

Identification of VCR, a Repeated Sequence Associated with a Locus Encoding a Hemagglutinin in *Vibrio cholerae* O1

ANDREW BARKER, CHRISTOPHER A. CLARK, AND PAUL A. MANNING*

*Microbial Pathogenesis Unit, Department of Microbiology and Immunology,
University of Adelaide, Adelaide, South Australia 5005, Australia*

Received 9 November 1993/Accepted 14 June 1994

We have determined the nucleotide sequence of a 6.3-kb *Bam*HI fragment of the chromosome of *Vibrio cholerae* 569B that includes the sequence of the mannose-fucose-resistant hemagglutinin reported previously (V. L. Franzon, A. Barker, and P. A. Manning, *Infect. Immun.* 61:3032–3037, 1993). This region contains nine copies of a 124-bp direct repeat, here named VCR, of imperfect dyad symmetry, that are shown by Southern hybridization to occur at least 60 to 100 times in the *V. cholerae* O1 chromosome. Large-scale chromosomal mapping suggests that the repeats are confined to about 10% of the chromosome. Related sequences are also found in non-O1 *V. cholerae* but not in other members of the family *Vibrionaceae*. However, VCR is unrelated to other previously described repetitive sequences.

Vibrio cholerae of the O1 serotype is the causative agent of the disease Asiatic cholera. Pathogenesis occurs as a number of sequential, interdependent steps: bacteria are ingested via contaminated food or water (38), colonize the intestinal epithelium, and elaborate a number of extracellular proteins, leading to profuse watery diarrhea and causing high mortality if untreated (34, 35, 45). Since colonization is a prerequisite to establishing a productive infection (48), the factors responsible have been the subject of much research. A number of potential colonization factors, including fimbriae such as TCP (54) and other hemagglutinins (24), have been identified, but to date only TCP has been shown to be a virulence determinant in humans (25). The mannose-fucose-resistant hemagglutinin (MFRHA) (18, 55, 56) has also been implicated as a virulence determinant (17), but its exact role in colonization remains to be determined.

Repetitive sequences have been identified in a wide variety of eubacterial chromosomes (33), the best characterized of these being the REP or PU sequences identified in members of the family *Enterobacteriaceae* (19, 52), with related sequences being detected in diverse bacterial species (57). There is considerable controversy as to whether such sequences have any function: identification of interactions in vitro with a number of nucleoid-associated proteins, such as DNA gyrase, the histone-like protein HU, and DNA polymerase I, have led to proposals that REP sequences might represent a site for the assembly of scaffold proteins on DNA (20, 59, 60). These repeats are similar to the eukaryotic matrix association regions, which contain repetitive sequence elements that bind topoisomerase II (4, 51). However, there are other repetitive sequences also involved in scaffold attachment that do not bind topoisomerase II (41). REP sequences often occur in clusters called REP elements (21, 26), at least 90 of which are found randomly distributed in the *Escherichia coli* chromosome (12). This approximates the number of supercoiled domains in the chromosome (20, 59, 60) and also the number of “high-affinity” gyrase cleavage sites (9). Consequently, it has been suggested that REP sequences are involved in the maintenance of chromosomal structure (20, 59, 60). Alternatively, the

proposal that REP sequences are selfish DNA (26) has not been discounted and remains a possibility (33).

Another role attributed to REP sequences is that they act as the endpoints of RecA-independent recombination events, including large-scale DNA rearrangements and deletions (49). This can have important implications for bacterial virulence, as chromosomally encoded virulence determinants, at least in pathogenic *E. coli*, are located in “pathogenicity islands” rather than being randomly distributed throughout the genome (23). It was also noted that these islands could be deleted, leading to a loss of virulence genes. Although no mechanism has been proposed to mediate this phenomenon, it is worth noting that large-scale DNA rearrangements can be mediated by DNA gyrase in the absence of both extensive sequence homology at the recombinational endpoints and *recA* function (40) and that large deletions whose end points are REP sequences have been characterized in *Salmonella typhimurium* (49). The high degree of genetic instability in *Streptomyces* species (32) and spontaneous deletions leading to loss of the pigmentation phenotype in *Yersinia pestis* (14) are also associated with deletions whose endpoints are repetitive sequences.

To aid our investigation of the role of the MFRHA in the virulence of *V. cholerae*, we have sequenced the region surrounding the corresponding genetic locus. We have shown that this section of the chromosome contains a large number of copies of a 124-bp direct repeat originally detected by van Dongen et al. (56), and we discuss the possible roles that these repeats could have in the expression of closely linked genes and on the structure and evolution of the *V. cholerae* chromosome.

MATERIALS AND METHODS

Bacterial strains and media. The *E. coli* strains used were JM101 ([F⁺ *traD36 proAB*⁺ *lacI*^qZΔM15] *supE thi* Δ[*lac-proAB*] λ⁻) (58) and DH5α [F⁻ φ80*dlacZ*ΔM15 Δ(*lac-argF*)U169 *recA1 endA1 hsdR17*(r_k⁻ m_k⁺) *supE44 thi-1 gyrA96 relA1* λ⁻] (Bethesda Research Laboratories). Other strains used are listed in Tables 1 through 4. Unless otherwise stated, the growth medium used was nutrient broth (Oxoid), except for phage propagation, for which M9 minimal medium and 2xTY broth (39) were used.

DNA sequencing. The nucleotide sequence of the 6.3-kb

* Corresponding author. Phone: 61 8 303 5974. Fax: 61 8 3034362. Electronic mail address: pmanning@microb.adelaide.edu.au.

TABLE 1. Characteristics of *V. cholerae* O1 strains

Strain	Biotype	Serotype	Source (yr of isolation)
569B	Classical	Inaba	India ^a (1946)
CA401	Classical	Inaba	Texas ^b (1953)
CA411	Classical	Ogawa	Texas ^b (1953)
C21	Classical	Ogawa	Holland ^c (1957)
AA14041	Classical	Ogawa	Bangladesh ^d (1985)
Z17561	Classical	Inaba	Bangladesh ^d (1985)
O17	El Tor	Ogawa	India ^a (pre-1965)
C5	El Tor	Ogawa	Holland ^c (1957)
C31	El Tor	Ogawa	Holland ^c (1957)
AA14073	El Tor	Ogawa	Bangladesh ^d (1985)
BM69	El Tor	Inaba	India ^e (1985)

^a K. Bhaskaran, Central Drug Institute, Lucknow, India.
^b J. Berry, University of Texas, Austin.
^c P. Guinée, Rijksinstituut voor Volksgezondheid en Milieuhygiene, Bilthoven, Holland (deceased).
^d B. Kay, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh.
^e S. Pal, National Institute of Cholera and Enteric Diseases, Calcutta, India.

*Bam*HI fragment was determined by subcloning small, overlapping restriction fragments derived from plasmid pPM471 (18) into the phage vectors M13 mp18 and mp19 (60). Sequence was determined with chain terminators (47), Sequenase or Sequenase II enzyme according to the manufacturer's specifications (United States Biochemical Corporation), and an "oligo-walk" strategy with primers generated from partial sequence and manufactured on an Applied Biosystems model 381A DNA synthesizer. When cloned restriction fragments did not overlap, the junction in pPM471 was sequenced by using supercoiled plasmid template (7) with primers binding near the restriction site (not shown). The nucleotide sequence from bp 2727 to 3380 has been reported previously (17). Sequences were compared with entries in the GenBank and EMBL databases with the BLAZE algorithm (6).

Other DNA manipulations. DNA was manipulated for subcloning by standard procedures (46) with restriction endonucleases and other DNA-modifying enzymes purchased from New England Biolabs, Bethesda Research Laboratories, Boehringer Mannheim, or Amersham. For Southern blots, chromosomal DNAs were prepared as described before (36), digested with *Bam*HI or *Hind*III, and run on standard 0.8 or 1% agarose gels, respectively. Samples for pulsed-field gels were prepared by a modification of published procedures (43, 50) as follows. A 20-ml culture of *V. cholerae* in M9 medium (39) was grown at 30°C to an *A*₆₅₀ of 0.6 and incubated with chloramphenicol (final concentration, 180 µg/ml) for 60 min. Cells were then pelleted by centrifugation in a bench centrifuge at 5,000 × *g* for 10 min, washed in 10 ml of 1 M NaCl-10

TABLE 2. *V. cholerae* non-O1 strains

Strain	Isolate type	Source
NCV165	Environmental	India ^a
N34	Human (septicemia)	Australia ^b
N41	Human (traveler's diarrhea)	Australia ^b
N50	River water	Australia ^b
N125	Wallaby	Australia ^b
BV7	Environmental	Maryland ^c
BV22	Environmental	Maryland ^c
BV41	Environmental	Maryland ^c

^a See Table 1, footnote a.
^b P. Desmarchelier, University of Queensland, Brisbane, Australia.
^c M. Voll, University of Maryland, College Park.

TABLE 3. Other *Vibrio* spp. strains

Species	Strain	Features	Source
<i>V. mimicus</i>	V800	None described	Australia ^a
<i>V. fluvialis</i>	V564	None described	Australia ^a
<i>V. parahaemolyticus</i>	NCTC 10884	Kanagawa ⁺	Australia ^b
	NCTC 10885	Kanagawa ⁻	Australia ^b

^a S. Attridge, University of Adelaide, Adelaide, Australia.
^b Institute of Medical and Veterinary Science (Adelaide, Australia) culture collection.

mM Tris-HCl (pH 8.0), resuspended in 4 ml of the same buffer, and warmed to 45°C. Agarose beads were formed by adding 5 ml of 1% LGT agarose (Ultrapure; Bethesda Research Laboratories) in 1 M NaCl-10 mM Tris-HCl (pH 8.0)-20 ml of sterile mineral oil (Primol 352; ESSO), both at 45°C, to the cells, emulsifying the mixture, pouring it into 100 ml of ice-cold 1 M NaCl-10 mM Tris-HCl (pH 8.0), and stirring the mixture for 5 min. Beads were then pelleted by centrifugation at 4°C and 500 × *g* in 50-ml tubes for 10 min, and the mineral oil was removed. The tubes were recentrifuged, and the beads in the pellet were combined in a single tube and then centrifuged as before. The beads were then resuspended in 15 ml of 1-mg/ml proteinase K-1% sodium *n*-lauroyl sarcosine-0.5 M EDTA (pH 9.2) warmed to 65°C and incubated at 65°C for at least 48 h. A lysozyme treatment step was omitted because of excessive nicking of the DNA during this step (1), presumably by secreted DNases (15). After proteinase K digestion, the beads were recovered by centrifugation as before and washed a minimum of five times for a total of 2 h in 10 mM Tris-HCl (pH 7.5)-0.1 mM EDTA at room temperature before storage at 4°C. Restriction endonuclease digestion of agarose beads was performed under standard conditions, and digests were separated on a 1% agarose gel with a Biorad CHEF II apparatus and run conditions of initial A time, 1 min, final A time, 10 min, and 200 V for 18 h at 14°C in 0.5× TBE.

All gels were transferred to nitrocellulose filters by standard conditions and were probed with radioactive or digoxigenin (DIG)-labeled probes as described below, with stringent washes and autoradiography done as described before (46). Stringent washes for filters probed with oligodeoxynucleotide probes were done with tetramethyl ammonium chloride (58). DIG-labeled probes were detected with anti-DIG-horseradish peroxidase conjugate followed by the addition of enhanced chemiluminescence substrate (Amersham) and exposure to X-ray film. Densitometer analysis of autoradiographs was performed on a Molecular Dynamics model 300A computing densitometer. Filters were stripped for reprobings with boiling water and checked for stripping by autoradiography before reprobings. The oligodeoxynucleotide probe used was 5'-CGCGGCTCAATGGGACTGGAAA-3' and was synthesized on an Applied Biosystems model 381A DNA synthesizer. Fragment probes were the 0.7-kb *Cl*aI-*X*baI fragment of

TABLE 4. *Aeromonas* spp. strains^a

Species	Strain
<i>A. hydrophila</i>	A006
<i>A. sobria</i>	A187
<i>A. sobria</i>	A191
<i>A. caviae</i>	V14
<i>Aeromonas</i> sp.	AB1

^a All strains were from M. Atkinson, University of South Australia, Adelaide.

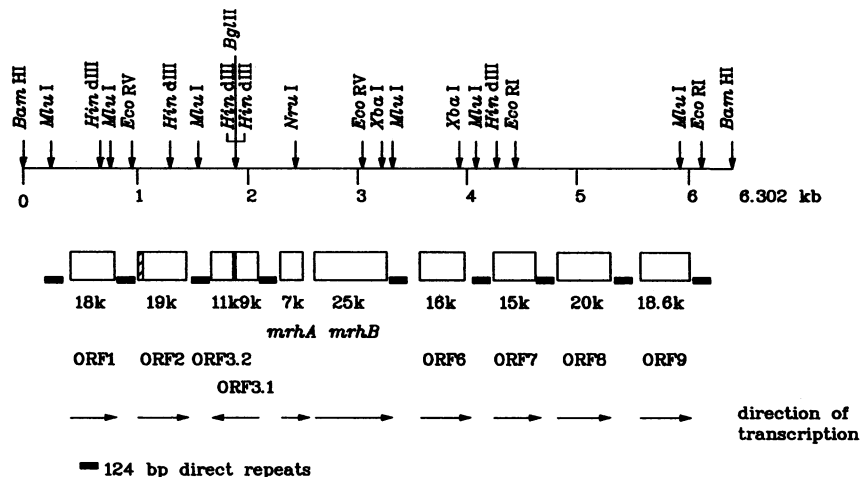


FIG. 1. Genetic organization of a 6.3-kb *Bam*HI fragment of *V. cholerae* 569B. ORFs are represented by open boxes, with the polarity of transcription indicated by arrows, and the approximate predicted molecular weights (in thousands) of the products are shown. The sequence of the MFRHA (*mrhB*) has been reported previously (17) and encompasses bp 2727 to 3380 of this sequence. The locations of nine copies of a 124-bp direct repeat (VCR) are indicated by solid boxes. Some restriction sites in the region are indicated.

plasmid pPM1127 (18), referred to in the text as the *mrhB* probe; the 0.2-kb *Xba*I-*Nsi*I fragment of plasmid pPM1631 (2), referred to in the text as the VCR probe; and the 0.2-kb *Apa*I-*Pst*I fragment of pPM1648 (1), referred to in the text as the ORF2 probe. Probes referred to in the text as to the proximal and distal portions of plasmid pPM471 were the 0.7-kb *Bam*HI-*Hind*III fragment and the 0.15-kb *Eco*RI-*Bam*HI fragments of pPM471 (18), respectively, subcloned into M13mp19 (61). Oligonucleotides were 5' end labeled with [γ - 32 P]ATP (BRESATEC) in a polynucleotide kinase reaction (44), restriction fragments were gel purified and labeled with [α - 32 P]dCTP (BRESATEC) in a random-primed reaction (13), and M13mp19 subclones were labeled by extension from the primer 5'-CACAAATCCACACAAC-3' (61), incorporating 11-DIG-dUTP (Boehringer Mannheim).

Nucleotide sequence accession number. The DNA sequence described in this study is available under GenBank and EMBL accession no. X64097.

RESULTS

Nucleotide sequence and analysis. We initially set out to determine the nucleotide sequence of a 2.4-kb *Hind*III fragment of *V. cholerae* 569B encoding the MFRHA (18). A report of the sequence of part of this region appeared recently (17). However, the presence of three copies of a 124-bp direct repeat within this region (see below) led us to perform Southern hybridizations (not shown), which indicated the presence of multiple copies of this repeat within the original MFRHA clone, pPM471 (18). Therefore, the sequence of the entire insert of pPM471, corresponding to a 6.3-kb *Bam*HI fragment of *V. cholerae* 569B, was determined and is represented schematically in Fig. 1.

Ten open reading frames (ORFs) are apparent within the sequenced region, eight of which are of the same predicted transcriptional polarity (Fig. 1). Two ORFs, encoding 7-kDa and 27-kDa proteins, respectively, are associated with MFRHA activity (17, 56), and we propose the gene names *mrhA* for the ORF encoding the 7-kDa protein and *mrhB* for the ORF encoding the 27-kDa protein (for mannose-fucose-

resistant hemagglutinin). The sequences of all ORFs were compared with entries in the PIR (Release 33) and SWISS-PROT (Release 22) databases for similar entries. Significant matches were obtained for ORF2 (1); ORF3.1, which shows 32% identity to the amino-terminal 72 amino acids of DnaT (primosomal protein i) of *E. coli* (36); ORF3.2, which shows 56% identity with the hypothetical RelE protein of *E. coli* (4); and *mrhB* (1). The alignments for ORF3.1 and ORF3.2 are shown in Fig. 2. Attempts to examine the functions of ORF3.1 and ORF3.2 by engineering mutations in them or by subcloning were unsuccessful (not shown).

A

	X	10	20	30	40	50	60
ORF3.1	MTTRILADVAAS	ITTEFRKANFMKV	ATSFAFGAPVAVL	LNRRNEPAPYCV	PASTYEIMMDK	LEDL	
DnaT	MSSRVLTPDVG	IDALVHDHQT	VLAKEAGGVAV	FANNAPAFYAV	TPA-----	RLA--	
X		10	20	30	40	50	
		70		X			
ORF3.1	ELLAI	AKERLSEDSV	SVNIDDL				
DnaT	ELLALE	-EKLARPGSD	VALDDQ				
		60	70X				

B

	X	10	20	30	40	50	60
ORF3.2	MTYKLEPKKS	SALKKEWK	LAVPLQQQF	KKLIERLEN	PHVPSAKLS	GAENIYKIK	LRSQGY
RelE	MAYFLDF	DERALK	KEWRKLG	STVREQLK	LVLEBSP	RIEANKLR	GMPCYKIKLRSSGY
X		10	20	30	40	50	60
		70	80	90	X		
ORF3.2	RLVYQV	ENDIIVVT	VLA	VGKR	RESEV	YTKAL	QLRD
RelE	RLVYQV	IDKVVVF	VISV	GKRE	SEV	SEAV	KRLL
		70	80	90	X		

FIG. 2. Pairwise alignments of homologous protein sequences. (A) Alignment between ORF3.1 and the amino-terminal 72 amino acids of DnaT (primosomal protein i) of *E. coli*. Full-length DnaT is 179 amino acids long (37). (B) Alignment between ORF3.2 and the hypothetical RelE protein of *E. coli* (3). Alignments were performed with the BLAZE algorithm (6). Identical amino acids are joined with a solid line; similar amino acids are joined with dots. The boundaries of the region of similarity are indicated by an X above and below the sequences.

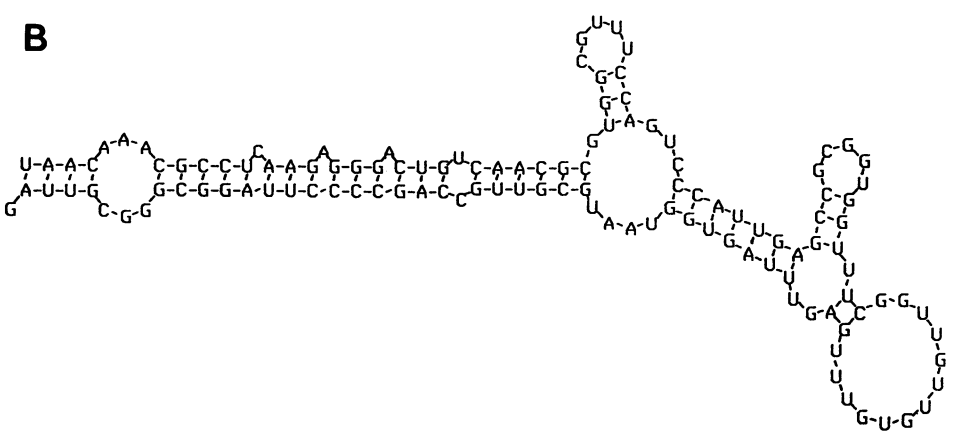
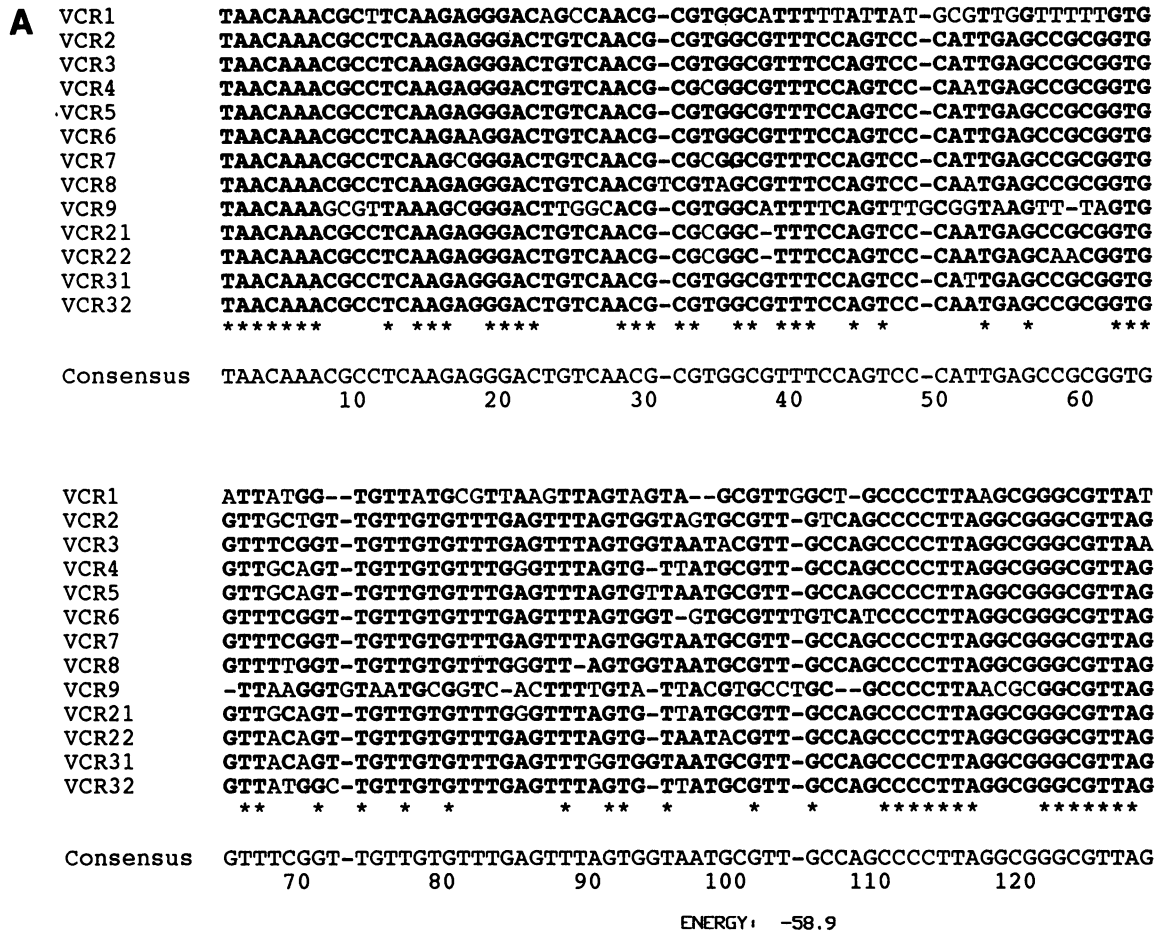


FIG. 3. Nucleotide sequence of the 124-bp element VCR. (A) Alignment of the nine copies of VCR sequenced in this study (numbered from left to right in Fig. 1 in ascending order), as well as two copies (numbered 21 and 22) sequenced by van Dongen et al. (56) and two copies (numbered 31 and 32) sequenced by Takeda et al. (53). A proposed consensus sequence based on this alignment is also shown. The alignment was performed with the multiple-alignment program CLUSTAL (27). Bases identical to the corresponding base in the consensus are in boldface type, and bases identical among all sequenced copies of VCR are indicated by an asterisk. (B) Proposed secondary structure for the consensus sequence, $\Delta G = -58.9$ kcal/mol. Secondary structure and free energy value were determined with the DNASIS package of software.

The most striking feature of the sequence presented is the occurrence of nine 124-bp repeats, all of which are in the same orientation relative to one another, and all of which are found only outside of predicted ORFs, that is, in an extragenic location (Fig. 1). The sequences of these repeats, along with

four repeats sequenced by others (53, 56), show an overall similarity of about 92% (Fig. 3A). Several stretches of dyad symmetry can be identified within the repeats, and the whole repeat itself shows imperfect dyad symmetry, which if transcribed would give a stable secondary structure. In the case of

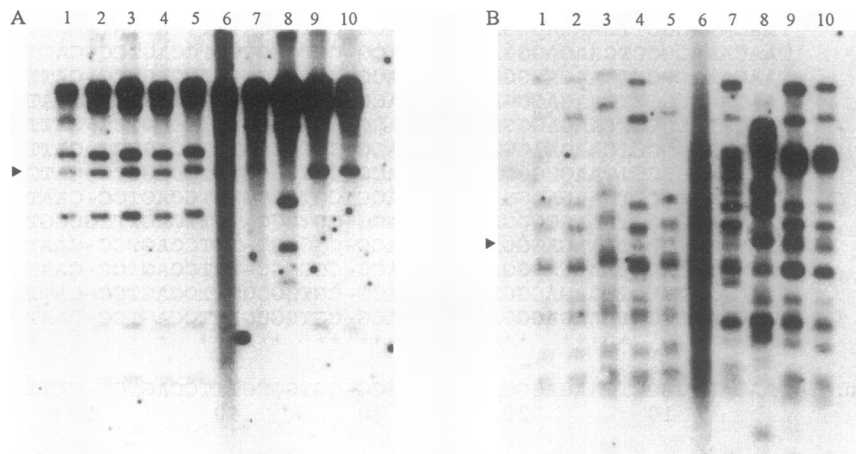


FIG. 4. Hybridization of a VCR-specific oligonucleotide probe to *V. cholerae* O1 chromosomal DNA. (A) *Bam*HI-digested DNA; (B) *Hind*III-digested DNA. *V. cholerae* O1 strains: lanes 1, 569B; lanes 2, CA401; lanes 3, CA411; lanes 4, AA14041; lanes 5, Z17561; lanes 6, O17; lanes 7, C5; lanes 8, C31; lanes 9, AA14073; lanes 10, BM69. A 6.3-kb *Bam*HI fragment (A) and 2.4-kb *Hind*III fragment (B) that correspond to the hybridizing fragments in Fig. 5 are indicated by an arrowhead.

the consensus sequence, a loop of free energy of -58.9 kcal/mol is predicted (Fig. 3B). We propose to name these repeats VCR, for *V. cholerae* repetitive DNA sequence.

Chromosomal distribution of VCR elements. Because of the number of repeats found in the region of the *V. cholerae* chromosome shown in Fig. 1, we investigated the possibility of the occurrence of further copies of VCR by probing 10 different *V. cholerae* O1 strains of both the classical and El Tor biotypes with a VCR-specific oligodeoxynucleotide probe. This test indicated the presence of a large number of copies of VCR (Fig. 4). To aid quantitation of VCR and to ensure that the large number of bands detected were not due to partial digestion of the chromosomal DNAs, the filters were stripped and reprobated with the *mrhB*-specific probe (Fig. 5). Except for strain C31, in which this region is deleted (55) (see below), this probe binds to a single fragment of the size expected from sequence analysis (Fig. 1). Given that the VCR-specific probe used will bind to six of the nine VCR sequences of pPM471 under conditions of high stringency (not shown), an estimate of the number of VCRs in the chromosome was obtained by

quantitating the density of the bands in Fig. 4A (not shown). This gives an estimate of 90 to 100 copies of VCR in strains of the classical biotype and 60 to 80 copies of VCR in strains of the El Tor biotype. The patterns obtained with other probes which could be expected to identify all copies of the repeat were not markedly different, and so these estimates represent a reasonable approximation.

The pattern of hybridization shown in Fig. 4B shows remarkable similarity to the pattern observed when a number of MFRHA-positive cosmid clones were probed with the 2.4-kb *Hind*III fragment indicated in Fig. 5B (16). As this fragment contains three copies of VCR, the cross-hybridization observed (16) was probably due to these copies of VCR, and this result suggests that the copies of VCR observed in the chromosome of *V. cholerae* (Fig. 4) are linked. This possibility was investigated by probing digested DNA separated on a pulsed-field gel with the VCR-specific oligodeoxynucleotide used above (Fig. 6A). Two hybridizing bands are observed for strain 569B, of 285 and 195 kb for *Not*I-digested DNA and of 265 and 205 kb for *Sfi*I-digested DNA. For strain O17,

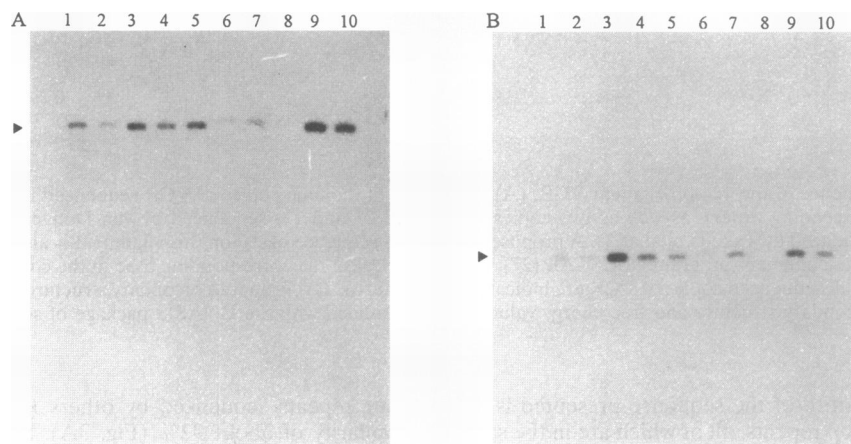


FIG. 5. Hybridization of an *mrhB* (MFRHA)-specific probe to *V. cholerae* O1 chromosomal DNA. (A) *Bam*HI-digested DNA; (B) *Hind*III-digested DNA. Lanes are the same as in Fig. 4. Hybridizing fragments are indicated by an arrowhead.

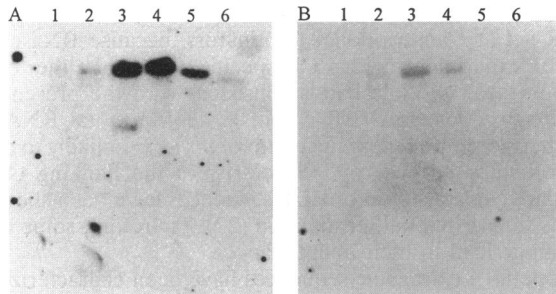


FIG. 6. Hybridization of a VCR-specific probe to *V. cholerae* O1 chromosomal DNAs separated on a pulsed-field gel. (A) Pulsed-field gel transferred to nitrocellulose and probed with the VCR-specific oligonucleotide probe. (B) Same filter after stripping and reprobing with the *mrhB*-specific probe. Strain and restriction endonuclease: lanes 1, 569B, *NotI*; lanes 2, 569B, *SfiI*; lanes 3, O17, *NotI*; lanes 4, O17, *SfiI*; lanes 5, C31, *NotI*; lanes 6, C31, *SfiI*.

two hybridizing bands are observed with *NotI*-digested DNA, of 265 and 125 kb, and a single band of 300 kb is observed with *SfiI*-digested DNA. Finally, strain C31 shows single bands of 245 and 235 kb with *NotI*- and *SfiI*-digested DNA, respectively (Fig. 6A). The filter was stripped and reprobed with the *mrhB*-specific probe used above, and the result of this is shown in Fig. 6B. In this case, bands of 285 and 195 kb are detected for 569B DNA digested with *NotI*, bands of 265 and 205 kb are detected for 569B DNA digested with *SfiI*, and bands of 265 and 300 kb are detected for O17 DNA digested with *NotI* and *SfiI*, respectively. As expected, no hybridizing fragments were detected for C31. The detection of two hybridizing bands for 569B with an *mrhB* probe (Fig. 6B) is unexpected, as a single hybridizing band is detected for both *BamHI*- and *HindIII*-digested 569B DNA when probed with the same probe (Fig. 5), indicating that the larger band is probably due to incomplete digestion of the samples. In case the oligodeoxynucleotide probe used in Fig. 6A was not hybridizing with some copies of VCR (see Discussion), leading to a biased result, the filter was again stripped and reprobed with the VCR-specific fragment probe. No additional bands were apparent (not shown), indi-

cating that the oligonucleotide probe has not "missed" significant bands through mismatch. Therefore, most copies of VCR lie within a region of approximately 200 kb in strain 569B, 300 kb in strain O17, and 240 kb in strain C31. Given a genome size of approximately 2,800 kb for *V. cholerae* O1 (30), this correlates with a region encompassing approximately 10% of the chromosome.

Specificity of VCR within the family *Vibrionaceae*. To see whether VCR-related sequences could be detected in other members of the family *Vibrionaceae*, a number of different species were probed with the VCR-specific oligodeoxynucleotide used above. Hybridization was observed in a number of non-O1 serotype *V. cholerae*, but no hybridization was observed for the other *Vibrio* or *Aeromonas* spp. tested (Fig. 7). These filters were stripped and reprobed with the *mrhB*-specific probe used above (Fig. 5), but a positive signal was obtained only for *V. cholerae* O1 strains 569B and C21 (not shown).

Is VCR involved in deletion events? In an attempt to define the deletion in strain C31, *BamHI*- and *HindIII*-digested chromosomal DNAs were probed with probes corresponding to the ends of plasmid pPM471. The distal probe hybridizes to a single band for 569B (6.3 kb for *BamHI*-digested DNA and 6.5 kb for *HindIII*-digested DNA) and no band for strain C31 (Fig. 8B), but the proximal fragment hybridizes to two bands for 569B (11 and 6.3 kb for *BamHI*-digested DNA and 2.8 and 1.2 kb for *HindIII*-digested DNA) and one band for strain C31 (10 kb for *BamHI*-digested DNA and 5.4 kb for *HindIII*-digested DNA) and a faint additional band of 3.9 kb for *HindIII*-digested C31 DNA (Fig. 8A). Next, the same samples were probed with an ORF2-specific probe, which again gives two bands of 6.3 and 4.5 kb for *BamHI*-digested 569B DNA and a single band of 11.5 kb for *BamHI*-digested C31 DNA (Fig. 8C). Bands were not apparent with *HindIII*-digested DNA in this case. This indicates that some sequences contained within pPM471 have been duplicated but that the deletion in C31 is larger than the cloned DNA. Therefore, further mapping of the deletion endpoints was not undertaken as part of this study.

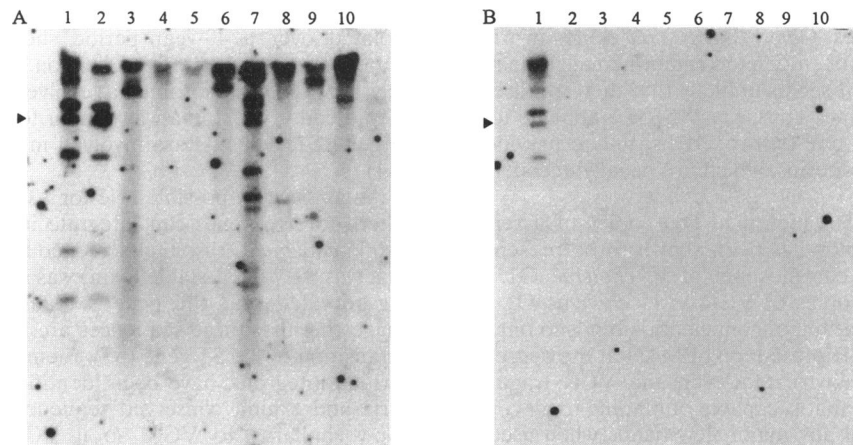


FIG. 7. Hybridization of a VCR-specific oligonucleotide probe to other members of the family *Vibrionaceae*. Chromosomal DNAs were digested with *BamHI*, transferred to nitrocellulose, and probed with the VCR-specific oligonucleotide probe. (A) DNAs isolated from *V. cholerae* 569B, C21, NCV165, BV7, BV22, BV41, N34, N41, N125, and N50 (lanes 1 to 10, respectively). (B) DNAs isolated from *V. cholerae* 569B, *Aeromonas hydrophila* A006, *Aeromonas* sp. AB1, *A. sobria* A187, *A. sobria* A191, *A. caviae* V14, *V. parahaemolyticus* NCTC 10884, *V. parahaemolyticus* NCTC 10885, *V. mimicus* V800, and *V. fluvialis* V564 (lanes 1 to 10, respectively). The position of the 6.3-kb band that hybridizes with the *mrhB* probe in *V. cholerae* O1 is indicated by an arrowhead.

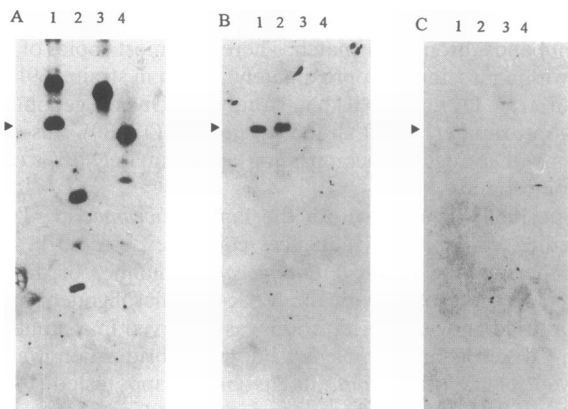


FIG. 8. Hybridization of probes derived from the ends of pPM471 to *V. cholerae* 569B and C31 DNAs. (A) DNA hybridized with a probe from the proximal end of pPM471; (B) DNA hybridized with a probe from the distal end of pPM471; (C) DNA hybridized with an ORF2-specific probe. Lanes in all panels: 1, 569B, *Bam*HI digested; 2, 569B, *Hind*III digested; 3, C31, *Bam*HI digested; 4, C31, *Hind*III digested. The position of a 6.3-kb band that hybridizes with the *mrhB* probe in *Bam*HI-digested DNA is indicated by an arrowhead.

DISCUSSION

The nucleotide sequence reported here extends that reported previously (17, 56) to include an additional eight ORFs linked to the genes encoding the MFRHA. The sequence of *mrhA* presented here is identical to that of van Dongen et al. (56), but that of *mrhB* shows six changes at the nucleotide level (17). Comparison of ORF sequences with entries in the databases revealed homology between ORF3.1 and the amino-terminal 72 amino acids of DnaT (primosomal protein i) of *E. coli* (38) and between ORF3.2 and the hypothetical *E. coli* RelE protein, presumed to be the translational inhibitor accumulated in *relB* mutants during amino acid starvation (3, 31). Attempts to engineer mutations in these ORFs by deleting portions of plasmid pPM471 (18) were unsuccessful, presumably because the desired products were lethal (not shown). Therefore, their functions remain unclear. Homology was also detected for ORF2 and *mrhB* and will be described elsewhere (1). None of the other ORFs reveal significant similarity to known protein sequences. Given that *E. coli* clones in which these other ORFs have been deleted retain hemagglutinating activity (17, 18, 55, 56), it seems unlikely that they are directly involved in this phenotype. Thus, the MFRHA appears to be an adhesin encoded by one or two ORFs, unlike previously characterized fimbrial adhesins, which are usually encoded by extensive operons (28).

Van Dongen et al. (56) identified two copies of a repeat sequence that in this study has been shown to be present in multiple copies in the chromosome of *V. cholerae* O1. An estimate of 60 to 100 copies of VCR in *V. cholerae* O1 was obtained. It is known that the oligonucleotide binds to only six of the nine copies of VCR present on pPM471 at the degree of stringency used (not shown). However, the VCR fragment probe (random primed) that is capable of binding to all copies of VCR does not reveal any major new bands when used to probe a filter identical to that shown in Fig. 3A.

VCR shows several regions of dyad symmetry, and the whole repeat shows imperfect symmetry that if transcribed would form a stable secondary structure. This raises the possibility that VCR might play a role in transcription termination in *V. cholerae*. Transcription has been shown to terminate within at

least one of the repeats, but it seems unlikely that all copies of VCR act as transcriptional terminators, because RNase protection experiments with VCR-specific probes of either orientation reveal multiple bands consistent with the presence of full-length copies of VCR within *V. cholerae* total RNA (2). Furthermore, the repeat upstream of *mrhA* is unlikely to act as a terminator because of the polarity of the flanking ORFs. Similarly, only a subset of REP sequences have been shown to act as transcriptional terminators (22). Therefore, some other role must lead to their maintenance.

Repetitive extragenic sequences have been characterized in a number of bacterial genera but do not show any recognizable sequence similarities (32). REP sequences found in *E. coli* and in diverse genera (57) have been shown to bind DNA polymerase 1 (20), DNA gyrase (59), and IHF (5, 42) as well as the histone-like protein HU (60). In *S. typhimurium*, REP sequences have been shown to be the endpoints of recombination events that occur in a *recA*-independent manner (49), and in *E. coli*, DNA gyrase (topoisomerase II) has been shown to catalyze illegitimate recombination (40). However, REP sequences are distributed more or less randomly throughout the *E. coli* chromosome (12), whereas VCR sequences are restricted to a region of approximately 10% of the *V. cholerae* chromosome. Another class of repetitive elements designated ERIC (29) have also been identified, with related sequences in *V. cholerae*. However, although they have a similar length and the potential to form a stable stem-loop structure, like VCR, there is no sequence homology.

The most striking feature of VCR is its localization to a limited region of the chromosome. An upper limit on the region occupied by VCR has been determined by pulsed-field gel electrophoresis, but the lower limit is still unknown. Assuming a density of VCR repeats equivalent to that seen on pPM471 (i.e., nine in 6 kb), a lower limit of about 42 to 70 kb (depending on the strain) can be predicted.

The orientation of the nine copies of VCR within pPM471 is such that recombination events between them will lead to deletion of the intervening DNA; the presence of multiple copies of VCR on multicopy plasmids and M13 vectors leads to a very high rate of deletion events in *recA*⁺ strains (1, 10, 16). Such a "deletogenic" activity would make VCR an effector sequence in *V. cholerae* for deletion events associated with pathogenicity islands. Despite this, the VCR region of *V. cholerae* O1 seems to be stable within biotypes, with strains isolated over a 40-year period showing a similar banding pattern on Southern hybridization. However, perhaps this region of the chromosome has evolved by tandem duplications mediated by VCR, in a way similar to that by which the much larger RS1 elements are involved in toxin gene amplification (44).

An alternative possible role for VCR is to provide a region where foreign genes can integrate into the chromosome. Of 500 *V. cholerae* O1 isolates screened by Takeda et al. (53), one that was ST⁺ (heat-stable toxin) was identified. DNA sequencing revealed that the gene is flanked by copies of VCR, suggesting that these sequences are somehow involved in the acquisition of the ST gene (53). Elements for gene integration, termed integrons, have been identified in gram-negative bacteria and exhibit conserved sequences (8), but these do not show similarity to VCR, so if VCR does fulfill a similar function, then it represents a different class of integron than those described previously (8). However, as sequencing has revealed that ST genes in virulent non-O1 *V. cholerae* are also flanked by copies of VCR (53), it is also possible to explain the acquisition of the ST gene by this strain as occurring by a conjugation or transduction event from a non-O1 strain with-

out invoking a novel integration mechanism. Another alternative is that the extreme clustering of VCR could be due to the entire region, including VCR sequences, originating in a foreign host, but the presence of genes such as *mrh*, exclusively in O1 *V. cholerae* (55), (not shown), and ST toxin genes, predominantly in non-O1 *V. cholerae* (53), makes this possibility seem less likely. Similarly, the suggestion that a single copy of VCR has integrated and is slowly spreading throughout the chromosome requires an explanation of why the spread is restricted to one region of the chromosome and is not occurring randomly. Nevertheless, there are insufficient data to discount any of these possibilities.

Although the biological function of VCR is not known, the extreme clustering and maintenance of these elements suggest that each fulfills the same or a similar role. An earlier proposal that REP sequences may be involved in bacterial speciation within the family *Enterobacteriaceae* (21) has been questioned, as REP-related sequences have recently been identified in diverse bacterial taxons (57). We have not detected VCR-related sequences in bacteria other than *V. cholerae* in the limited survey reported here but are loath to make a similar suggestion for VCR, as Takeda et al. (53) have identified cross-reacting sequences in *Vibrio mimicus* by colony blot. Identification of proteins (if any) that interact specifically with VCR should provide some clues as to their function and hopefully help explain how they have persisted in what is potentially a deleterious arrangement without being lost from the *V. cholerae* genome.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Health and Medical Research Council of Australia, the Diarrhoeal Diseases Programme of the World Health Organization, and the Clive and Vera Ramaciotti Foundations. A.B. is the recipient of an Australian Postgraduate Research Award priority scholarship.

We are grateful to Sue Williams for comments on the manuscript, Chris Cursaro for invaluable assistance with oligonucleotide synthesis and photography, and Steve Attridge for suggesting the acronym VCR. We are also grateful to M. W. Heuzenroeder for allowing us access to a Biorad CHEF II apparatus and to S. Bawden for allowing access to a densitometer.

REFERENCES

- Barker, A., and P. A. Manning. Unpublished data.
- Barker, A., S. G. Williams, and P. A. Manning. Submitted for publication.
- Bech, F. W., S. T. Jørgensen, B. Diderichsen, and O. H. Karlström. 1985. Sequence of the *relB* transcription unit from *Escherichia coli* and identification of the *relB* gene. *EMBO J.* 4:1059-1066.
- Blasquez, V. C., A. O. Sperry, P. N. Cockerill, and W. T. Garrard. 1989. Protein: DNA interactions at chromosomal loop attachment sites. *Genome* 31:503-509.
- Boccard, F., and P. Prentki. 1993. Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units. *EMBO J.* 12:5019-5027.
- Brutlag, D. L., J.-P. Dautricourt, R. Diaz, J. Fier, and R. Stamm. Unpublished data.
- Chen, E. Y., and P. H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* 4:165-170.
- Collis, C. M., and R. M. Hall. 1992. Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J. Bacteriol.* 174:1574-1585.
- Condemine, G., and C. L. Smith. 1990. Transcription regulates oxolinic acid-induced DNA gyrase cleavage at specific sites on the *E. coli* chromosome. *Nucleic Acids Res.* 18:7389-7396.
- DeGraaf, F. K. (Vrije Universiteit, Amsterdam, Holland). 1989. Personal communication.
- Desmarchelier, P. M. (University of Queensland, Brisbane, Queensland, Australia). 1982. Personal communication.
- Dimri, G. P., K. E. Rudd, M. K. Morgan, H. Bayat, and G. F.-L. Ames. 1992. Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. *J. Bacteriol.* 174:4583-4593.
- Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.
- Fetherston, J. D., P. Shuetze, and R. D. Perry. 1992. Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. *Mol. Microbiol.* 6:2693-2704.
- Focareta, T., and P. A. Manning. 1991. Distinguishing between the extracellular DNases of *Vibrio cholerae* and development of a plasmid transformation system. *Mol. Microbiol.* 5:2547-2555.
- Franzon, V. L. 1988. Ph.D. thesis, University of Adelaide, Adelaide, Australia.
- Franzon, V. F., A. Barker, and P. A. Manning. 1993. Nucleotide sequence and the construction of a mutant in the mannose-fucose-resistant hemagglutinin (MFRHA) of *Vibrio cholerae* O1. *Infect. Immun.* 61:3032-3037.
- Franzon, V. L., and P. A. Manning. 1986. Molecular cloning and expression in *Escherichia coli* K-12 of the gene for a hemagglutinin from *Vibrio cholerae*. *Infect. Immun.* 52:279-284.
- Gilson, E., J. M. Clément, D. Brutlag, and M. Hofnung. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J.* 3:1417-1421.
- Gilson, E., D. Perrin, and M. Hofnung. 1990. DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organisation. *Nucleic Acids Res.* 18:3941-3952.
- Gilson, E., D. Perrin, W. Saurin, and M. Hofnung. 1987. Species specificity of bacterial palindromic units. *J. Mol. Evol.* 25:371-373.
- Gilson, E., J. P. Rousset, J. M. Clément, and M. Hofnung. 1986. A subfamily of *E. coli* palindromic units implicated in transcription termination? *Ann. Inst. Pasteur Microbiol.* 137B:259-270.
- Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre, and W. Goebel. 1990. Deletions of chromosomal regions coding for fimbriae and haemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. *Microb. Pathogen.* 8:213-225.
- Hanne, L. F., and R. A. Finkelstein. 1982. Characterization and distribution of the hemagglutinins produced by *Vibrio cholerae*. *Infect. Immun.* 36:209-214.
- Herrington, D. A., R. H. Hall, G. Lososky, J. J. Mekalanos, R. K. Taylor, and M. M. Levine. 1988. Toxin, toxin coregulated pili and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J. Exp. Med.* 168:1487-1492.
- Higgins, C. F., R. S. McLaren, and S. F. Newbury. 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion?—a review. *Gene* 72:3-14.
- Higgins, D. G., and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Cabios* 5:151-153.
- Hultgren, S. J., S. Abraham, M. Caparon, P. Falk, J. W. St. Gene III, and S. Normark. 1993. Pilus and non-pilus bacterial adhesins: assembly and function in cell recognition. *Cell* 73:887-901.
- Hultgren, C. S. J., C. F. Higgins, and P. M. Sharp. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol. Microbiol.* 5:825-834.
- Kaper, J. B., and M. M. Baldini. 1992. Genetics, p. 69-94. In D. Barua and W. B. Greenough III (ed.), *Cholera*. Plenum Medical Book Company, New York.
- Karlström, O. H. (University of Copenhagen, Copenhagen, Denmark). 1992. Personal communication.
- Leblond, P., P. Demuyter, J. M. Simonet, and B. Decaris. 1990. Genetic instability and hypervariability in *Streptomyces ambofaciens*: towards an understanding of a mechanism of genome plasticity. *Mol. Microbiol.* 4:707-714.
- Lupski, J. R., and G. M. Weinstock. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J. Bacteriol.* 174:4525-4529.

34. Manning, P. A. 1987. Involvement of cell envelope components in the pathogenesis of *Vibrio cholerae*: targets for cholera vaccine development. *Vaccine* **5**:83–87.
35. Manning, P. A. 1994. Surface-associated and soluble components of *Vibrio cholerae* involved in bacteria-host interactions. *Curr. Top. Microbiol. Immunol.* **192**:265–281.
36. Manning, P. A., M. W. Heuzenroeder, J. Yeadon, D. I. Leavesley, P. R. Reeves, and D. Rowley. 1986. Molecular cloning and expression in *Escherichia coli* K-12 of the O antigens of the Inaba and Ogawa serotypes of the *Vibrio cholerae* O1 lipopolysaccharides and their potential for vaccine development. *Infect. Immun.* **53**:272–277.
37. Masai, H., M. W. Bond, and K.-I. Arai. 1986. Cloning of the *Escherichia coli* gene for primosomal protein i: the relationship to *dnaT*, essential for chromosomal replication. *Proc. Natl. Acad. Sci. USA* **83**:1256–1260.
38. Miller, C. J., R. G. Feachem, and B. S. Drasar. 1985. Cholera epidemiology in developed and developing countries: new thoughts on transmission, seasonality and control. *Lancet* **i**:261–263.
39. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. Naito, A., S. Naito, and H. Ikeda. 1984. Homology is not required for recombination mediated by DNA gyrase of *Escherichia coli*. *Mol. Gen. Genet.* **193**:238–243.
41. Neuer-Nitsche, B., X. Lu, and D. Werner. 1988. Functional role of a highly repetitive DNA sequence in anchorage of the mouse genome. *Nucleic Acids Res.* **16**:8351–8360.
42. Oppenheim, A. B., K. E. Rudd, I. Mendelsohn, and D. Teff. 1993. Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. *Mol. Microbiol.* **10**:113–122.
43. Overhauser, J., and M. Z. Radic. 1987. Encapsulation of cells in agarose beads for use with pulsed-field gel electrophoresis. *Focus* **9**(3):8–9.
44. Pearson, G. D. N., A. Woods, S. L. Chang, and J. J. Mekalanos. 1993. CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor. *Proc. Natl. Acad. Sci. USA* **90**:3750–3754.
45. Rabbani, G. H. 1986. Cholera. *Clin. Gastroenterol.* **15**:507–528.
46. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
47. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
48. Schrank, G. D., and W. F. Verwey. 1976. Distribution of cholera organisms in experimental *Vibrio cholerae* infections: proposed mechanisms of pathogenesis and antibacterial immunity. *Infect. Immun.* **13**:195–203.
49. Shyamala, V., E. Schneider, and G. F.-L. Ames. 1990. Tandem chromosomal duplications: role of the REP sequences in the recombination event at the join-point. *EMBO J.* **9**:939–946.
50. Smith, C. L., and C. R. Cantor. 1987. Purification, specific fragmentation and separation of large DNA molecules. *Methods Enzymol.* **155**:449–467.
51. Sperry, A. O., V. C. Blasquez, and W. T. Garrard. 1989. Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. *Proc. Natl. Acad. Sci. USA* **86**:5497–5501.
52. Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37**:1015–1026.
53. Takeda, T., A. Ogawa, H. Abe, H. Kasuga, B. G. Nair, S. C. Pal, and H. Watanabe. 1991. A gene encoding a heat-stable enterotoxin (O1-ST) of *Vibrio cholerae* O1 El Tor Inaba is flanked by direct repeats, p. 82–87. *In Proceedings of the 27th Joint Conference on Cholera and Related Diarrheal Diseases*. National Institutes of Health, Bethesda, Md.
54. Taylor, R. K., V. L. Miller, D. B. Furlong, and J. J. Mekalanos. 1987. Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc. Natl. Acad. Sci. USA* **84**:2833–2837.
55. van Dongen, W. M. A. M., and F. K. DeGraaf. 1986. Molecular cloning of a gene coding for a *Vibrio cholerae* haemagglutinin. *J. Gen. Microbiol.* **132**:2225–2234.
56. van Dongen, W. M. A. M., M. A. M. van Vlerken, and F. K. DeGraaf. 1987. Nucleotide sequence of a DNA fragment encoding a *Vibrio cholerae* haemagglutinin. *Mol. Gen. (Life Sci. Adv.)* **6**:85–91.
57. Versalovic, J., T. Koeuth, and J. R. Lupski. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.* **19**:6823–6831.
58. Wood, W. I., J. Gitschier, L. A. Lasky, and R. Lawn. 1985. Base composition independent hybridization in tetramethyl ammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. USA* **82**:1585–1588.
59. Yang, Y., and G. F.-L. Ames. 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc. Natl. Acad. Sci. USA* **85**:8850–8854.
60. Yang, Y., and G. F.-L. Ames. 1990. The family of repetitive extragenic palindromic sequences: interaction with DNA gyrase and histonelike protein HU, p. 211–225. *In* M. Riley and K. Drlica (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.
61. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103–119.