

Cloning and Sequencing of the Genes Downstream of the *wbf* Gene Cluster of *Vibrio cholerae* Serogroup O139 and Analysis of the Junction Genes in Other Serogroups

SHANMUGA SOZHAMANNAN,^{1*} YING KANG DENG,¹ MANRONG LI,¹ ALEXANDER SULAKVELIDZE,¹
JAMES B. KAPER,² JUDITH A. JOHNSON,^{3,4} G. BALKRISH NAIR,⁵ AND J. GLENN MORRIS, JR.,^{1,4}

Division of Hospital Epidemiology, Department of Medicine,¹ Department of Microbiology and Immunology,² and Department of Pathology,³ School of Medicine, University of Maryland at Baltimore, and Department of Veterans Affairs, Veterans Affairs Medical Center,⁴ Baltimore, Maryland 21201, and National Institute of Cholera and Enteric Diseases, Calcutta, India⁵

Received 12 April 1999/Returned for modification 25 June 1999/Accepted 7 July 1999

The DNA sequence of the O-antigen biosynthesis cluster (*wbf*) of a recently emergent pathogen, *Vibrio cholerae* serogroup O139, has been determined. Here we report the sequence of the genes downstream of the O139 *wbfX* gene and analysis of the genes flanking the *wbf* gene cluster in other serogroups. The gene downstream of *wbfX*, designated *rjg* (right junction gene), is predicted to be not required for O-antigen biosynthesis but appears to be a hot spot for DNA rearrangements. Several variants of the *rjg* gene (three different insertions and a deletion) have been found in other serogroups. DNA dot blot analysis of 106 *V. cholerae* strains showed the presence of the left and right junction genes, *gmhD* and *rjg*, respectively, in all strains. Further, these genes mapped to a single *I-CeuI* fragment in all 21 strains analyzed by pulsed-field gel electrophoresis, indicating a close linkage. The insertion sequence element *ISI358*, found in both O1 and O139 *wb** regions, is present in 61% of the strains tested; interestingly, where present, it is predominantly linked to the *wb** region. These results indicated a cassette-like organization of the *wb** region, with the conserved genes (*gmhD* and *rjg*) flanking the divergent, serogroup-specific *wb** genes and *ISI358*. A similar organization of the *wb** region in other serogroups raises the possibility of the emergence of new pathogens by homologous recombination via the junction genes.

Vibrio cholerae, the etiological agent of clinical cholera, has more than 155 serogroups based on the heat stable somatic O-antigen types (33). Of all of these serogroups, only O1 has traditionally been associated with cholera (4). Although other serogroups have not been recognized as having epidemic potential, occasional outbreaks caused by other serogroup strains have been recorded (44). In 1992, *V. cholerae* belonging to serogroup O139 emerged as an epidemic strain (1, 2, 29, 32). Molecular epidemiological analyses such as zymovar analysis, ribotyping, and pulsed-field gel electrophoresis (PFGE) revealed that *V. cholerae* O139 Bengal strains are virtually identical to Asian seventh-pandemic O1 El Tor strains (5, 28). Furthermore, *V. cholerae* O139 strains had all of the standard virulence factors of O1 El Tor. Clinically, *V. cholerae* O1 and O139 Bengal cause cholera of comparable severity in infected persons (6, 27). However, in striking contrast to O1 strains, O139 strains are encapsulated (20). The *V. cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide (LPS) (41, 45). The LPS of serogroup O139 does not contain any long O-antigen side chain, whereas O1 strains have a core substituted with an average of 17 repeat units of 4-NH₂-4,6-dideoxymannose, each substituted with 3-deoxy-L-glycero-tetronic acid (23). The O139 LPS appears to be an efficiently substituted core polysaccharide (CPS), albeit with only one or two additional sugar moieties (45). These changes have rendered the O139 organism immunologically distinct

from the O1 El Tor strains, as evidenced by the susceptibility of the individuals preexposed to O1 strains. In cholera-endemic areas such as Bangladesh and India, cholera is most common among children (16). In contrast, the majority of O139 infections occurred in adults when O139 first emerged (11, 19).

The major genetic differences accounting for the phenotypically distinct surface polysaccharide of O1 El Tor and O139 Bengal have been determined. The genes responsible for the synthesis of O antigen are present in a cluster designated *wb** (*rfb*) region. It was shown that a large portion of DNA corresponding to the *wbe* (*rfb*) region of O1 strains is missing in O139 strains and that O139 has acquired a unique DNA region (7, 8, 42). Specifically, it was demonstrated that the serogroup O139 resulted from a 22-kb deletion of the *wbe* (*rfb*) region of O1 and replacement with a 35-kb *wbf* region encoding the O139 O-antigen (12). Several groups collectively sequenced the O139-specific *wbf* (*rfb*) region. The 14.363-kb sequence of the left part of the *wbf* region, *gmhD* to *gmd'* (originally designated *otn*), was reported by Bik et al. (8), the 12.938-kb right part of the *wbf* region, *wbfQ* to *wbfX*, was reported by Comstock et al. (12), and the intermediate region was reported by Stroehrer et al. (38). The complete DNA sequence of the O1 *wbe* region was previously determined by Stroehrer et al. (35). The sequenced *wb** regions of *V. cholerae* O1 and O139 and the *V. anguillarum* serogroups O1 and O2 all have the *gmhD* gene at the left junction (8, 40). The sequenced right junction of the O1 *wbe* cluster has a 30-bp overlap with that of the O139 *wbf* right junction (14). The intervening regions are divergent in the two serogroups except for the presence of an insertion sequence (IS) element, *ISI358*. The genes downstream of the right junction were not sequenced, and it was assumed that

* Corresponding author. Mailing address: Division of Hospital Epidemiology, Department of Medicine, University of Maryland School of Medicine, 934-MSTF, 10 S. Pine St., Baltimore, MD 21201. Phone: (410) 706-5157. Fax: (410) 706-4581. E-mail: ssozhama@medicine.umaryland.edu.

TABLE 1. Bacterial strains, plasmids, and primers

Strain, plasmid, or primer	Description	Source, reference, or GenBank accession no.
<i>E. coli</i>		
DH5 α	F ⁻ ϕ 80 <i>dlacZM15</i> Δ (<i>lacZYA-argF</i>) <i>U169 endA1 recA1 hsdR17</i> (r _k ⁻ m _k ⁺) <i>deoR thi-1 supE44 gyrA96 relA1</i> λ ⁻	Lab collection
DH5 Δ <i>lac</i> Rif ^r	Spontaneous mutant	This work
NLC51	F ⁻ <i>araD139</i> Δ (<i>argF-lac</i>) <i>U169 rpsL150 relA1 ffb5301 deoC1 ptsF25 rbsR Val^r recA56</i>	N. L. Craig
CW51	F ⁻ <i>ara arg lacXIII</i> Nal ^r Rif ^r <i>recA56</i>	N. L. Craig
<i>V. cholerae</i>		
AI1837	Wild-type serogroup O139	M. J. Albert
395	O1 classical	Lab collection
E7946	O1 El Tor	Lab collection
C6706	O1 El Tor	Lab collection
N16961	O1 El Tor	Lab collection
Arg-3	O139 Argentinian isolate	Lab collection
Non-O1/non-O139	43 strains	Lab collection and 31
Smith serogroup	39 strains	34
Plasmids		
pBluescript SK(-)	Cloning vector, Amp ^r	Stratagene
pCR2.1	PCR fragment cloning vector, Amp ^r	Invitrogen
pQE70	Expression vector, Amp ^r	Qiagen
pWSK29	Low-copy-number plasmid (pSC101 replicon), Amp ^r	43
pLC40	21.5-kb <i>EcoRI</i> fragment in cosmid pCVD301	12
pLC <i>Sst</i> - <i>Kpn</i>	pBluescript with 5-kb <i>Sst</i> I- <i>Kpn</i> I fragment	This work
pREP4	pACYC184 <i>lacI</i> ^q	Qiagen
pOX38-Gen ^r	F ^r derivative with Gen ^r	N. L. Craig
pQE <i>tnp</i>	pQE70 containing the <i>tnp</i> gene of IS1358 Amp ^r	This work
pWSK IS::Tet ^r	pWSK29 containing IS1358 Tet ^r Tet ^r Amp ^r	This work
PCR primers		
<i>gmhD</i>	5'-TTACTTACGATTAATCAGCGCCAT 5'-GGCGGCGCTGGCATGATTGGCAGC	X90547
<i>rjg</i>	5'-CATGGAAGTGGTTTCATCACGGAGG 5'-GTGGACGCGTTCAAAGCACCGAATATCCGAGTT	AF090685
IS1358	5'-CTAAATGCATGCATGAGCGAGTTAATCAACCC 5'-CAAGATCTCGTAAGGCTTTCAAGAACCCTTACT	X90547
<i>orf2</i>	5'-GGTGTATGCCACTAGTGAGGTAAT 5'-GGTGACATCAAAGGGACCACTTTTTTC	X59554
ISalg	5'-GGCCACGGTTTTAGAGTTTTCC 5'-GTGAATTTGGTCACTTAATTAG	This work
ISO22	5'-TAAGCGTTTACCAGTAGCATG 5'-TAAACGTTTACCAGTACCAGCA	This work and AB012957

they were not involved in O-antigen biosynthesis. Favre et al. showed that the cloned O139 DNA can express O139-specific antigens in *Escherichia coli*, although characterization of the cloned DNAs, with respect to the size of the insert and the genes present, was not reported (15).

We were interested in determining the role of the genes downstream of *wbfX* in LPS/capsular polysaccharide (CPS) biosynthesis and whether non-O1/non-O139 serogroups are similar with respect to genetic organization of the *wb** region. The broad objectives of this study were to decipher the mechanism underlying the origin of O139 and to explore the possibility of the future emergence of other pathogenic *V. cholerae* strains by a similar mechanism. Two hypotheses have been proposed to explain the emergence of *V. cholerae* O139. The first hypothesis proposes that a transposition event mediated by the IS element IS1358 resulted in the replacement of the O1 *wbe* genes with the O139 *wbf* genes (26, 37–39). The second hypothesis involves a homologous recombination event resulting in the replacement of the O1 *wbe* region en masse by the O139 *wbf* region (12, 26, 40). Here, we present the sequence of the genes downstream of *wbfX* and the analysis of the genes

flanking the O-antigen region (*gmhD* and *rjg*) in other serogroups. We show that all the serogroups we analyzed have an organization of the *wb** region that is similar to that of O1 and O139 strains. This result favors the latter hypothesis and raises the possibility that pathogenic strains belonging to non-O1/non-O139 serogroups can emerge by homologous recombination via the junction genes.

MATERIALS AND METHODS

Bacterial strains, plasmids, media, and growth conditions. The various bacterial strains, plasmids, and primers used in this study are listed in Table 1. Growth of bacterial cultures was done following standard laboratory practices. Cultures were grown in Luria-Bertani broth (LB; pH 6.5) at 37°C, and frozen stocks were maintained at -70°C in LB containing 50% glycerol. For short-term storage, strains were maintained on LB plates at room temperature. Antibiotics were used at the following concentrations: ampicillin, 100 μ g/ml; chloramphenicol, 25 μ g/ml; tetracycline, 12.5 μ g/ml; rifampin, 100 μ g/ml; streptomycin, 100 μ g/ml; and gentamicin, 5 μ g/ml. Nucleic acid manipulations were carried out by standard molecular biological methods (22). Plasmid DNA purification was done by using a QIAprep Spin Miniprep kit (Qiagen, Inc.). DNA fragment purification was done with a Gene Clean Spin kit (Bio 101, Inc.).

Sequence of the genes downstream of *wbfX*. The DNA sequence of the genes downstream of *wbfX* was obtained from a subclone of pLC40, which is a cosmid

TABLE 2. Identities of putative ORFs downstream of the *wbfX* gene with products in other bacteria

O139 ORF	Size (bp)	% G+C content	Homologous gene ^a	Identity (%) / positive (%) / Total ^b	Gene product	GenBank accession no.
<i>wbfX</i>	1,941	45.4	<i>Pseudomonas aeruginosa wbpM</i>	41/60/563	Epimerase/dehydratase	X90711
<i>rjg</i>	1,341	53.2	<i>Methanococcus jannaschii</i> MJ1236	29/50/444	Putative mRNA 3'-end processing factor	U67564
<i>ybdG</i>	804	47.3	<i>Escherichia coli ybdG</i>	52/69/221	Hypothetical transmembrane protein/transport protein?	AE000163
<i>ycdW</i> (partial)	516	51.1	<i>E. coli ycdW</i>	34/52/177	Putative 2-hydroxyacid dehydrogenase	AE000205

^a Only the gene encoding the protein with the highest homology score is listed.

^b Total amino acids compared by using NCBI BLAST server.

clone and has a 21.5-kb *EcoRI* insert of the *V. cholerae* O139 *wbf* region (12). pLC40 DNA was digested with *SstI*, a 7.8-kb fragment was gel purified and redigested with *KpnI*, and the resulting two fragments were cloned into pBlue-script SK(-). The clone with the larger fragment (5 kb) was sequenced. DNA analysis was carried out with the DNASIS (Hitachi Software Corporation) program, and protein homology search was done using the National Center for Biotechnology Information BLAST server program (3).

Dot blot analysis. Distribution of the *gmhD*, IS1358, and *rjg* genes in various *V. cholerae* strains was analyzed in dot blots. Minichromosomal DNA preparations were made from various strains by using a Wizard genomic DNA purification kit (Promega Corporation). Approximately 1 to 2 µg of DNA was blotted onto a Zeta-Probe membrane by using a dot blot apparatus (Bio-Rad Laboratories). The blot was hybridized with nonradioactive probes prepared by an enhanced chemiluminescence using the ECL labeling kit (Amersham Inc.). All probes were prepared by PCR amplification of the genes, using the specific primers listed in Table 1.

PFGE. PFGE was carried out by using published protocols (25). Agarose plugs were prepared with the bacterial cells and then treated with lysis buffer followed by protease treatment and restriction endonuclease digestion overnight with *I-CeuI* at 37°C. The digested plugs were then run on a 1% agarose gel in the CHEF DR11 system (Bio-Rad Laboratories) at 14°C in 0.5× Tris-borate-EDTA. The following electrophoretic conditions were used: block 1, 6 V/cm² for 12 h with 15- and 30-s (initial and final switch times) pulses; block 2, 6 V/cm² for 18 h with 13'-13" pulses.

Construction of a marked IS1358 element and in vivo transposition assay. IS1358 was PCR amplified from the O139 chromosomal DNA and cloned into pCR2.1 vector. In the next step, a deletion of the *NdeI* fragment within the IS element was created and a Tet^r fragment from pBR322 (*AvaI-EcoRI*) was introduced in the *tnp* gene of IS1358. This inactivated the transposase gene. The IS1358::Tet^r was subcloned into a pSC101 replicon (*KpnI-NotI* fragment of pCR2.1 IS::Tet^r cloned into pWSK29 [43]). The *tnp* gene was amplified from the O139 chromosome by appropriate primers (Table 1) and cloned into the expression vector pQE70. In this vector, the transposase is expressed from the *p_{lac}* promoter, and a compatible replicon pREP-4 (Kan^r), provides the *lac* repressor. Transposase gene expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside (IPTG). The donor strain carried four plasmids: pSC101 IS::Tet^r, pQE70::*tnp*, pREP4 (*lac^r*), and an F' derivative (pOX38-Gen^r). The recipient strain was marked with Rif^r (DH5 *Δlac*) or Str^r (NLC51). A mating-out assay, in which donor and recipient cells were mixed and mated for 3 h and plated on selective plates, was used to measure the transposition of IS1358. The transfer of pOX38-Gen^r was measured as Gen^r Rif^r or Gen^r Str^r colonies, and transposition was measured as the frequency of Tet^r colonies among the Gen^r exconjugants.

Nucleotide sequence. The sequence of the 4-kb DNA downstream of *wbfX* is available as GenBank accession no. AF090685, and the sequence of IS1358 is available as GenBank accession no. AF133213.

RESULTS

DNA sequence of the genes downstream of *wbfX* of *V. cholerae* O139. It was shown previously that the O139 *wbf* region is about 35 kb in length and that at the molecular level, O139 probably arose by the replacement of a 22-kb *wbe* region of the serogroup O1 biotype El Tor strain. The entire sequences of the O1 and O139 *wb** regions have been determined (8, 12, 35, 38). The sequenced O1 and O139 *wb** regions contained only a 30-bp overlap at the right junction. The genes downstream of *wbfX* were not known, and it was presumed that *wbfX* is the last gene of the *wbf* cluster. In an effort to understand the mechanism of acquisition of this unique O139 *wbf* region, i.e., the presence of IS elements or phage attachment sites, we ex-

tended the sequence 4 kb further downstream of the *wbfX* gene, into the region shared by O1 and O139. For convenience, we have compiled all published sequences with the new sequence and refer to the various sites with a new numbering system. According to this numbering system, the entire sequence is 41,221 bp in length. The O139-specific DNA (35,807 bp) begins at position 1127, at the start of *gmhD*, and it ends at position 36934, after *wbfX*.

Resequencing of the *wbfX* and *wbfY* region (open reading frames [ORFs] formerly designated *orf10* and *orf11* [12]) revealed an additional 233-bp *SstI* fragment at the *SstI* site at position 11459 (35525 in the new numbering) in *wbfX*. This additional sequence creates a new single ORF, *wbfX*, combining the former *wbfX* and *wbfY*. The presence of the *SstI* fragment was confirmed by PCR (11a). This ORF has extensive homology to the epimerase/dehydratase found in *Pseudomonas aeruginosa* and other bacterial species (reference 10 and Table 2). Figure 1A shows the revised map of the ORFs of the O139 *wbf* region.

The DNA downstream of *wbfX* has been shown in hybridization experiments to be present in both O1 and O139 strains (12). The overall G+C content of the newly sequenced region downstream of *wbfX* is about 49%, which is the chromosomal average for *V. cholerae* (21). This is greater than the average G+C content of the *wbf* region, which is 42%, indicating that the downstream DNA is not part of the *wbf* region. In the DNA shared by O1 and O139 serogroups, three ORFs were identified. Immediately downstream of *wbfX* we have identified a gene, designated *rjg* (right junction gene), the putative product of which has considerable homology to proteins in other bacterial species (Table 2). Interestingly, these proteins have a highly conserved VDALILTHAHIDH motif whose significance is unknown. The *rjg* product is predicted to be an mRNA 3'-end processing factor. Several variants of the *rjg* gene that contain IS elements and an internal deletion have been found (see below). These *rjg* mutants presumably are inactive due to IS insertion and deletion, thus indicating the nonessential nature of this gene. Downstream of *rjg* and transcribed in the opposite orientation is an ORF whose product has homology to the product of a gene designated *ybdG* in the *pheP-nfnB* intergenic region of *E. coli* K-12. The next ORF, for which only a partial sequence was obtained, encodes a product that has homology to putative dehydrogenases from various sources (Table 2). There is a 51-bp almost perfect (50/51) direct repeat separating a 978-bp region (TTGGTGTGAAAT ACCCCCTCCCGTCCTCCCCCTAGAAGGGGGAGGGGT AAG [variant base is underlined]) at the end of *rjg*. The O1-O139 homology begins 17 bp into the *rjg* gene. Hence, the predicted N-terminal amino acid sequences of the *rjg* product are different in the O1 and O139 serogroups (Fig. 2).

Genes flanking the *wbf* region of O139 are conserved in other serogroups. We were interested in determining whether

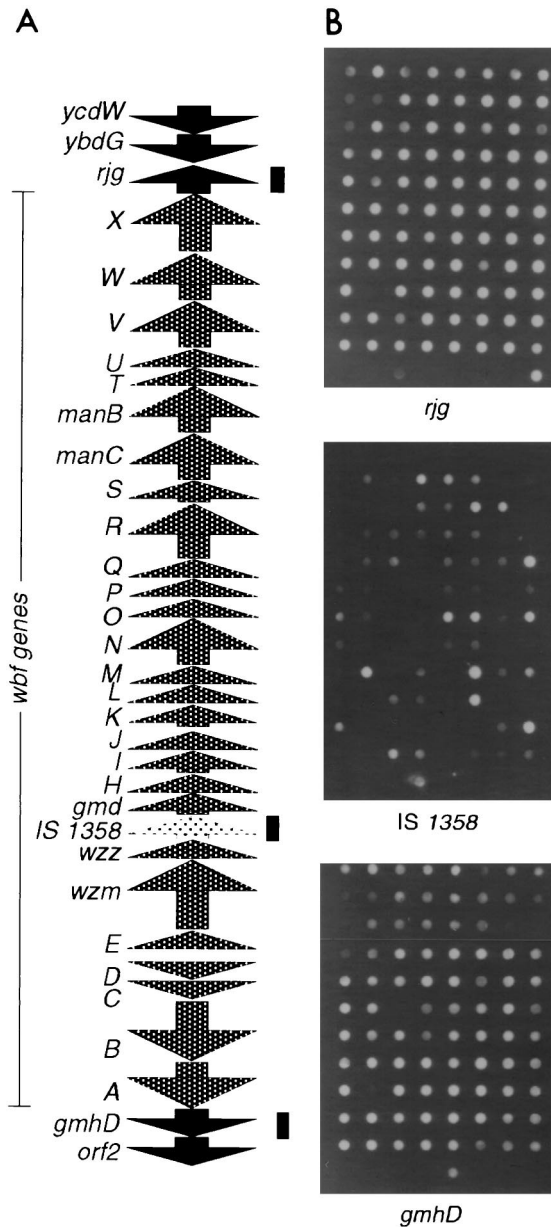


FIG. 1. (A) Revised ORF map of the *V. cholerae* O139 *wbf* region. The solid arrows at the top (*rjg*, *ybdG*, and *ycdW*) and bottom (*gmhD* and *orf2*) represent genes common to all serogroups. The entire region is 41,221 bp in length, and the *wbf* region (*wbfA* through *wbfX*) is 35,807 bp in length starting from *wbfA* to *wbfX*. The map is not drawn to scale. (B) DNA dot blot analysis of various serogroups of *V. cholerae* for the distribution of the three genes (*gmhD*, *IS1358*, and *rjg*) common to serogroups O1 and O139. The bars in panel A indicate positions of the hybridization probes. The intensities of some of the hybridization spots in the *gmhD* panel are weak, and these strains were reprobated separately to show that they carry *gmhD*. All strains tested were found to contain *gmhD* and *rjg*; only 61% of the strains had *IS1358*.

non-O1/non-O139 *V. cholerae* strains have a cassette with a genetic organization similar to that of the O1/O139 *wb** cluster, i.e., where *gmhD* and *rjg* genes flank the *wb** genes and *IS1358*. As a first step toward this goal, we looked for the presence of *gmhD*, *IS1358*, and *rjg* in other serogroups. Clinical and environmental isolates belonging to various serogroups (67 Shimada serogroup strains consisting of 24 O1/O139 and 43 non-O1/non-O139 strains and 39 Smith serogroup strains)

were examined by dot blot analysis (Fig. 1B). All strains had *gmhD* and *rjg*, and *IS1358* was present in 61% of the strains. The intensity of the hybridization spots for *gmhD* and *IS1358* varied in many strains, reflecting either sequence variations or the presence of more than one copy of the IS element. Consistent with this observation, *gmhD* could not be amplified from all the strains by using common PCR primers, most probably due to sequence variation (data not shown). *orf2*, a gene downstream of *gmhD* and *rjg*, could be amplified from all of the *V. cholerae* strains examined and showed uniform intensities in all strains, indicating a high degree of conservation.

***gmhD*, *IS1358*, and *rjg* are located in the same I-CeuI fragment.** We concluded from the above results that *gmhD* and *rjg* genes are present in all of the *V. cholerae* strains examined. This result does not indicate whether they are closely linked as in the case of O1/O139 strains. We performed PFGE and

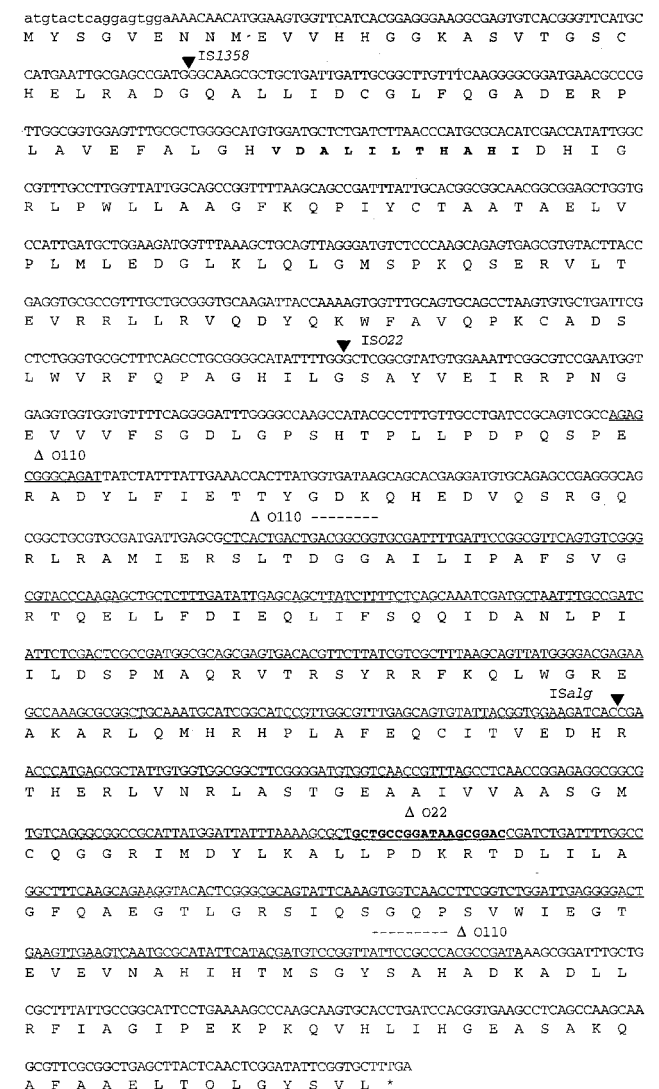


FIG. 2. DNA sequence of the *rjg* gene (1,341 bp). The first 17 bp (in lowercase) are different in O1 and O139 strains. The highly conserved VDALILTH AHI motif in *rjg* is indicated in boldface. The insertion sites of *IS1358* (at bp 90) in Smith serogroup O48, *ISO22* (at bp 469) in O22, and *ISalg* (at bp 933) in O103 are indicated by filled arrowheads. Deletions of 13 and 539 bp in Smith serogroup O110 are underlined; an 18-bp deletion found in an O22 strain is indicated in boldface.

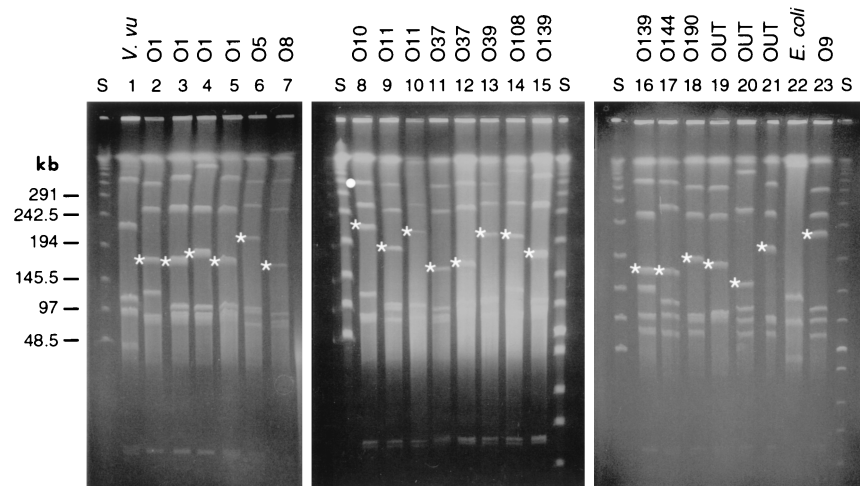


FIG. 3. PFGE of I-CeuI-digested chromosomal DNAs of various *V. cholerae* strains. Lanes: S, λ ladder; 1, *Vibrio vulnificus* (*V. vu*); 2, *V. cholerae* serogroup O1 (classical 395); 3, O1 (El Tor E7946); 4, O1 (El Tor C6706); 5, O1 (El Tor N16961); 6, O5 (CO545); 7, O8 (CO845); 8, O10 (AS12/1); 9, O11 (CO639); 10, O11 (AM124); 11, O37 (S21); 12, O37 (Y276); 13, O39 (AM25); 14, O108 (CO603B); 15, O139 (AI1837); 16, O139 (Arg-3); 17, O144 (AM107); 18, O190 (AS67); 19 to 21, O-antigen untyped (OUT) (AS416, CO668, and AS119, respectively); 22, *E. coli* DH5 Δ lac; 23, O9 (AM2). The bands that hybridized with the *gmhD* and *rjg* probes are indicated by asterisks. The IS1358 probe also hybridized to the same fragment in strains where present except in an O10 strain where a different fragment had IS1358 (indicated by a dot in lane 8).

hybridization analysis to determine whether these genes map in the same fragment in non-O1/non-O139 strains as they do in O1 and O139 strains. Chromosomal DNAs were digested with the rarely cutting I-CeuI enzyme and hybridized with *gmhD* and *rjg* probes. In all 21 strains examined, a single I-CeuI fragment hybridized with the two probes (Fig. 3). The size of the I-CeuI fragments varied from 150 to 250 kb. IS1358 was present in 15 strains, and in 14 of these strains it was located on the same fragment as *gmhD* and *rjg*. The ISalq probe (see below) hybridized to different fragments (data not shown). These results indicated that *gmhD* and *rjg* genes are conserved in all the strains and are probably arranged similarly flanking the O-antigen gene cluster.

Identification of novel IS elements in the *rjg* gene of non-O1/non-O139 *V. cholerae* strains. To confirm the hybridization data, we analyzed the presence of the *rjg* gene in a subset of 24 *V. cholerae* strains of various serogroups by using PCR. One strain, belonging to the O103 serogroup, exhibited a PCR product longer than the rest. Cloning and sequencing of this PCR product identified the presence of an IS element, designated ISalq, in the *rjg* gene of this strain. This element showed extensive homology (86% at the DNA level) to an IS element found in *Vibrio alginolyticus* which was shown earlier to be homologous to IS911 of *Shigella dysenteriae* (17). ISalq has a 22-bp imperfect (16/22) inverted repeat at its ends with no obvious target duplication at the site of insertion in the *rjg* gene. We examined the presence of this element in other serogroups by DNA dot blot analysis. This analysis showed that it is distributed in 69% (56 of 81) of the strains tested (data not shown). We examined whether ISalq is present in the same locus in all of these strains by PCR amplification of the *rjg* locus. It appeared that only in an O103 strain was ISalq inserted in *rjg*. During this analysis, we found three PCR products that differed in size from the rest. Sequencing of these PCR products indicated the presence of two other IS elements as well as a deletion in the *rjg* gene in three different serogroups (Fig. 4). In serogroup O22 there was an IS element designated ISO22 (48); it was distributed in 6% (5 of 81) of the *V. cholerae* strains tested (data not shown). In a Smith serogroup O48 strain, a copy of IS1358 (37) was present. In a Smith

serogroup O110 strain, the *rjg* gene had a large internal deletion. The DNA sequence of the *rjg* ORF and the various insertion sites and the deletion endpoints in O110 strain are shown in Fig. 2. These results suggested that *rjg* is not an essential gene for cell viability and that *rjg* is a hot spot for genetic rearrangements.

IS1358 present in *V. cholerae* O139 strain appears to be an inactive element in *E. coli*. The majority (61%) of the *V. cholerae* strains analyzed contained IS1358 most likely in the *wb** region. It was recently proposed that there are two copies of IS1358 in the O139 *wbf* region, thus giving this region the

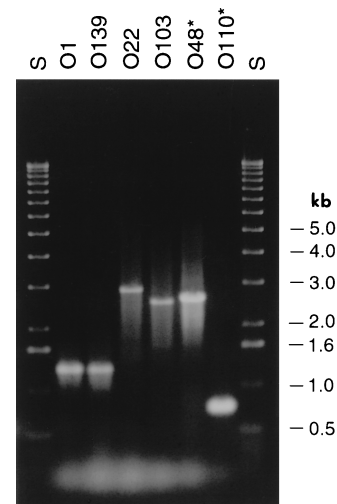


FIG. 4. PCR analysis of the *rjg* gene in various serogroups. The *rjg* genes from strains of various serogroups were amplified by using *rjg* primers. The PCR products from O1 and O139 strains were of the same size (about 1.3 kb). In three other strains (O22, O103, and O48) there is an insertion, as seen by the increase in the size of the amplified product. The PCR product from the O110 strain is shorter, which indicated an internal deletion. The asterisks in O110 and O48 are shown to indicate that they belong to O serogroups based on the Smith typing system (34). Lane S, 1-kb ladder.

structure of a compound transposon and that at least this part of the O139 *wbf* cluster was transferred by an IS1358-based compound transposon (37, 38). One of the prerequisites of this hypothesis is the ability of IS1358 to transpose. The observation that the IS1358 present in the O139 *wbf* region has an apparently intact transposase ORF (37) suggested that it might be an active element. We were interested in characterizing the transposition mechanism exhibited by IS1358 in order to understand the origin of the O139 *wbf* region. We constructed a derivative of IS1358 containing a Tet^r marker and an inducible source of the transposase (see Materials and Methods). Using a conventional mating-out assay and F' derivative pOX38-Gen^r as a target for transposition, we measured the transposition of IS1358::Tet^r onto pOX38. The frequency of pOX38 transfer ranged from 30 to 100%, and that of Tet^r colonies among the pOX38 exconjugants was less than 10⁻⁷ to 10⁻⁸, which was the background level (i.e., the frequency observed in the absence of transposase expression). On further analysis, these Tet^r colonies were found to be the donor mutants that acquired the counterselection resistance marker. Hence, we could not detect any true transposition of IS1358. The same *tnp* gene fragment in a T7 vector overexpressed the transposase protein, as seen in sodium dodecyl sulfate-polyacrylamide gels (data not shown). We conclude from these experiments that the IS1358 is defective for transposition in *E. coli* and presumably in *V. cholerae* as well. The requirement for specific *V. cholerae* factors for transposition of this element or the possibility that any *cis* element of IS1358 necessary for transposition may have been inactivated due to the insertion of Tet^r marker cannot be ruled out at this time. However, these results are consistent with a recent study showing that IS1358 by itself is transposition defective (cointegrate formation), although a composite transposon with two IS1358 copies is capable of direct transposition (13).

DISCUSSION

In this work we have defined the right junction of the *wb** gene cluster involved in the synthesis of O antigen. The gene at the left junction of the *wb** cluster, *gmhD*, has been previously shown to be essential since inactivation of this gene was lethal (36). Several lines of evidence indicates that the right junction gene, *rjg*, is not essential for cell viability or for LPS/CPS synthesis in *V. cholerae*. Four variants of the *rjg* gene have been found that contain IS elements and an internal deletion. These *rjg* mutants are presumably inactive, thus indicating the non-essential nature of this gene. The G+C content of the *rjg* gene (53%) is well above the average G+C content of the *wbf* region (42%), which indicates that it is not part of the *wbf* region. Generally, it is known that the G+C content of the *wb** regions is less than that of the chromosomal average (30). Based on sequence analysis, *rjg* does not resemble any polysaccharide biosynthesis gene but rather seems to be an mRNA 3'-end processing factor. Clones lacking the genes downstream of *wbfX* produce essentially the same LPS/CPS structures in *E. coli* (47). From these results, we conclude that the *gmhD* and *rjg* genes form the left and right boundaries, respectively, of the O-antigen gene cluster. Analysis of *rjg* in various serogroups suggests that it is a hot spot for DNA rearrangements, since we have found three independent IS elements and a deletion allele in this gene in a set of 81 strains. On the other hand, the left junction genes *gmhD* and *orf2* are intact. Hence, *rjg* seems to function as an anchor region for a variety of DNA transfer events, most likely in creating diversity in the *wb** region.

The emergence of *V. cholerae* serogroup O139 as an epidemic strain represents the first known example of horizontal

transfer of a polysaccharide biosynthetic gene cluster in *V. cholerae*. Based on sequence analysis, three serogroups (O1, O22, and O139) appear to have similar cassette-like organization of the *wb** cluster. Earlier it was shown that yet another pathogenic strain (serogroup O37), which caused a local outbreak of cholera in Sudan in 1962, has a very similar genetic organization of *wb** genes (9). Data presented here suggest that many strains of *V. cholerae* serogroups may possess a similar organization.

The O-antigen cluster resembles pathogenicity islands (PAIs) in many respects; such a proposal has been advanced previously (24). They both have a specific location on the chromosome. The *wb** cluster is located between *gmhD* (*rfaD*) and *rjg*. The *wb** genes have varying G+C content, suggesting a horizontal transfer, although a regulatory mechanism based on codon usage cannot be excluded. Like PAIs, *wb** gene clusters also have IS elements, which suggests genetic rearrangements and horizontal transfer. The presence or absence of pathogenicity islands determines the virulent or avirulent nature of the organism. *V. cholerae* strains without a "*wb** island" have not yet been found. However, the *wb** cluster makes a major contribution to virulence since mutations in this region results in attenuation of virulence in animal models (18). The different O-antigen clusters seem to provide serogroup specificity and a mechanism to evade host immune detection. For example, the majority of O139 infections occurred in adults, who were presumably immune to O1 when O139 first emerged. As the epidemic continued, the majority of the cases shifted back to persons <15 years of age as people became immune (11, 19).

The absence of phage *att*-like sites in the O139 *wbf* cluster suggests that the mechanism and the mode of horizontal transfer may be different from that of PAIs, which are thought to be transferred as phages. Two mechanisms have been proposed for the acquisition of O-antigen clusters: IS-mediated transposition (26, 38, 39) and homologous recombination (11, 26, 38). It has been proposed that two copies of IS1358 are present in the O139 *wbf* region (a degenerate copy just upstream of IS1358) and that this region was transferred as a transposable element (38). However, we find no evidence of a compound transposon or transposition of IS1358 in an *E. coli* background. Recently, a composite transposon with two copies of IS1358 has been shown to transpose, although the individual IS element has not been shown to transpose successfully (13). A preponderance of evidence suggests that the latter mechanism most likely is operative in the transfer of large DNA segments, although the vector for such a transfer is unknown. It could be a generalized transducing phage or a conjugative plasmid (26). Since the genes flanking the *wb** region in O1/O139 strains are conserved in all of the strains that we analyzed, we conclude that O139 arose from an O1 El Tor strain by a double homologous recombination event with DNA from a nonpathogenic donor that had the O139 *wbf* cluster. The IS elements probably are involved in local genetic rearrangements by transposing small gene segments. It is also probable that IS elements provide portable regions of homology within nonhomologous regions, in order to facilitate genetic rearrangements (26). Such an event has been shown in the case of the *wba* genes of *Salmonella enterica* (46). The preferential linkage of IS1358 to the *wb** region in the majority of the *V. cholerae* strains that we analyzed further supports the idea that IS1358 was acquired by homologous recombination rather than by transposition. A random transposition event is expected to deliver the IS element to any part of the chromosome. A major hurdle to any horizontal transfer mechanism is that the transferred DNA has to overcome barriers such as restriction and recombination in

order to be established in the recipient strain. Whatever the mechanism may be, horizontal transfer of the *wb** cluster led to the emergence of a new pathogen from an already existing pathogen in a single step by evading the host immune system. Our future studies will focus on these mechanisms and the barriers to horizontal gene transfer in *V. cholerae*.

ACKNOWLEDGMENTS

We thank Nick Ambulos and Lisa Sadzewicz (UMAB Biopolymer Laboratory) for automated sequencing. We thank De Qi Xu and Laurie Comstock for sharing unpublished data and Nancy L. Craig and Joseph Eric Peeters (Johns Hopkins University) for providing the *E. coli* strains used in the transposition assays. We thank Arnold Kreger and Rick Blank for comments on the manuscript. *V. cholerae* AI1837 was kindly provided by M. John Albert (International Center for Diarrheal Disease Research, Bangladesh).

This work was supported by PHS grants AI135729 to J.A.J., AI19716 to J.B.K., and AI28856 to J.G.M. and by a VA/DOD grant on emerging infectious diseases to J.G.M.

REFERENCES

- Albert, M. J., A. K. Siddique, M. S. Islam, A. S. G. Faruque, M. Ansaruz-zaman, S. M. Faruque, and R. B. Sack. 1993. Large outbreak of clinical cholera due to *Vibrio cholerae* non-O1 in Bangladesh. *Lancet* **341**:704.
- Albert, M. J., M. Ansaruzzaman, P. K. Bardhan, A. S. G. Faruque, S. M. Faruque, M. S. Islam, D. Mahalanabis, R. B. Sack, M. A. Salam, A. K. Siddique, M. D. Yunus, and K. Zaman. 1993. Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. *Lancet* **342**:387-390.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389-3402.
- Barua, D. 1992. History of cholera, p. 1-36. In D. Barua and W. B. Greenough III (ed.), *Cholera*. Plenum Publishing Corp., New York, N.Y.
- Berche, P., C. Poyart, E. Abachin, H. Lelievre, J. Vandepitte, A. Dodin, and J.-M. Fournier. 1994. The novel epidemic strain O139 is closely related to pandemic strain O1 of *Vibrio cholerae*. *J. Infect. Dis.* **170**:701-704.
- Bhattacharya, S. K., M. K. Bhattacharya, G. B. Nair, D. Dutta, A. Deb, T. Ramamurthy, S. Garg, P. K. Saha, P. Dutta, A. Moitra, B. K. Mandel, T. Shimada, Y. Takeda, and B. C. Deb. 1993. Clinical profile of acute diarrhoea cases infected with the new epidemic strain of *V. cholerae* O139: designation of disease as cholera. *J. Infect. Dis.* **27**:11-15.
- Bik, E. M., A. E. Bunschoten, R. D. Gouw, and F. R. Mooi. 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. M., A. E. Bunschoten, R. J. L. Willems, A. C. Y. Chang, and F. R. Mooi. 1996. Genetic organization and functional analysis of the *otn* DNA essential for cell-wall polysaccharide synthesis in *Vibrio cholerae* O139. *Mol. Microbiol.* **20**:799-811.
- Bik, M., R. D. Gouw, and F. R. Mooi. 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.
- Burrows, L. L., D. F. Charter, and J. S. Lam. 1996. Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAI) B-band lipopolysaccharide gene cluster. *Mol. Microbiol.* **22**:481-495.
- Cholera Working Group. 1993. Large epidemics of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. *Lancet* **342**:387-390.
- Comstock, L. E. Personal communication.
- Comstock, L. E., J. A. Johnson, J. M. Michalski, J. G. Morris, Jr., and J. B. Kaper. 1996. Cloning and sequence of a region encoding a surface polysaccharide of *Vibrio cholerae* O139 and characterization of the insertion site in the chromosome of *Vibrio cholerae* O1. *Mol. Microbiol.* **19**:815-826.
- Dumontier, S., P. Trieu-Cuot, and P. Berche. 1998. Structural and functional characterization of *IS1358* from *Vibrio cholerae*. *J. Bacteriol.* **180**:6101-6106.
- Fallarino, A., C. Mavrangelos, U. H. Stroecher, and P. A. Manning. 1997. Identification of additional genes required for O-antigen biosynthesis in *Vibrio cholerae* O1. *J. Bacteriol.* **179**:2147-2153.
- Favre, D., S. J. Cryz, Jr., and J.-F. Viret. 1996. Construction and characterization of a potential live oral carrier-based vaccine against *Vibrio cholerae* O139. *Infect. Immun.* **64**:3565-3570.
- Glass, R. I., S. Becker, M. I. Huq, B. J. Soll, M. U. Khan, M. H. Merson, J. V. Lee, and R. E. Black. 1982. Endemic cholera in rural Bangladesh, 1966-1980. *Am. J. Epidemiol.* **116**:959-970.
- Hayashi, M., J. Fujii, and T. Unemoto. 1997. Genetic defect of the sodium pump-defective mutant nap-1 from the marine *Vibrio alginolyticus*. *Biochem. Mol. Biol. Intl.* **41**:41-47.
- Iredell, J. R., U. H. Stroecher, H. M. Ward, and P. A. Manning. 1998. Lipopolysaccharide O-antigen expression and the effect of its absence on virulence in *rfb* mutants of *Vibrio cholerae* O1. *FEMS Immunol. Med. Microbiol.* **20**:45-54.
- Jesudason, M. V., and T. J. John. 1996. The Vellore vibrio watch. *Lancet* **347**:1493-1494.
- Johnson, J. A., C. A. Salles, P. Panigrahi, M. J. Albert, A. C. Wright, R. J. Johnson, and J. G. Morris, Jr. 1994. *Vibrio cholerae* O139 synonym Bengal is closely related to *Vibrio cholerae* El Tor but has important differences. *Infect. Immun.* **62**:2108-2110.
- Kaper, J. B., J. G. Morris, Jr., and M. M. Levine. 1995. Cholera. *Clin. Microbiol. Rev.* **8**:48-86.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Manning, P. A., U. H. Stroecher, and R. Morona. 1994. Molecular basis for O-antigen biosynthesis in *Vibrio cholerae* O1: Ogawa-Inaba switching, p. 77-94. In I. K. Wachsmuth, P. A. Blake, and O. Olsvik (ed.), *Vibrio cholerae* and cholera: molecular to global perspectives. American Society for Microbiology, Washington, D.C.
- Mecasas, J., and E. J. Strauss. 1996. Molecular mechanisms of bacterial virulence: type III secretion and pathogenicity islands. *Emerg. Infect. Dis.* **2**:271-288.
- Monaco, A. P. 1995. *Pulse field gel electrophoresis: a practical approach*. Oxford University Press, Oxford, England.
- Mooi, F. R., and E. M. Bik. 1997. The evolution of epidemic *Vibrio cholerae* strains. *Trends Microbiol.* **4**:161-165.
- Morris, J. G., Jr., G. E. Losonsky, J. A. Johnson, C. O. Tacket, J. P. Nataro, P. Panigrahi, and M. M. Levine. 1995. Clinical and immunological characteristics of *Vibrio cholerae* O139 Bengal infection in North American volunteers. *J. Infect. Dis.* **171**:903-908.
- Popovic, T., P. I. Fields, O. Olsvik, J. G. Wells, G. M. Evins, D. N. Cameron, J. J. Farmer III, C. A. Bopp, K. Wachsmuth, R. B. Sack, M. J. Albert, G. B. Nair, T. Shimada, and J. C. Feeley. 1995. Molecular subtyping of toxigenic *Vibrio cholerae* O139 causing epidemic cholera in India and Bangladesh, 1992-1993. *J. Infect. Dis.* **171**:122-127.
- Ramamurthy, T., S. Garg, R. Sharma, S. K. Bhattacharya, G. B. Nair, T. Shimada, T. Takeda, T. Karasawa, H. Kurazano, A. Pal, and Y. Takeda. 1993. Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* **341**:703-704.
- Schnaitman, C. A., and J. D. Klena. 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**:655-682.
- Sharma, C., M. Thungapathra, A. Ghosh, A. K. Mukhopadhyay, A. Basu, R. Mitra, I. Basu, S. K. Bhattacharya, T. Shimada, T. Ramamurthy, T. Takeda, S. Yamasaki, Y. Takeda, and G. B. Nair. 1998. Molecular analysis of non-O1, non-O139 *Vibrio cholerae* associated with an unusual upsurge in the incidence of cholera-like disease in Calcutta, India. *J. Clin. Microbiol.* **36**:114-117.
- Shimada, T., G. B. Nair, B. C. Deb, M. J. Albert, R. B. Sack, and Y. Takeda. 1993. Outbreak of *Vibrio cholerae* non-O1 in India and Bangladesh. *Lancet* **341**:1347.
- Shimada, T., E. Arakawa, K. Itoh, T. Okitsu, A. Matsushima, Y. Asai, S. Yamai, T. Nakazato, G. B. Nair, M. J. Albert, and Y. Takeda. 1994. Extended serotyping scheme for *Vibrio cholerae*. *Curr. Microbiol.* **28**:175-178.
- Smith, H. L. 1979. Serotyping of non-cholera vibrios. *J. Clin. Microbiol.* **10**:85-90.
- Stroecher, U. H., L. E. Karageorgos, R. Morona, and P. A. Manning. 1992. Serotype conversion in *Vibrio cholerae* O1. *Proc. Natl. Acad. Sci. USA* **89**:2566-2570.
- Stroecher, U. H., L. E. Karageorgos, R. Morona, and P. A. Manning. 1995. In *Vibrio cholerae* serogroup O1, *rfaD* is closely linked to the *rfb* operon. *Gene* **155**:67-72.
- Stroecher, U. H., K. E. Jedani, B. K. Dredge, R. Morona, M. H. Brown, L. E. Karageorgos, J. M. Albert, and P. A. Manning. 1995. Genetic rearrangement of the *rfb* regions of *Vibrio cholerae* O1 and O139. *Proc. Natl. Acad. Sci. USA* **92**:10374-10378.
- Stroecher, U. H., G. Parasivam, B. K. Dredge, and P. A. Manning. 1997. Novel *Vibrio cholerae* O139 genes involved in lipopolysaccharide biosynthesis. *J. Bacteriol.* **179**:2740-2747.
- Stroecher, U. H., and P. A. Manning. 1997. *Vibrio cholerae* serotype O139: swapping genes for surface polysaccharide biosynthesis. *Trends Microbiol.* **5**:178-180.
- Stroecher, U. H., K. E. Jedani, and P. A. Manning. 1998. Genetic organization of the regions associated with surface polysaccharide synthesis in *Vibrio cholerae* O1, O139 and *Vibrio anguillarum* O1 and O2: a review. *Gene* **223**:269-282.
- Waldor, M. K., R. Colwell, and J. J. Mekalanos. 1994. The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants. *Proc. Natl. Acad. Sci. USA* **91**:11388-11392.
- Waldor, M. K., and J. J. Mekalanos. 1994. *Vibrio cholerae* O139 specific gene sequences. *Lancet* **343**:1366.
- Wang, R. F., and S. R. Kushner. 1991. Construction of versatile low-copy

- number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* **100**:195–199.
44. **World Health Organization**. 1969. Outbreak of gastro-enteritis by nonagglutinable (NAG) *Vibrio*. *Weekly Epidemiol. Rev.* **44**:1–28.
 45. **Weintraub, A., G. Widmalm, P.-E. Jansson, M. Jansson, K. Hultenby, and M. J. Albert**. 1994. *Vibrio cholerae* O139 Bengal possesses a capsular polysaccharide which may confer increased virulence. *Microb. Pathog.* **16**:235–241.
 46. **Xiang, S.-H., M. Hobbs, and P. R. Reeves**. 1994. Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J. Bacteriol.* **176**:4357–4365.
 47. **Xu, D. Q.** Personal communication.
 48. **Yamasaki, S., T. Shimizu, K. Hoshino, S. Ho, T. Shimada, G. B. Nair, and Y. Takeda**. 1998. *Vibrio cholerae* genes for O-antigen synthesis, strain O22, complete cds. GenBank accession no. ABO12957.

Editor: D. L. Burns