The *Vibrio* seventh pandemic island-II is a 26·9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43·4 kb genomic island in *V. vulnificus*

Yvonne A. O'Shea, Shirley Finnan, F. Jerry Reen, John P. Morrissey, Fergal O'Gara and E. Fidelma Boyd

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

Correspondence

E. Fidelma Boyd

f.boyd@ucc.ie

Vibrio cholerae is the aetiological agent of the deadly diarrhoeal disease cholera. In this study the 7·5 kb *Vibrio* seventh pandemic island-II (VSP-II) that is unique to *V. cholerae* El Tor and O139 serogroups was analysed and it was found to be part of a novel 26·9 kb genomic island (GEI) encompassing VC0490–VC0516. The low-GC-content VSP-II encompassed 24 predicted ORFs, including DNA repair and methyl-accepting chemotaxis proteins, a group of hypothetical proteins and a bacteriophage-like integrase adjacent to a tRNA gene. Interestingly, *V. cholerae* ORFs VC0493–VC0498, VC0504–VC0510 and VC0516, which encodes an integrase, were homologous to *Vibrio vulnificus* strain YJ016 ORFs VV0510–VV0516, VV0518–VV0525 and VV0560, which also encodes an integrase, respectively. Some ORFs showed amino acid identities greater than 90 % between the two species in these regions. In *V. vulnificus* strain YJ016, a 43·4 kb low-GC-content (43 %) GEI encompassing ORFs VV0509–VV0560 was identified and named *V. vulnificus* island-I (VVI-I). The 52 ORFs of VVI-I included a phosphotransferase system gene cluster, genes required for sugar metabolism and transposase genes. There was synteny and homology between the 5' region of *V. cholerae* VSP-II and the 5' region of *V. vulnificus* VVI-I; however, VVI-I contained an additional 31·5 kb of DNA between VV0526 and VV0560 in strain YJ016. A second *V. vulnificus* strain, CMCP6, did not contain the 43·4 kb VVI-I; in this strain two ORFs were found between the 5' and 3' flanking genes VV10636 and VV10632, showing 100 % identity to VV0508 and VV0561, respectively, which flank VVI-I.

Received 16 March 2004

Revised 16 July 2004

Accepted 5 October 2004

INTRODUCTION

The genus *Vibrio*, which belongs to the Gram-negative γ -*Proteobacteria*, is ubiquitous in marine and estuarine environments. There are 71 recognized species within the genus *Vibrio*, of which 12 species are pathogenic to humans, in the current List of Bacterial Names with Standing in Nomenclature (Euzéby, 1997; Farmer *et al.*, 2003). *Vibrio cholerae* is the most important human pathogen belonging to the genus since it is the aetiological agent of the deadly diarrhoeal disease cholera. Cholera is endemic in many areas of the world, particularly those areas where clean drinking water is absent. *V. cholerae* pathogenic isolates contain two main virulence factors, cholera toxin (CT) and toxin co-regulated pilus (TCP), both of which help facilitate cholera spread and pathogenesis (Sears & Kaper, 1996; Taylor *et al.*, 1987). CT, the main cause of the

explosive diarrhoeal disease, is encoded on a filamentous bacteriophage CTX ϕ (Waldor & Mekalanos, 1996). The receptor for CTX ϕ on the *V. cholerae* cell, TCP, which is an essential intestinal colonization factor, is also encoded on a mobile genetic element named the *Vibrio* pathogenicity island (VPI) (Waldor & Mekalanos, 1996; Kovach *et al.*, 1996; Karaolis *et al.*, 1998; Taylor *et al.*, 1987). To date only two *V. cholerae* serogroups, O1 and O139, are responsible for epidemic and pandemic cholera (Kaper *et al.*, 1995). The O1 serogroup is divided into two biotypes, classical and El Tor. Of the seven recorded pandemics of cholera, the classical biotype was responsible for the sixth pandemic and the El Tor biotype is responsible for the seventh and ongoing pandemic, which began in 1961 (Kaper *et al.*, 1995). The *V. cholerae* O139 serogroup emerged in 1992 as a major cause of epidemic cholera and is related to O1 serogroup strains (Albert *et al.*, 1993; Cholera Working Group, 1993; Waldor & Mekalanos, 1994). Several studies have shown that O139 serogroup strains arose by the acquisition of the O139 biosynthesis genes by an O1 El Tor

Abbreviations: GEI, genomic island; VPI, *Vibrio* pathogenicity island; VSP-II, *Vibrio* seventh pandemic island-II; VVI-I, *Vibrio vulnificus* island-I.

biotype strain, which was probably mediated by a bacteriophage (Berche *et al.*, 1994; Bik *et al.*, 1995, 1996; Waldor & Mekalanos, 1994; Strocher *et al.*, 1997).

The complete genome sequence of *V. cholerae* strain N16961 is available and has been shown to contain two circular chromosomes of 2.96 and 1.07 Mb (Trucksis *et al.*, 1998; Heidelberg *et al.*, 2000). Comparative genomic analysis of *V. cholerae* using a DNA microarray that displayed over 93% of the predicted genes of strain N16961, determined differences in gene content between sixth (classical biotype) and seventh (El Tor biotype) pandemic isolates (Dziejman *et al.*, 2002). These authors identified two genomic regions designated the *Vibrio* seventh pandemic island-I (VSP-I) and VSP-II that were unique to seventh pandemic El Tor isolates (Dziejman *et al.*, 2002). VSP-I and VSP-II showed several characteristics of pathogenicity islands. VSP-I spans a 16 kb region (VC0175–VC0185) encompassing 11 ORFs, which include a deoxycytidylate deaminase-related protein, a transcriptional regulator, a patatin-related protein, a transposase and a number of hypothetical proteins, with a GC content of 40 mol%, in contrast to 47 mol% for the entire genome (Dziejman *et al.*, 2002). The 7.5 kb VSP-II region encompasses eight ORFs (VC0490–VC0497), which encode a transcriptional regulator and a ribonuclease H1; however, the chromosomal boundaries of the region were not clearly defined (Dziejman *et al.*, 2002).

Vibrio vulnificus is a human pathogen that is highly invasive, causing fulminate pulmonary septicaemia, with mortality rates as high as 60%, one of the highest death rates of any food-borne disease (Linkous & Oliver, 1999; Strom & Paranjpye, 2000). In addition to septicaemia, *V. vulnificus* can also cause wound infections and gastroenteritis (Linkous & Oliver, 1999; Strom & Paranjpye, 2000). Similar to *V. cholerae*, *V. vulnificus* is associated with the estuarine environment and occurs in high numbers in molluscan shellfish (DePaola *et al.*, 1994; Kaysner *et al.*, 1987; Oliver *et al.*, 1982; Chiavelli *et al.*, 2001; Huq *et al.*, 1984; Hood & Winter, 1997; Montanari *et al.*, 1999). However, phylogenetic analysis based on 16S rRNA and *hsp60* gene sequences indicate that *V. vulnificus* is more closely related to *Vibrio parahaemolyticus* than to *V. cholerae* (Kwok *et al.*, 2002). A number of potential virulence factors have been identified in *V. vulnificus* clinical isolates, but these are also found associated with oyster isolates (DePaola *et al.*, 2003). The *V. vulnificus* strain YJ016 genome sequence has been published recently (Chen *et al.*, 2003). *V. vulnificus* strain YJ016 is a biotype 1 hospital isolate from Taiwan. The genome is 5.3 Mb, consisting of two circular chromosomes of 3.4 Mb (3262 ORFs) and 1.9 Mb (1697 ORFs), respectively, and a 49 kb plasmid (Chen *et al.*, 2003). Of the 4959 genes identified, 1688 (34%) encode hypothetical proteins, which accounts for most of the genes that are unique to the *V. vulnificus* genome.

In this paper, we show that the 7.5 kb VSP-II region (VC0490–VC0497) encompasses a 26.9 kb region

(VC0490–VC0516) in *V. cholerae* biotype El Tor and O139 serogroup isolates.

The 5' region of *V. cholerae* VSP-II, ORFs VC0493–VC0498 and VC0504–VC0510, shows homology to a region in *V. vulnificus* strain YJ016, ORFs VV0510–VV0516 and VV0518–VV0525, respectively. We determine that in *V. vulnificus* strain YJ016, ORFs VV0510–VV0525 are part of a 43.4 kb genomic island (GEI) (VV0509–VV0560) named *Vibrio vulnificus* island-I (VVI-I) which encompasses a number of transport and sugar metabolism genes.

METHODS

Bacterial isolates. A total of 21 *V. cholerae* isolates were examined in this study (Table 1). The 21 *V. cholerae* isolates belonged to 4 different serogroups; 17 isolates belonged to serogroup O1, of which 2 isolates were of the classical biotype and 15 isolates were of the El Tor biotype. The O139 serogroup was represented by two isolates, the O37 serogroup by one isolate and the O141 serogroup by one isolate. Of the 21 strains examined, 15 were clinical isolates and 2 were environmental isolates (Table 1), and 11 contained the 7.5 kb VSP-II region (Table 1) (Dziejman *et al.*, 2002; O'Shea *et al.*, 2004). All *Vibrio* strains were grown in Luria–Bertani broth (LB) and stored at -70°C in LB broth with 20% (v/v) glycerol.

DNA isolation. Chromosomal DNA was extracted from each *V. cholerae* isolate using the G-nome DNA isolation kit from Bio 101. Briefly, a single colony of each isolate was inoculated into 3 ml LB broth and incubated overnight at 37°C with shaking at 150 r.p.m. The bacterial cells were pelleted at 3000 r.p.m. for 5 min, supernatant was discarded and the pellet was brought to a final volume of 1.85 ml in cell suspension solution. The cells were lysed and treated with RNase and protease. DNA was extracted with ethanol and resuspended in TE buffer.

PCR analysis. PCR was used to assay 21 *V. cholerae* isolates for the presence of VSP-II. PCR analysis was carried out using 10 primer pairs (Table 2), the location of which can be seen in Fig. 1. Gene fragments were amplified from chromosomal DNA isolated from the 21 *V. cholerae* strains. PCR was performed in a 20 μl reaction mixture by using the following cycles: an initial denaturation step at 96°C for 1 min followed by 30 cycles of denaturation at 94°C for 30 s, 30 s of primer annealing at (49.6 – 58°C) and 1–4 min of primer extension at 72°C .

Southern hybridization analysis. To confirm negative PCR results, Southern hybridization analysis was carried out with four DNA probes (490, 493, 502, 514) generated by PCR using the reference *V. cholerae* strain N16961 as template (Fig. 1). The 2.3, 3.6, 0.5 and 2.1 kb PCR products were purified using the ConceRT-PCR purification kit (Gibco-BRL) and 100 ng probe DNA was conjugated to horseradish peroxidase using the ECL direct nucleic acid labelling system (Amersham Pharmacia Biotech). DNA from each strain of interest was digested with the restriction enzyme *EcoRI* (Roche) and separated by electrophoresis in 0.6% (w/v) $1\times$ TBE agarose. Separated DNA fragments were transferred to a nitrocellulose membrane for Southern hybridization. Southern hybridization was carried out using the ECL direct nucleic acid labelling and detection system according to manufacturer's instructions (Amersham Pharmacia Biotech).

Nucleotide sequence and bioinformatic analysis. A region spanning 26.9 kb from position 523 156 to 550 021 of the *V. cholerae* genome from strain N16961 and a 43.4 kb region spanning 498 755 to 542 138 of the *V. vulnificus* genome of strain YJ016 was analysed

Table 1. Strains used in this study

<i>V. cholerae</i> strain	Serogroup (biotype)	VSP-II	Source	Country of isolation	Year of isolation
Sixth pandemic classical					
O395	O1(classical)	—	Clinical	India	1964
569B	O1(classical)	—	Clinical	India	1948
Seventh pandemic El Tor					
GP33	O1(El Tor)	+	Unknown	USSR	1971
GP43	O1(El Tor)	—	Unknown	Australia	1972
GP155	O1(El Tor)	—	Unknown	Australia	1979
N16961	O1(El Tor)	+	Clinical	Bangladesh	1975
3038	O1(El Tor)	+	Clinical	Vietnam	Unknown
F1873	O1(El Tor)	+	Clinical	Rwanda	1994
F1875	O1(El Tor)	+	Clinical	Goma, Zaire	199
F1939	O1(El Tor)	+	Clinical	Rwanda	1994
SM115	O1(El Tor)	+	Clinical	Bahrain	1978
2125-98	O1(El Tor)	+	Clinical	Bangladesh	1998
2164-78	O1(El Tor)	—	Environmental	US Gulf Coast	1978
Pre-seventh pandemic El Tor					
RV79	O1(El Tor)	—	Clinical	Indonesia	1937
C5	O1(El Tor)	+	Unknown	Indonesia	1957
Non-toxicogenic El Tor					
2740-80	O1(El Tor)	—	Environmental	USA Gulf Coast	1980
468-83	O1(El Tor)	—	Clinical	USA	1983
O139 serogroup					
MO2	O139	+	Clinical	Madras, India	1992
MO10	O139	+	Clinical	Madras, India	1992
Non-O1 and non-O139 serogroup					
V46	O141	—	Clinical	USA	1978
V52	O37	—	Clinical	Sudan	1968

for sequence similarities using the Basic Local Alignment Search Tool (BLAST) at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/COG/>). Bioinformatic analysis was performed using the software Artemis (Rutherford *et al.*, 2000).

RESULTS

A 7.5 kb *V. cholerae* island VSP-II (ORFs VC0490–VC0497) that was confined to *V. cholerae* El Tor isolates and O139 serogroup isolates was described by Dziejman *et al.* (2002). Two genes 18.2 kb downstream of VSP-II, VC0514 and VC0516, were also identified as being present only in seventh pandemic El Tor O1 and O139 strains (Dziejman *et al.*, 2002). Interestingly, ORF VC0516 encodes a bacteriophage-like integrase and was inserted adjacent to a tRNA gene (Fig. 1). To determine whether the original VSP-II region ORFs VC0490–VC0497 were linked to VC0516, we carried out PCR analysis with a set of 10 primer pairs that spanned the entire region from VC0489 (on the core chromosome), which marked the 5' flanking region, to VC0517 (on the core chromosome), which marked the 3' flanking region (Fig. 1 and Table 2). First, we investigated whether the region directly downstream of VC0497 was present in a range of *V. cholerae* El Tor VSP-II-positive and -negative strains (Dziejman *et al.*, 2002; O'Shea *et al.*, 2004). With primer pair VC0502F/VC0502R

V. cholerae strains C5, 3038, F1873, F1875, F1939, 2125-98, GP33 and N16961, which are VPI-II-positive, gave an expected 0.5 kb PCR band (Fig. 2). The remaining six *V. cholerae* isolates, which are VSP-II-negative, gave no PCR product (Fig. 2). Next, we examined the distribution of VC0516 (integrase) among *V. cholerae* VSP-II-positive and -negative O1 and non-O1 serogroup isolates by PCR analysis with primer pair VC0516F/VC0516R. All four of the VSP-II-positive *V. cholerae* El Tor and O139 strains (N16961, SM115, MO2 and MO10) tested gave an expected 965 bp band, indicating the presence of VC0516 in these strains. In contrast, PCR analysis of five *V. cholerae* VSP-II-negative strains (569B, O395, V46, V52 and 468-83) failed to amplify a PCR band. Thus, both VC0502 and VC0516 only appear to be associated with strains that contained the previously described 7.5 kb VSP-II region (VC0490–VC0497).

To determine whether the region between VC0497 and VC0516 was confined to VSP-II-positive isolates and absent amongst VSP-II-negative *V. cholerae* isolates, we carried out PCR analysis with four primer pairs that spanned this region (Table 2, Fig. 1, and Fig. 3). Primer pair VC0498F/VC0498R gave an expected 4.1 kb PCR product band from the four VSP-II-positive *V. cholerae* isolates (N16961, SM115, MO2 and MO10), whereas no

Table 2. PCR primers used in this study encompassing VC0489–VC0517

Primer name	Sequence (5'–3')	Predicted PCR product size (bp)	Annealing temp. (°C)
VC0489F	AGATCAACTACGATCAAGCC	1677	54.2
VC0489R	AGAGGTTGCTATGCTTTCTAC		
VC0490F	CGTGAAGGGATATAGGAG	2337	49.6
VC0490R	TGCAGTTGTTGAATGGAC		
VC0493F	AATGCTTCTCAGGGGGTCTT	3600	57
VC0493R	CGCTCTTCTTTCCACGCTTCA		
VC0498F	AGGTGGTATCGGGCTGGT	4140	58
VC0498R	TGCGGCTGGAATGGAGTCTG		
VC0502F	TCATCAGTTAGCACACGAAC	476	52
VC0502R	GCTATCGTTATACTTGGCG		
VC0504F	CAGCAAAGGCGGAAGAGGTAG	3240	55
VC0504R	AGCCCCGAAATGAATCCCCAAA		
VC0512F	CAGTGGCTTCGCAGAGGA	3900	53
VC0512R	CCCTCCACTGCTATTCCG		
VC0514F	TTATGATCCAAGGAGTAGGG	2089	52
VC0514R	AGGCTGAAAAACAACCTGAG		
VC0516F	GTTTTCTGCGTTGTTTCGAG	965	52
VC0516R	TCCTGATGTCTCTCTTGCCG		
VC0517F	CCCACTTCTTCCAGAGTATG	1753	54.3
VC0517R	CGCAGTCACAGCTTAAACAAC		

product was amplified from the five *V. cholerae* VSP-II-negative isolates (569B, O395, V46, V52 and 468-83) tested. Similarly, PCR primer pairs VC0504F/VC0504R, VC0512F/VC0512R and VC0514F/VC0514R gave expected 3.2, 3.9 and 2.1 kb PCR product bands with the four VSP-II-positive *V. cholerae* isolates and, as expected, the five VSP-II-negative strains tested gave no PCR product (Fig. 3). The absence of the 19.4 kb region between VC0497 and VC0516 in VSP-II-negative *V. cholerae* isolates was verified by Southern hybridization analysis using two DNA probes (502, 514) derived from PCR fragments generated using primer pairs from Table 2 which spanned the region of interest (Fig. 1). No hybridization fragments were obtained with the DNA probes for 11 *V. cholerae* VSP-II-negative strains and, as expected, a positive hybridization band was obtained for *V. cholerae* VSP-II-positive strains, which contain this region (Fig. 4 and data not shown).

We found by PCR analysis with primer pair VC0489F/VC0489R that chromosomal DNA tested from all *V. cholerae* strains gave a 1.7 kb PCR band, indicating that this region is present in these isolates (Fig. 3). Thus, the VC0489 gene marks the 5' flanking region of VSP-II as shown by Dziejman *et al.* (2002). The gene directly downstream of VC0516 (integrase), VC0517, which encodes an RNA polymerase sigma factor, was assayed by PCR to determine whether it marked the 3' insertion site of the VSP-II region. Primer pair VC0517F/VC0517R was used to amplify a 1.8 kb PCR product from all strains tested (Fig. 3), indicating the presence of this region in all strains. Thus, VC0517 marks the 3' flanking region of VSP-II. By PCR analysis we have shown that VSP-II encompasses a

26.9 kb region, which is confined to *V. cholerae* El Tor and O139 serogroup isolates.

To investigate whether the region between VC0490 and VC0516 is empty or contains unique DNA in *V. cholerae* VSP-II-negative strains, we used primers VC0489F and VC0517R from core chromosome-specific DNA, lying respectively in the left and right chromosomal junction fragments of the 26.9 kb VSP-II. *V. cholerae* strains that lacked the 26.9 kb VSP-II region gave a 4 kb PCR product, while PCR analysis of *V. cholerae* VSP-II-positive strains did not produce a PCR product as expected since these strains contain the 26.9 kb region, which could not be amplified under the PCR conditions employed.

Bioinformatic analysis of the 26.9 kb VSP-II and VVI-I regions

The 26.9 kb region in *V. cholerae* El Tor and O139 isolates contained a total of 24 ORFs (Table 3). The overall GC content of VSP-I I (43 mol%) was much lower than the overall GC content of the *V. cholerae* genome (47 mol%) (Fig. 1). Each of the ORFs was analysed using the BLAST program (Altschul *et al.*, 1997). Of the 24 ORFs identified, three ORFs (VC0490–VC0492) showed homology to three ORFs from *Agrobacterium tumefaciens*. Two ORFs (VC0513 and VC0514) showed homology to methyl-accepting chemotaxis proteins from the sequenced *V. vulnificus* strain YJ016. Of the 24 ORFs identified, 11 ORFs in this region were hypothetical proteins (VC0493–VC0510). Interestingly, VC0493–VC0498 and VC0504–VC0510 showed high sequence homology to a region in the sequenced

Fig. 1. (a) Regional variation in the mean proportion of GC content of the VSP-II region from *V. cholerae* N16961 based on a sliding window of 500 bp for *V. cholerae*. (b) Schematic representation of the organization of the VSP-II region in *V. cholerae* N16961. The position and direction of transcription of the ORFs are indicated by the direction of the open arrows. The black arrowheads above ORFs indicate the position of the PCR primers used in this study. The black bold horizontal lines indicate the position of the DNA probes used in Southern hybridization analyses in this study. The numbers refer to the genetic organization of the genes along the genome (Heidelberg *et al.*, 2000). Genes are pattern-coded according to their function.

V. vulnificus strain YJ016 (Chen *et al.*, 2003) from VV0510–VV0516 and VV0518–VV0525, respectively. Eight ORFs in this region showed over 90 % amino acid sequence

Fig. 2. PCR analysis of gene VC0502, which encodes a type IV pilin on the VSP-II island in biotype El Tor *V. cholerae* isolates. Lanes: 1, 1 kb marker; 2–14, *V. cholerae* strains RV79, C5, 2740-80, 468-83, 3038, F1873, F1875, F1939, 2164-78, 2125-98, GP33, GP43, GP155; 15, negative control; 16, positive control.

identity between the two species. In addition, *V. cholerae* ORF VC0516, which encodes an integrase, showed 93 % homology to VV0560 on the *V. vulnificus* genome. *V. vulnificus* ORF VV0560, similar to VC0516, was adjacent to a tRNA gene identified using Artemis. Next, we examined the region between VV0509 and VV0560 on the *V. vulnificus* genome in strain YJ016. In *V. vulnificus* strain YJ016, a clinical isolate from the blood of a patient with primary septicaemia isolated in Taiwan, 52 ORFs were identified (Fig. 5, Table 4). A second *V. vulnificus* genome sequence has recently become available, *V. vulnificus* strain CMCP6, a clinical isolate from South Korea (Chen *et al.*, 2003; Kim *et al.*, 2003). We examined the equivalent region, VV10636–VV10632, in strain CMCP6. Only two ORFs (VV10634 and VV10635) were present between VV10636 and VV10632 and these ORFs showed no homology to ORFs between VV0509 and VV0560 in strain YJ016 (Table 4, Fig. 5). Therefore, VV0509–VV0560 marked a 43.4 kb GEI in *V. vulnificus*, which we named VVI-I, that was only present in strain YJ016 and absent from strain CMCP6.

Fig. 3. PCR analysis of VSP-II in *V. cholerae* O1 and non-O1 serogroup isolates. (a) Gene VC0489, which indicates the 5' flanking region, is found in all *V. cholerae* isolates examined. (b–d) The presence of genes VC0490, VC0512 and VC0514 encoded on VSP-II were investigated by PCR analysis. (e) The 3' flanking region of gene VC0517. Lanes: 1, 1 kb marker; 2–10, *V. cholerae* strains 569B, 0395, SM115, N16961, MO2, MO10, V46, V52, 468-83. NT, Non-toxicogenic.

Among the 52 ORFs identified between VV0509 and VV0560 in *V. vulnificus* strain YJ016, were VV0510–VV0516 and VV0518–VV0525, which as stated previously showed homology of between 25 and 96% at the amino acid level to *V. cholerae* VC0493–VC0498 and VC0504–VC0510, respectively (Tables 3 and 4). Within *V. vulnificus* ORFs VV0526–VV0560 there were a number of gene clusters; VV0528–VV0530 showed homology to sugar metabolism genes from *Salmonella* Typhimurium (STM3793–STM3791), VV0537–VV0540 showed homology to a phosphotransferase system from *Yersinia pestis* strain CO92 and ORFs VV0543, VV0554 and VV0555 showed homology to a sugar isomerase, sugar kinase and 2-keto-3-deoxy-6-phosphogluconate aldolase, respectively

(Table 4). Six transposase genes were also identified in this region, VV0531, VV0545, VV0546, VV0556, VV0558 and VV0559. Twenty-six ORFs were hypothetical or conserved hypothetical proteins. Similar to the 26.9 kb VSP-II region in *V. cholerae* El Tor and O139 serogroup isolates, the 43.4 kb VVI-I region in *V. vulnificus* strain YJ016 showed aberrant GC contents: 43 mol% as opposed to 46 mol% for the entire *V. vulnificus* genome (Fig. 5). So VVI-I is the first GEI to be identified in *V. vulnificus* strain YJ016.

DISCUSSION

Our study has demonstrated that the 7.5 kb VSP-II region is part of a 26.9 kb island that encompasses ORFs

Fig. 4. Southern hybridization analysis of *V. cholerae* VSP-II negative isolates. Lanes: 1–11, VSP-II-negative isolates confirmed by Southern hybridization by probing for VC0502; 12–13, VSP-II positive controls.

VC0490–VC0516 (Table 3 and Fig. 1). In addition, we identified a 7.9 kb region within VSP-II that was homologous to a region on the *V. vulnificus* strain YJ016 genome. The 7.9 kb region of homology on the *V. cholerae* genome is located in the equivalent position on the *V. vulnificus* genome between a gene encoding a haemolysin (VC0489 and VV0508) and a gene encoding an RNA polymerase sigma factor (VC0517 and VV0561) (Fig. 5). In *V. vulnificus* strain YJ016 we examined the region between ORFs VV0508 and VV0561 and identified a 43.4 kb GEI which we named VVI-I. Zhang & Zhang (2004) examined *V. vulnificus* strain CMCP6 for anomalous sequence features and identified three regions that had GEI characteristics; none of the three GEIs corresponded to VVI-I. We found that VVI-I was absent in *V. vulnificus* strain CMCP6 where the site (VV10636–VV10632) is empty (Fig. 5).

In comparison to other known PAIs and GEIs, *V. cholerae* VSP-II and *V. vulnificus* VVI-I possessed most of the characteristics of laterally acquired genomic regions (Dobrindt *et al.*, 2004; Hacker & Kaper, 2000). VSP-II

Table 3. The 24 ORFs encompassing the 26.9 kb VSP-II island in *V. cholerae* N16961

<i>V. cholerae</i> ORF	Length (aa)	Homologue*	Strain	Gene designation	Amino acid identity (%)
VC0489	586	Haemolysin	<i>V. parahaemolyticus</i> RIMD 2210633	VP0372	68
			<i>V. vulnificus</i> YJ016	VV0508	67
			<i>V. vulnificus</i> CMCP6	VV106	36–67
VC0490	653	Yme	<i>A. tumefaciens</i> C58	ATYme	34
VC0491	178	Ymf	<i>A. tumefaciens</i> C58	ATYmf	24
VC0492	388	Ymg	<i>A. tumefaciens</i> C58	ATYmf	22
VC0493	291	HP	<i>V. vulnificus</i> YJ016	VV0510	25
VC0494	214	CHP	<i>V. vulnificus</i> YJ016	VV0511	39
VC0495	224	CHP	<i>V. vulnificus</i> YJ016	VV0512	91
VC0496	199	HP	<i>V. vulnificus</i> YJ016	VV051	72
VC0497	66	TR	<i>V. vulnificus</i> YJ016	VV0515	92
VC0498	146	Ribonuclease	<i>V. vulnificus</i> YJ016	VV0516	95
VC0502	174	Type IV pilin	<i>V. vulnificus</i> YJ016	VV2940	35
VC0503	426	CHP	<i>V. vulnificus</i> YJ016	VV21633	50
VC0504	75	HP	<i>V. vulnificus</i> YJ016	VV0518	86
VC0505	122	HP	<i>V. vulnificus</i> YJ016	VV0519	77
VC0506	244	HP	<i>V. vulnificus</i> YJ016	VV0520	91–96
VC0507	58	HP	<i>V. vulnificus</i> YJ016	VV0522	96
VC0508	147	HP	<i>V. vulnificus</i> YJ016	VV0523	94
VC0509	147	HP	<i>V. vulnificus</i> YJ016	VV0524	93
VC0510	157	RadC	<i>V. vulnificus</i> YJ016	VV0525	96
VC0511	39	HP	No hits	–	–
VC0512	529	MACP	<i>E. coli</i>	ECORF1	65
VC0513	271	AraC	<i>V. vulnificus</i> YJ016	VV2233	31
VC0514	626	MACP	<i>V. vulnificus</i> YJ016	VV2386	46
VC0515	410	CHP	<i>V. vulnificus</i> YJ016	VV11139	34
VC0516	413	Int	<i>V. vulnificus</i> YJ016	VV0560	93
VC0517	625	RpoD	<i>V. vulnificus</i> YJ016	VV0561	89

*HP, Hypothetical protein; CHP, conserved hypothetical protein; TR, transcriptional regulator; MACP, methyl-accepting chemotaxis protein; Int, phage integrase.

Fig. 5. (a) Regional variation in the mean proportion of GC content of VVI-I from *V. vulnificus* strain YJ016 based on a sliding window of 500 bp (Chen *et al.*, 2003). (b–d) Schematic representation of the organization of VSP-II from *V. vulnificus* strain YJ016 (b), *V. cholerae* strain N16961 (c) and *V. vulnificus* strain CMCP6 (d). The position and direction of transcription of the ORFs are indicated by the direction of the arrows. The numbers refer to the genetic organization of the genes. Grey bars link homologous ORFs. Genes are pattern-coded according to their function.

and VVI-I have an atypical GC content of 43 mol% compared with 47 mol% for the entire *V. cholerae* genome and 46 mol% for *V. vulnificus*, they are inserted adjacent to a tRNA gene and each island contains a gene encoding a bacteriophage-like integrase (VC0516 and VV0560) (Fig. 5). The *V. cholerae* VSP-II region encodes transcriptional regulators, methyl-accepting chemotaxis proteins, a putative ribonuclease, a putative type IV pilin, a DNA repair protein, an integrase and a number of hypothetical proteins (Table 3, Fig. 1). The *V. vulnificus* VVI-I region encompasses genes homologous to VC0493–VC0510, with eight ORFs showing greater than 90% amino acid identity (Table 3). This is significantly higher than the overall amino acid sequence similarity between the two species: approximately 80% based on the housekeeping gene *mdh*. In addition VVI-I encodes a PTS mannose/fructose gene cluster, a number of sugar metabolism genes, an integrase and several transposase genes (Table 4). The role of these proteins in *V. cholerae* and *V. vulnificus* is unknown, but it

is expected that GEIs contribute in some way to the fitness and survival of these species. The ecological niche of both *V. cholerae* and *V. vulnificus* is the aquatic ecosystem where conditions such as nutrient availability, temperature and pH are continuously changing. In this setting GEIs offer an opportunity for the organism to acquire a unique set of genes that may increase the chances of survival. For example, the acquisition of the ability to transport and metabolize additional sugars in the bacterial cell is a single-step gain of function event that could increase the fitness of a particular bacterium in a nutrient-limited environment.

VVI-I of *V. vulnificus* strain YJ016 appears to consist of three distinct regions. Region I is comprised of VV0510–VV0530 and is flanked by a transposase, VV0531. Region II from VV0532 to VV0544 is flanked by two transposase genes, VV0545 and VV0546. Region III from VV0547 to VV0557 is homologous to a sugar metabolism gene cluster on chromosome II of *V. vulnificus* strain CMCP6

Table 4. The 52 ORFs encompassing the 43.4 kb VVI-I island in *V. vulnificus* YJ016

<i>V. vulnificus</i> ORF	Length (aa)	Homologue*	Strain	Gene designation	Amino acid identity (%)
VV0508	586	Haemolysin	<i>V. vulnificus</i> CMCP6	VV10636	99
VV0509	239	CHP	<i>V. vulnificus</i> CMCP6	VV10634	98
VV0510–16	1213	Homologous to ORFs in <i>V. cholerae</i>		VC0493–VC0498	25–95
VV0517	414	HP	No hits	–	–
VV0518–25	922	Homologous to ORFs in <i>V. cholerae</i>		VC0504–VC0510	77–96
VV0526	306	TR	<i>V. cholerae</i> N16961	VC0486	22
VV0527	220	DERA	<i>V. vulnificus</i> CMCP6	VC11725	36
			<i>V. vulnificus</i> YJ016	VC2679	36
VV0528	310	Sugar kinase	<i>S. typhimurium</i> LT2	STM3793	72
VV0529	443	Fucose permease	<i>S. typhimurium</i> LT2	STM3792	75
VV0530	348	Cytoplasmic protein	<i>S. typhimurium</i> LT2	STM3791	61
VV0531	353	Transposase	<i>V. cholerae</i> N16961	VC0282	84
VV0532	901	HP	No hits	–	–
VV0533	797	CHP	<i>V. vulnificus</i> YJ016	VC0534	46
VV0534	807	CHP	<i>V. vulnificus</i> YJ016	VC0533	46
VV0535	314	HP	<i>E. coli</i> CFT073	EC0325	26
VV0536	273	Sugar regulator	<i>V. vulnificus</i> YJ016	VV1518	25
VV0537	163	PTS system	<i>Y. pestis</i> CO92	YP0834	45
VV0538	258	PTS system	<i>Y. pestis</i> CO92	YP0835	46
VV0539	295	PTS system	<i>Vibrio furnissii</i> SR1514	VFmazZ	41
VV0540	147	PTS system	<i>Y. pestis</i> CO92	YP3222	52
VV0541	347	CHP	<i>Y. pestis</i> CO92	YP0840	70
VV0542	241	EP	<i>Thermoanaerobacterium thermosulfurigenes</i> EM1	TTxynA	28
VV0543	214	Sugar isomerase	<i>V. parahaemolyticus</i> RIMD 2210633	VP0080	68
VV0544	496	CHP	<i>Y. pestis</i> CO92	YP0843	59
VV0545	77	Transposase	<i>V. cholerae</i> 1322-69	VCtnp	96
VV0546	296	Transposase	<i>Pseudomonas syringae</i> pv. Tomato	PStnpA	59
VV0547	53	HP	No hits	–	–
VV0548	171	CHP	<i>Y. pestis</i> CO92	YP1874	37
VV0549	126	CHP	<i>Rhizobium</i> sp. NGR234	Y4J0	52
VV0550	270	Regulator	<i>V. vulnificus</i> CMCP6	VV20915	79
VV0551	49	HP	No hits	–	–
VV0552	254	Dehydrogenase	<i>V. vulnificus</i> CMCP6	VV20914	88
VV0553	186	CHP	<i>V. vulnificus</i> CMCP6	VV20913	59
VV0554	311	Sugar kinase	<i>V. vulnificus</i> CMCP6	VV20905	65
VV0555	208	2K3D6P	<i>V. vulnificus</i> CMCP6	VV20904	84
VV0556	147	Tn5 transposase	<i>E. coli</i>	ECtnpA	57
VV0557	57	HP	<i>V. vulnificus</i> YJ016	VV1358	80
			<i>V. vulnificus</i> CMCP6	VV12912	80
			<i>V. parahaemolyticus</i> RIMD 2210633	VP1145	80
VV0558	92	Transposase A	<i>Y. pestis</i> CO92	YPIInsA	55
VV0559	142	Transposase B	<i>Y. pestis</i> CO92	YPIInsB	61
VV0560	410	Int	<i>V. cholerae</i> N16961	VC0516	93
VV0561	644	RpoD	<i>V. vulnificus</i> CMCP6	VV12912	100
			<i>V. cholerae</i> N16961	VC 0517	89

*CHP, conserved hypothetical protein; HP, hypothetical protein, TR, transcriptional regulator; DERA, deoxyribose-phosphate aldolase; PTS, phosphotransferase system; EP, endoxylanase precursor; 2K3D6P, 2-keto-3-deoxy-6-phosphogluconate; Int, phage integrase.

(VV20914–VV20904), with amino acid identities of between 59 and 89 % (Table 4). This region also showed homology to a gene cluster in *V. parahaemolyticus* VP0073–VP0076 on chromosome 1 (Makino *et al.*, 2003). Region III was

flanked on each end by two IS elements, one adjacent to VV0560 (integrase) and the other flanking region II. Furthermore, if one examined the GC content across the VVI-I region, differences between the regions can be

observed (Fig. 5). Thus, given the large number of transposase genes in VVI-I, it appears that different parts of the island may have been acquired at different times and from different sources. The extensive mosaic structure of PAIs has been revealed in a number of pathogenic bacteria (Ko *et al.*, 2003; Welch *et al.*, 2002; Dobrindt *et al.*, 2002). It was suggested that VPI-2 (Jermyn & Boyd, 2002), which is unique to *V. cholerae* O1 serogroup isolates, is a mosaic, since the 5' region showed homology to a type 1 restriction modification system, the central region showed homology to amino sugar utilization genes and the 3' region had homology to numerous phage-like genes. Only the 3' region of the island was present in *V. cholerae* serogroup O139 isolates, indicating deletion of most of VPI-2 from these isolates (Jermyn & Boyd, 2002). Indeed, it has been shown recently that in *Vibrio mimicus* isolates, only ORFs VC1773–VC1784 of VPI-2 are present, suggesting that this region is highly unstable (Jermyn & Boyd, 2005).

To date three *V. vulnificus* biotypes have been described; biotype 1 is most frequently associated with pathogenicity in humans, whereas biotype 2 isolates cause disease in humans and eels, and biotype 3 has only been isolated in Israel from patients who handled St Peter's fish. Population genetic studies have identified a distinct eel-pathogenic clone; however, no clones showed any association with human infection (Gutacker *et al.*, 2003). The significance of VVI-I to *V. vulnificus* virulence and survival needs to be examined, first by determining the occurrence of the island among isolates and examining the distribution pattern of the region among the various biotypes and genotypes.

ACKNOWLEDGEMENTS

Research in E. F. B.'s laboratory is funded under the National Development Plan by the Higher Education Authority PRTL1-3 grant and Enterprise Ireland basic research grants. Research in J.M.'s and F.O.G.'s groups is supported by grants from the European Union (QLK3-2000-31759, QLK3-2001-00101 and QLK5-CT-2002-00914), the HEA PRTL1 programmes, Enterprise Ireland (SC/02/0520 and SC/02/517) and the Science Foundation of Ireland (02/IN.1/B1261).

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.
- 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.
- Chen, C. Y., Wu, K. M., Chang, Y. C. & 12 other authors (2003). Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res* **13**, 2577–2587.
- Chiavelli, D. A., Marsh, J. W. & Taylor, R. K. (2001). The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl Environ Microbiol* **67**, 3220–3225.
- Cholera Working Group (1993). Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* non-O139 Bengal. *Lancet* **342**, 387–390.
- DePaola, A., Capers, G. M. & Alexander, D. (1994). Densities of *Vibrio vulnificus* in the intestines of fish from the US gulf coast. *Appl Environ Microbiol* **60**, 984–988.
- DePaola, A., Nordstrom, J. L., Dalsgaard, A., Forslund, A., Oliver, J., Bates, T., Bourdage, K. L. & Gulig, P. A. (2003). Analysis of *Vibrio vulnificus* from market oysters and septicemia cases for virulence markers. *Appl Environ Microbiol* **69**, 4006–4011.
- Dobrindt, U., Blum-Oehler, G., Nagy, G., Schneider, G., Johann, A., Gottschalk, G. & Hacker, J. (2002). Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. *Infect Immun* **70**, 6365–6372.
- Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**, 414–424.
- 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.
- Euzéby, J. P. (1997). List of bacterial names with standing in nomenclature: a folder available on the internet. *Int J Syst Bacteriol* **47**, 590–592 (<http://www.bacterio.cict.fr/>).
- Farmer, J. J., III, Janda, J. M. & Birkhead, K. (2003). *Vibrio*. In *Manual of Clinical Microbiology*, 8th edn, pp. 706–718. Edited by P. R. Murray, E. J. Baron, M. A. Pfaller, J. H. Tenover & R. H. Tenover. Washington, DC: American Society for Microbiology.
- Gutacker, M., Conza, N., Bengali, C., Pedroli, A., Bernasconi, M. V., Permin, L., Aznar, R. & Piffaretti, J. C. (2003). Population genetics of *Vibrio vulnificus*: identification of two divisions and a distinct eel-pathogenic clone. *Appl Environ Microbiol* **69**, 3203–3212.
- Hacker, J. & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641–679.
- 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.
- Hood, M. A. & Winter, P. A. (1997). Attachment of *Vibrio cholerae* under various environmental conditions and to selected substrates. *FEMS Microbiol Ecol* **22**, 215–223.
- Huq, A., West, P. A., Small, E. B., Huq, M. I. & Colwell, R. R. (1984). Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Appl Environ Microbiol* **48**, 420–424.
- Jermyn, W. S. & Boyd, E. F. (2002). Characterization of a novel *Vibrio* pathogenicity island (VPI-2) encoding neuraminidase (*nanH*) among toxigenic *Vibrio cholerae* isolates. *Microbiology* **148**, 3681–3693.

- Jermyn, W. S. & Boyd, E. F. (2005). Molecular evolution of *Vibrio* pathogenicity island (VPI-2): mosaic structure among *Vibrio cholerae* and *V. mimicus* isolates. *Microbiology* 151 (in press).
- 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.
- Kaysner, C. A., Tamplin, M. L., Wekell, M. W., Scott, R. F. & Colburn, K. G. (1987). Virulent strains of *Vibrio vulnificus* isolated from estuaries of the United States west coast. *Appl Environ Microbiol* 53, 1349–1351.
- Kim, Y. R., Lee, S. E., Kim, C. M. & 9 other authors (2003). Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect Immun* 71, 5461–5471.
- Ko, K. S., Lee, H. K., Park, M. Y. & Kook, Y. H. (2003). Mosaic structure of pathogenicity islands in *Legionella pneumophila*. *J Mol Evol* 57, 63–72.
- 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.
- Kwok, A. Y., Wilson, J. T., Coulthart, M., Ng, L. K., Mutharia, L. & Chow, A. W. (2002). Phylogenetic study and identification of human pathogenic *Vibrio* species based on partial hsp60 gene sequences. *Can J Microbiol* 48, 903–910.
- Linkous, D. A. & Oliver, J. D. (1999). Pathogenesis of *Vibrio vulnificus*. *FEMS Microbiol Lett* 174, 207–214.
- Makino, K., Oshima, K., Kurokawa, K. & 14 other authors (2003). Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet* 361, 743–749.
- Montanari, M. P., Pruzzo, C., Oane, L. & Colwell, R. R. (1999). Vibrios associated with plankton in a coastal zone of the Adriatic Sea (Italy). *FEMS Microbiol Ecol* 29, 241–247.
- Oliver, J. D., Warner, R. A. & Cleland, D. R. (1982). Distribution and ecology of *Vibrio vulnificus* and other lactose-fermenting marine vibrios in coastal waters of the southeastern United States. *Appl Environ Microbiol* 44, 1404–1414.
- O'Shea, Y. A., Reen, F. J., Quirke, A. M. & Boyd, E. F. (2004). Evolutionary genetic analysis of the emergence of epidemic *Vibrio cholerae* isolates based on comparative nucleotide sequence analysis and multilocus virulence gene profiles. *J Clin Microbiol* 42, 4657–4671.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945.
- Sears, C. L. & Kaper, J. L. (1996). Enteric bacterial toxins: mechanism of action and linkage to intestinal secretion. *Microbiol Rev* 60, 167–215.
- 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.
- Strom, M. S. & Paranjpye, R. N. (2000). Epidemiology and pathogenesis of *Vibrio vulnificus*. *Microb Infect* 2, 177–188.
- 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-coordinately regulated with cholera toxin. *Proc Natl Acad Sci U S A* 84, 2833–2837.
- Trucksis, M., Michalski, J., Deng, Y. K. & Kaper, J. B. (1998). The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc Natl Acad Sci U S A* 95, 14464–14469.
- Waldor, M. K. & Mekalanos, J. J. (1994). Emergence of a new cholera pandemic: molecular analysis of virulence determinants in *Vibrio cholerae* O139 and development of a live vaccine prototype. *J Infect Dis* 170, 278–283.
- Waldor, M. K. & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1914.
- Welch, R. A., Burland, V., Plunkett, G., 3rd & 16 other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 99, 17020–17024.
- Zhang, R. & Zhang, C. T. (2004). A systematic method to identify genomic islands and its application in analyzing the genomes of *Corynebacterium glutamicum* and *Vibrio vulnificus* CMCP6 chromosome 1. *Bioinformatics* 20, 612–622.