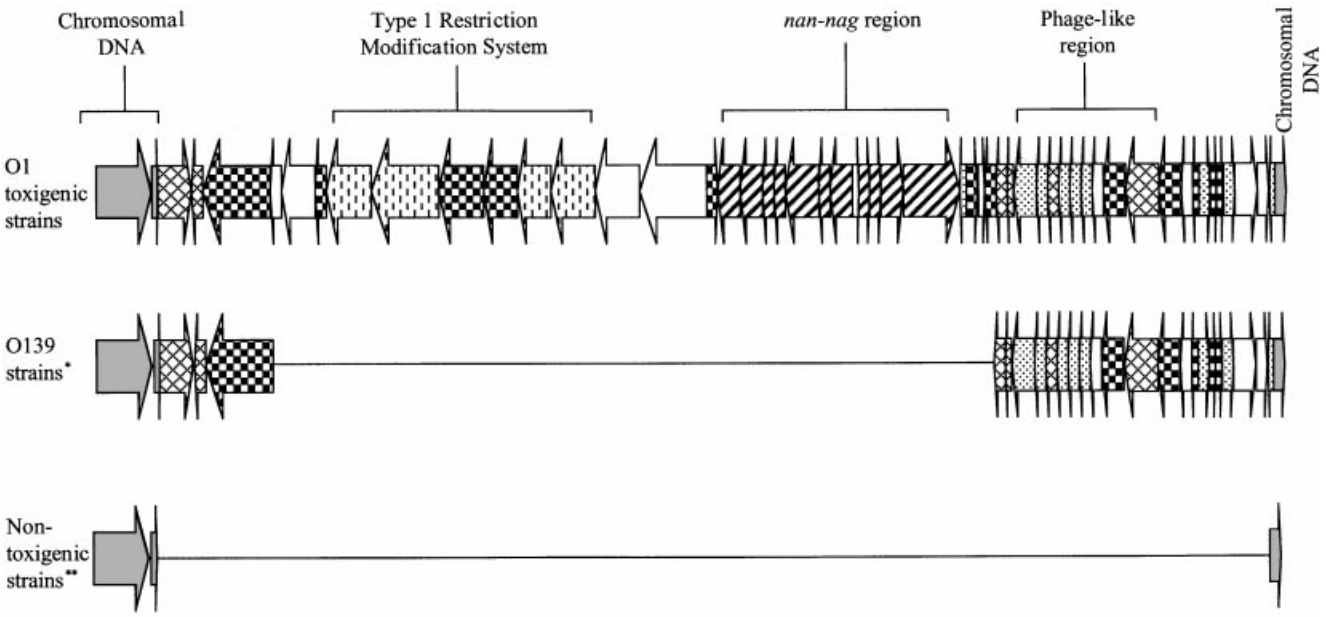

Fig. 4. The structure of VPI-2 among *V. cholerae* strains. An arrow denotes the presence of each individual VPI-2 ORF, whereas a thin black line denotes absence. The grey arrow at either end of VPI-2 denotes chromosomal DNA that is present in all strains. Three specific regions that are discussed in the text are labelled. *One O139 strain, MO2, isolated from Madras, India, in 1992 contained a full VPI-2, whereas all other O139 strains examined contained only a truncated VPI-2 containing 24 ORFs shown. **Includes both O1 and non-O1/non-O139 serogroup isolates. All non-toxigenic isolates examined lacked VPI-2 except for three strains: 468-83, 2740-80 and E714. No significant additional DNA has replaced the ORFs that are absent in both the O139 strains and non-toxigenic strains.

VC1757 immediately 5′ of the serine tRNA gene (Table 3). An identical 2·9 kb PCR band was obtained for all *V. cholerae* isolates examined (Fig. 3a). Therefore, it appears that the serine tRNA gene marks the point of insertion of the 39·5 kb region (*int-nanH*) that is only associated with toxigenic *V. cholerae* isolates.

We then tested whether the flanking DNA 3′ of *nanH* was present in all *V. cholerae* isolates. A PCR primer walking method was used; eight primer pairs (12–19) were designed downstream of *nanH* from the genome sequence of the strain N16961 (Heidelberg *et al.*, 2000) and used to PCR-amplify the corresponding regions in *V. cholerae* isolates (Fig. 1 and Table 2). This analysis revealed the presence of an additional 17·8 kb region that is unique to *nanH*-positive *V. cholerae* isolates. Thus, the 45 *nanH*-positive *V. cholerae* strains gave positive PCR bands with seven sets of primer pairs (12–18), whereas no PCR products were obtained for the *nanH*-negative isolates with the exception of the 13 *nanH*-negative O139 serogroup strains; these strains gave positive PCR bands with six of the primer pairs (13–18), indicating the presence of this region in these isolates (Table 3). Two DNA probes (vc1A and vc3A)

Table 4. ORFs within the *V. cholerae* VPI-2 pathogenicity island

ORF	Length (aa)	Homologous protein (organism)*	Amino acid identity (%)	*E* value
VC1758	411	Integrase (*V. cholerae*)	51	e-113
VC1759	159	Integrase (*V. cholerae*)	89	e-123
VC1760	940	Helicase (*Bacillus subtilis*)	29	1e-41
VC1761	202	No significant match	–	–
VC1762	483	No significant match	–	–
VC1763	244	Chemotaxis protein MotB (*Aquifex aeolicus*)	29	0·013
VC1764	706	Methyl-accepting subunit (*Deinococcus radiodurans*)	21	1e-04
VC1765	1022	Restriction enzyme helicase subunit (*Xylella fastidiosa*)	38	e-178
VC1766	663	Hypothetical protein (*P. multocida*)	26	8e-36
VC1767	470	Hypothetical protein NMA2230 (*Neisseria meningitidis*)	41	2e-04
VC1768	462	Type I restriction enzyme (*Xylella fastidiosa*)	51	1e-47
VC1769	793	Type I restriction DNA methylase (*Xylella fastidiosa*)	42	2e-78
VC1770	687	No significant match	–	–
VC1771	1220	No significant match	–	–
VC1772	286	Hypothetical protein NMA1157 (*Neisseria meningitidis*)	29	7e-16
VC1773	356	Hypothetical protein HI0148 (*Haemophilus influenzae*)	29	2e-36
VC1774	384	Hypothetical protein HI0148 (*Haemophilus influenzae*)	36	4e-58
VC1775	278	Hypothetical protein HI0143 (*Haemophilus influenzae*)	28	4e-22
VC1776	298	*N*-Acetylneuraminidase lyase HI0142 (*Haemophilus influenzae*)	27	9e-17
VC1777	427	C4-Dicarboxylase transporter HI0147 (*Haemophilus influenzae*)	48	9e-87
VC1778	173	C4-Dicarboxylase transporter HI0147 (*Haemophilus influenzae*)	33	1e-13
VC1779	321	C4-Dicarboxylase binding protein HI0146 (*Haemophilus influenzae*)	51	7e-77
VC1780	97	No significant match	–	–
VC1781	240	*N*-Acetylmannosamine-6-phosphate 2-epimerase HI0145 (*Haemophilus influenzae*)	52	1e-55
VC1782	287	*N*-Acetylmannosamine amide kinase HI0144 (*Haemophilus influenzae*)	37	7e-40
VC1783	378	*N*-Acetylamide glucosamide-6-phosphate deacetylase HI0140 (*Haemophilus influenzae*)	35	7e-42
VC1784	807	Neuraminidase (*Bacteroides fragilis*)	25	5e-06
VC1785	68	DNA binding protein (P4 bacteriophage)	54	4e-08
VC1786	158	RadC DNA repair (*Xylella fastidiosa*)	53	3e-32
VC1787	45	No significant match	–	–
VC1788	231	Hypothetical protein chromosome II (*V. cholerae*)	100	0
VC1789	290	IS*911*	66	1e-106
VC1790	114	IS*911*	72	5e-31
VC1791	346	Mu-like gp42 (Mu prophage)	35	4e-48
VC1792	119	Mu-like gp41 (Mu prophage)	33	1e-05
VC1793	125	Transposase (*Caenorhabditis elegans*)	32	3·4
VC1794	192	gp12 protein (PSA bacteriophage)	25	4·0
VC1795	106	MOR protein (Mu phage)	28	0·26
VC1796	124	MOR protein (Mu phage)	29	3e-05
VC1797	153	No significant match	–	–
VC1798	383	Eha protein (*Salmonella typhi*)	43	6e-37
VC1799	585	Integrase (*Salmonella typhimurium* LT2)	21	0·002
VC1800	323	Plasmid replication protein C (*Clostridium butyricum*)	27	4·0
VC1801	120	No significant match	–	–
VC1802	78	Transcriptional regulator (*Salmonella typhi*)	37	3·0
VC1803	153	Repressor protein (CP-933O prophage)	33	0·56
VC1804	104	Hypothetical protein VC0509 (*V. cholerae*)	55	1e-28
VC1805	148	Hypothetical protein VC0508 (*V. cholerae*)	55	8e-47
VC1806	328	Replicase (Fr bacteriophage)	38	0·35
VC1807	214	No significant match	–	–
VC1808	281	No significant match	–	–
VC1809	76	Vis protein (P4 bacteriophage)	50	3e-11

*Homology based on BLASTP database analysis.

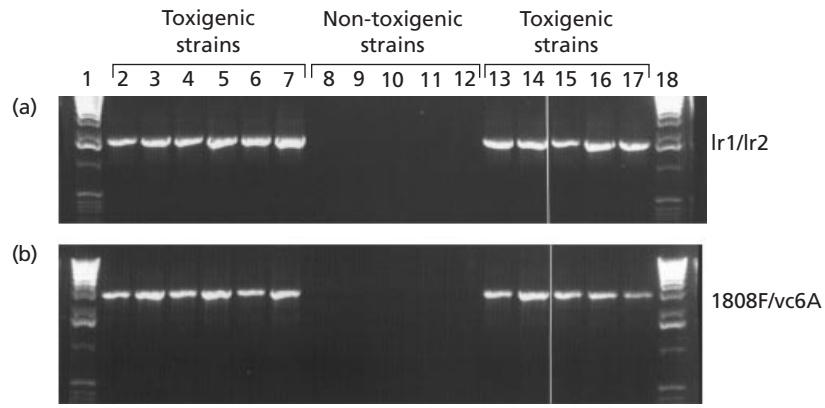

Fig. 5. PCR analysis of the insertion site of VPI-2. Lanes: 1, 18, 1 kb molecular mass marker; 2, 569B; 3, SM115; 4, C5; 5, N16961; 6, MO2; 7, V52; 8, 1528-79; 9, SG3; 10, SG6; 11, SG7; 12, SG10; 13, AS207; 14, AS209; 15, MO10; 16, MO45; 17, O395. (a) A 1·6 kb band containing the left junction of VPI-2 was obtained from all VPI-2-positive strains, using the primer pair lr1 and lr2 (lanes 2–7 and 13–17). (b) A 2·8 kb band containing the right junction of VPI-2 was obtained from all VPI-2-positive strains, using the primer pair 1808F and vc6A (lanes 2–7 and 13–17). For both primer pairs no PCR product was obtained from VPI-2-negative strains (lanes 8–12).

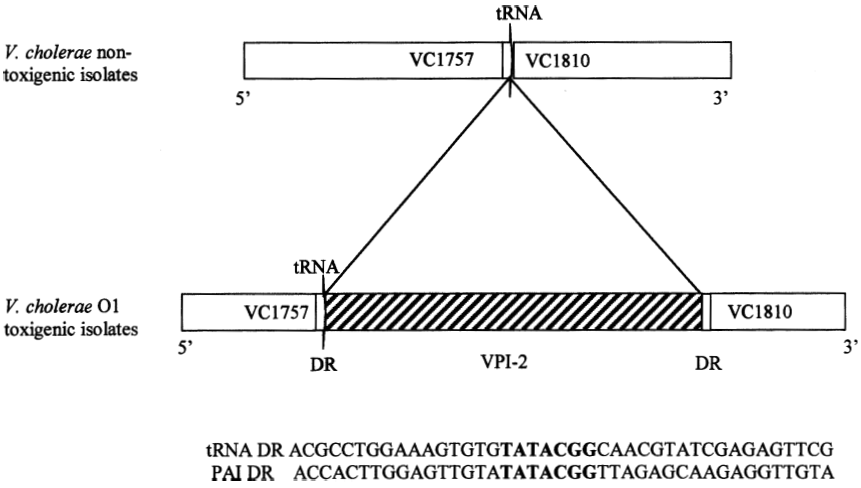

Fig. 6. Schematic diagram and nucleotide sequence of VPI-2 direct repeats (DR). In VPI-2-negative strains, a single copy of the 7 bp repeat is located at the 3′ end of the serine tRNA gene. Integration of the island produces a second copy of the repeat at the 3′ end of the pathogenicity island. The DNA sequence at the 3′ end of the serine tRNA gene and the 3′ end of the pathogenicity island (PAI) is shown, with the DR sequence given in bold type.

were used to verify the distribution of this region by Southern hybridization analysis (Fig. 1). Thus, hybridization analysis with these probes produced the expected 11·9 and 14·6 kb hybridization fragments, respectively, in all *nanH*-positive strains and the O139 serogroup strains, whereas no hybridization fragments were obtained for the *nanH*-negative *V. cholerae* strains examined (Fig. 2c and data not shown). The eighth primer pair, vc5B and vc6B, which amplifies gene VC1810, produced a 1·2 kb band with all toxigenic and non-toxigenic isolates examined, indicating the presence of this region in all isolates, and therefore marks the end of the 3′ region unique to *nanH*-positive isolates (Fig. 3b).

Among the 13 *V. cholerae* O139 serogroup isolates that contained a deletion within VPI-2, a primer pair (fk3 and fk4) was used to determine whether the region between the *int* gene (VC1758) and the IS911-like element (VC1789) was 'empty' in these isolates. Primer fk3 was designed 300 bp from the end of *int* (VC1758) and primer fk4 was designed within 100 bp from the start of IS*911* (VC1789) (Fig. 1). A 3·8 kb PCR product was amplified from all 13 O139 serogroup isolates, indicating the absence of additional DNA at this site. The 3·8 kb PCR product was sequenced and analysed using the BLAST program (Altschul *et al.*, 1997) to reveal the presence of the VC1759 gene and a truncated VC1760 gene (Fig. 4). Sequencing showed that the VC1760 gene has a 665 bp deletion in the 3′ region of the gene and this partial VC1760 sequence is adjacent to the IS911-like sequence in the 13 O139 isolates (data not shown).

Overall, it appears that *V. cholerae* toxigenic isolates contain a unique 57·3 kb region, which we named VPI-2, that is absent from most non-toxigenic isolates. Only one O139 strain, MO2, which was isolated from Madras in 1992 contains the entire 57·3 kb. In 13 O139 strains examined most of the 5′ region of VPI-2 is deleted with a 20·0 kb region remaining, which encompasses genes VC1758, VC1759, the partial sequence of VC1760 and VC1789 to VC1809 (Fig. 4). The VC1789 to VC1809 genes show homology to a number of bacteriophage genes (Table 4). The 13 O139 serogroup isolates that contained the truncated VPI-2 include two strains (MO10 and MO45) isolated from Madras in 1992, eight strains (AS207, AS209, AS210, AS212, AS213, AS231, AS259 and AS260) isolated from Calcutta in 1996, two strains (36054/98, 35636/97) isolated from Bangladesh in 1997 and 1998, and one strain (SG20) isolated from Calcutta.

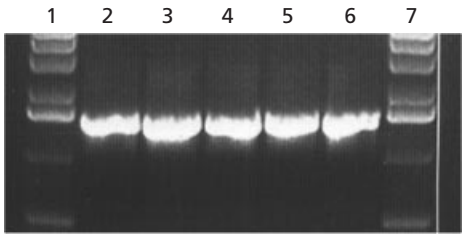

Fig. 7. PCR analysis of VPI-2-negative strains to determine whether the region between the 5′ serine tRNA gene and the 3′ VC1810 gene is 'empty' or contains unique DNA. Lanes: 1, 7, 1 kb molecular mass marker; 2, 1528-79; 3, SG3; 4, SG6; 5, SG7; 6, SG10. A 1·6 kb band was obtained from all VPI-2-negative strains, using the primer pair fk1 and fk2, indicating that no significant additional DNA was inserted into the region.

Characterization of the insertion site of VPI-2

To determine the insertion site of VPI-2 in toxigenic *V. cholerae* isolates, the left junction of VPI-2 was investigated using primer lr1 designed from part of the chromosome-specific DNA of VC1757 and primer lr2 designed from specific VPI-2 island DNA (VC1758) (Fig. 1 and Table 2). As expected a 1·6 kb PCR fragment was obtained with all 45 VPI-2 island-positive strains and the 13 O139 serogroup isolates, whereas no PCR band was obtained with VPI-2-negative strains (Fig. 5a). The insertion site of the right junction of VPI-2 was studied using the primer pair 1808F, designed from specific VPI-2 DNA (VC1808), and vc6A, designed from part of the chromosomal DNA of VC1810 (Fig. 1 and Table 2). These primers generated an expected 2·8 kb PCR band in all VPI-2-positive strains and the 13 O139 serogroup isolates. VPI-2-negative *V. cholerae* strains gave no PCR band (Fig. 5b). Sequence analysis of the insertion site revealed the presence of a 7 bp direct repeat at the 3′ end of the serine tRNA gene (Fig. 6). This 7 bp repeat sequence could mark the potential chromosomal integration site of VPI-2 in the *V. cholerae*.

V. cholerae VPI-2-negative strains were further analysed to determine whether the region between the 5′ serine tRNA gene and the 3′ VC1810 gene was 'empty' or contained unique DNA. A primer pair fk1 and fk2 was designed from chromosome-specific DNA of VC1757 and VC1810, which lie respectively in the left and right chromosomal junction fragment of VPI-2. VPI-2-negative strains produced a 1·6 kb product, which is slightly larger than the expected 1·0 kb product, but indicates that this region does not contain any significant additional DNA (Fig. 7). Analysis of VPI-2-positive *V. cholerae* strains did not yield a PCR product as expected, since these strains contain the entire VPI-2 or the truncated island which is beyond the capabilities of PCR amplification under the conditions employed.

Sequence analysis of the VPI-2 pathogenicity island

The first gene within VPI-2, *int*, lies immediately downstream of serine tRNA gene (Fig. 1). The *int* gene shows similarity (51% amino acid identity) to the *int* gene described in the VPI encoding TCP (Karaolis *et al.*, 1998) (Table 4). The *int* gene is preceded by a 247 bp non-coding region carrying a 35 bp sequence that is 88% identical to the region between the *int* and *ssrA* tRNA-like gene of the VPI (data not shown). However, the homology between VPI-2 and the VPI ends immediately 3′ of the *int* genes. This indicates that the mechanism of insertion of the VPI and VPI-2 may have common features, but the genes within each island are distinct.

VPI-2 contains a total of 52 ORFs and is 57·3 kb in length. The overall G + C content of VPI-2 (42 mol%) is lower than that of the entire *V. cholerae* genome (47–49 mol%). Of the 52 ORFs (VC1758–VC1809), 13 ORFs showed similarity to bacteriophage genes, 29 ORFs were homologous to genes of known function and 10 ORFs showed no significant matches in the database (Table 4). Among the 29 ORFs of known function, a number of gene clusters were observed. For example, downstream of the *int* gene (VC1758), a restriction modification system (VC1764–VC1769), containing a type 1 restriction modification gene (*hsdR*) and its associated DNA methylase gene (*hsdM*) was identified, which showed sequence similarity to a restriction modification system from *Xylella fastidiosa* (Simpson *et al.*, 2000). Adjacent to this region is a cluster of 11 genes (VC1773–VC1783) that showed considerable homology (27–52% amino acid identity) to an equivalent gene cluster in the *Haemophilus influenzae* genome, encoding enzymes involved in the utilization of amino sugars (*nan-nag* region) (Fleischmann *et al.*, 1995). The gene order in *H. influenzae* (*nanE-nanK*-HI0143-*nanA-nagB-nagA*) is somewhat different to that in *V. cholerae* (*nanA*-VC1777-VC1778-VC1779-VC1780-*nanE-nanK-nagC*), but it suggests a remarkable grouping of ORFs of related function. The *nan-nag* gene cluster is located immediately upstream of *nanH* and may potentially be involved in the utilization of sialic acid released by the enzymic action of this neuraminidase.

An IS-like element (VC1789–VC1790), located downstream of *nanH*, showed significant identity (66 and 72%) to the *Shigella flexneri* IS911 element (Prere *et al.*, 1991). Adjacent to this IS*911*-like element within a region that encompasses 19 ORFs (VC1791–VC1809), lie 10 ORFs that exhibit striking amino acid similarities (identities ranging from 21 to 50%) to ORFs of bacteriophage origin (Table 4). These 10 ORFs include genes encoding phage regulatory and tail functions. For example, the amino acid sequence of the ORFs VC1791 and VC1792 share similarity (identity 35%, *E* value 4e-48; identity 33%, *E* value, 1e-05) with the gp42 protein and gp41 protein of Mu phage (Table 4). Furthermore, the ORFs VC1793, VC1794, VC1795, VC1796 and VC1799 all exhibit sequence similarity to bacteriophage proteins and two of these (VC1793 and VC1799) show similarity to a transposase and integrase protein, respectively, suggesting that these genes may be involved in the mobilization and integration of this region (Table 4). Additionally, the ORF VC1803 exhibits 33% similarity with a repressor protein from phage CP-933O and

the amino acid sequence of ORF VC1809 has a 50% identity (*E* value 3e-11) with the Vis protein of P4. The IS*911*-like element is the first ORF that is present following the deletion of VC1761–VC1788 in the *nanH*-negative O139 serogroup strains.

DISCUSSION

In this study, we have identified a 57·3 kb VPI, which we named VPI-2, that exhibits all the characteristics of a pathogenicity island. Thus, VPI-2 is present in pathogenic *V. cholerae* isolates and is absent from non-pathogenic isolates. VPI-2 encodes a neuraminidase, a type I restriction modification system, a gene cluster involved in the utilization of amino sugars, an *eha* homologue and several conserved ORFs of unknown function. The significance of neuraminidase in cholera pathogenesis remains unclear. However, given our incomplete knowledge of *V. cholerae* pathogenesis it is possible that genes within VPI-2 may play a role in virulence. VPI-2 has a G+C content of 42 mol% that is considerably lower than that of the entire *V. cholerae* genome (47–49 mol%) and encodes a bacteriophage-like integrase (*int*), as well as a number of potential mobility genes, such as a transposase-like gene and an IS*911*-like insertion sequence.

In *V. cholerae*, VPI-2 is located within the 3′ end of a serine tRNA gene. The tRNA loci serve as conserved genomic landmarks for the insertion of mobile genetic elements (bacteriophages and pathogenicity islands) in a range of bacterial pathogens. For example, among pathogenic *E. coli* isolates five pathogenicity islands, PAI-1 and LEE, PAI-2, PAI-4 and PAI-5 have insertion sites in *selC*, *leuX*, *pheR* and *pheV*, respectively (Blum *et al*., 1994; Hacker *et al*., 1997; McDaniel & Kaper, 1997). These same tRNA genes also serve as integration sites for different bacteriophages in *E. coli* K-12 and O157 strain EDL933 (Blattner *et al*., 1997; Perna *et al*., 2001). Furthermore, the serine tRNA locus in particular serves as the insertion site for the *vap* region from *Dichelobacter nodosus* (Cheetham *et al*., 1995), SPI-5 in *Salmonella enterica* serovar Typhimurium (Wood *et al*., 1998), an 84 kb pathogenicity island in *E. coli* O157 strain EDL933 (Perna *et al*., 2001) and bacteriophage T12 of *Streptococcus pyogenes* (McShan *et al*., 1997).

An additional characteristic of some pathogenicity islands is their instability and the VPI-2 region between VC1760 (helicase) and VC1789 (IS*911*-like element), which includes the restriction modification system and the *nan-nag* regions, is deleted from 13 of the 14 *V. cholerae* O139 serogroup strains examined. Of the 13 O139 strains with the truncated VPI-2 most were isolated after 1992. Only one O139 strain MO2, which was isolated in 1992, contains the entire 57·3 kb VPI-2. A recent study by Dziejman *et al*. (2002) analysing differences in gene content between endemic and pandemic cholera isolates identified the same deletion of the region spanning VC1761–VC1786 in O139 serogroup strain MO10, consistent with our findings. In 1992 CT-producing *V. cholerae* O139 serogroup isolates emerged as the first non-O1 serogroup isolates to cause epidemic cholera in the Indian subcontinent, replacing the seventh pandemic *V. cholerae* O1 El Tor biotype strains (Albert *et al*., 1993; Cholera Working Group, 1993; Ramamurthy *et al*., 1993). However, the El Tor biotype, which was the progenitor for the O139 epidemic clone, soon re-emerged as the dominant cause of cholera in these areas (Faruque *et al*., 1997). A possible explanation for the re-emergence of El Tor cholera could be the presence of VPI-2 in the El Tor biotype and the deletion of most of VPI-2 from O139 serogroup strains.

The presence of *nanH* on a pathogenicity island suggests that it was acquired by horizontal transfer. Interestingly, it has been suggested that *nanH* from a number of bacterial pathogens was acquired by horizontal gene transfer (Hoyer *et al*., 1992; Roggentin *et al*., 1993). On the *S. enterica* serovar Typhimurium LT2 genome the *nanH* gene is encoded within a prophage Fels-1 and shows a G+C content (47 mol%) that differs from the host genome (50 mol%) (McClelland *et al*., 2001; Figueroa-Bossi, 2001). Similarly, in *Clostridium perfringens*, the *nanH* gene is located near a phage attachment site on the chromosome and has a G+C content of 31·9 mol% compared to the chromosomal DNA G+C content of 24–27 mol% (Canard & Cole, 1989). In *M. viridifaciens* it has been observed that a region flanking the *nanH* gene shows significant homology to a transposase (Sakurada *et al*., 1992).

In *V. cholerae*, it is hypothesized that the function of neuraminidase, encoded by *nanH*, is to act synergistically with CT by increasing the binding and penetration of the toxin to host enterocytes (Galen *et al*., 1992). However, these authors demonstrated only a modest effect of *nanH* on CT function *in vitro* and no significant effect *in vivo* in the suckling mouse model. Therefore, the true extent of the role of *nanH* in virulence remains unclear. We speculate that VPI-2 is likely to be important in pathogenesis, either directly in cholera virulence or indirectly in the transfer and integration of the island. The VPI-2 region may contribute to the survival of the bacterium in different ecological niches. For example, the product of *nanH*, neuraminidase, acts on higher order gangliosides converting them to GM_1 gangliosides, with the subsequent release of sialic acid. The *nan-nag* region encodes proteins involved in the utilization of both *N*-acetylglucosamine and sialic acid. The ability to utilize the sialic acid by-product as an alternative nutrient source could convey a significant competitive advantage to pathogenic *V. cholerae* strains. VPI-2 also encodes a number of ORFs of unknown function and we are now in the process of examining these regions for a possible role in *V. cholerae* pathogenesis.

ACKNOWLEDGEMENTS

We thank Yvonne O'Shea, Anne-Marie Quirke and Keith Neligan for helpful discussion. This research was funded by Enterprise Ireland and Higher Education Authority grants to E.F.B.

REFERENCES

Albert, M. J., Siddique, A. K., Islam, M. S., Faruque, A. S., Ansaruzzaman, M., Faruque, S. M. & Sack, R. B. (1993). Large outbreak of clinical cholera due to *Vibrio cholerae* non-O1 in Bangladesh. *Lancet* **341**, 704.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Barua, D. (1992). History of cholera. In *Cholera*, pp. 1–36. Edited by D. Barua & W. B. Greenough 3rd. New York: Plenum.

Berche, P., Poyart, C., Abachin, E., Lelievre, H., Vandepitte, J., Dodin, A. & Fournier, J. M. (1994). The novel epidemic strain O139 is closely related to the pandemic strain O1 of *Vibrio cholerae*. *J Infect Dis* **170**, 701–704.

Bik, E. M., Bunschoten, A. E., Gouw, R. D. & Mooi, F. R. (1995). Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J* **14**, 209–216.

Bik, E., Gouw, R. & Mooi, F. (1996). DNA fingerprinting of *Vibrio cholerae* strains with a novel insertion sequence element: a tool to identify epidemic strains. *J Clin Microbiol* **34**, 1453–1461.

Blattner, F. R., Plunkett, G. I., Bloch, C. A. & 14 other authors (1997). The complete genome sequence of *Escherichia coli* K12. *Science* **277**, 1453–1474.

Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschape, H. & Hacker, J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* **62**, 606–614.

Canard, B. & Cole, S. T. (1989). Genome organization of the anaerobic pathogen *Clostridium perfringens*. *Proc Natl Acad Sci U S A* **86**, 6676–6680.

Cheetham, B., Tattersall, D., Bloomfield, G., Rood, J. & Katz, M. (1995). Identification of a gene encoding a bacteriophage-related integrase in a *vap* region of the *Dichelobacter nodosus* genome. *Gene* **162**, 53–58.

Cholera Working Group. (1993). Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* non-O139 Bengal. *Lancet* **342**, 387–390.

Comstock, L. E., Maneval, D., Jr, Panigrahi, P., Joseph, A., Levine, M. M., Kaper, J. B., Morris, J. G., Jr & Johnson, J. A. (1995). The capsule and O antigen in *Vibrio cholerae* O139 Bengal are associated with a genetic region not present in *Vibrio cholerae* O1. *Infect Immun* **63**, 317–323.

Dziejman, M., Balon, E., Boyd, D., Fraser, C. M., Heidelberg, J. F. & Mekalanos, J. J. (2002). Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc Natl Acad Sci U S A* **99**, 1556–1561.

Faruque, S. M., Ahmed, K. M., Abdul Alim, A. R. M., Qadri, F., Siddique, A. K. & Albert, M. J. (1997). Emergence of a new clone of toxigenic *Vibrio cholerae* O1 biotype El Tor displacing *V. cholerae* O139 Bengal in Bangladesh. *J Clin Microbiol* **35**, 624–630.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D. & Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol Microbiol* **39**, 260–271.

Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Galen, J. E., Ketley, J. M., Fasano, A., Richardson, S. H., Wasserman, S. S. & Kaper, J. B. (1992). Role of *Vibrio cholerae* neuraminidase in the function of cholera toxin. *Infect Immun* **60**, 406–415.

Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**, 1089–1097.

Heidelberg, J. F., Eisen, J. A., Nelson, W. C. & 23 other authors (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483.

Hoyer, L. L., Hamilton, A. C., Steenbergen, S. M. & Vimr, E. R. (1992). Cloning, sequencing and distribution of the *Salmonella typhimurium* LT2 sialidase gene, *nanH*, provides evidence for interspecies gene transfer. *Mol Microbiol* **6**, 873–884.

Kaper, J. B., Morris, J. G., Jr & Levine, M. M. (1995). Cholera. *Clin Microbiol Rev* **8**, 48–86.

Karaolis, D. K., Johnson, J. A., Bailey, C. C., Boedeker, E. C., Kaper, J. B. & Reeves, P. R. (1998). A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc Natl Acad Sci U S A* **95**, 3134–3139.

Karaolis, D. K., Somara, S., Maneval, D. R., Jr, Johnson, J. A. & Kaper, J. B. (1999). A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375–379.

Kovach, M. E., Shaffer, M. D. & Peterson, K. M. (1996). A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio cholerae*. *Microbiology* **142**, 2165–2174.

McClelland, M., Sanderson, K., Spieth, J. & 23 other authors (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**, 852–856.

McDaniel, T. & Kaper, J. (1997). A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. *Mol Microbiol* **23**, 399–407.

McShan, W., Tang, Y. & Ferretti, J. (1997). Bacteriophage T12 of *Streptococcus pyogenes* integrates into the gene encoding a serine tRNA. *Mol Microbiol* **23**, 719–728.

O'Shea, Y. A. & Boyd, E. F. (2002). Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated via CP-T1 generalized transduction. *FEMS Microbiol Lett* (in press).

Perna, N. T., Plunkett, G., III, Burland, V. & 25 other authors (2001). Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533.

Prere, M., Chandler, M. & Fayet, O. (1991). Transposition in *Shigella dysenteriae*: isolation and analysis of IS*911*, a new member of the IS*3* group of insertion sequences. *Res Microbiol* **142**, 489–498.

Ramamurthy, T., Garg, S., Sharma, R. & others (1993). Emergence of a novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* **341**, 703–704.

Roggentin, P., Schauer, R., Hoyer, L. L. & Vimr, E. R. (1993). The sialidase superfamily and its spread by horizontal gene transfer. *Mol Microbiol* **9**, 915–921.

Sakurada, K., Ohta, T. & Hasegawa, M. (1992). Cloning, expression and characterization of the *Micromonspora viridifaciens* neuraminidase in *Streptomyces lividans*. *J Bacteriol* **174**, 6896–6903.

Sears, C. L. & Kaper, J. L. (1996). Enteric bacterial toxins: mechanism of action and linkage to intestinal secretion. *Microbiol Rev* **60**, 167–215.

Simpson, A., Reinach, F. C., Arruda, P. & 113 other authors (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature* **406**, 151–157.

Stroeher, U. H., Parasivam, G., Dredge, B. K. & Manning, P. A. (1997). Novel *Vibrio cholerae* O139 genes involved in lipopolysaccharide biosynthesis. *J Bacteriol* **179**, 2740–2747.

Taylor, R. K., Miller, V. L., Furlong, D. B. & Mekalanos, J. J. (1987). Use of *phoA* gene fusions to identify a pilus colonization factor co-ordinately regulated with cholera toxin. *Proc Natl Acad Sci USA* **84**, 2833–2837.

Waldor, M. K. & Mekalanos, J. J. (1994). *Vibrio cholerae* O139 specific gene sequences. *Lancet* **343**, 1366.

Waldor, M. K. & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914.

Wood, M., Jones, M., Watson, P., Hedges, S. & Wallis, T. G. E. (1998). Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. *Mol Microbiol* **29**, 883–891.

Received 22 April 2002; revised 2 July 2002; accepted 30 July 2002.