The Vibrio cholerae O1 chromosomal integron

Christopher A. Clark, Leanne Purins, Pranom Kaewrakon, Tony Focareta and Paul A. Manning†

Author for correspondence: Paul A. Manning. Tel: +1 781 839 4000. Fax: +1 781 839 4500.
e-mail: paul.manning@astrazeneca.com

INTRODUCTION

We have previously described a chromosomal repetitive sequence element, designated VCR for Vibrio cholerae repeat, nine of which were associated with a locus determining mannose-fucose-resistant haemagglutination in V. cholerae O1 (Barker et al., 1994; Franzon et al., 1993). Like other repetitive chromosomal elements, such as REP (Higgins et al., 1982) and ERIC sequences (Hulton et al., 1991), the consensus of the 129 bp VCR has extensive, but imperfect, dyad symmetry. This can potentially form a stem–loop structure with a predicted free energy of \(-59.8\) kcal mol\(^{-1}\) (Barker et al., 1994; Hulton et al., 1991). Southern hybridization of the V. cholerae O1 strain 569B chromosome has suggested that there are between 90 and 100 copies of VCR (Barker et al., 1994). In contrast to other repetitive elements (REP) (Higgins et al., 1982), these were localized by PFGE to a single 200 kb region of the chromosome (Barker et al., 1994; Clark et al., 1997; Manning et al., 1999).

Nucleotide sequencing of the original clone (Barker et al., 1994) revealed that each individual gene, with two minor exceptions, was oriented in the same direction and flanked by directly repeated copies of VCR. Both exceptions encoded small operons. This pattern is reminiscent of antibiotic-resistance gene cassette arrays of integrons, as recognized by Recchia & Hall (1995, 1997) who observed that the outer ends of the VCR consensus are related to the core and inverse core sites of...
59-base elements (59-be) found in integron cassette arrays. Integron is a term applied to a genetic unit encoding an integrase gene intI1, and immediately upstream, an attachment site attI with the core site, G1[TTRRRY, defining its 3’ end. IntI1-mediated site-specific recombination occurs at the nick site between GT residues of the core sites of attI and the 59-be in gene cassettes (reviewed by Collis & Hall, 1995; Recchia & Hall, 1995, 1997). The mobile elements termed gene cassettes, each of which carries a 59-be containing a core site, are defined as the region between two successive core nick sites within an array or as closed circles containing a single 59-be (Collis et al., 1993).

We and others have recently reported the presence in V. cholerae of a new integrase (Clark et al., 1997; Manning et al., 1999; Mazel et al., 1998), homologous to the IntI1 family of integrases. It is divergently transcribed from the genes contained within the VCR region, in the manner of antibiotic-resistant integrons. The VCR-associated region was the first example of such a chromosomal locus containing all the components of an integron and was dubbed the Vibrio cholerae mega-integron (Clark et al., 1997), but will now be referred to as the Vibrio cholerae chromosomal integron (VCCI). The sequence of an intI4 was described along with the presence of VCR cassettes in Vibrio metchnikovii and the ability of a VCR cassette to be integrated into an intI1 integron (Mazel et al., 1998). In this study we present a general means of cloning the ends of this and similar arrays, and define the extent of the region. The extent of the region is defined and further evidence of its chromosomal variation is shown using both Southern blot and DNA sequence analysis in a variety of Vibrio spp. isolates. Together with evidence demonstrating that VCRs act as 59-bes in recombination assays, we show that the region functions as a new class of integron.

METHODS

Bacterial strains. The V. cholerae O1 strains used included 569B (Classical, Inaba), AA14073 (El Tor, Inaba), O17 (El Tor, Ogawa), H1 (El Tor, Ogawa), C5 (El Tor, Ogawa), C31 (El Tor, Ogawa), AA14073 (El Tor, Ogawa), BM69 (El Tor, Inaba). Non-O1 serotype V. cholerae strains include O41, O46, O48, O50, O110, O111, O128, O132, O133, O134, O137, O139, O140, O142, O150, O152 and O155 (M. J. Albert, ICDDR,B, Dhaka, Bangladesh). Other Vibrio species used included Vibrio mimicus strain V800, Vibrio parahaemolyticus strains NCTC 10884 (Kanagawa positive) and NCTC 10885 (Kanagawa negative), and Vibrio anguillarum strains ATCC 43305 and ATCC 43306.

PCR. Amplification using AmpliTaq DNA polymerase (Hoffman-La Roche) was carried out by standard protocols with the oligo-deoxynucleotide primers 343 (5’-GTGGTTDDCAAGGTTTTTGTGTA-3’), 922 (5’-CCCCCCATAGGGCGGTGTCT-3’), 923 (5’-CCCCCCATAGGGCGGTGTATAA-3’), 2355 (5’-TAAGCAGGCCTGTTTAAGGCT-3’) and 2583 (5’-GTAACCCGTAATTTTTCCTTAG-3’). Dye-labelled terminators were used in the sequencing reactions. The accession numbers for the individual sequences are listed in the legend to Fig. 3.

Southern hybridization. DNA from Vibrio strains was prepared as described by Manning et al. (1986). Transfer of DNA from agarose gels to Hybond-N+ filters (Amersham) was performed as described by Southern (1975) with modifications described by Sambrook et al. (1989), except that two stringent washes were performed at 4°C below Tm for all oligo prober. Detection was by enhanced chemiluminescence and autoradiography.

Mating-out assay for site-specific recombination. Donor strains were constructed by introducing all three plasmids at once into Escherichia coli S17-1 by electroporation. These were mated with E. coli DH5α as previously described (Hanssen et al., 1997) and plated on agar plates supplemented with nalidixic acid and chloramphenicol or nalidixic acid and kanamycin.

RESULTS

Detection of VCR in Vibrio isolates by PCR and Southern hybridization

We examined 65 V. cholerae strains belonging to different O serotypes, by using VCR5, the VCR immediately downstream of mraAB (Barker et al., 1994), as a probe in Southern hybridization. All strains revealed multiple reacting DNA fragments totalling in each instance more than 40 kb. A representative selection is shown in Fig. 1. The probe also hybridized to V. mimicus and V. anguillarum DNA, but no reaction was detected with V. parahaemolyticus even after longer exposures. A striking feature of these profiles generated by HindIII digestion is that each has a unique fingerprint.

![Fig. 1. Hybridization of the VCR5 (2) probe to HindIII-digested chromosomal DNA of Vibrio isolates. Non-O1 serotype V. cholerae strains: lanes A, O128; B, O132; C, O133; D, O134; E, O137; F, O139; G, O140; H, O142; I, O150; J, O151; K, O152; L, O155. O1 V. cholerae strains: lanes M, H1; N, O17. Lane O, V. parahaemolyticus NCTC 100844; lane P, V. mimicus V800; V. anguillarum strains: lanes Q, ATCC 43305; lane R, ATCC 43306. Arrows indicate DNA molecular mass marker positions, sizes in kb.](image-url)
However, closely related isolates such as the *V. cholerae* El Tor strains O17 and H1 have a number of bands in common (see Fig. 1). The hybridization pattern is somewhat reminiscent of a partial digest but this is a reflection on the ability of the probe to bind to different fragments determined by both the degree of homology to VCR5 and also the number of VCR copies on the particular fragment. The VCR-containing regions appear to be quite variable.

Each of the strains in the remainder of the *V. cholerae* O1 serotype reference collection (obtained from M. J. Albert, ICDDR,B, Dhaka, Bangladesh) had at least six gene cassettes as determined by the number of discrete bands detected following PCR amplification with oligodeoxynucleotides 922 and 923, which are directed outwards from the conserved ends of VCR (data not shown).

To confirm that the VCR elements are in fact all direct repeats, as with all 59-bes studied so far, single-primer PCR was performed with either oligodeoxynucleotide 922 or 923 using DNA from a variety of O1 and non-O1 serotype strains. The inability to obtain a PCR product with the individual primers suggests that in each case the VCR sequences are present as direct repeats within the arrays.

**Identification of the 5′ restriction fragment flanking the VCR region by Southern hybridization**

The conserved nature of the VCR sequence and the apparent organization into one contiguous region (Barker *et al.*, 1994) provided the means of identifying sequences immediately flanking this region. The VCR sequences contain four 6 bp restriction sites: *Hin*II, *Ksp*I, *Mlu*I and *Bsu*36I, which are conserved in 28, 26, 21 and 28, respectively, of the 32 VCR elements identified and sequenced. Considering that the mean distance between these VCRs is 0.7 kb, the size of these restriction fragments from within the array would be expected on average to be significantly smaller than those from the immediate region flanking the array. This is particularly true of *Ksp*I as it has a G + C content of 100 mol %, whereas that of the *V. cholerae* genome is 47–49 mol % (Baumann *et al.*, 1984). Genomes with a similar G + C content to that of *V. cholerae*, such as *E. coli*, have a mean *Ksp*I restriction fragment size of greater than 20 kb. Thus, the 5′ restriction fragment of the array could be expected to be large and carrying only the 5′ part of the first VCR. Using each enzyme separately, the genomic DNA from a variety of *V. cholerae* non-O1, and El Tor and Classical O1 serotype strains was digested and probed with oligodeoxynucleotide 923, which is derived from the 5′ end of the VCR. In Fig. 1, hybridization of the probe to the 5′ end of the VCR in chromosomal *Ksp*I digests reveals a series of bands, which we have divided into two groups of bands—one corresponding to fragments above 7 kb and the other to fragments below 5 kb. In the range 4.4–0.72 kb there is a ladder of more than 20 discrete bands in each track, but when higher percentage agarose gels were used, numerous bands down to 0.3 kb were observed (data not shown), each probably corresponding to an individual VCR cassette. Throughout the size range there are a number of conserved restriction fragments among the O139 and all of the O1 serotype strains except for C31, which is known to have a major deletion in this region of the chromosome (Barker *et al.*, 1994; Fig. 2).

In the group of larger DNA fragments, probing with oligodeoxynucleotide 923 revealed one or two *Ksp*I restriction fragments between 7.2 and 20 kb (Fig. 2). Since 923 binds 5′ of the *Ksp*I site, only one large fragment was expected, corresponding to the 5′ flanking fragment and the first cassette. Only strains O41, O46, O48, O111 and O134 showed the anticipated pattern. The remainder had a second smaller fragment of 7.2–8.5 kb, except for strain O128. The imperfect dyad symmetry of VCRs, particularly at the ends, suggested the possibility that the probe was binding to both ends. Even though this should have been excluded by stringent washes, the filter was stripped and reprobed with oligodeoxynucleotide 343, which binds 3′ to *Ksp*I within the loop region potentially formed by the VCR. The absence, in most cases, of hybridization to the largest *Ksp*I fragment in each lane to probe 343 (data not shown) suggested that it contains the 5′ end of the array. The second fragment, in most cases, reacted with the probe, indicating it is unlikely to contain the 5′ end. A similar pattern was also seen for the *Bsu*36I Southern blot probed with 923, whereas the discrimination between fragment groups was less clear in several *Mlu*I and most *Hin*II digests (data not shown).

![Image](https://www.microbiologyresearch.org/content/...)}
Cloning of the \textit{V. cholerae} integrase gene \textit{intI4} and defining the class 4 integron

After identifying the potential 5' end of the VCR region, the presence of an integrase immediately upstream of the first was sought. Genomic DNAs of \textit{V. cholerae} strains 569B and H1 were digested with \textit{KspI} and fragments of approximately 20 kb were isolated by agarose-gel purification and further digested with 20 different six-base recognition restriction enzymes found within the poly-linker of pBluescript KS. A bank for each strain was constructed from 20 separate chromosomal digests ligated to pBluescript KS cut with the corresponding enzyme. Screening each ligation by vector-anchored PCR amplification with oligo-deoxynucleotide 923 and pBluescript reverse primer revealed a 1 kb product in the \textit{ClaI} ligations with both 569B and H1, with an additional 2 kb fragment for H1. Both fragments were cloned. Oligo-deoxynucleotide 2583 probe, derived from the 569B 1 kb clone sequence, hybridized to a cloned 2-24 kb \textit{HindIII} fragment from a 2–3 kb 569B subgenomic library. After sequencing, an integrase identical to IntI4 (Mazel et al., 1998) of the \textit{IntI} family was found in a transcriptionally divergent direction to the first VCR cassette in 569B and the first three cassettes in H1. This is characteristic of integrons. The two sequenced VCRs of H1 do not possess \textit{KspI} sites, thus explaining the two amplicons as well as some of the potential 5' array end fragments still hybridizing with oligo-deoxynucleotide 343. Comparison of \textit{intI4} from 569B and H1 showed nucleotide sequence identity from the \textit{ClaI} site within the gene to 218 nt upstream of the \textit{intI4} initiation site. The point of divergence appeared to be the 5' boundary of the first cassette of the array. PCR and sequence analyses of \textit{V. cholerae} O134 and \textit{V. mimicus} using primers 2583 and 2355 were conducted. The DNA of strain O134 was 98% identical up to the same point of divergence in both sequences. The point of divergence from the 5' conserved sequence (5'-CS) in all three \textit{V. cholerae} strains immediately follows 8 bp that conform to the core site consensus (see Fig. 4), thus defining the 3' end of the \textit{attI4} site and thereby defining the class 4 integron of \textit{V. cholerae}. Southern hybridization to the \textit{intI4} probes, oligo-deoxynucleotides 2593 and 2594, shows that the strains surveyed for array 5' flanking fragments can be divided into two groups with a \textit{KspI} fragment of either 20 kb or 13-5 kb in size (Fig. 2). Furthermore, in all but one instance the largest VCR-hybridizing \textit{BstI} fragments contained the 5' end of the array (data not shown).

The region containing the integrase and first gene cassette from \textit{V. mimicus} strain V800 was cloned and sequenced. This revealed a level of nucleotide identity between \textit{V. mimicus} and \textit{V. cholerae} of 75% and 65% in the coding and non-coding regions, respectively, while there was no significant identity with other \textit{intI} family members. The degree of variation in the coding region is sufficient to designate this as a fifth class of integrase, IntI5. Again the point at which homology to the previous \textit{V. cholerae} DNA ends is at a core-like site and is tentatively designated the 3' end of \textit{attI5}.

Mapping the extent of the VCR region in 569B

In addition to the lipocalin-encoding gene \textit{vlpA}, which we have cloned and sequenced (Barker & Manning, 1997), a second copy was identified on a 4.5 kb \textit{BamHI} fragment, which contains an array of seven VCRs. Southern data, probing with oligo-deoxynucleotide 922, indicated the presence of three \textit{KpnI} restriction fragments each with a similar size to the whole bacteriophage lambda genome. A 43 kb fragment was cloned into the cosmid vector cZMXCS (Reilly & Silva, 1993) from a \textit{KpnI} genomic library and screened by PCR with oligo-deoxynucleotide 922 and 923 primers. The 1.3 kb and 6-2 kb \textit{KpnI}/SacI ends of the insert were subcloned and sequenced (Fig. 3). The remainder of the cosmid insert had VCR-specific signals in a number of different digests in a Southern hybridization probed with 922, suggesting a continuous VCR cassette array (data not shown). Southern blots of 569B genomic DNA using both single and double restriction enzyme combinations were probed with oligo-deoxynucleotide 923, then stripped and reprobed with the probes indicated in Fig. 3. Combining these sources, a restriction map of the region showing all the \textit{KpnI} fragments hybridizing with VCRs and the position of sequenced regions was constructed (Fig. 3). The first two \textit{KpnI} fragments could not be linked. The two regions of the cassette array total 120 kb of VCR-containing DNA in the \textit{V. cholerae} 569B chromosome.

Analysis of gene cassettes

Nucleotide sequence analysis and examination of the encoded ORFs showed that the average cassette of 693 nt generally contained very little non-coding DNA in addition to the VCR (Fig. 3). Of the 28 cassettes with an identifiable ORF in the same orientation as the VCR, 20 had insufficient space for a promoter between the 5' end of the VCR and the translational initiation site. Typical of 59-bes (Reilly & Silva, 1993), most ORFs finish within 20 nt of a VCR, and in fact, many use the TAA of the inverse core as the stop codon. We have given ORFs within cassettes that do not have clear database homologues the name \textit{Vco} for \textit{Vibrio} cassette ORF. The associated numbers are the same as previously published ORF numbers where possible; otherwise they have been numbered in the order in which they appeared in the NCBI database. Subsequently identified \textit{vco} homologues with greater than 95% identity are numbered in the order of discovery; for example, \textit{vco13.2}.

An unusual feature is the presence of ORFs transcribed in the opposite direction to that of the array such as the gene pairs \textit{vco3AB}, \textit{vco21AB} and \textit{vco27AB}, as well as \textit{vco24B} which is convergent with \textit{vco24A} (Fig. 3). All of these ORFs have sufficient space for a promoter between the proposed start and the VCR upstream of it. \textit{IS1360} is a new \textit{IS} element with 77% nucleotide identity with \textit{IS5} (accession no. X13668) and was found inserted into the first complete cassette on the cosmid clone C169. \textit{IS1359} apparently disrupts ORF23, the gene after the last cassette (Fig. 3), and is a new member of the IS3 family.
with 64% identity at the nucleotide level to the functional IS911 of Shigella dysenteriae. IS1359 appears to mark the end of the chromosomal VCR array, as the last 4-8 kb of the cosm id is devoid of VCR elements.

mccF1, a new cassette-associated ORF of 344 residues, has 55% identity at the nucleotide level over its full length with the E. coli microcin C7 immunity protein MccF (Gonzalez-Pastor et al., 1995). A second homologue was found to lie on a 7-3 kb BamHI fragment within the second KpnI fragment by probing with mccF1 (data not shown). The vco13 cassette is yet another with homologues: the vco20 cassette is 67% identical. The vlpA and vlpa2 genes were nearly identical.

**DISCUSSION**

*Vibrio cholerae* is a diverse species, as shown by the more than 165 different O serotypes, which also represent temporal, geographical and ecological diversity. PCR and Southern analysis of all these strains (shown here) and work by others on a limited subset (Shangkuan and Southern analysis of all these strains (shown here) predates the speciation of *Vibrio* and work by others on a limited subset (Shangkuan and Southern analysis of all these strains (shown here).

The HindIII and KspI restriction polymorphisms that were detected imply that the VCR array in *Vibrio* is plastic and potentially provides a rapid means of demonstrating whether strains are closely related. The biotype El Tor strains are considered to be a clonal group for some loci (Byun et al., 1999). However, no two strains had identical Southern blot profiles, although most did have many fragments in common. The absence of common fragments in strain C31 confirms the previous suggestions that it has undergone a major
deletion within the region of the chromosome encoding the mrhAB locus (Barker et al., 1994; Van Dongen et al., 1987).

Restriction enzyme sites within the VCR elements also provided a means of identifying the 5'-CS of the region in many strains. The approach presented here for cloning the 5' end of the array may be generally applicable to cloning the ends of other large arrays of highly conserved repetitive elements. Cosmid C169 appears to contain the 3' end of the array since 8.8 kb of VCR-free DNA is located downstream of its last VCR. Furthermore, the sequence up to the end of IS1359 of region E from 569B (Fig. 3) is nearly identical to a similar region devoid of VCR on the chromosome of the El Tor strain N16961 sequenced by The Institute for Genomic Research (http://www.tigr.org) (Fig. 4). The sequence in 569B after IS1359 is not homologous to that of N16961, and preliminary PCR analysis indicates that El Tor and Classical strains differ at this point, suggesting that IS1359 may be a recombinational hotspot. The location of IS elements at the 3' ends of arrays is quite typical of other classes of integrons (Hall, 1995).

The size of the integron array here is defined as approximately 120 kb, consistent with a previous estimate in which strain 569B was estimated to have 90–100 copies of VCR (Barker et al., 1994), suggesting that there could be in excess of 150 gene cassettes, vastly greater than the few genes detected in the average antibiotic-resistance integron. Recent data from the complete genome sequence of V. cholerae El Tor strain N16961 have identified 147 VCR elements within a 147 kb chromosomal integron (Collis et al., 1998). Using a cointegration assay, we have been able to demonstrate that the VCR functions as a 59-be and that it can act as a substrate for site-specific recombination using IntI1 (unpublished data). The presence of IntI1 led to a 35-fold higher recombination rate, which was confirmed by sequence analysis to correspond to site-specific recombination events that are consistent with typical recombination involving a 59-be core, as has been shown previously for a 59-be (Collis et al., 1993; Martinez & de La Cruz, 1990).

The comparison of strains 569B, H1 and O134 implies that the 5'-CS satisfies all of the definitional requirements for the fourth class of integron and also provides the first direct evidence that attI core-site-specific recombination typical of integrons has occurred in V. cholerae. Recently, it has been shown that a VCR cassette can be incorporated into the attI site of the class 1 integron on R388 (Mazel et al., 1998). Using a cointegration assay, we have been able to demonstrate that the VCR functions as a 59-be and that it can act as a substrate for site-specific recombination using IntI1 (unpublished data). The presence of IntI1 led to a 35-fold higher recombination rate, which was confirmed by sequence analysis to correspond to site-specific recombination events that are consistent with typical recombination involving a 59-be core, as has been shown previously for a 59-be (Collis et al., 1993; Martinez & de La Cruz, 1990).

There are three main differences between the VCCI and the plasmid-borne antibiotic-resistant integrons: size, mobility and homogeneity of the 59-be. The expression of the antibiotic-resistance cassettes within integrons derived solely from the Pant promoters located in the 5'-CS (Lesesque et al., 1994) seems to be one of the limitations on their size unless most of the array is silent. However, this arrangement would be inadequate within the VCCI. This is of particular concern since some 59-bes may act as weak terminators (Collis & Hall, 1995) as
has been shown for VCRs within region C (A. Barker, S. G. Williams & P. A. Manning, unpublished). This implies therefore that either large regions of the VCCI are silent or there are numerous internal promoters.

While the presence of promoters within VCRs is an intriguing possibility, this activity, as yet, has not been demonstrated. An alternative is that promoters are provided by some of the cassettes with enough space 5′ of the start codon to encode a promoter, as has been shown for mrhAB (Barker et al., 1994). Clearly, the presence of four cassettes, with ORFs in the opposite orientation with respect to the array, and which has not been seen in other integrons, further indicates the necessity for endogenous promoters, and each of these cassettes has sufficient coding capacity. This ad hoc arrangement might well explain the need for multiple copies of genes, such as in the case of the promoterless vlpA, which is present in up to five copies in some strains (P. Kaewrakon, P. A. Manning and T. Focareta, unpublished).

It has been suggested (Mazel et al., 1998) that the high degree of homology of VCR elements indicates that each of the previously variant 59-be represents a single member of similar vast arrays found in many different bacterial genera. It is indeed likely that there are other chromosomal integrons and that they are an important means of gene capture and chromosomal building. However, there are a number of factors that may contribute to drift in 59-bes outside these integrons. For example, the permanent residency of the integron provides a stable genetic environment for co-evolution with the host not likely to be duplicated, as in the mobile plasmid-associated integrons.

To date, there appears to be a significant difference in the types of genes acquired by the VCCI, with the possible exception of mccF1 should it prove to act as a microcin-immunity protein. This difference is likely to reflect the antibiotic-resistance bias in sampling of the plasmid-based integrons examined, rather than any fundamental differences between the two systems. The pool for gene capture is likely to be the converse. That is, the VCCI samples genes from epismes passing through. For example, the large plasmid carrying mccF (Gonzalez-Pastor et al., 1995) might be the ancestral source for mccF1. In this way the VCCI captures useful genes and eliminates the burden of maintaining the epismes. It is tempting to speculate that the chromosomal integrons, at least within the genus Vibrio, act as a large gene pool for horizontal gene transfer between Vibrios. This could potentially occur via natural transformation of VCR cassette circles or by plasmid-integron intermediums such as that found recently in an El Tor isolate (Fablo et al., 1999).

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of the National Health and Medical Research Council of Australia and Ruth Hall for many invigorating discussions. We would also like to thank Chris Cursaro for his outstanding technical assistance over many years.

REFERENCES


Received 5 January 2000; revised 29 June 2000; accepted 6 July 2000.