INFECTION AND IMMUNITY, May 2002, p. 2441–2453 Vol. 70, No. 5
0019-9567/02/$04.00+0 DOI: 10.1128/IAI.70.5.2441–2453.2002

Evidence for the Emergence of Non-O1 and Non-O139 *Vibrio cholerae* Strains with Pathogenic Potential by Exchange of O-Antigen Biosynthesis Regions

Manrong Li,[1,2]† Toshio Shimada,[3] J. Glenn Morris, Jr.,[1,2] Alexander Sulakvelidze,[1] and Shanmuga Sozhamannan[1,2]*

Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine,[1] and VA Maryland Health Care System,[2] Baltimore, Maryland 21201, and Laboratory of Enteric Infection 1, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan[3]

Received 8 November 2001/Returned for modification 20 December 2001/Accepted 18 January 2002

The novel epidemic strain *Vibrio cholerae* O139 Bengal originated from a seventh-pandemic O1 El Tor strain by antigenic shift resulting from homologous recombination-mediated exchange of O-antigen biosynthesis (*wb) clusters. Conservation of the genetic organization of *wb** regions seen in other serogroups raised the possibility of the existence of pathogenic non-O1 and non-O139 *V. cholerae* strains that emerged by similar events. To test this hypothesis, 300 *V. cholerae* isolates of non-O1 and non-O139 serogroups were screened for the presence of virulence genes and an epidemic genetic background by DNA dot blotting, IS*1004* fingerprinting, and restriction fragment length polymorphism (RFLP) analysis. We found four non-O1 strains (serogroups O27, O37, O53, and O65) with an O1 genetic backbone suggesting exchange of *wb** clusters. DNA sequence analysis of the O37 *wb** region revealed that a novel ~23.4-kb gene cluster had replaced all but the ~4.2-kb right junction of the 22-kb O1 *wbe* region. In sharp contrast to the backbones, the virulence regions of the four strains were quite heterogeneous; the O53 and O65 strains had the El Tor vibrio pathogenicity island (VPI) cluster, the O37 strain had the classical VPI cluster, and the O27 strain had a novel VPI cluster. Two of the four strains carried CTXϕ; the O27 strain possessed a CTXϕ with a recently reported immune specificity (*rstR-4*** allele) and a novel *ctxB* allele, and the O37 strain had an El Tor CTXϕ ($rstR^{ET}$ allele) and novel *ctxAB* alleles. Although the O53 and O65 strains lacked the *ctxAB* genes, they carried a pre-CTXϕ (i.e., $rstR^{cla}$). Identification of non-O1 and non-O139 serogroups with pathogenic potential in epidemic genetic backgrounds means that attention should be paid to possible future epidemics caused by these serogroups and to the need for new, rapid vaccine development strategies.**

Cholera is a diarrheal disease caused by the gram-negative bacterium *Vibrio cholerae,* and an estimated 120,000 deaths from cholera occur globally every year (58). Cholera is both an endemic and epidemic disease and is the only bacterial pandemic disease known in modern times. One of the etiological agent's major protective antigens appears to be the O antigen (27), and the enormous serological diversity of *V. cholerae* is shown by the fact that more than 200 O-antigen serogroups have been identified (43). Interestingly, only the O1 and O139 serogroups are known to cause epidemic and pandemic disease, although occasional outbreaks caused by non-O1 and non-O139 strains have been reported in the past. For example, strains of serogroup O37 were responsible for localized outbreaks in 1965 in Czechoslovakia (2) and in 1968 in the Sudan (26). Seven pandemics have been recorded in the history of cholera; the sixth pandemic was caused by O1 strains of the classical biotype, and the seventh pandemic, which started in 1961 and continues to the present time, is caused by O1 El Tor biotype strains (27). The only other serogroup known to cause epidemic cholera (O139) emerged in 1992 in the Bay of Bengal region and has remained endemic to this region (1, 41). The O139 Bengal strains cause disease with severity comparable to the severity of the disease caused by O1 strains, and prior exposure to O1 strains does not provide protection against O139 infections (7, 37). Molecular epidemiological studies (ribotyping, fingerprinting, multilocus enzyme electrophoresis, etc.) have indicated that the O139 strains have genetic backbones very similar to those of the O1 El Tor Asian seventh-pandemic strains (6, 25, 56).

The O1 O-antigen biosynthesis genes of *V. cholerae* are organized in a cluster (*wbe* cluster) on chromosome I, between open reading frames (ORFs) VC0240 (*gmhD*) and VC0264 (*rjg*) (22). DNA sequence analyses of the *wb** clusters of two other serogroups (O22 and O139) revealed a similar organization of this region; i.e., serogroup-specific genes are flanked by *gmhD* (which encodes D-glycero-D-manno-heptose 1-phosphate guanosyltransferase, involved in lipopolysaccharide core biosynthesis) at the left junction and by *rjg* (which encodes a conserved hypothetical protein with similarities to mRNA 3′ end processing factor) at the right junction (9, 14, 18, 33, 48, 59). These data led to the idea that the *V. cholerae* O139 Bengal strain originated from an O1 strain by homologous recombination-mediated replacement of the *wbe* region of an O1 strain with the O139 *wbf* region (36, 47, 49). However, the donor or the vehicle for this horizontal transfer event is not yet known. An O22 serogroup strain has been proposed to be a

* Corresponding author. Present address: Intralytix, Inc., The Columbus Center, 701 E. Pratt St., Room 4016, Baltimore, MD 21201. Phone: (410) 625-2422. Fax: (410) 625-2506. E-mail: ssozhamannan@intralytix.com.

† Present address: Intralytix, Inc., Baltimore, MD 21201.

TABLE 1. Characterization of non-O1 and non-O139 *V. cholerae* strains with pathogenic potential[a]

Strain	Country	Year	Source	Sakazaki serogroup	CTXϕ			VPI	
					ctxA	*ctxB*	*rstR*	Cluster[b]	*tcpA*
395	India	1966	Diarrhea	O1 cla	WT	cla	cla	cla	cla
NIH35A3	India	1941	Diarrhea	O1 cla	WT	cla	cla	cla	cla
5011	Unknown		H. Smith collection	O1 cla (O333)[c]	WT	cla	cla	cla	cla
N16961	Bangladesh	1975	Diarrhea	O1 ET	WT	ET	ET	ET	ET
E7946	Bahrain	1978	Diarrhea	O1 ET	WT	ET	ET	ET	ET
365-96	Japan	1996	Prawn, import from Thailand	O27	WT	NT	*rstR-4***[d]	NT	NT[e]
1322-69	India	1969	Diarrhea	O37	NT[f]	NT	ET	cla	cla
8585	Iraq	1966	Diarrhea	O53 (O340)[c]			cla	ET	ET
981-75	India	1975	Diarrhea	O65			cla	ET	ET
63-93 (MO45)	India	1992	Diarrhea	O139	WT	ET	ET	ET	ET
AM2	India	1995	Diarrhea	O9					
AM107	India	1996	Diarrhea	O144					
NRT36-S	Japan	1990	Diarrhea	O31					

[a] Abbreviations: wt, wild type; cla, classical; ET, El Tor; NT, novel type.
[b] Presence of the entire VPI cluster based on restriction mapping and hybridization.
[c] O333 and O340 are Smith serogroups.
[d] *rstR-4*** = SCE223 (38).
[e] Differs from the *tcpA-env* allele described by Mukhopadhyay et al. (38) at one position ($^{O27}V9D^{env}$).
[f] There is a single amino acid substitution ($^{wt}S46N^{O37}$).

possible donor since its *wb** region shares extensive homology with the O139 *wbf* region (59), and a generalized transducing phage or a conjugative plasmid is the speculated vector (36, 47).

DNA fingerprinting and phylogenetic analyses of *V. cholerae* strains have established that there is a lack of correlation between serogroup and phylogeny (8, 46); i.e., strains belonging to various serogroups appear to fall in the same phylogenetic clade, and strains belonging to the same serogroup have been found in many different clades. These data support the hypothesis that there are frequent horizontal transfers of O-antigen clusters among non-O1 and non-O139 *V. cholerae* strains. However, such transfers into epidemic strains seem to have been limited, since O139 is the only known example of O-antigen transfer into an epidemic strain.

Two critical virulence factors have been associated with epidemic strains. Cholera toxin is the primary virulence factor responsible for the severe diarhheal symptoms (27), and the toxin coregulated pilus (TCP) is the primary factor responsible for efficient colonization of the human intestinal tract (52, 53). In a landmark study, Waldor and Mekalanos (57) showed that the cholera toxin genes (*ctxAB*) are carried on a filamentous, f1-like, single-stranded DNA phage, designated CTXϕ. Also, it has been demonstrated (42, 57) that TCP serves not only as a colonization factor but also as the receptor for CTXϕ. Recently, the *tcpA* gene has been shown (28) to be located on a pathogenicity island designated the vibrio pathogenicity island (VPI) and has been reported to be predominantly associated with epidemic and pandemic strains. Interestingly, the VPI has also been proposed (29) to be a filamentous phage, designated VPIϕ. Despite the lack of further evidence of the existence of VPIϕ, this idea raises the interesting possibility that there is phage-phage interaction in horizontal gene transfer (54).

Several *Vibrio mimicus* strains carrying VPI and CTXϕ have been identified (12), and the remarkable identity at the sequence level of some of the *V. mimicus* and El Tor VPI genes (*aldA* and *toxT*) suggested that there was recent interspecies horizontal transfer of these factors between *V. cholerae* and *V. mimicus*. Identification of several non-O1 and non-O139 serogroup *V. cholerae* strains containing the *tcpA* gene (13, 39, 40) suggests that these strains represent the environmental reservoirs of this virulence factor. Recently, extensive analysis (38) of the VPI and CTX prophage regions of several environmental *V. cholerae* strains has been described, although the mechanism(s) of the origin of these strains has not been addressed.

Conservation of the genetic organization of the *wb** region raises the possibility that non-O1 and non-O139 *V. cholerae* strains with an epidemic genetic background may have arisen by exchange of O-antigen biosynthesis regions. In order to evaluate this hypothesis, we analyzed 300 *V. cholerae* strains in all of the 194 known serogroups, and we found several non-O1 and non-O139 strains possessing *ctxAB* and *tcpA* genes. Four of these strains appeared to have a genetic background similar to that of the epidemic strains. DNA sequencing of the O-antigen cluster in one of the strains (O37 serogroup) revealed that most of the O1 *wbe* region had been replaced by a novel *wb** cluster. Thus, homologous recombination-mediated O-antigen shift appears to be a general mechanism for the emergence of novel virulent strains of *V. cholerae*.

MATERIALS AND METHODS

Bacterial strains, growth conditions, and primers. The bacterial strains used in this study included strains in the Shimada type culture collection (194 serotypes) (43), 36 clinical *V. cholerae* isolates, and 70 clinical and environmental strains from the Smith collection (44). The strains characterized in this study are listed in Table 1, and the primers used in various analyses are listed in Table 2. Media and culture conditions have been described previously (45).

DNA dot blotting and PFGE. DNA dot blot analysis and pulsed-field gel electrophoresis (PFGE) were carried out as described previously (45).

Long-range PCR. In order to determine the lengths of the *wb** regions in various serogroup strains, long-range PCR was performed with an XL-PCR kit (Perkin-Elmer Cetus Corp., Foster City, Calif.). PCRs were performed by using 1 μg of genomic DNA as the template, primers J 101 and J 103, and the protocol and conditions recommended by the manufacturer. The PCR products were electrophoresed in 0.5% agarose gels and stained with ethidium bromide.

Sequencing of the O37 *wb region.** The entire DNA sequence of the region between the *gmhD* and *rjg* genes of an O37 serogroup strain was obtained from the XL-PCR product. The XL-PCR fragment amplified from an O37 serogroup

TABLE 2. List of primers

Primer	Gene	Sequence	Reference(s)
J 101	*gmhD*$_1$	5′-GCCATCCCACTCTGTGGTCGCAGAGCAAGCTCC-3′	14
J 103	*rjg*$_2$	5′-CCCGTGACACTCGCCTTCCCTCCGTGATGAACC-3′	14
M 177	*gmhD*$_1$	5′-TTACTTACGATTAATCAGCGCCAT-3′	45
M 178	*gmhD*$_2$	5′-GGCGGCGCTGGCATGATTGGCAGC-3′	45
M 179	*rjg*$_1$	5′-CATGGAAGTGGTTCATCACGGAGG-3′	45
J 414	*rjg*$_2$	5′-GTGGACGCGTTCAAAGCACCGAATATCCGAGTT-3′	45
M 310	*orf2*$_1$	5′-GGTGACATCAAAGGGACCACTTTTTC	22; this study
M 311	*orf2*$_2$	5′-GGTGTATGCCACTAGTGTAGGTAAT	22; this study
M 459	IS*1004*$_1$	5′-CCCCAGCTTTTGACGCTTATTGTGAACGT-3′	22; this study
M 460	IS*1004*$_2$	5′-GATCGATATCTTTCTAACTTCTGTATAAGG-3′	22; this study
M 277	*ctxA*$_1$	5′-ACAGAGTGAGTACTTTGACC-3′	22; this study
M 278	*ctxA*$_2$	5′-ATACCATCCATATATTTGGGAG-3′	22; this study
S 86	*ctxAB*$_1$	5′-GGCTGTGGGTAGAAGTGAAACGG-3′	22; this study
S 87	*ctxAB*$_2$	5′-CTAAGGATGTGGAATAAAAACATC-3′	22; this study
M 279	*tcpA*$_1$	5′-AAAACCGGTCAAGAGGG-3′ (same as KAR 24)	28
M 280	*tcpA*$_2$	5′-CAAAAGCTACTGTGAATGG-3′ (same as KAR 25)	28
M 281	*tcpA*$_3$	5′-CAAATGCAACGCCGAATGG-3′ (same as KAR 82)	28
M 590	*tcpA*$_{L1}$	5′-GATCGCATGCCAGAGTTCTATCTTTCGTC-3′	22; this study
M 591	*tcpA*$_{L2}$	5′-GATCGTCGACATAGTGATAAGAGTCTTACCC-3′	22; this study
M 668	*smt*$_1$ (*smt*-VCA0198)	5′-CCGAAATACGGTCATTACTTGGGC-3′	22; this study
M 669	*smt*$_2$	5′-CACTTCATTATTCCCGTAAGCAGC-3′	22; this study
M 680	*smt*$_{1.1}$ (*nupC*-VCA0179)	5′-AATAGCCAATCACGCACCAAG-3′	22; this study
M 681	*smt*$_{1.2}$	5′-TAATCGCACTGCGGCTTTCAG-3′	22; this study
M 682	*smt*$_{2.1}$ (*hmpA*-VCA0183)	5′-TGACCCACCAGAAAACCGGAC-3′	22; this study
M 683	*smt*$_{2.2}$	5′-GCGCCTTATCCACACCAAGCG-3′	22; this study
M 684	*smt*$_{3.1}$ (*rhlE*-VCA0204)	5′-CGCTCAATCGCAAATAATTCC-3′	22; this study
M 685	*smt*$_{3.2}$ (*dcuB*-VCA0205)	5′-TGCTCTCTCTCCCCAAATGAC-3′	22; this study
M 686	*smt*$_{4.1}$ (VCA0206)	5′-GTATTGTCGGATTTCATTTGC-3′	22; this study
M 687	*smt*$_{4.2}$ (VCA0208)	5′-AGTGACGGCCTCTGGCGGAGC-3′	22; this study
M 688	*smt*$_{5.1}$ (*hlyA*-VCA0219)	5′-GGGTTCCGCGACACCGGATGC-3′	22; this study
M 689	*smt*$_{5.2}$	5′-TGTTTAATGGCTATGTTGACG-3′	22; this study
M 698	*ctxrgn*$_{1.1}$ (VC1444)	5′-TAATCTGCTATTTCACTGAAG-3′	22; this study
M 699	*ctxrgn*$_{1.2}$	5′-TTCCTGAGTGATCCCCAATCC-3′	22; this study
M 700	*ctxrgn*$_{2.1}$ (VC1451-*rtxA*)	5′-GCGGAAAAGCTGAAAGGCACC-3′	22; this study
M 701	*ctxrgn*$_{2.2}$	5′-ACCTTCATGGTGTGAAATCAC-3′	22; this study
M 702	*ctxrgn*$_{3.1}$ (VC1465)	5′-CCGCTGTCTCAATAGAACCTG-3′	22; this study
M 703	*ctxrgn*$_{3.2}$	5′-GGACATCATACAAGAGAAGAC-3′	22; this study
M 704	*ctxrgn*$_{4.1}$ (VC1470)	5′-GAACATGAACCTTAATGCGAG-3′	22; this study
M 705	*ctxrgn*$_{4.2}$	5′-CACGTCATTTATGAATTACGG-3′	22; this study
M 706	*ctxrgn*$_{5.1}$ (VC1476-VC1477)	5′-GGTATCAGCATGAGACTTTTTTG-3′	22; this study
S 122	*ctx* core (*orfU*)	5′-CGTCACACCAGTTACTTTTCG-3′	22; this study
S 123	*ctx* core (*zot*)	5′-AACCCCGTTTCACTTCTAC-3′	22; this study
M 707	*ctxrgn*$_{5.2}$	5′-CCAATAGTGATAACTACTTCG-3′	22; this study
M 452	*ald*$_2$	5′-TTTTCTTGATTGTTAGGATGC-3′	12
M 453	*ald*$_1$	5′-ATTCTTCTGAGGATTGCTGAT-3′	12
M 644	*tagD*$_1$	5′-GCGGTGACACTAAAGTAGTGTTTG-3′	22; this study
M 645	*tagD*$_2$	5′-GATGGTCAGATAAAAGAACGCAGG-3′	22; this study
M 664	*tcpAdn*$_1$	5′-TTCGCAATTACAGTCGGTGGCTTG-3′	22; this study
M 665	*tcpAdn*$_2$	5′-AGCCAACTCAGTTAAAACTTGTTC-3′	22; this study
M 448	*toxT*$_2$	5′-CTTGGTGCTACATTCATGG-3′	12
M 449	*toxT*$_1$	5′-AGGAGATGGAAGTGGTGTG-3′	12
M 646	*vpi0845*$_1$	5′-ATCATTCCAGATAAAGTTACGCAGA-3′	22; this study
M 647	*vpi0845*$_2$	5′-TCTACTTCCGGCTTCCCTGCCACG-3′	22; this study
M 407	*rstR1*	5′-GACGTAGCGTGCGGAGTCGCGTTG-3′	22; this study
M 408	*rstR2*	5′-TGAAGCATAAGGAACCGACCAAGC-3′	22; this study
M 573	*rstA1*	5′-ACTCGATACAAACGCTTCTC-3′	22; this study
M 574	*rstA2*	5′-AGAATCTGGAGGTTGAGTG-3′	22; this study

strain, 1322-69, was digested with *Pst*I, and the resulting fragments were gel purified, cloned in the pBluescript vector (Stratagene Corp., La Jolla, Calif.), and sequenced by primer walking. The fragments were aligned using restriction maps of the XL-PCR fragment. The order of the *Pst*I sites was confirmed by targeted PCR and sequencing of the PCR fragments. The XL-PCR end fragments were PCR amplified, cloned in pCR2.1, and sequenced. The final aligned sequence of the O37 *wb** region was analyzed by the DNASIS program (Hitachi Software Engineering Co., Ltd., South San Francisco, Calif.) in order to identify the ORFs, and the individual ORFs were searched by using the National Center for Biotechnology Information Blast program (3) for identifying protein similarities.

Isolation of *wb regions.** In order to isolate the entire *wb** region on a single restriction fragment, a unique *Not*I site in the *rjg* gene at the right junction of the *wb** region was utilized. A second *Not*I site was introduced at the left junction. In order to accomplish this, the genes *orf-2* (VC0239) and *gmhD* (VC0240) were cloned on either side of the *Not*I site of the pBluescript vector, and a Kan^r cassette was introduced between *orf-2* and *gmhD*. A *Sac*I-*Eco*RV fragment con-

taining *orf-2*–*Not*I–Kan^r–*gmhD* was blunt ended and cloned into *Sal*I-digested and blunt-ended suicide vector pCVD442. The *Not*I site in the derivative of pVCD442 was introduced into the chromosomes of selected strains, as previously described (17), by the sucrose selection procedure. The resulting strains were subjected to PFGE after *Not*I digestion of their DNAs in agarose plugs as previously described (45). Southern analysis was performed by transfer of the restriction enzyme-digested chromosomal DNA fragments onto a nitrocellulose membrane, followed by hybridization with probes prepared by enhanced chemiluminescence (Amersham Biosciences, Piscataway, N.J.).

IS*1004* fingerprinting. Genomic DNAs were digested with *Hpa*II, transferred onto a Zeta-probe membrane (Bio-Rad Laboratories, Hercules, Calif.), and hybridized with an IS*1004* probe. Direct amplification of the IS*1004* element from genomic DNA was unsuccessful, and hence, the following procedure was used to clone the IS*1004* element. Primers M 459 and M 460 were used to PCR amplify a fragment (2.2 kb) encompassing an IS*1004* copy and the neighboring sequences, and the fragment was cloned into the pCR2.1 vector (Invitrogen Life Technologies, Carlsbad, Calif.). The resulting plasmid was digested with *Eco*RI, and this was followed by purification of the *Eco*RI fragment containing the genomic sequences and not the plasmid sequences. This fragment was further digested with *Hae*III and *Avr*II, and a 0.5-kb fragment containing the IS*1004* sequences was cloned into an *Sma*I-*Xba*I-digested pBluescript vector. An *Acc*I-*Sac*II fragment of the resulting plasmid containing the insertion (IS) sequences was used as the probe.

RFLP analysis. (i) *smt* region. The region analyzed by restriction fragment length polymorphism (RFLP) spanned 48,759 bp on chromosome I. The left end of the region was at coordinate 190747 within ORF VCA0174, and the right end was at coordinate 239506 within ORF VCA0219 (*hlyA*). This region contains nine *Sph*I fragments. They are, in the order in which they are arranged on the chromosome, 5,201, 5,755, 705, 6,993, 11,899, 1,755, 972, 839, and 14,640 bp long. Five different probes (*smtrgn 1* to *5*, prepared by PCR using primers M 680 to M 689) were used to detect the various fragments. The probes were designed in such a way as to detect two fragments with a single probe. For example, the 5,201- and 5,755-bp *Sph*I fragments are adjacent to each other, and a probe designed centrally at the *Sph*I site could detect these two fragments.

(ii) *ctx* region. The *ctx* region spanned a 54,715-bp region starting within ORF VC1443 (at the 5′ end of *ccoN*; coordinate 1539558) and ending with coordinate 1594273 within VC1488 (encoding a hypothetical protein). It included sequences upstream of RTX, the RTX cassette, RS1, and the CTX core and sequences downstream of the CTX prophage (*cri*, *tlc*, transposase, *fabA*, *rmf*, and several hypothetical proteins). The *Sph*I fragments detected were 1,106, 13,818, 2,862, 13,387, 3,255, 1,365, 2,651, 2,180, 2,651, and 11,400 bp long, in the order in which they are arranged on the chromosome. Five probes (*ctxrgn 1* to *5*, prepared by PCR using primers M 698 to M 707) were used to detect these fragments.

(iii) VPI region. The VPI region analyzed included ORFs VC0815 to VC0850 (coordinates 871612 to 915843; 44,231 bp). The VPI cluster spans coordinates 873020 (between VC0816 and VCO817) to 914296 (between VC0847 and VC0848). The *Xmn*I fragments that were detected included 11,064-, 3,994-, 837-, 3,273-, 1,070-, 12,332-, 3,976-, and 7,685-bp fragments, and five different probes (*ald*, *tagD*, *tcpAdn*, *toxT*, *vpi0845*) were used.

Nucleotide sequence accession numbers. The nucleotide sequences of the *rstR*, *tcpA*, and *ctxAB* genes of strain 365-96 (serogroup O27) and the *wb** region of strain 1322-69 (serogroup O37) determined in this study have been deposited in the GenBank database under accession no. AF390570, AF390571, AF390572, and AF390573, respectively.

RESULTS

Conservation of the genetic organization of the *wb regions.** The *gmhD* and *rjg* genes have previously been reported (45) to be present in all of the *V. cholerae* serogroups examined, and the *wb** regions of these genes are believed to be organized in a cluster. To further investigate the genetic organization of this region in non-O1 and non-O139 serogroups, long-range PCRs were performed using primers designed from the *gmhD* and *rjg* genes. In several serogroups (O37, O53, O65, O77, O80, and O89), the region between *gmhD* and *rjg* was amplified, and the resulting XL-PCR fragments varied in size from 18 to 28 kb (Fig. 1a). In serogroup O105, amplification of this region was not possible, perhaps because of limitations of the PCR. Perhaps the *gmhD* and *rjg* genes flank a *wb** cluster that is too long to be PCR amplified, or perhaps they do not flank the *wb** cluster and are present at different locations on the chromosome in this strain.

Thus, an alternative strategy was devised in order to investigate the organization of the *wb** region in an O105 strain and to isolate the entire *wb** region on a single restriction fragment. PCR and restriction analyses of the *rjg* gene in many serogroup strains indicated that a rare restriction enzyme site, *Not*I, was present in the gene. If *rjg* and *gmhD* are close to one another, introduction of a *Not*I site at the left junction would yield a unique *Not*I fragment containing the entire *wb** cluster. Hence, a *Not*I site was introduced between *orf-2* and *gmhD* at the left boundary of the *wb** region in three strains (serogroups O1, O139, and O105), as described in Materials and Methods. Genomic DNAs of the resulting strains were digested with *Not*I in agarose plugs, and the fragments were separated by PFGE. A fragment that was not present in the parent strains was detected in the strains with an engineered *Not*I site at the left boundary (Fig. 1b, lanes 2, 4, and 6). As expected based on previously published DNA sequence data (9, 14, 18, 33, 48), the O1 and O139 *Not*I fragments were estimated to be 22 and 35 kb long, respectively. In the O105 strain, a 45-kb fragment was released (Fig. 1b, lane 6), indicating that the sizes of the serogroup-specific regions varied. Digestion of the chromosomal DNAs by various restriction enzymes and hybridization with *gmhD* and *rjg* probes revealed fragments of various sizes in different serogroups, indicating that the *wb** regions contain unique sequences (Fig. 1c). Taken together, these data demonstrated that the *wb** region is organized in a cluster.

Identification of the junctions of the O37 *wb region.** Based on the data described above, we hypothesized that the organization of the *wb** regions of the non-O1 and non-O139 strains is similar to that of epidemic strains; i.e., the *gmhD* and *rjg* genes flank serogroup-specific O-antigen biosynthesis genes. In order to test this hypothesis, restriction enzyme *Pst*I-digested O37 XL-PCR fragments (Fig. 1a) were subcloned and sequenced to see if these fragments encode polysaccharide biosynthesis genes. We chose to sequence the O37 region instead of the regions of other serogroups because this serogroup has clinical relevance and strains of this serogroup were implicated in localized cholera outbreaks in the past (2, 26). The entire O37 *wb** region is 27,552 bp long, and the O37-specific sequence is 23,388 bp long. A comparison of this sequence with the sequences of the previously described *wb** clusters (9, 14, 18, 33, 48, 59) revealed that there was an exchange of the *wb** region, since the O37 region had some remnants of the O1 *wbe* region (Fig. 2). The homology breakpoints (i.e., the DNA sequences where the common backbone sequence ends and the serogroup-specific sequence starts) in O37 strains were different from those of the O1 and O139 junctions (Fig. 2 and 3). In the O37 serogroup, the left and right junctions are at *gmhD* and *wbeV*, respectively, instead of at *gmhD* and *rjg* as they are in O139 (Fig. 2). The left junction in O37 is at the *gmhD* locus, and the O37-specific sequence begins about 120 bp to the right of initiation codon ATG of the *gmhD* ORF. In contrast, the left junction in the O139 strain is at the start codon (ATG) of the *gmhD* locus (Fig. 3A). In the case of O37, the right junction is well within the O1 *wbeV* gene; i.e., part of the O1 *wbe* region is retained in the O37 *wb** region (Fig. 3B). Previously, it was shown that the right junction in the O139

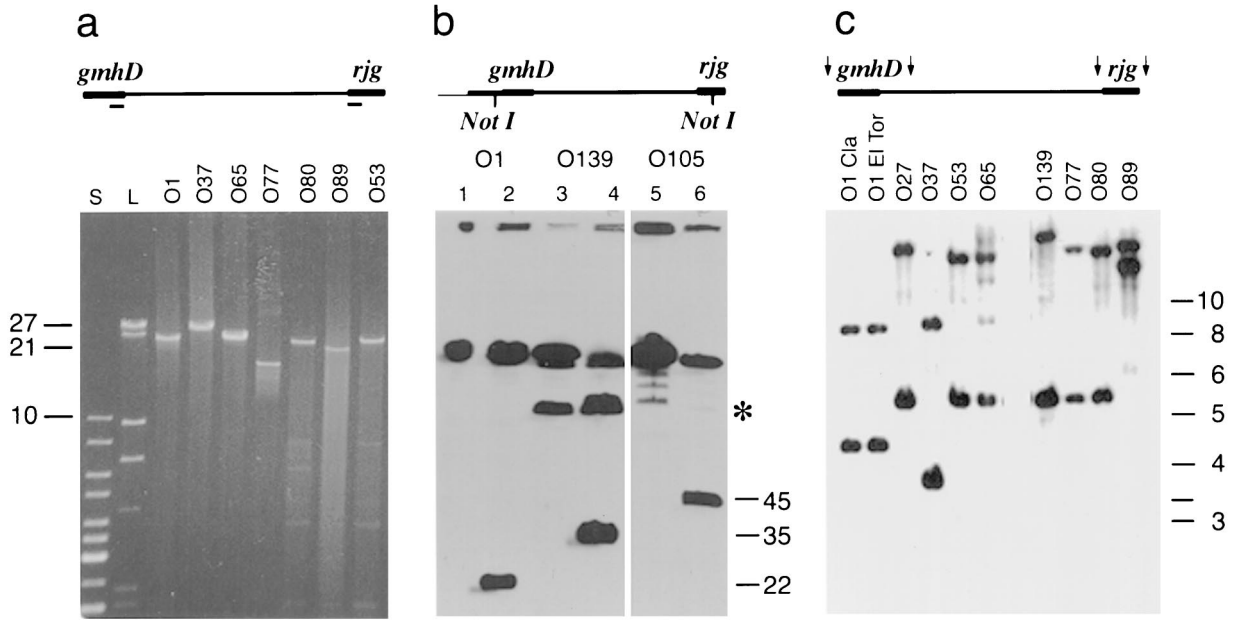

FIG. 1. Genetic organization of the *wb** regions. (a) Agarose gel electrophoresis of XL-PCR amplification products of the O-antigen biosynthesis regions (*wb**) from different serogroup strains. The line at the top is a schematic diagram of the *wb** region with *gmhD* and *rjg* at the ends; the primers used in XL-PCR (indicated by the small bars below the line) are situated at the 5′ ends of the *gmhD* and *rjg* genes. Lane S contained a 1-kb ladder, and lane L contained a *Hin*dIII digest of λ DNA. The positions (in kilobases) of the markers are indicated on the left. (b) PFGE and Southern analysis of the *Not*I fragments of O1, O139, and O105 serogroup strains. The blot was hybridized with a *gmhD* probe. The asterisk indicates the position of unexplained recombination intermediates or partial digestion products. Lanes 1, 3, and 5 contained parent strains without the *Not*I site at the left junction, and lanes 2, 4, and 6 contained the strains with the *Not*I sites flanking the *wb** regions. Lanes 2, 4, and 6 also contained partially digested parent fragments of the *wb** region in addition to the newly generated *wb** fragments. (c) Southern blot analysis of the *gmhD* and *rjg* junction fragments in various strains. The genomic DNAs were digested with *Eco*RI and were simultaneously hybridized with the *gmhD* and *rjg* probes. The upper and lower bands in each lane are the *rjg*- and *gmhD*-specific fragments, respectively.

strain is at the *rjg* gene (right after the *wbeW* gene and the *wbfX* gene of the O1 and O139 strains, respectively) (Fig. 3C); i.e., the entire O1 *wbe* region has been replaced by the *wbf* cluster (14).

O37 *wb region.** The O37 *wb** cluster is unique, since we did not observe any significant DNA homology between the O37-specific sequences and the previously published *V. cholerae* *wb** cluster sequences except at the right junction (see above).

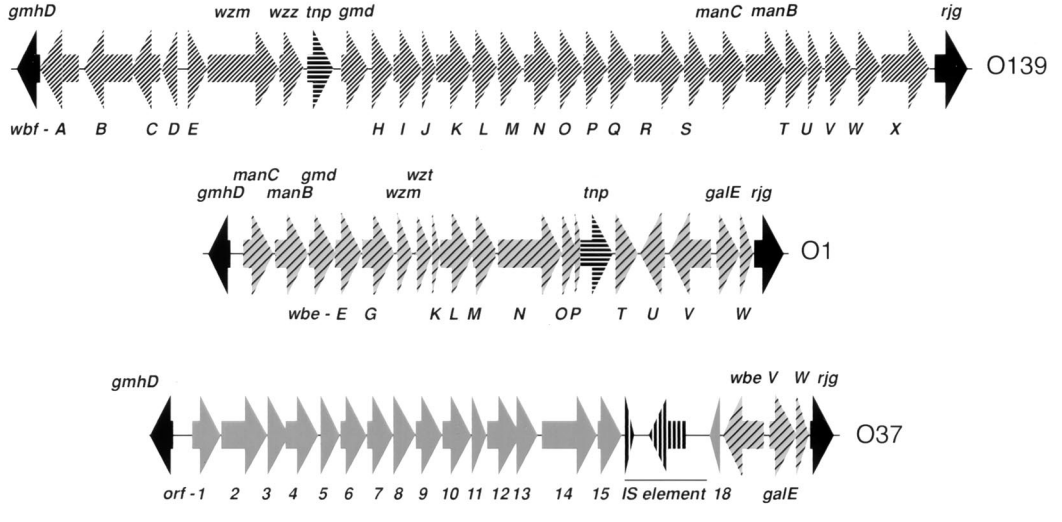

FIG. 2. Comparison of the O37 *wb** region with the O1 *wbe* and O139 *wbf* regions. The genetic organization of the *wb** region of the O37 strain is compared to the genetic organizations reported previously (49) for the O1 and O139 strains. The O139, O1, and O37 regions are 35, 22, and 27 kb long, respectively. The ORFs and the directions of transcription are indicated by the arrows, and the common ORFs in the three serogroups are indicated by the same types of arrows. The common junction genes (*gmhD* and *rjg*) of the three serogroups are indicated by solid arrows. While the left junction is at *gmhD* in all three serogroups, the right junction in O37 is different from the right junction in O139. The O1 right junction genes, *wbeV* through *rjg*, are conserved in the O37 *wb** region.

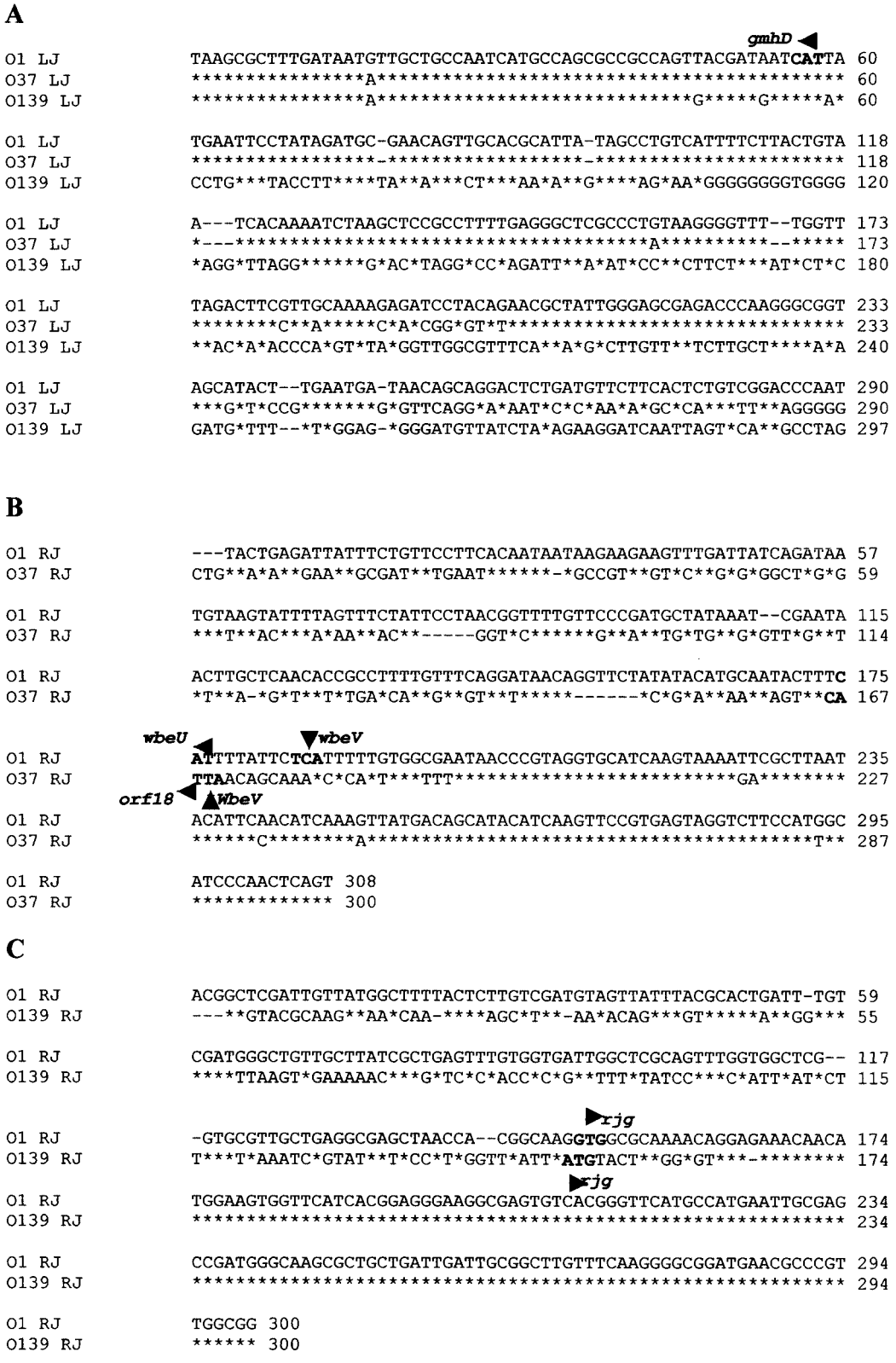

FIG. 3. Analyses of the left junctions (LJ) and the right junctions (RJ) of O1, O139, and O37 *wb** regions performed using the ClustalW program (http://www.ebi.ac.uk/index.html). The asterisks indicate identical bases in the sequences compared. (A) Left junctions of the O1, O37, and O139 *wb** regions. The solid arrowhead and its orientation indicate the *gmhD* ORF start codon and the direction of transcription, respectively. The homology breakpoints between O1 and O37 are ca. 120 bp upstream of the start codon of *gmhD* since the sequences diverge near this position. The homology breakpoints of the O1 and O139 sequences are at the start codon of the *gmhD* ORF, since the sequences after *gmhD* completely diverge. (B) Right junctions of the O1 and O37 sequences. The homology breakpoint is in the *wbeV* ORF, which is preceded by *wbeU* in O1 and

Twenty-three ORFs were identified in the O37 *wb** region, and as expected, many of the *orf* genes (*orf-1* to *-13* and *orf-18*) (Fig. 2) encode enzymes involved in polysaccharide biosynthesis. *orf-14* and *orf-15* encode hypothetical proteins of unknown functions. *orf-14* has very weak homology (26% identical and 46% positive in a 64-amino-acid region) to *yhfO* (which encodes a hypothetical protein) of *Bacillus subtilis*. A 1,549-bp promoter region separates *gmhD* and *orf-1*, and this region contains a putative promoter and *ops* elements found in known polysaccharide biosynthesis regions (4, 23). As seen in other *wb** regions, there is an IS element in the interval between the region that is unique to O37 and the right junction, and this element, which appears to contain an insertion of another fragment encoding a transposase, spans 3,580 bp and has an 18-bp inverted repeat at its ends. The original IS element is 1,054 bp long and is virtually identical (95% identical at the DNA level; the transposase is 90% identical [278 of 306 amino acids]) to the IS elements found in *Vibrio parahaemolyticus* (ISV-3L, ISV-5R, ISV-5L, ISV-4R) (55). Interestingly, the *tnp* gene of the *V. cholerae* element is interrupted after the 76th amino acid residue by a 2,527-bp DNA fragment. This insert encodes an ORF (*orf-17*) that is transcribed in the opposite direction and has extensive homology to transposases of various IS elements found in other bacterial species. The region downstream of the IS element has three ORFs almost identical to the O1 *wbe* cluster: *wbeV*, *galE*, and *wbeW* followed by *rjg*. The *wbeV* gene in the O37 strain appears to consist of three small ORFs. Thus, these data clearly demonstrate that (i) the region between *gmhD* and *rjg* in the O37 serogroup encodes O-antigen biosynthesis genes and (ii) the O37 *wb** region originated by homologous recombination-mediated exchange of *wb** gene clusters.

Identification of non-O1 and non-O139 strains containing virulence genes. Based on the data described above, we reasoned that a homologous recombination event flanking the *gmhD* and *rjg* genes may result in non-O1 and non-O139 pathogens similar to O139 strains. In order to identify O139 Bengal-like strains that originated by exchange of *wb** clusters, 300 *V. cholerae* strains were screened for the presence of the *ctxAB* and *tcpA* genes, which are known to be present in epidemic and pandemic strains. DNA dot blot analysis was performed by using the *ctxAB* and *tcpA* genes as the probes. Fifteen non-O1 and non-O139 strains were found to be $tcpA^+$. All of these strains also carried three other genes (*aldA*, *toxT*, and *int*) of the VPI cassette, which indicated that the entire VPI may be present in these strains. Thirteen strains carried the *rstR* and *rstA* genes; nine of these strains also carried the *ctxAB* genes; and two strains did not carry any of these three genes. Of the four *ctxAB* mutant strains, one carried just the *rstR* and *rstA* genes, while the three other strains carried the *rstR* and *rstA* genes and the genes of the core region except the *ctxAB* genes (data not shown). Further characterization of 4 of the 15 tcp^+ strains (serogroups O27, O37, O53, and O65) indicated that they contained the entire VPI and a pre-CTXϕ (CTXϕ without the *ctx* genes) or a CTXϕ. These four strains had genetic backgrounds very similar to those of the epidemic O1 strains (see below), thus indicating that there was an O-antigen shift in an O1 strain.

Genetic relatedness of the non-O1 and non-O139 strains to the epidemic strains as determined by IS*1004* fingerprinting and RFLP analysis. In order to determine the genetic relatedness among the various strains and to determine which of these strains resulted from an O-antigen shift in an epidemic strain, IS*1004* fingerprinting and RFLP analyses were performed.

IS*1004* is an IS element present in multiple copies in O1 classical and El Tor strains, and it has been used previously (8) for typing *V. cholerae* strains. In agreement with a previous report (8), O1 classical and El Tor strains exhibited unique fingerprints; i.e., the classical and El Tor strains had six and four IS*1004* copies, respectively (Fig. 4a, lanes O1 cla and O1 El Tor). In the present analysis, the O27 and O139 strains were identical to the El Tor strains (Fig. 4a, lanes O139 and O27), and the O37 strain was similar to the classical strains (Fig. 4a, lane O37). The O53 and O65 strains exhibited some similarity to both O1 El Tor and O1 classical strains and thus appeared to have diverged early from the progenitor of the O1 strains (Fig. 4a, lanes O53 and O65).

Further evidence of genetic similarity among the strains was obtained by RFLP analyses of two regions of chromosomes I and II. For this analysis, the *vps* (vibrio polysaccharide synthesis) (22) operon from chromosome I and a region on chromosome II that contains a site-specific methyl transferase gene (*smt*-VCA 0198) were chosen. Restriction enzyme *Sph*I-digested genomic DNAs were hybridized with various probes (described in Materials and Methods). The RFLP patterns of the *vps* regions of the O1 classical and El Tor strains and the four non-O1 and non-O139 strains (serogroups O27, O37, O53, and O65) were identical, whereas many other non-O1 and non-O139 strains had extensive variations (data not shown).

The *smt* region in the classical and El Tor strains exhibited unique RFLP patterns (Fig. 4b). The RFLP patterns of the *smt* region of the O139 and O27 strains were identical to those of the El Tor strains (Fig. 4b, compare lanes O1 El Tor to lanes O139 and O27), and the RFLP patterns of the O37, O53, and O65 strains were similar but not identical to the pattern of the *smt* region of either the classical or El Tor strains. Furthermore, the *smt* gene was found only in strains that have an O1 backbone (Fig. 4c). None of the non-O1 and non-O139 strains used in this analysis (O9, O31, and O144) have an O1-like backbone, as determined either by IS*1004* fingerprinting (Fig. 4a, lanes O144, O9, and O31) or by RFLP analyses (Fig. 4b, lanes O144, O9, and O31), and they lacked the *smt* gene. Taken together, IS*1004* fingerprinting and RFLP analysis results indicated that the four non-O1 and non-O139 strains

by *orf-18* in O37. The stop codon of *wbeV* (indicated by the arrowhead pointing upward) overlaps with the start codon of *orf-18* in O37. The stop codon of *wbeV* of O1 is indicated by the arrowhead pointing downward. (C) Right junctions of the O1 and O139 sequences published previously (14, 45). The homology breakpoint is at the start of the *rjg* ORF. The *rjg* genes of O1 and O139 strains have different N terminus and start codons. In all these cases, the actual recombination crossover sites could be anywhere within homologous segments far away from the homology breakpoints and not necessarily at the junctions.

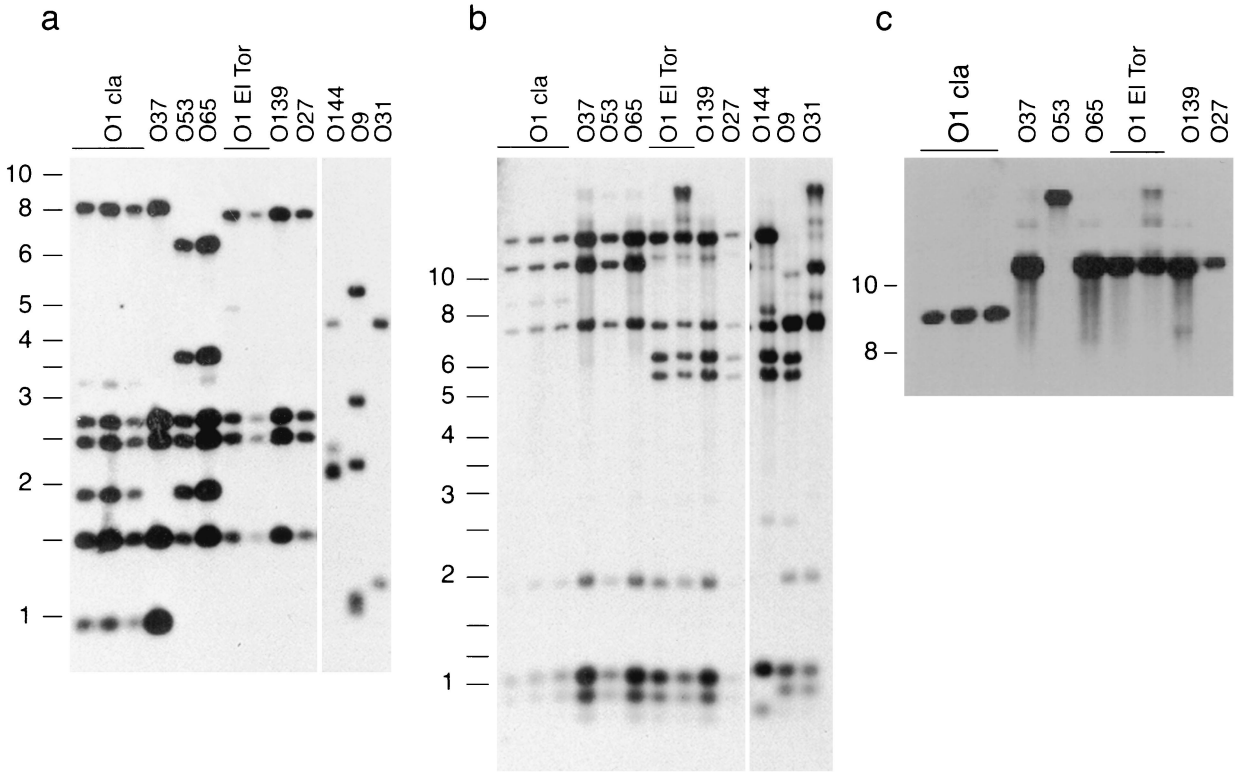

FIG. 4. IS*1004* fingerprinting and RFLP analyses of the various serogroup strains. (a) IS*1004* fingerprints of *Hpa*II-digested chromosomal DNAs of various *V. cholerae* strains. Since *Hpa*II does not cut within IS*1004*, each band represents one copy of the element. (b) RFLP analysis of the *smt* region of chromosome II of various *V. cholerae* strains. The genomic DNAs were digested with *Sph*I and hybridized simultaneously with multiple probes (*smtrgn 1* to *5*). The identities of the bands were deduced from their predicted sizes, based on the published *V. cholerae* genome sequence (22), and they were confirmed by separately hybridizing the same blot with individual probes. The sizes of the molecular weight markers (in kilobases) are indicated on the left. (c) Blot used in panel b rehybridized with an *smt* probe (primers M 668 and M 669). Lanes O144, O9, and O27 were omitted since they did not show any hybridizing band with this probe.

were derived from an epidemic strain by *wb** cluster exchange and subsequently diverged. A DNA sequence analysis of one of the *wb** regions (serogroup O37) determined in this study (see above) supports this conclusion.

Genetic organization of the CTX prophage and VPI regions of the non-O1 and non-O139 strains. We reasoned that exchange of the *wb** region in an O1 strain would result in a non-O1 and non-O139 strain containing either the classical or El Tor *ctx* and VPI regions. In order to determine the structural organization of the *ctx* region in the non-O1 and non-O139 strains examined in this study, chromosomal DNAs were digested with *Eco*RI (an enzyme that does not cut within the CTXϕ genome), and the fragments were hybridized with *rstA*, *rstR*, *ctx* core, and *ctxAB* gene probes. As expected, there was diversity in the arrangement and location of the CTX prophage genomes in various strains. The sizes of the fragments of the second CTX prophage copy (chromosome II) varied in the three classical strains analyzed [Fig. 5a, upper panel, lanes 395 (O1 Cla), NIH35-A3 (O1 Cla), and 5011 (O1 Cla)]. Like O1 El Tor and O139 strains, the O37 strain had a chromosome I copy(ies) of CTXϕ [Fig. 5a, upper panel, lanes N16961 (O1 El Tor), E7946 (O1 El Tor), O37, and O139], and the O27 strain also contained a single copy (Fig. 5a, lane O27). The same fragment was detected when the blot was hybridized with *rstA* or *rstR* or *ctx* core probes in all of the strains except the O53 and O65 strains (Fig. 5a, lower panel). These two strains produced a single band when they were hybridized with *rstA* or *rstR* or *ctx* core probes (Fig. 5a, lower panel, lanes O53 and O65) but did not hybridize with any band when the *ctxAB* probe was used.

To further elucidate the genetic organization, the *ctx* region on chromosome I was analyzed by RFLP. The region analyzed encompassed about 54,715 bp, which included the RTX cassette, CTXϕ, and the sequences upstream and downstream of them. Figure 5b shows that the O37 and O139 strains have an El Tor-like organization (compare lane O1 El Tor to lanes O37 and O139). The O27, O53, and O65 strains resembled El Tor strains in the RTX cassette segment (bands a and b were produced only by El Tor strains and not by a classical strain [Fig. 5b]) but differed from epidemic strains by lacking the pTLC element (Fig. 5b, compare lanes O1 Cla, O1 El Tor, O37, and O139 to lanes O27, O53, and O65, bands f/g, h, and i).

We analyzed the VPI region in the non-O1 and non-O139 strains examined in this study. Digestion of the genomic DNAs with *Xmn*I, followed by hybridization of the blot with five different probes spanning the entire VPI region, resulted in identification of seven and eight fragments in the O1 classical and El Tor strains, respectively (Fig. 5c, lanes O1 Cla and O1 El Tor). Among the other strains, the O37 strain had a classical VPI cluster (Fig. 5c, compare lanes O1 Cla and O37), the O53,

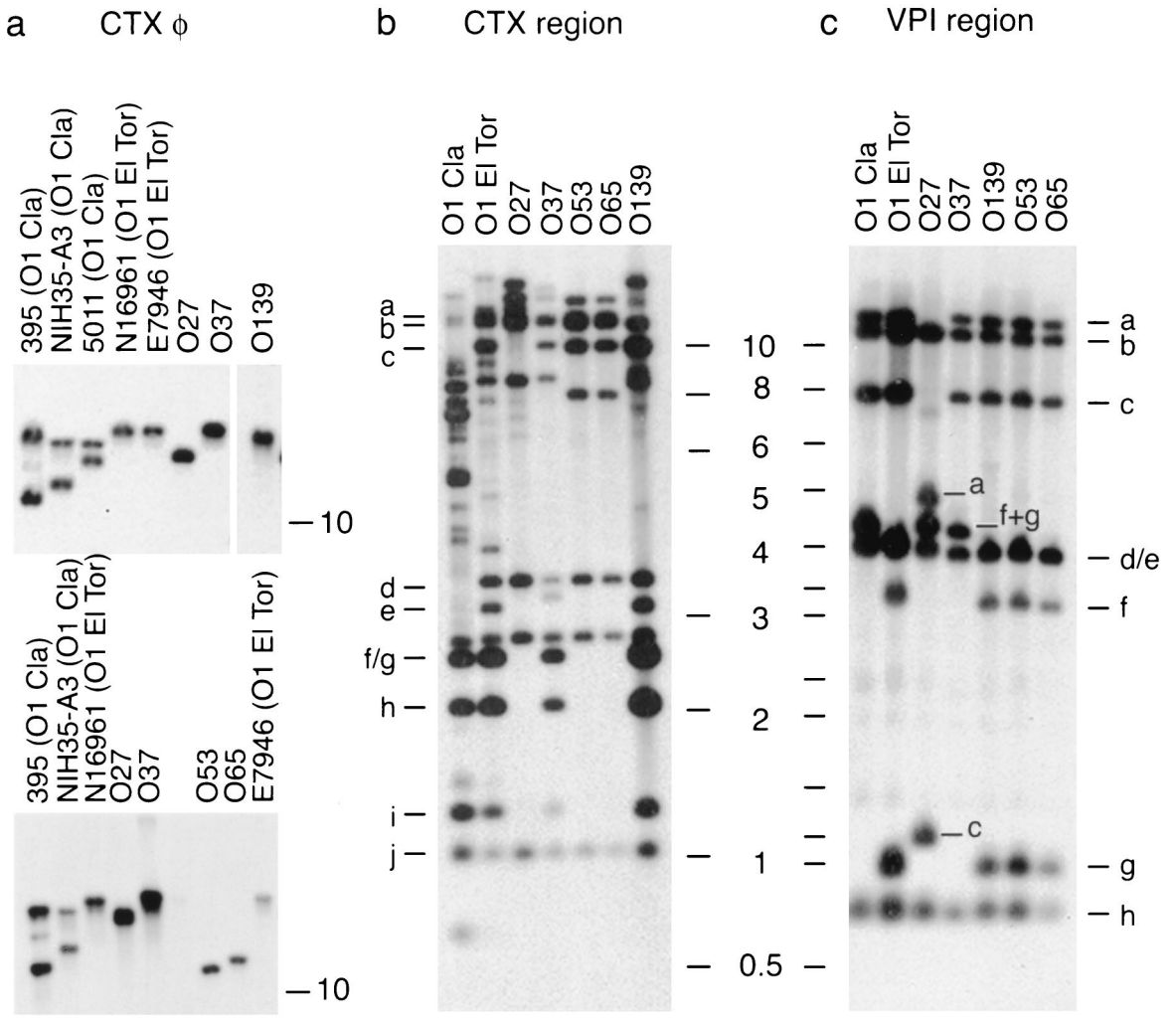

FIG. 5. RFLP analysis of the virulence regions of the various strains. (a) *Eco*RI-digested genomic DNAs were probed with *ctxAB* gene probes (upper panel) and the *ctx* core probe (lower panel). *Eco*RI does not cut within the CTXϕ genome, and each hybridizing band therefore represents one copy of CTX present in a strain. (b) *Sph*I-digested genomic DNAs of the strains were probed with five probes spanning a 45-kb region surrounding the CTXϕ genome integrated on chromosome I. The probes and the sizes of the fragments and the corresponding bands are as follows: *ctxrgn 1*, 1,106 and 13,818 bp (bands a and j); *ctxrgn 2*, 2,862 and 13,387 bp (bands e and b); *ctxrgn 3*, 3,255 and 1,365 bp (bands d and i); *ctxrgn 4*, 2,651 and 2,180 bp (bands f and h); and *ctxrgn 5*, 2,651 and 11,400 bp (bands g and c). (c) Genomic DNAs digested with *Xmn*I and probed simultaneously with five different probes from the VPI region. The identities of the bands were confirmed by separate hybridization with individual probes. The probes and the sizes of the fragments and the corresponding bands are as follows: *ald*, 11,064 bp (band b); *tagD*, 3,994 bp (band e) and 837 bp (band h); *tcpAdn*, 3,273 bp (band f) and 1,070 bp (band g); *toxT*, 12,332 bp (band a); and *vpi0845*, 3,976 bp (band d) and 7,685 bp (band c).

O65, and O139 strains had an El Tor VPI cluster (compare lane O1 El Tor to lanes O139, O53, and O65), and the O27 strain appeared to contain a novel VPI cluster (Fig. 5c, lane O27). Further hybridization analyses with additional probes indicated that the O27 strain had the entire VPI cluster with two additional *Xmn*I sites (data not shown).

DNA sequence of the *rstR*, *ctxAB*, *tcpA*, and *aldA* genes of the non-O1 and non-O139 strains. In order to understand further the origin of the *ctx* and VPI regions, the *rstR* gene (which encodes the repressor of CTXϕ and determines the phage immune specificity) and the *ctxAB* genes of CTXϕ and the *tcpA* and *aldA* genes of the VPI were sequenced. These genes are known to be different in classical and El Tor biotypes (15, 30, 31, 32). The results are summarized in Table 1.

The *rstR* genes were PCR amplified, cloned into the pCR2.1 vector, and sequenced. Strains of serogroups O37 and O139 had an *rstR* allele with El Tor specificity, the O53 and O65 strains had classical immune specificity and the O27 strain had novel specificity. The $rstR^{O27}$ allele, whose specificity differs from that of the classical, El Tor, and Calcutta CTXϕs, is identical to the recently reported *rstR-4*** gene (38) encoding an RstR repressor protein that is 86 amino acids long.

We sequenced the *ctxAB* genes of the two non-O1 and non-O139 (O27 and O37) strains in order to identify the variations in the *ctx* genes in various serogroups. The O53 and O65 strains did not possess *ctx* genes. The O27 and O37 strains possessed highly conserved *ctxAB* genes; i.e., the 1,152 bp of their *ctxAB* genes differed by less than three or four nucleotides, and their proteins differed by two or three amino acids (CtxA, $^{cla/ET}S46N^{O37}$; CtxB, $^{cla/ET}Q24H^{O27}$, $^{cla/ET}D28A^{O27}$,

$^{cla/ET}F46L^{O37}$, and $^{cla/ET}K55N^{O37}$, where cla indicates classical and ET indicates El Tor). In addition to these changes, the $CtxB^{O27,O37}$ alleles were identical to the classical alleles at two other positions, $^{ET}Y39H^{cla/O27/O37}$ and $^{ET}I68T^{cla/O27/O37}$.

A 1.4-kb fragment encompassing the *tcpA* gene was PCR amplified, and the *tcpA* gene was sequenced. The $tcpA^{O37}$ allele was identical to the $tcpA^{cla}$ allele, the $tcpA^{O53,O65}$ alleles were identical to the $tcpA^{ET}$ allele, and the O27 strain had a novel allele. The $tcpA^{O27}$ allele is identical to the recently reported *tcpA-env* allele (38) except for a single amino acid substitution ($^{cla/ET}V9D^{tcpA\text{-}env}$). In contrast to the *tcpA* alleles, the *aldA* genes of the strains exhibited very few variations (data not shown).

DISCUSSION

Antigenic shift and emergence of novel pathogens. Exchange of polysaccharide biosynthesis clusters resulting in novel pathogens has been well documented in many bacterial pathogens, such as *Streptococcus pneumoniae* and *Neisseria meningitidis* (16, 24, 51). Previous phylogenetic studies of *V. cholerae* (8, 46) indicated that exchange of *wb** regions may be common in this species. Moreover, the O-antigen diversity and conservation of the structural architecture of the *wb** regions seen in *V. cholerae* provide ample opportunity for O-antigen shifts resulting from exchange of *wb** regions. Serogroup changes in nonpathogenic *V. cholerae* may not have serious consequences and may go unnoticed, whereas in pathogenic backgrounds, such as O1, antigen shifting may result in novel pathogens (e.g., *V. cholerae* O139 Bengal, which causes cholera epidemics even in populations immune to *V. cholerae* O1).

The non-O1 and non-O139 strains identified in this study further expand the repertoire of *V. cholerae* strains with epidemic potential and underscore the idea that the emergence of the O139 serogroup was not a unique event. The genetic relatedness of the four non-O1 strains, based on IS*1004* fingerprinting and RFLP analyses, was supported by an extensive multilocus sequence typing analysis in which these four strains always clustered with the epidemic strains (data not shown). Interestingly, the four strains diverged further by acquiring different virulence cassettes or parts of cassettes. For example, the O37 serogroup strain with a classical backbone and a classical VPI acquired an El Tor CTXϕ, the O53 and O65 strains with an El Tor backbone acquired a preclassical CTXϕ (without the *ctx* genes), and the O27 strain with an El Tor backbone acquired a novel CTXϕ and a VPI cluster. Thus, it appears that the genetic backbone, *wb** cluster, VPI, and CTX regions have evolved as independent units.

The O37 *wb** region DNA sequence determined in our study revealed that the O37 strain resulted from an exchange of the O1 *wbe* cluster with a *wb** O37 cluster, similar to the event that occurred in the O139 serogroup. However, the homology breakpoints in O37 are different from the breakpoints of the O139 junctions. Unlike O139, in which the entire *wbe* region has been replaced by the *wbf* region, in O37 some of the O1 *wbe* genes closer to the right junction have been retained, which provides further support for the idea that the O37 strain emerged from an O1 strain by O-antigen shifting. Although the precise crossover points in this recombination event cannot be predicted from the sequence, phage-mediated transfer of the donor *wb** cluster would place the crossover sites closer to the *wb** junctions, due to the large size of the *wb** clusters (25 to 45 kb) and the general packaging limits of transducing phages. On the other hand, conjugational transfer does not impose such constraints, and the actual crossover could occur anywhere, even kilobases away from the junctions in the homologous segments flanking the divergent *wb** clusters. It is also possible that the *wb** junctions are hyperrecombinogenic. Such a proposal has been advanced before (50), and chi-like sequences have been found at the junctions of *wb** clusters in *Escherichia coli* and *Klebsiella* spp.

The sequence of the O37 *wb** region has also revealed the putative genes responsible for synthesis of the receptor of an O1-specific generalized transducing phage, CP-T1. The O1 O-antigen has been postulated to be the receptor for the phage (21), and CP-T1 has recently been reported (10) to be able to infect O37 serogroup strains as well. Our sequence analysis suggests that the three genes (*wbeV*, *galE*, and *wbeW*) shared by the O1 and O37 *wb** regions are directly involved in a step in the synthesis of the phage receptor component of the O-antigen.

Based on the XL-PCR and hybridization data (Fig. 1), we predict that the O27, O53, O65, O77, O80, and O139 strains probably have similar left junctions. The left and right junctions in the O53 and O65 strains may be similar since the sizes of the *Eco*RI junction fragments are the same (Fig. 1c) and the sizes of the *wb** regions of these two strains are also similar, indicating that the strains may have minor variations in their *wb** regions. These hybridization data further suggest that the three other strains (serogroups O27, O53, and O65) which have backbones very similar to those of O1 strains probably arose by exchange of *wb** clusters.

Heterogeneity in the genetic organization of CTX prophage. Heterogeneity in the genetic organization of the CTX prophage region in various strains of *V. cholerae* has been documented previously. For example, O1 classical strains have two copies of the CTXϕ, one located on each of the two chromosomes. Also, some El Tor strains contain a single copy, while many others have two or more copies arranged in tandem on chromosome I (5, 12, 15, 31, 32, 34). *V. cholerae* O139 strains are similar to O1 El Tor strains in that they have two copies of the CTXϕ arranged in tandem on chromosome I (5, 56). There was remarkable diversity in the CTXϕs of the four strains identified in this study with respect to the arrangement of the copies and their gene sequences.

VPI is not unique to epidemic serogroups. Unlike the *ctx* region, the VPI has not been analyzed in great detail in any strain other than O1 classical and El Tor strains (30) and, recently, several environmental strains (38). A few studies have examined the *tcpA* gene from non-O1 and non-O139 serogroup strains, and they have identified several *tcpA* variants (13, 20, 39, 40). These studies, together with our results, clearly demonstrate that the VPI is not unique to the epidemic and pandemic O1 and O139 Bengal strains, as was originally reported (28). While some of the non-O1 and non-O139 pathogens (O27, O37, and O139) have been derived from O1 strains, independent acquisition of these virulence factors via phage-mediated transfer in multiple O serogroup strains has also been observed, and many of these non-O1 and non-O139

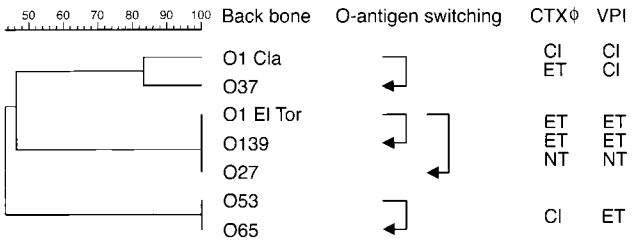

FIG. 6. Model for emergence of the non-O1 and non-O139 strains with epidemic potential by O-antigen switching. A dendrogram of the genetic relatedness of the various serogroups, based on IS*1004* fingerprinting analysis, is shown on the left. The O37 strain appears to have been derived from an O1 preclassical strain (classical VPI cluster) by O-antigen shifting, divergence, and subsequent acquisition of an El Tor CTXϕ. The El Tor strains diverged from classical strains, acquired a distinct CTXϕ, and probably acquired a *tcpA* allele by recombination (30). The O139 strain appears to have been derived from an El Tor strain by O-antigen switching, and the O27 strain appears to have been derived from an El Tor progenitor and to have acquired a distinct CTXϕ and a distinct VPI. It is also probable that O27 emerged from an El Tor strain by O-antigen switching and subsequently changed its *rstR* and *tcpA* genes by allelic exchange. The O53 and O65 strains have an El Tor VPI cluster and are identical to each other. They have genetic backbones resembling those of both classical and El Tor strains, and hence, they probably emerged from an O1 progenitor, diverged, and subsequently acquired a classical pre-CTXϕ and and an El Tor VPI. These two strains probably underwent O-antigen switching events independently or sequentially, from O1 to O53 and O65. Cl, classical; ET, El Tor; NT, novel type (38).

strains have a full complement of the VPI cluster (unpublished data).

Model for the emergence of virulent *V. cholerae* strains. Since TCP has been demonstrated to be the receptor for CTXϕ, a two-step model for the origin of virulent *V. cholerae* strains has emerged (10, 19, 35). In the first step, VPI is horizontally acquired by a nontoxigenic strain, and the TCP produced by the VPI-containing strain serves as the receptor for CTXϕ, which leads to the second step, in which the CTXϕ carrying *ctxAB* genes is acquired. This model is supported by the findings that (i) CTXϕ and VPI are associated with a majority of the O1 and O139 strains and (ii) $CTX\phi^-$ VPI^+ non-O1 and non-O139 strains have been occasionally found, whereas $CTX\phi^+$ VPI^- non-O1 and non-O139 strains are rare (19). However, to account for these rare strains, a TCP-independent mechanism of acquisition of CTXϕ has been proposed, which involves a generalized transducing phage, CP-T1 (10). Data obtained during our studies not only support the two-step model (none of the 300 non-O1 and non-O139 strains screened by us was $CTX\phi^+$ VPI^-) but also delineate some of the additional steps involving O-antigen exchange (Fig. 6). As proposed before (11), classical and El Tor biotype strains originated, diverged, and subsequently acquired the classical and El Tor VPI and CTXϕs from a nonpathogenic *V. cholerae* progenitor (O1 serogroup) strain that did not possess either VPI or CTXϕ. The pre-CTX phages evolved independently, as revealed by their distinct *rstR* alleles (11; unpublished data), and *ctxAB* genes were subsequently acquired by a horizontal transfer event, as revealed by the different *ctx* alleles in O27 and O37 strains.

Based on our data, the O37 strain seems to have arisen from the classical strain progenitor (classical backbone and VPI) by changing its O-antigen, diverging, and subsequently acquiring an El Tor CTXϕ. The O53 and O65 strains seem to have originated from an O1 progenitor (backbone showing similarity to both O1 classical and El Tor) by changing O-antigens (sequentially or independently) and then diverging and acquiring an El Tor VPI and a classical pre-CTXϕ. The O27 strain originated from an El Tor progenitor (El Tor backbone) and acquired a novel VPI (novel *tcpA* allele) and a novel CTXϕ. The *tcpA* gene and presumably the *rstR* gene evolved by recombination from a common VPI and CTXϕ rather than as different phages. The idea of a mosaic pattern of similar and divergent genes within VPI was proposed recently based on analyses of the VPI of several environmental *V. cholerae* strains (38). A comparison of the complete VPI sequences of a classical strain and an El Tor strain also supports the role of recombination in the evolution of VPI (30). Thus, a change in the O-antigen of the O27 strain may have occurred in the progenitor prior to acquisition of the VPI and CTX phages, or the O27 strain may have originated directly from an El Tor strain and its *rstR* and *tcpA* genes may have subsequently been altered by allelic exchange.

Our data demonstrate that genetic switching of O-antigen biosynthesis regions resulted in the emergence of at least some non-O1 and non-O139 *V. cholerae* strains having pathogenic potential (i.e., containing the known *V. cholerae* virulence regions VPI and CTX prophage). The nonrandom distribution of these virulence markers in various *V. cholerae* serogroups (not all 200 serogroups have these virulence factors) suggests that horizontal transmission of virulence genes may not be similarly effective or frequent in various *V. cholerae* strains. The underlying mechanisms for this phenomenon are not clear at this time. However, the data presented in this paper improve our understanding of the evolution of the species and provide insight into the possible mechanisms for emergence of epidemic *V. cholerae* strains and serogroups from nonepidemic *V. cholerae* strains and serogroups. From a public health standpoint, our data raise the possibility that existing *V. cholerae* vaccines may provide little or no protection against the newly identified pathogenic strains, and they suggest that there is a need for novel strategies for developing vaccines against *V. cholerae*.

ACKNOWLEDGMENTS

We thank Nick Ambulos and Lisa Sadzewicz (UMAB Bioploymer Laboratory) for DNA sequencing and primer synthesis. Thanks are also due to Rick Blank, Arnold Kreger, and an anonymous reviewer for helpful comments on the manuscript and to Judy Johnson for providing access to the Smith strain collection.

Funding for this study was provided by a grant from the Department of Veterans Affairs (to J.G.M.), by grant RO1 GM60791 from the National Institutes of Health (to J.G.M.), by a University of Maryland intramural grant (to S.S.), and by BREF intramural support from the Department of Veterans Affairs (to S.S.).

REFERENCES

1. **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.
2. **Aldova, E., K. Laznickova, E. Stepankova, and J. Lietava.** 1968. Isolation of nonagglutinable vibrios from an enteritis outbreak in Czechoslovakia. J. Infect. Dis. **118:**25–31.

3. **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.
4. **Bailey, M. J. A., C. Hughes, and V. Koronakis.** 1997. RfaH and the *ops* element, components of a novel system controlling bacterial transcription elongation. Mol. Microbiol. **26:**845–851.
5. **Basu, A., A. K. Mukhopadhyay, C. Sharma, J. Jyot, N. Gupta, A. Ghosh, S. K. Bhattacharya, Y. Takeda, A. S. Faruque, M. J. Albert, and G. B. Nair.** 1998. Heterogeneity in the organization of the CTX genetic element in strains of *Vibrio cholerae* O139 Bengal isolated from Calcutta, India and Dhaka, Bangladesh and its possible link to the dissimilar incidence of O139 cholera in the two locales. Microb. Pathog. **24:**175–183.
6. **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.
7. **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.
8. **Bik, E. 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.
9. **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.
10. **Boyd, E. F., and M. K. Waldor.** 1999. Alternative mechanism of cholera toxin acquisition by *Vibrio cholerae*: generalized transduction of CTXϕ by bacteriophage CP-T1. Infect. Immun. **67:**5898–5905.
11. **Boyd, E. F., A. J. Heilpern, and M. K. Waldor.** 2000. Molecular analyses of a putative CTXϕ precursor and evidence for independent acquisition of distinct CTXϕs by toxigenic *Vibrio cholerae*. J. Bacteriol. **182:**5530–5538.
12. **Boyd, E. F., K. E. Moyer, L. Shi, and M. K. Waldor.** 2000. Infectious CTXϕ and the vibrio pathogenicity island prophage in *Vibrio mimicus*: evidence for recent horizontal transfer between *V. mimicus* and *V. cholerae*. Infect. Immun. **68:**1507–1513.
13. **Chakraborty, S., A. K. Mukhopadhyay, R. K. Bhadra, A. N. Ghosh, R. Mitra, T. Shimada, S. Yamasaki, S. M. Faruque, Y. Takeda, R. R. Colwell, and G. B. Nair.** 2000. Virulence genes in environmental strains of *Vibrio cholerae*. Appl. Environ. Microbiol. **66:**4022–4028.
14. **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.
15. **Davis, B. M., H. H. Kimsey, W. Chang, and M. K. Waldor.** 1999. The *Vibrio cholerae* O139 Calcutta bacteriophage CTXϕ is infectious and encodes a novel repressor. J. Bacteriol. **181:**6779–6787.
16. **Dillard, J. P., M. Caimano, T. Kelly, and J. Yother.** 1995. Capsules and cassettes: genetic organization of the capsule locus of *Streptococcus pneumoniae*. J. Dev. Biol. Stand. **85:**261–265.
17. **Donnenberg, M. S., and J. B. Kaper.** 1991. Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive selection suicide vector. Infect. Immun. **59:**4310–4317.
18. **Fallarino, A., C. Mavrangelos, U. H. Stroeher, and P. A. Manning.** 1997. Identification of additional genes required for O-antigen biosynthesis in *Vibrio cholerae* O1. J. Bacteriol. **179:**2147–2153.
19. **Faruque, S. M., M. J. Albert, and J. J. Mekalanos.** 1998. Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. Microbiol. Mol. Biol. Rev. **62:**1301–1314.
20. **Ghosh, C., R. K. Nandy, S. K. Dasgupta, G. B. Nair, R. H. Hall, and A. C. Ghose.** 1997. A search for cholera toxin (CT), toxin coregulated pilus (TCP), the regulatory element ToxR and other virulence factors in non-O1/non-O139 *Vibrio cholerae*. Microb. Pathog. **22:**199–208.
21. **Guidolin, A., and P. A. Manning.** 1985. Bacteriophage CP-T1 of *Vibrio cholerae*. Identification of the cell surface receptor. Eur. J. Biochem. **153:**89–94.
22. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406:**477–483.
23. **Hobbs, M., and P. R. Reeves.** 1994. The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. Mol. Microbiol. **12:**855–856.
24. **Jiang, S. M., L. Wang, and P. R. Reeves.** 2001. Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters. Infect. Immun. **69:**1244–1255.
25. **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.
26. **Kamal, A. M.** 1971. Outbreak of gastro-enteritis by non-agglutinable (NAG) vibrios in the republic of the Sudan. J. Egypt. Public Health Assoc. **XLVI:**125–159.
27. **Kaper, J. B., J. G. Morris, Jr., and M. M. Levine.** 1995. Cholera. Clin. Microbiol. Rev. **8:**48–86.
28. **Karaolis, D. K., J. A. Johnson, C. C. Bailey, E. C. Boedeker, J. B. Kaper, and P. R. Reeves.** 1998. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. Proc. Natl. Acad. Sci. USA **95:**3134–3139.
29. **Karaolis, D. K., S. Somara, D. R. Maneval, Jr., J. A. Johnson, and J. B. Kaper.** 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature **399:**375–379.
30. **Karaolis, D. K., R. Lan, J. B. Kaper, and P. R. Reeves.** 2001. Comparison of *Vibrio cholerae* pathogenicity islands in sixth and seventh pandemic strains. Infect. Immun. **69:**1947–1952.
31. **Kimsey, H. H., and M. K. Waldor.** 1998. CTX ϕ immunity: application in the development of cholera vaccines. Proc. Natl. Acad. Sci. USA **95:**7035–7039.
32. **Kimsey, H. H., G. B. Nair, A. Ghosh, and M. K. Waldor.** 1998. Diverse CTX ϕs and evolution of new pathogenic *Vibrio cholerae*. Lancet **352:**457–458.
33. **Manning, P. A., U. H. Stroeher, and R. Morona.** 1994. Molecular basis for O-antigen biosyntheis 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.
34. **Mekalanos, J. J.** 1983. Duplication and amplification of toxin genes in *Vibrio cholerae*. Cell **35:**253–263.
35. **Mekalanos, J. J., E. J. Rubin, and M. K. Waldor.** 1997. Cholera: molecular basis for emergence and pathogenesis. FEMS Immunol. Med. Microbiol. **18:**241–248.
36. **Mooi, F. R., and E. M. Bik.** 1997. The evolution of epidemic *Vibrio cholerae* strains. Trends Microbiol. **4:**161–165.
37. **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.
38. **Mukhopadhyay, A. K., S. Chakraborty, Y. Takeda, G. B. Nair, and D. E. Berg.** 2001. Characterization of VPI pathogenicity island and CTXϕ prophage in environmental strains of *Vibrio cholerae*. J. Bacteriol. **183:**4737–4746.
39. **Nandi, B., R. K. Nandy, A. C. Vicente, and A. C. Ghose.** 2000. Molecular characterization of a new variant of toxin-coregulated pilus protein (TcpA) in a toxigenic non-O1/non-O139 strain of *Vibrio cholerae*. Infect. Immun. **68:**948–952.
40. **Novais, R. C., A. Coelho, C. A. Salles, and A. C. Vicente.** 1999. Toxin-coregulated pilus cluster in non-O1, non-toxigenic *Vibrio cholerae*: evidence of a third allele of pilin gene. FEMS Microbiol. Lett. **171:**49–55.
41. **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.
42. **Rhine, J. A., and R. K. Taylor.** 1994. TcpA pilin sequences and colonization requirements for O1 and O139 *Vibrio cholerae*. Mol. Microbiol. **13:**1013–1020.
43. **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.
44. **Smith, H. L.** 1979. Serotyping of non-cholera vibrios. J. Clin. Microbiol. **10:**85–90.
45. **Sozhamannan, S., Y. K. Deng, M. Li, A. Sulakvelidze, J. B. Kaper, J. A. Johnson, G. B. Nair, and J. G. Morris, Jr.** 1999. 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. Infect. Immun. **67:**5033–5040
46. **Stine, O. C., S. Sozhamannan, Q. Gou, S. Zheng, J. G. Morris, Jr., and J. A. Johnson.** 2000. Phylogeny of *Vibrio cholerae* based on *recA* sequence. Infect. Immun. **68:**7180–7185.
47. **Stroeher, U. H., and P. A. Manning.** 1997. *Vibrio cholerae* serotype O139: swapping genes for surface polysaccharide biosynthesis. Trends Microbiol. **5:**178–180.
48. **Stroeher, 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.
49. **Stroeher, 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.
50. **Sugiyama, T., N. Kido, Y. Kato, N. Koide, T. Yoshida, and T. Yokochi.** 1997. Evolutionary relationship among *rfb* gene clusters synthesizing mannose homopolymer as O-specific polysaccharides in *Esherichia coli* and *Klebsiella*. Gene **198:**111–113.

51. **Swartley, J. S., A. A. Marfin, S. Edupuganti, L.-J. Liu, P. Cieslak, B. Perkins, J. D. Wenger, and D. S. Stephens.** 1997. Capsule switching of *Neisseria meningitidis*. Proc. Natl. Acad. Sci. USA **94:**271–276.
52. **Tacket, C. O., R. K. Taylor, G. Losonsky, Y. Lim, J. P. Nataro, J. B. Kaper, and M. M. Levine.** 1998. Investigation of the roles of toxin-coregulated pili and mannose-sensitive hemagglutinin pili in the pathogenesis of *Vibrio cholerae* O139 infection. Infect. Immun. **66:**692–695.
53. **Taylor, R. K., V. L. Miller, D. B. Furlong, and J. J. Mekalanos.** 1987. The use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc. Natl. Acad. Sci. USA **84:**2833–2837.
54. **Taylor, R. K.** 1999. Virus on virus infects bacterium. Nature **399:**312–313.
55. **Terai, A., K. Baba, H. Shirai, O. Yoshida, Y. Takeda, and M. Nishibuchi.** 1991. Evidence for insertion sequence-mediated spread of the thermostable direct hemolysin gene among *Vibrio* species. J. Bacteriol. **173:**5036–5046.
56. **Waldor, M. K., and J. J. Mekalanos.** 1994. Emergence of a new cholera pandemic: molecular analysis of virulence determinants in *Vibrio cholerae* O139 and development of a live vaccine prototype. J. Infect. Dis. **170:**278–283.
57. **Waldor, M. K., and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272:**1910–1914.
58. **World Health Organization.** 2001. Cholera vaccines. Wkly. Epidemiol. Rec. W. H. O. **76:**117–124.
59. **Yamasaki, S., T. Shimizu, K. Hoshino, S. T. Ho, T. Shimada, G. B. Nair, and Y. Takeda.** 1999. The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are closely related to those of *Vibrio cholerae* O22. Gene **237:**321–332.

Editor: D. L. Burns