Genetic analyses of the putative O and K antigen gene clusters of pandemic Vibrio parahaemolyticus

Okura, Masatoshi ; Osawa, Ro ; Tokunaga, Akihiko ; Morita, Masatomo ; Arakawa, Eiji ; Watanabe, Haruo

(Citation)
Microbiology and Immunology,52(5):251-264

(Issue Date)
2008-06

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000685

Genetic analyses of the putative O- and K-antigen gene clusters of pandemic *Vibrio parahaemolyticus*

Masatoshi Okura[1], Ro Osawa[1,2*], Akihiko Tokunaga[1], Masatomo Morita[3], Eiji Arakawa[3], Haruo Watanabe[3]

[1]Department of Bioresource Science, Graduate School Agricultural Science, Kobe University, Rokko-dai 1-1, Nada-ku, Kobe 657-8501, Japan

[2] Research Center for Food Safety and Security, Graduate School Agricultural Science, Kobe University, Rokko-dai 1-1, Nada-ku, Kobe 657-8501, Japan

[3] Department of Bacteriology, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo 162-8640, Japan

Specified field: Bacteriology/Genetics

Running head: O:K antigen genes of pandemic *V. parahaemolyticus*

Key words: Pandemic *Vibrio parahaemolyticus* O3:K6; Pandemic *Vibrio parahaemolyticus* O4:K68; Serotype conversion

* Corresponding author: Masatoshi Okura

Mailing address: Department of Bioresource Science, Graduate School Agricultural Science, Kobe University, Rokko-dai 1-1, Nada-ku, Kobe 657-8501, Japan

Tel & Fax: +81-78-803-5804

E-mail address: mokura@nih.go.jp

***Abstract*:**

Pandemic *Vibrio parahaemolyticus* strains have rapidly changed their serotypes, but its determinants, especially K-antigen, and the genes involved in serotype have been an open question. The purpose of this study was to gain insights into these points. Although *V. parahaemolyticus* is known to be lacking O-side chain on lipopolysaccharide, and O-antigens are thought to be represented by core oligosaccharides (core OS), the genome sequence of *V. parahaemolyticus* O3:K6 strain RIMD2210633 suggested that this bacterium potentially synthesizes O-side chain. To reveal relatedness between this O-side chain biosynthesis gene cluster, which shows similarity with them of *Vibrio cholerae*, and serotype of *V. parahaemolyticus*, we amplified both core OS and O-side chain gene clusters from various serotype of *V. parahaemolyticus* strains by long PCR and performed PCR restriction fragment lengths polymorphism (RFLP) analyses. The results of RFLP analyses suggested that the core OS biosynthesis gene cluster is related to the O-antigens of pandemic *V. parahaemolyticus*, and putative O-side chain gene cluster do not correlated with O-antigens, but it correlate with K-antigens of pandemic *V. parahaemolyticus*. We then determined the sequence of these regions of a pandemic O4:K68 strain, and compared with the sequence of RIMD2210633. In addition, PCR analysis showed the putative O4 and K68 antigen gene clusters are unique to the strains belonging to the O4 serogroup and K68 serovar, respectively. The data implies that the pandemic O4:K68 *V. parahaemolyticus* strain emerged from the pandemic O3:K6 strain by replacement of the putative O- and K-antigen gene clusters.

INTRODUCTION

Vibrio parahaemolyticus is a major agent of seafood-borne gastroenteritis, which is often associated with the consumption of raw or undercooked seafood (2). A group of strains carrying the *tdh* gene, encoding thermostable direct hemolysin (TDH); the *trh* gene, encoding TDH-related hemolysin (TRH); or both genes are known to cause the gastroenteritis (24). These virulent *V. parahaemolyticus* strains are presently classified into 11 O-serogroups and 65 K-serovars by commercial antisera (16). Although various serotypes of the bacterium can cause infections, a group of strains belonging to O3:K6 and several other serotypes producing TDH have been recognized to be the most prevalent groups in outbreaks occurring worldwide since 1996 (22). Past molecular epidemiological studies (7, 21, 25, 26, 31) revealed that this group of strains showed almost identical genotypic profiles, suggesting them to be clonally related. Thus, these strains have been collectively referred to as pandemic strains. However, it remains unclear why the pandemic strains have been so strongly implicated in epidemics.

To date, the pandemic strains of 14 different serotypes (O1:K25, O1:K41, O1:K56, O1:KUT, O3:K6, O3:K58, O3:K68, O3:K75, O4:K8, O4:K12, O4:K68, O4:KUT, O5:KUT, and OUT:K6) have been reported (4, 6, 12, 18, 22, 25). An O4:K68 serotype strain was initially isolated in 1997 from international travelers (21). Chowdhury et al. (6) suggested that the O4:K68 and other non-O3:K6 pandemic strains had diverged from the pandemic O3:K6 strain with serotype transition. In *Vibrio cholerae*, O139 serogroup strains are genetically very similar

to O1 El Tor strains and the O1 to O139 serogroup conversion has been postulated to be responsible for the emergence of O139 from an O1 El Tor strain (10, 15). The evidence for this conversion is the presence in O139 strains of a locus encoding the O139 antigen biosynthesis genes that has partially replaced the region found in the chromosome 1 of O1 El Tor that encodes the O1 antigen (9, 27)

In general, different serotypes occur as a result of variation in the sugar composition and linkage specificity of the surface-associated polysaccharides. Lipopolysaccharides (LPS) represent O-antigens and usually consist of three distinct regions: lipid A, core oligosaccharide (OS), and O-side chain. The LPS of *V. parahaemolyticus*, however, consists of only lipid A and a short-chain polysaccharide moiety corresponding to the core OS; moreover, the core OS is recognized as the determinant of O-serogroups (14, 16). Meanwhile, capsular polysaccharides (CPS) generally represent the K-antigens that mask the underlying O-antigens. Guvener and McCarter (13) identified a CPS locus of the *V. parahaemolyticus* BB22 strain, which correlated with phase variation-displaying opaque (OP) and translucent (TR) colony morphologies, and the capsule appears as a thick layer surrounding the cell in electron micrographs (11). However, limited information has been available about the determinant of K serovars and the molecular genetics of the O- and K-antigens of *V. parahaemolyticus*. In *V. cholerae*, the putative core OS gene cluster and O antigen (O-side chain) gene clusters were identified, and their nucleotide sequences are available in the database (5, 17, 20, 23, 28, 32).

In this study, in order to identify the genes related to the O- and K- serotype

change of pandemic *V. parahaemolyticus,* we first examined the relationship between putative core OS and O-side chain loci that have similarity to those of *V. cholerae,* and serotype of *V. parahaemolyticus* and we found a genetic region that is presumably related to the O- and K-antigens of pandemic *V. parahaemolyticus*. We subsequently determined the sequence of the putative O- and K-antigen-associated region of a pandemic *V. parahaemolyticus* O4:K68 strain (putative O4:K68 antigen gene clusters) and compared with the corresponding sequence of a pandemic *V. parahaemolyticus* O3:K6 strain RIMD2210633 (19) and other serotypes of strains. From the results obtained, we discuss the possibility that entirely replacing these gene clusters is required for serotype conversions of pandemic *V. parahaemolyticus* strains.

Materials and Methods

Bacterial strains and media

A total of 20 isolates and 43 serotype reference strains of *V. parahaemolyticus* were used in this study, and they are described in Table 1. The serotype reference strains have been used for the manufacture of commercial antisera and were deposited by the National Institute of Infectious Diseases. The strains were classified into 11 O-serogroups and 53 K-serovars, and included 9 pandemic strains of various serotypes (i.e., O3:K6, O4:K68, O1:K25, and O1:KUT), which had been identified by a pandemic strain specific-multiplex PCR assay described previously (26). Genetic profiles (i.e., the presence or absence of *tdh*, *trh*, and a pandemic strain specific *toxRS* sequence [21]) of the strains

are described in Table 1. Bacterial cultures were grown at 37° C in HI broth (25 g of heart infusion broth [Nissui, Tokyo, Japan] and 15 g of NaCl per liter) or on heart infusion (HI) plates (25 g of HI broth [Nissui], 15 g of NaCl, and 15 g of agar per liter). The O and K serotypes were determined by slide agglutination by using a commercial set of O and K antisera (Denka Seiken, Tokyo. Japan). *Escherichia coli* DH5α and pUC119 were used for cloning. *E. coli* strains containing pUC119 were grown in LB broth (Difco, MI) supplemented with 50 μg/ml ampicillin or LB plates supplemented with 100 μg/ml ampicillin.

Identification of the putative O- and K-antigen genes in the genome sequence

In *V. parahaemolyticus,* a CPS locus has been identified (GenBank accession no. AY217749); therefore, we thought this locus to be a genetic region related to K-serovar. In *V. cholerae*, the putative core OS biosynthesis (*wav*) gene cluster has been identified in the available genome sequence of the O1 El Tor strain N16961, and it has been found to reside between *coaD* and *gmhD* (23). The O-antigen (O-side chain) genes of *V. cholerae* were also sequenced for 5 serogroups (O1, O139, O22, O37 and O31) and found to be located between *gmhD* and *rjg* (5, 17, 20, 28, 32). We thus searched for any homolog of the above loci in the entire genome sequence of a pandemic O3:K6 strain RIMD2210633 (19) by using the BLAST programs (1).

PCR-RFLP analysis

As will be described in the results section, we considered that the gene

clusters comprising ORFs VP0190–VP0214, VP0214–VP0238, and VPA1402–VPA1412 were possibly associated with the O- and K-antigen of the sequenced O3:K6 strain RIMD2210633. PCR-RFLP analyses were performed on 56 serotypes of strains to evaluate whether the gene clusters were involved in serotype of *V. parahaemolyticus*. Long and accurate PCR (LA-PCR) assays were carried out with *TaKaRa LA Taq* (Takara Shuzo, Shiga, Japan) to amplify the gene clusters.

In our preliminary study, the regions of VP0190–VP214 and VP0214–VP0238 were found to be too large for successful PCR amplification, and the ORFs, namely, VP0189–VP0195, VP0214, VP0215, VP0217, and VP0239–VP0244, were conserved among 9 pandemic strains. We therefore performed LA-PCR assays using a primer pair to amplify VP0195–VP0214 (VP0195F, GGCTATACATCTTGGCCAGTCGCTACCTA and VP0214R, GTGTTTCTGGCATCTTCAACTGCGGT) and another primer pair to amplify VP0217–VP0239 (VP0217F, CCCTTTCAGAATCGTCAATTCAACAGGC and VP0239R, CAGTATGTGCACGTCAAATCAACGCAG). LA-PCR assay targeting the region corresponding to the ORFs VPA1402–VPA1412 was performed with the primer pair—VPA1402F (CGTAGGTCCATCTGGTTGTGGTAAATC) and VPA1412R (AATACTGACGATCAGACTCGCCAGCA). The primer pairs for the LA-PCR assays were designed based on the genome sequence of RIMD2210633 (GenBank accession no. BA000032).

LA-PCR conditions comprised an initial denaturation at 94° C for 1 min, 35 cycles of denaturation at 98° C for 10 sec, and annealing and extension at 68° C for 15 min (for VP0195—VP0214), 25 min (for VP0217—VP0239) or 10 min

(for VPA1402—VPA1412); and a final extension at 72° C for 10 min.

The LA-PCR products were digested for 6 h by either 2 U of *Hin*dIII or *EcoR*I (Takara Shuzo) with the buffers supplied by the manufacturers and analyzed by 2.0% agarose gel electrophoresis in 0.5 x TBE (45 mM Tris-HCl, 45 mM boric acid, and 1.0 mM EDTA [pH 8.0]) buffer for 70 min, followed by ethidium bromide staining.

DNA sequencing and sequence analyses of putative O4:K68 antigen genes

The sequence of putative O- and K-antigen gene clusters of the O4:K68 strain NIID242-200 was determined based on the PCR products. PCR products less than 20 kbp in length were purified with High Pure PCR Product Purification Kit (Roche, Penzberg, Germany) and sequenced by primer walking. PCR product more than 20 kbp in length was fragmented by sonication, blunt-ended, and size-selected prior to subcloning into the pUC119 vector. Recombinant pUC119 clones were individually picked, amplified, and purified in 96-well plates as single-stranded templates. The sequence was determined using cycle sequencing with BigDye Cycle Sequencing Kit (Applied Biosystems, Foster City, Calif.) and an automated sequencer. Assembly of the DNA sequences and ORF searches were carried out using Sequencher DNA Sequencing Software (Hitachi Software Engineering, Tokyo, Japan). The ORFs were subsequently subjected to a database search by using the BlastX program (1).

Based on the obtained sequence, we designed a total of 20 primer pairs (No.1—No.20) successively covering ORF genes as described in Fig. 3B. In addition, the primer pairs (No. 21) were designed from the genome sequence of

RIMD2210633 to amplify the genes which are located downstream of the putative O- and K-antigen gene clusters. In order to investigate the distribution of the genes in the putative O4:K68 antigen gene clusters among the various serotypes of strains, PCR assays using the above mentioned primer pairs were performed for the selected strains as described in Table 3 under appropriate PCR conditions with *TaKaRa EX Taq* (Takara Shuzo).

Nucleotide sequence accession number

The nucleotide sequence of the putative O4:K68 antigen gene cluster has been deposited in the DNA Data Bank of Japan (DDBJ). The accession number is AB353134.

RESULTS

Candidate gene clusters involved in O- and K-antigen of pandemic *V. parahaemolyticus* O3:K6

The CPS locus of *V. parahaemolyticus* BB22 was highly homologous with the ORFs VPA1402—VPA1412 located on chromosome 2 of *V. parahaemolyticus* RIMD2210633 (homology was over 98%). The ORFs VP0190—VP0214 on chromosome 1 of RIMD2210633 showed partial homology with the core OS biosynthesis locus of *V. cholerae* O1 El Tor N16961 (Fig. 1). Meanwhile, the ORFs VP0214—VP0238 on chromosome 1 of RIMD2210633 also showed partial homology with the O-side chain loci of *V. cholerae* strain serogroups O1 N16961, O139 MO45, and O31 NRT36S (Fig. 1), although *V.*

parahaemolyticus has no O-side chain. According to the genome sequence database of *V. parahaemolyticus* RIMD2210633, the ORFs VP0190—VP0238 contained many genes encoding glycosyl transferase, and for synthesis of the nucleotide sugar precursor for formation of polysaccharides. On the basis of these results, we assumed that the regions located on chromosome 1, comprising ORFs VP0190—VP0238, and that on chromosome 2, comprising ORFs VPA1402—VPA1412 are involved in the O- and K-antigen of the pandemic O3:K6 strain.

PCR-RFLP analyses of the putative O- and K-antigen gene clusters for various serotype of strains

The results of the PCR-RFLP analysis with *Hin*dIII restriction enzyme are presented in Fig.2. Seven of the pandemic strains displayed almost identical RFLP patterns for the CPS locus (corresponding to VPA1402—VPA1412 of RIMD 2210633) (Fig. 2A), suggesting that this locus of *V. parahaemolyticus* is not responsible for the differences in the K-serovar. RFLP profiles of the putative core OS locus (corresponding to VP0195—VP0214) showed a high level of similarity among the strains belonging to the same O-serogroup (Fig. 2B). Meanwhile, RFLP analyses on the putative O-side chain locus (corresponding to VP0217—VP0239) revealed that strains expressing the same O-antigen exhibited distinct patterns, but the strains belonging to the same K-serovar (K25, K4 and K68) showed very similar patterns (Fig. 2C, lanes 1 and 4, 9 and 23, and 22 and 32). Similar results were obtained for RFLP analyses with *Eco*RI restriction enzyme (data not shown). These results implied that the regions

equivalent to the putative core OS locus of *V. cholerae* may be involved in the O-serogroup of *V. parahaemolyticus* and the regions equivalent to the O-side chain locus of *V. cholerae* are related to the K-serovar of *V. parahaemolyticus*. However, all of the strains belonging to O2, O6, and O9 serogroup and 16 K serovars of the strains (K3, K5, K10, K15, K17, K19, K28, K36, K40, K50, K51, K52, K59, K60, K63, and K70) failed to yield the PCR products targeting VP0195—0214 and VP0217—0239, respectively (Table 1). With regard to these strains, both the PCR products targeting VP0195 and VP0214 and those targeting VP0217 and VP0239 were not amplified (data not shown); therefore, we assumed that LA-PCR products of these strains could not be yielded probably because of a less similarity or the absence of genes containing primer annealing sequences.

Putative O4:K68 antigen gene clusters

The region containing of the putative O- and K-antigen gene clusters explained above of an O4:K68 strain NIID242-200 was sequenced and its length was approximately 62 kbp. The GC content of this region was 40.4%, which is lower than that of the entire genome sequence of RIMD2210633 (45.4%). There were 62 ORFs, and the annotation for each ORF is listed in Table 2; it shows the extent of homology of the ORFs with the genes of an O3:K6 strain and other bacterial species. An approximately 9-kbp DNA sequences contained 8 ORFs (ORF54–ORF61) inserted in the *rjg* (ORF53) region (Fig. 3A). Many of these ORFs encoded glycosyl transferase, and enzymes involved in the pathway for synthesis of the nucleotide sugar precursor for external polysaccharides (Table

2). 4 ORFs (ORF52, ORF58, ORF59 and ORF60) encoded transposases (Table 2, Fig. 3). ORF1-ORF7, and ORF62 were found to have more than 98% homology with those of the O3:K6 strain (VP0190–VP0196, and VP0239, respectively)(Fig. 3A). ORF23–ORF24, ORF26-ORF29, ORF31-ORF33, and ORF50-ORF51 had 80–95% homology with those of the O3:K6 strain (VP0212–VP0213, VP0214–VP0217, VP0219–VP0221, and VP0236, respectively) (Fig. 3A). The other ORFs had no significant sequence similarity with any genes in the putative O3:K6 antigen gene cluster.

Distribution of the putative O4:K68 antigen genes in the strains belonging to various serotypes

The results of the PCR analysis using the primer pairs described in Fig. 3B are presented in Table 3. Ten of the non-pandemic strains belonging to O4 serogroup were positive in all the PCR assays targeting the putative O-antigen genes of the O4:K68 strain (primer pairs 1–9) except the assays using primer pairs 3 and 5 (Table 3). In contrast, the strains expressing other O-antigens were negative in all the assays except those using primer pairs 1, 2, 3, and 4. PCR assays targeting the putative K-antigen genes of the O4:K68 strain (primer pairs 10–20) indicated that the O5:K68 strain was positive in all the assays but the product size of the assay using primer pair 17 was about 1.3-kbp smaller than that of O4:K68 strains. The pandemic O1:K25 strains were negative for the assays using primer pairs 11–16 but positive for the assays using primer pairs 17–20, although the fragment sizes of the PCR assay with primer pair 17 were smaller than those of the O4:K68 strains as seen with an O5:K68 strain (Table 3).

The strains belonging to other K-serovar were all negative for the assays using primer pairs 11–20.

Analyses of the left and right junctions of the putative O- and K-antigen region of pandemic strains

Based on the above results, we defined the left and right junctions (i. e., the nucleotide sequences where the common backbone sequence ends and the serotype-specific sequence starts) of the putative O- and K-antigen region of the pandemic O4:K68 strain, which compared with that of the O3:K6 strain. The left junction was located at ORF7 and the right junction was located at ORF62 (Fig. 3A). In our preliminary study, pandemic O1:K25 and O1:KUT strains had the genes corresponding to ORF1–ORF6 and ORF62 (data not shown); thus we assumed that left and right junctions of the putative O- and K-antigen region of the pandemic O1:K25 strains and those of the pandemic O1:KUT strains, which compared with those of the pandemic O3:K6 strain, are at the region inside of the genes corresponding to ORF6–ORF62. The results of multiple-sequence alignment are described in Fig. 4A and 4B. The left junctions were identical among pandemic non-O3:K6 strains (Fig. 4A). Meanwhile, the right junction of the pandemic O1:KUT strain AN-16000 was found to be different from those of the pandemic O4:K68 strain NIID242-200 and O1:K25 strain AP-18000 (Fig. 4B). We performed further sequence analysis in order to determine the right junction of the putative O- and K-antigen region between the pandemic O4:K68 strain and pandemic and non-pandemic O1:K25 strains. On the basis of results of this sequence analysis, we assumed that the genetic alignment of right boundary

regions of 4 serotype of pandemic strains (belonging to O3:K6, O4:K68, O1:K25 and O1:KUT) and non-pandemic O1:K25 strain are represented in Fig. 4C. The pandemic O1:KUT strain AN-16000 had lost the entire *rjg* sequence (Fig. 4B and 4C), while the pandemic O1:K25 strain AP-18000 had genes corresponding to ORF51–ORF61 of the O4:K68 antigen gene cluster except for a 1303-bp region which contained an IS element (ORF52) (Fig 3A, Fig.4C), and which possessed identical 9-bp repeats (CTGTTTTAA) at their extremities (data not shown). In contrast, the right junction of the non-pandemic O1:K25 strain REF-K25 was different from that of the pandemic O1:K25 strain AP-18000 and its right boundary region showed close similarity to that of the pandemic O3:K6 strain RIMD2210633 (Fig. 4C). These results pointed to the possibility that pandemic O1:K25 strains were diverged from the pandemic O4:K68 strain.

Discussion

We found that the pandemic O3:K6 strain had gene clusters that showed some similarity with core OS and O-side chain regions of the *V. cholerae* strains (Fig. 1), and CPS locus of a *V. parahaemolyticus* strain. Since the CPS loci of pandemic *V. parahaemolyticus* strains displayed almost identical PCR-RFLP patterns regardless of serotypes (Fig. 2A), this locus may not be directly involved in the K-serovar of *V. parahaemolyticus*. According to the results of the RFLP analyses of the regions equivalent to the core OS gene cluster of *V. cholerae*, these regions correlated with the O serogroup of *V. parahaemolyticus* (Fig. 2B). These findings are consistent with previously reported results (14, 16) that

recognized the core OS as the O antigen in *V. parahaemolyticus*. Meanwhile, the results of the RFLP analyses of the regions equivalent to the O-side chain gene cluster of *V. cholerae* suggests that this region is not related to the O-antigen of *V. parahaemolyticus*, but may be involved in the K-antigen of *V. parahaemolyticus* (Fig. 2C), although *V. parahaemolyticus* strains have no O-side chain. Unlike O1 serogroup strains, O139 and O31 isolates of *V. cholerae* are known to have capsules that are a capsular form of the O-side chain unlinked to lipid A and core OS (29, 30), and the biosynthesis genes of the capsules of the O139 and O31 strains are encoded in the genetic locus of the O-side chain (5, 32). In addition, we could not find any homolog of the gene encoding O-antigen ligase, which is involved in the linkage between core OS and O-side chain, in these gene clusters of pandemic O3:K6 and O4:K68 strains. Thus, *V. parahaemolyticus* strains have the potential to produce such capsules. However, it remains unconfirmed whether the gene clusters identified in this study were related to the expression of O and K antigens and the possibility that other genetic regions are associated with O and K antigen differences is left open; therefore further work is required to clarify them.

Comparison of the putative O- and K-antigen gene clusters between the pandemic O3:K6 and O4:K68 strains suggested that dynamic rearrangement of a large DNA sequence was essential for the O3:K6 to O4:K68 serotype conversion (Fig. 3A). In addition, other non-pandemic strains belonging to O4 serogroup were likely to share genetic alignments of the O antigen gene cluster similar to those of the pandemic O4:K68 strains and the O5:K68 serotype strain were considered to have genetic alignments of the K antigen gene cluster similar

to those of the pandemic O4:K68 strains (Table 3). These data suggest that the pandemic O4:K68 strain originated from a pandemic O3:K6 strain by a one- or two-step recombination-mediated replacement of the approximately 50-kbp long putative O3:K6 antigen gene cluster (VP0190–VP0238) with an approximately 60-kbp gene cluster in O4:K68 (ORF1–ORF61).

To determine the breakpoints of the gene cluster among different serotypes of pandemic strains, we performed sequence analyses of the boundary region. Pandemic strains belonging to non-O3:K6 serotypes shared almost identical sequence alignments in the region outside of the left break points (Fig. 4A). In contrast, the sequences outside of the right break points were different (Fig. 4B). Pandemic O1:K25 strains as well as pandemic O4:K68 strains possessed the insertion sequences in *rjg*, although the non-pandemic O1:K25 strain did not have the insertion sequence (Table 3, Fig. 4C). This result suggests that pandemic O1:K25 strains are closely related to the pandemic O4:K68 strain. Moreover, the pandemic non-O3:K6 strains used in this study showed deletion of *rjg* (Fig. 4C). Since the *rjg* has been reported to be a hot spot for DNA rearrangement and seems to function as an anchor region for a variety of DNA transfer events in *V. cholerae* (17), *rjg* deletion may play a critical role in transitions of O- and K-antigens of the pandemic *V. parahaemolyticus.*

In accordance with the above observations, it is highly likely that genetic exchange events at this gene clusters identified in this study through horizontal transfer of the genetic elements resulted in the emergence of a new serotype of pandemic *V. parahaemolyticus.* However, it was not obvious in this study whether the serotype conversions were actually caused among pandemic

strains and how the serotype conversions were caused, but this is the first report to determine the sequence of the gene cluster related to the O and K antigens of *V. parahaemolyticus* and provide insight into the K antigen of *V. parahaemolyticus* and the transitions of O- and K-antigens of pandemic strains.

This work was supported by grants-in-aid for science research from the Ministry of Education, Science and Culture of Japan, the Ministry of Health and Welfare of Japan, and the Japan Health Science Foundation.

References

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403-410.
2. Blake, P.A., Weaver, R.E., and Hollis, D.G. 1980. Disease of humans (other than cholera) caused by vibrios. Annu. Rev. Microbiol. **34**: 341-367.
3. Blokesch, M., and Schoolnik, G.K. 2007. Serogroup conversion of *Vibrio cholerae* in aquatic Reservoirs. Plos. Pathogens. **3**: 733-742.
4. Bhuiyan, N.A., Ansaruzzaman, M., Kamruzzaman, M., Alam, K., Chowdhury, N.R., Nishibuchi, M., Faruque, S.M., Sack, D.A., Takeda, Y., and Nair, G.B. 2002. Prevalence of the pandemic genotype of *Vibrio parahaemolyticus* in Dhaka, Bangladesh, and significance of its distribution across different serotypes. J. Clin. Microbiol. **40**: 284-286.
5. Chen, Y., Bystricky, P., Adeyeye, J., Panigrahi, P., Ali, A., Johnson, J.A., Bush, C.A., Morris, J.G., and Stine, O.C. 2007. The capsule polysaccharide structure and biogenesis for non-O1 *Vibrio cholerae* NRT36S: genes are

embedded in the LPS region. BMC Microbiol. **7**: 20.

6. Chowdhury, A., Ishibashi, M., Thiem, V.D., Tuyet, D.T., Tung, T.V., Chien, B.T., Seidlein-Lv, L., Canh, D.G., Clemens, J., Trach, D.D., and Nishibuchi, M. 2004. Emergence and serovar transition of *Vibrio parahaemolyticus* pandemic strains isolated during a diarrhea outbreak in Vietnam between 1997 and 1999. Microbiol. Immunol. **48**: 319-327.
7. Chowdhury, N.R., Stine, O.C., Morris, J.G., and Nair, G.B. 2004. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequencing typing. J. Clin. Microbiol. **42**: 1280-1282.
8. Cholera Working Group, International Center for Diarrhoeal Disease Research, Bangladesh. 1993. Large epidemic of cholera like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. Lancet **342**: 387-390.
9. Comstock, L. E., Johson, J. A., Michalski, J. M., Morris, J. G., and Kaper, J. B. 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.
10. Dziejman, M., Balon, E., Boyd, D., Fraser, C. M., Heidelberg, J. F., and Mekalanos, J. J. 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. Proc. Natl. Acad. Sci. U S A **88**: 5403-5407.
11. Enos-Berlage, J.L., and McCarter, L.L. Relation of capsular polysaccharide production and colonial cell organization to colony morphology in *Vibrio parahaemolyticus*. 2000. J. Bacteriol. **182**: 5513-5520.

12. Gil, A.I., Miranda, H.C., Lanata, F., Prada, A., Hall, E.R., Barreno, C.M., Nusrin, S., Bhuiyan, N.A., Sack, D.A., and Nair, G.B. 2007. O3:K6 Serotype of *Vibrio parahaemolyticus* identical to the global pandemic clone associated with diarrhea in Peru. Int. J. Infect. Dis. **11**: 324-328.
13. Guvener, Z.T., and McCarter, L.L. 2003. Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*. J. Bacteriol. **185**: 5431-5441.
14. Han, T.J., and Chai, T.J. 1992. Electrophoretic and chemical characterization of lipopolysaccharides of *Vibrio parahaemolyticus*. J. Bacteriol. **174**: 3140-3146.
15. Higa, N., Honma, Y., Albert, M. J., and Iwanaga, M. 1993. Characterization of *Vibrio cholerae* O139 synonym Bengal isolated from patients with cholera-like disease in Bangladesh. Microbiol. Immunol. **37**: 971-974.
16. Iguchi, T., Kondo, S., and Hisatune, K. 1995. *Vibrio parahaemolyticus* O serotypes from O1 to O13 all produce R-type lipopolysaccharide: SDS-PAGE and compositional sugar analysis. FEMS Microbiol. Lett. **130**: 287-292.
17. Li, M., Shimada, T.J., Morris Jr., G., Sulakvelidze, A., and Sozhamannan, S. 2002. Evidence for the emergence of non-O1 and non-O139 *Vibrio cholerae* strains with pathogenic potential by exchange of O-antigen biosynthesis regions. Infect. Immun. **70**: 2441-2453.
18. Laohaprertthisan, V., Chowdhury, A., Kongmuang, U., Kalnauwakul, S., Ishibashi, M., Matsumoto, C., and Nishibuchi, M. 2003. Prevalence and serodiversity of the pandemic clone among the clinical strains of *Vibrio*

19. Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, N., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, Y., Honda, T., Shinagawa, H., Hattori, M., and Iida, T. 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet. **361**: 743-749.

20. Manning, P.A., Heuzenroeder, M.W., Yeadon, J., Leavesley, D.I., Reeves, P.R., and Rowley, D. 1986. Molecular cloning and expression in *Escherichia coli* K-12 of the O antigens of the Inaba and Ogawa serotypes of the *Vibrio cholerae* O1 lipopolysaccharides and their potential for vaccine development. Infect. Immun. **53**: 272-277.

21. Matsumoto, C., Okuda, J., Ishibashi, M., Iwanaga, M., Garg, P., Rammamurthy, T., Wong, H.C., DePaola, A., Kim, Y.B., Albert, M.J., and Nishibuchi, M. 2000. Pandemic spread of an O3:K6 clone of *Vibrio parahaemolyticus* and emergence of related strains evidenced by arbitrarily primed PCR and toxRS sequence analyses. J. Clin. Microbiol. **38**: 578-585.

22. Nair, G.B., Ramamurthy, T., Bhattacharya, S.K., Dutta, B., Takeda, Y., and Sack, D.A. 2007. Global Dissemination of *Vibrio parahaemolyticus* Serotype O3:K6 and Its Serovariants. Clin. Microbiol. Rev. **20**: 39-48.

23. Nesper, J., Kraiss, A., Schild, S., Blass, J., Klose, K.E., Bockemuhl, J., and Reidl, J. 2002. Comparative and Genetic Analyses of the Putative *Vibrio cholerae* Lipopolysaccharide Core Oligosaccharide Biosynthesis (*wav*) Gene Cluster. Infect. Immun. **70**: 2419-2433.

24. Nishibuchi, M., and Kaper, J.B. 1995. Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. Infect. Immun. **63**: 2093-2099.

25. Okuda, J., Ishibashi, M., Hayakawa, E., Nishino, T., Takeda, Y., Mukhopadhyay, A.K., Garg, S., Bhattacharya, S.K., Nair, G.B., and Nishibuchi, M. 1997. Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. J. Clin. Microbiol. **35**: 3150-3155.

26. Okura, M., Osawa, R., Iguchi, A., Arakawa, E., Terajima, J., and Watanabe, H. 2003. Genotypic analyses of *Vibrio parahaemolyticus* and development of a pandemic group-specific multiplex PCR assay. J. Clin. Microbiol. **41**: 4676-4682.

27. Pajni, S., Sharma, C., Bhasin, N., Ghosh, A., Ramamurthy, T., Nair, G. B., Ramajayam, S., Das, B., Kar, S., Roychowdhury, S., et al. 1995. Studies on the genesis of *Vibrio cholerae* O139: identification of probable progenitor strains. J. Med. Microbiol. **42**: 20-25.

28. Sozhamannan, Y., Deng, K., Li, M., Sulakvelidze, A., Kaper, J.B., Johnson, J.A., Nair, G.B., and Morris Jr, J.G. 1999. Cloning and sequencing of the genes downstream of the *wbf* gene cluster of *Vibrio cholerae* serotype O139 and analysis of the junction genes in other serotypes. Infect. Immun. **67**: 5033-5040.

29. Stroeher, U. H., Jedani, K. E., and Manning, P. A. 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.

30. Waldor, M. K., Colwell, R., and Mekalanos, J. J. 1994. The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants. Proc. Natl. Acad. Sci. USA. **91**: 11388-11392.

31. Wong, H.C., Liu, S.H., Wang, T.K., Lee, C.L., Chiou, C.S., Liu, D.P., Nishibuchi, M., and Lee, B.K. 2000. Characteristics of *Vibrio parahaemolyticus* O3:K6 from Asia. Appl. Environ. Microbiol. **66**: 3981-3986.

32. Yamasaki, S., Shimizu, T., Hoshino, K., Ho, S.T., Shimada, T.G., Nair, B., and Takeda, Y. 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.

Figure Legends

Fig. 1. **Schematic representation of the putative core OS and O-side chain biosynthesis genes of *V. cholerae* O1 El Tor N16961 (A), putative O-side chain biosynthesis genes of *V. cholerae* O139 MO45 and O31 NRT36S (B), and the homologous region of the sequenced *V. parahaemolyticus* RIMD2210633.**

Arrows represent ORFs; black arrows indicate junction genes of putative core OS (*coaD* and *gmhD*) and O-side chain region (*gmhD* and *rjg*) of *V. cholerae*. The grey zones represent the homologous region between two strains. The sequence data were obtained from the DDBJ/EMBL/GenBank database under accession numbers AE003852 (*V. cholerae* O1 El Tor N16961), AB012956 (*V. cholerae* O139 MO45), DQ915177 (*V. cholerae* O31 NRT36S), and BA000031 (*V. parahaemolyticus* RIMD2210633). Gene names were adopted from previous reports (5, 17, 23, 32).

Fig. 2. **PCR-RFLP analyses of putative O and K antigen gene clusters by restriction with *Hin*d III for representative strains.**

K-serovars of strains are indicated on the lane numbers and numbers in parentheses are corresponding to experimental numbers in Table 1. Bars indicate the strains belonging to same O-serogroup. (A) CPS locus. Lanes: M, 1-Kb PLUS ladder marker; 1–2, pandemic O3:K6 strain; 3–5, pandemic O4:K68 strains; 6, pandemic O1:K25 strain; and 7, pandemic O1:KUT strain. (B) The locus corresponding to the putative core OS region of *V. cholerae.* Lanes: M, 1-Kb PLUS ladder marker; 1–5, O1 strains; 6–10, O3 strains; 11–15, O4 strains;

16–19, O5 strains; 20, O7 strains; 21–24, O8 strains; 25–27, O10 strains; and 28–30, O11 strains (C) The locus corresponding to the putative O-side chain region of *V. cholerae*. Lanes: M, 1-Kb PLUS ladder marker; 1–7, O1 strains; 8–21, O3 strains; 22–31, O4 strains; 32, O5 strains; 33–34, O6 strains; 35–38, O8 strains; 39, O9 strains; and 40, O10 strains. K serovars of the strains of Lanes 1 and 4, 9 and 23, and 22 and 32 are K25, K4, and K68, respectively.

Fig. 3. **Schematic representation of the putative O4:K68 antigen region of NIID242-200, and the homologous region of the sequenced *V. parahaemolyticus* RIMD2210633.**
(A) Comparative nucleotide sequence analysis between a pandemic O4:K68 strain NIID242-200 and a pandemic O3:K6 strain RIMD2210633. Grey zones represent the homologous region between two strains. (B) Order and orientation of the putative O4:K68 antigen gene cluster and location of the designed primers (No. 1–21) for PCR amplification. Arrows represent the ORFs; Black arrows indicate the homologous genes of *coaD*, *gmhD* or *rjg* of *V. cholerae*; Arrows indicated by black triangle encode transposases.

Fig. 4. **Comparison of nucleotide sequences, showing genetic alignments of the left and right breakpoints of putative O and K antigen gene clusters.**
(A) Multiple alignments of the left junctions from NIID242-200 (O4:K68), AP-18000 (O1:K25), and AN-16000 (O1:KUT) compared with RIMD2210633 (O3:K6). The asterisks show identical nucleotides. The solid arrowhead and its

orientation indicate the VP0196 and ORF7 stop codon (underlined). (B) Multiple alignments of the right junctions from NIID242-200 (O4:K68), AP-18000 (O1:K25), and AN-16000 (O1:KUT) compared with RIMD2210633 (O3:K6). The asterisks show identical nucleotides. The hollow arrowhead and its orientation indicate the VP0237 stop codon (underlined). The solid arrowhead and its orientation indicate the VP0239 stop codon (underlined). (C) Schematic drawing of gene alignments of the right boundary region of RIMD210633 (O3:K6), NIID242-200 (O4:K68), AP-18000 (O1:K25), AN-16000 (O1:KUT) and REF-K25 (O1:K25). Grey zones indicate regions with high homology. Black arrows show the genes corresponding to the *rjg* gene of putative O-side chain loci of *V. cholerae,* and shaded arrows indicate insertion sequences.

Abbreviations: aa, amino acid(s); CPS, capsularpolysaccharide; IS, insertion sequence; LA, long and accurate; LPS, lipopolysaccharide; orf, open reading frame; OS, oligosaccharide; p, plasmid; PCR, polymerase chain reaction; and UT, untypable

Table 1. *V. parahaemolyticus* strains used in this study and results of LA-PCR targeting the putative O and K antigen genes

Strain	O:K Serotype	Year of isolation	Country of isolation	Source	Results of the multiplex PCR[b]			Results of LA-PCR targeting following region[b] ;		Experimental no.[c]
					tdh	*trh*	*toxRS/ new*	VP0195-VP0214	VP0217-VP0239	
Pandemic strains										
RIMD2210633	O3:K6	1996	Japan	human	+	-	+	+	+	1
NIID K7	O3:K6	1998	Japan	human	+	-	+	+	+	2
NIID 59-99	O3:K6	1999	Thailand	human	+	-	+	+	+	3
NIID 242-200	O4:K68	2000	Korea	human	+	-	+	+	+	4
KIH 03-57	O4:K68	2003	Japan	human	+	-	+	+	+	5
DMST 17875	O4:K68	2003	Thailand	human	+	-	+	+	+	6
AO-24491	O1:K25	1999	Bangladesh	human	+	-	+	+	+	7
AP-18000	O1:K25	2000	Bangladesh	human	+	-	+	+	+	8
AN-16000	O1:KUT	1998	Bangladesh	human	+	-	+	+	+	9
Non-pandemic strains[a]										
ATCC17802	O1:K1	Unknown	Japan	human	-	+	-	+	+	10
REF-K25*	O1:K25	Unknown	Unknown	Unknown	+	-	-	+	+	11
DMST17876	O1:K56	2003	Thailand	human	+	-	-	+	+	12
REF-K64*	O1:K64	Unknown	Unknown	Unknown	+	-	-	+	+	13
REF-K69*	O1:K69	Unknown	Unknown	Unknown	+	+	-	+	+	14
REF-K3*	O2:K3	Unknown	Unknown	Unknown	+	-	-	-	-	
REF-O2*	O2:K28	Unknown	Unknown	Unknown	-	-	-	-	-	
REF-K4*	O3:K4	Unknown	Unknown	Unknown	+	-	-	+	+	15
REF-K5*	O3:K5	Unknown	Unknown	Unknown	-	-	-	+	-	16
KE9967	O3:K6	1981	Japan	human	+	-	-	+	+	
REF-K7*	O3:K7	Unknown	Unknown	Unknown	+	-	-	+	+	17
REF-K29*	O3:K29	Unknown	Unknown	Unknown	+	-	-	+	+	18
REF-K30*	O3:K30	Unknown	Unknown	Unknown	+	-	-	+	+	19
REF-K31*	O3:K31	Unknown	Unknown	Unknown	-	-	-	+	+	20
REF-K33*	O3:K33	Unknown	Unknown	Unknown	+	-	-	+	+	21
REF-K37*	O3:K37	Unknown	Unknown	Unknown	-	-	-	+	+	22
REF-K43*	O3:K43	Unknown	Unknown	Unknown	+	-	-	+	+	23
REF-K45*	O3:K45	Unknown	Unknown	Unknown	-	-	-	+	+	24
KE10542	O3:K48	1999	Thailand	human	-	-	-	+	+	25
REF-K54*	O3:K54	Unknown	Unknown	Unknown	+	-	-	+	+	26
REF-K58*	O3:K58	Unknown	Unknown	Unknown	+	-	-	+	+	27
REF-K59*	O3:K59	Unknown	Unknown	Unknown	+	-	-	+	-	
REF-K65*	O3:K65	Unknown	Unknown	Unknown	+	-	-	+	+	28
KIH VP19	O4:K4	Unknown	Unknown	human	-	+	-	+	+	29
KE10538	O4:K8	1999	Thailand	human	+	-	-	+	+	30
DMST17873	O4:K9	2003	Thailand	human	+	-	-	+	+	31
REF-K10*	O4:K10	Unknown	Unknown	Unknown	+	-	-	+	-	32
KIH 03-60	O4:K11	2003	Japan	food	-	-	-	+	+	33
RIMD 2210587	O4:K12	1996	unknown	unknown	+	-	-	+	+	34
REF-K13*	O4:K13	Unknown	Unknown	Unknown	+	-	-	+	+	35
REF-K42*	O4:K42	Unknown	Unknown	Unknown	-	-	-	+	+	36
REF-K49*	O4:K49	Unknown	Unknown	Unknown	-	+	-	+	+	37
REF-K53*	O4:K53	Unknown	Unknown	Unknown	-	-	-	+	+	38
REF-K63*	O4:K63	Unknown	Unknown	Unknown	+	-	-	+	-	
REF-K15*	O5:K15	Unknown	Unknown	Unknown	+	-	-	+	-	39
REF-O5*	O5:K17	Unknown	Unknown	Unknown	-	-	-	+	-	40
REF-K60*	O5:K60	Unknown	Unknown	Unknown	+	-	-	+	-	41
REF-K68*	O5:K68	Unknown	Unknown	Unknown	+	-	-	+	+	42
REF-O6*	O6:K18	Unknown	Unknown	Unknown	+	+	-	-	+	43
REF-K46*	O6:K46	Unknown	Unknown	Unknown	+	+	-	-	+	44
REF-O7*	O7:K19	Unknown	Unknown	Unknown	+	-	-	+	-	45
REF-O8*	O8:K20	Unknown	Unknown	Unknown	+	-	-	+	+	46
REF-K22*	O8:K22	Unknown	Unknown	Unknown	+	-	-	+	+	47
REF-K39*	O8:K39	Unknown	Unknown	Unknown	+	-	-	+	+	48
KE10541	O8:K41	1999	Thailand	human	+	-	-	+	+	49
REF-K70*	O8:K70	Unknown	Unknown	Unknown	+	-	-	+	-	
REF-O9*	O9:K23	Unknown	Unknown	Unknown	+	-	-	-	+	50
KIH VP28	O10:K19	Unknown	Japan	human	-	-	-	+	-	51
REF-K52*	O10:K52	Unknown	Unknown	Unknown	+	-	-	+	-	52
REF-K71*	O10:K71	Unknown	Unknown	Unknown	+	+	-	+	+	53
REF-K36*	O11:K36	Unknown	Unknown	Unknown	+	-	-	+	-	54
REF-O11*	O11:K40	Unknown	Unknown	Unknown	-	+	-	+	-	55
REF-K50*	O11:K50	Unknown	Unknown	Unknown	-	-	-	+	-	56
REF-K51*	O11:K51	Unknown	Unknown	Unknown	+	-	-	+	-	

[a] Asterisks indicate O-serogroup or K-serovar reference strains.

[b] +, present; -, absent.

[c] Experimental numbers corresponding to the representative strains for RFLP analyses in Fig.2.

Table 2. Putative O- and K-antigen biosynthesis gene products of *V. parahaemolyticus* O4:K68 NIID242-200

ORF	No. of amino acids	Species and locus tag of homologous protein	Accession no.	% aa Identity (/ aa overlap)	Predicted function[a)]
1	167	VP0190 of *V. parahaemolyticus*	BAC58453	99/ 167	Lipopolysaccharide kinase
2	311	VP0191 of *V. parahaemolyticus*	BAC58454	98/ 311	Putative lipopolysaccharide-modifying enzyme
3	262	VP0192 of *V. parahaemolyticus*	BAC58455	100/ 262	Glycosyl transferases
4	255	VP0193 of *V. parahaemolyticus*	BAC58456	99/ 255	Glycosyl transferases
5	351	VP0194 of *V. parahaemolyticus*	BAC58457	98/ 351	ADP-heptose-LPS heptosyltransferase
6	135	VP0195 of *V. parahaemolyticus*	BAC58458	99/ 135	Diacylglycerol kinase
7	236	VP0196 of *V. parahaemolyticus*	BAC58459	97/ 225	3-deoxy-D-manno-octulosonic-acid kinase
8	354	V12G01 11448 of *V. alginolyticus*	EAS75328	96/ 354	dTDP-D-glucose 4;6-dehydratase
		VP0222 of *V. parahaemolyticus*	BAC58485	73/ 351	dTDP-D-glucose 4;6-dehydratase
9	288	V12G01 11453 of *V. alginolyticus*	EAS75329	96/ 288	Glucose-1-phosphate-thymidylyltransferase
		VP0223 of *V. parahaemolyticus*	BAC58486	71/ 288	Glucose-1-phosphate-thymidylyltransferase
10	137	V12G01 11458 of *V. alginolyticus*	EAS75330	95/ 131	WblP protein
11	173	AIQ_0139 of *V. harveyi*	EDL68286	58/ 163	Putative butyryltransferase
12	368	V12G01 11468 of *V. alginolyticus*	EAS75332	96/ 368	WblQ protein
13	309	V12G01 11473 of *V. alginolyticus*	EAS75333	87/ 307	Glycosyl transferase
14	309	Noc 1947 of *Nitrosococcus oceani*	ABA58409	32/ 234	Putative integral membrane protein
15	137	V12B01 01087 of *Vibrio splendidus*	EAP95344	66/ 133	Hypothetical protein
16	283	V12B01 01082 of *V. splendidus*	EAP95343	63/ 272	Prenyltransferase
17	133	Maqu_1646; *Marinobacter aquaeolei*	ABM18730	54/ 128	Hypothetical protein
18	427	MCA2566; *Methylococcus capsulatus*	AAU91361	54/ 425	Putative oxidoreductase, FAD-binding
19	245	MCA2567; *M. capsulatus*	AAU91362	49/ 245	Putative oxidoreductase, short-chain dehydrogenase/reductase
20	540	SCB49_01497; unidentified eubacterium	EDM42937	28/ 535	Hypothetical protein
21	426	VP0211 of *V. parahaemolyticus*	BAC58474	66/ 423	3-deoxy-D-manno-octulosonic-acid transferase
22	289	BACCAC_00463; *Bacteroides caccae*	EDM22091	42/ 278	Glycosyl transferases
23	351	VP0212 of *V. parahaemolyticus*	BAC58475	88/ 344	ADP-heptose-LPS heptosyltransferase II
24	240	V12G01 11503 of *V. alginolyticus*	EAS75339	90/ 239	Glycosyl transferase, Lex2B
		VP0210 of *V. parahaemolyticus*	BAC58473	37/ 239	Glycosyl transferase, Lex2B
25	328	V12G01 11508 of *V. alginolyticus*	EAS75340	89/ 328	Lipid A biosynthesis lauroyl acyltransferase
		VP0213 of *V. parahaemolyticus*	BAC58476	86/ 328	Lipid A biosynthesis lauroyl acyltransferase
26	313	V12G01 11513 of *V. alginolyticus*	EAS75341	98/ 312	ADP glyceromanno-heptose 6-epimerase, GmhD
		VP0214 of *V. parahaemolyticus*	BAC58477	95/ 312	ADP glyceromanno-heptose 6-epimerase, GmhD
27	732	VP0215 of *V. parahaemolyticus*	BAC58478	88/ 732	Bacterial putative lipoprotein
28	253	VP0216 of *V. parahaemolyticus*	BAC58479	92/ 253	Hypothetical protein
29	225	VP0217 of *V. parahaemolyticus*	BAC58480	93/ 225	Putative regulator
30	115	Unknown			Hypothetical protein
31	171	VP0219 of *V. parahaemolyticus*	BAC58482	91/ 171	Hypothetical protein

a) Deduced from database with blastP program (1).

Table 2 (continued). Putative O- and K-antigen biosynthesis gene products of *V. parahaemolyticus* O4:K68 NIID242-200

ORF	No. of amino acids	Species and locus tag of homologous protein	Accession no.	% aa Identity (/ aa overlap)	Predicted function[a)]
32	890	VP0220 of *V. parahaemolyticus*	BAC58483	95/ 890	OtnA protein, putative capsule transport protein
33	322	VP0221 of *V. parahaemolyticus*	BAC58484	81/ 319	OtnB protein, chain length determinant
34	126	VCV52_0225 of *V. cholerae* O141	EAX61094	77/ 126	Hypothetical protein
35	528	VC0231 of *V. cholerae* O1	AAF93407	59/ 528	Hypothetical protein
36	245	VC0230 of *V. cholerae* O1	AAF93406	60/ 241	Hypothetical protein
37	774	OM2255_20756 of alpha proteobacterium HTCC2255	EAU50128	30/ 413	probable polysaccharide biosynthesis protein
38	373	CD2775 of *Clostridium difficile*	CAJ69663	29/ 366	putative minor teichoic acid biosynthesis protein
39	384	SCH_2098 of *Salmonella enterica* subsp. *enterica* serovar Choleraesuis	AAX66004	25/ 312	Hypothetical protein
40	332	CbeiDRAFT_1436 of *Clostridium beijerincki*	EAP58827	31/ 236	Hypothetical protein
41	287	Unknown			Hypothetical protein
42	368	VAS14_05908 of *Vibrio angustum*	EAS65231	87/ 368	GDP-D-mannose dehydratase
43	319	VV0350 of *Vibrio vulnificus*	BAC93114	81/ 317	GDP-fucose synthetase
44	151	VV0351 of *V. vulnificus*	BAC93115	64/ 150	GDP-mannose mannosylhydrolase
45	480	P3TCK_26602 of *Photobacterium profundum*	EAS44612	68/ 427	Mannose-1-phosphate guanylyltransferase
46	475	VAS14_05883 of *V. angustum*	EAS65226	66/ 472	Phosphomannomutase
47	405	PBPRA2716 of *P. profundum*	CAG21094	57/ 399	Putative mannose-6-phosphate isomer
		VPA1425 of *V. parahaemolyticus*	BAC62768	57/ 399	Putative mannose-6-phosphate isomer
		VPA0967 of *V. parahaemolyticus*	BAC62310	53/ 399	Putative mannose-6-phosphate isomer
48	269	AmsE of *Erwinia amylovora*	CAA54886	46/ 265	Amylovoran biosynthesis glycosyltransferase amsE
		V12G01_11588 of *V. alginolyticus*	EAS75356	41/ 265	WcaA protein
49	335	V12G01 _11593 of *V. alginolyticus*	EAS75357	82/ 335	Exopolysaccharide biosynthesis protein
50	388	VP0236 of *V. parahaemolyticus*	BAC58499	92/ 388	Nucleotide sugar dehydrogenase
51	292	VP0237 of *V. parahaemolyticus*	BAC58500	98/ 292	UTP-glucose-1-phosphate uridylyltransferase
52	400	*Shewanella putrefaciens*	EAY53136	77/ 400	Transposase, mutator type
53	112	VP0238 of *V. parahaemolyticus*	BAC58501	94/ 112	Hypothetical protein (Rjg)
54	158	A1Q_2005 of *Vibrio harveyi*	EDL69109	95/ 142	Hypothetical protein
		VP1395 of *V. parahaemolyticus*	BAC59658	25/ 133	Hypothetical protein
55	943	A1Q_2006 of *V. harveyi*	EDL69124	98/ 940	Hypothetical protein
56	247	A1Q_2007 of *V. harveyi*	EDL69093	100/ 247	Hypothetical protein
57	206	A1Q_2008 of *V. harveyi*	EDL69111	100/ 206	Hypothetical protein
58	346	A1Q_4987 of *V. harveyi*	EDL68386	97/ 346	Transposase, ISSod13
59	219	VP01787 of *V. parahaemolyticus*	BAC60050	219/ 91	Transposase, IS3 transposase
60	94	VP01788 of *V. parahaemolyticus*	BAC60051	90/ 94	Transposase, IS3 INSE transposition protein
61	112	Unknown			Hypothetical protein
62	256	VP0239 of *V. parahaemolyticus*	BAC58502	100/ 256	Triosephosphate isomerase (TIM)

[a)] Deduced from database with blastP program (1).

Table 3. Verification of the presence of O4:K68 antigen genes in representative strains by PCR assays

Stain	O:K Serotype	Results of PCR assays with following primer pairs[a)] :																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Pandemic strains																						
RIMD2210633	O3:K6	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
NIID 242-200	O4:K68	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
KIH 03-57	O4:K68	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DMST 17875	O4:K68	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
AO-24491	O1:K25	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+[b)]	+	+	+	+
AP-18000	O1:K25	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+[b)]	+	+	+	+
AN-16000	O1:KUT	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
Non-pandemic strains																						
ATCC17802	O1:K1	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-K25	O1:K25	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-K3	O2:K3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF-O2	O2:K23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF-K29	O3:K29	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
KIH VP19	O4:K4	+	+	+	+	-	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	+
KE10538	O4:K8	+	+	-	+	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	+
DMST17873	O4:K9	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
REF-K10	O4:K10	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+
KIH 03-60	O4:K11	+	+	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
RIMD 2210587	O4:K12	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
REF-K13	O4:K13	+	+	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
REF-K42	O4:K42	+	+	+	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
REF-K63	O4:K63	+	+	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
REF-K15	O5:K15	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF-O5	O5:K17	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF-K68	O5:K68	-	+	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+[b)]	+	+	+	+
REF-O6	O6:K18	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-K46	O6:K46	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-O7	O7:K19	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF-O8	O8:K20	+	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
KE10541	O8:K41	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-O9	O9:K23	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-K52	O10:K52	+	+	+	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-K71	O10:K71	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
REF-K36	O11:K36	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+
REF-K50	O11:K50	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

[a)] Primer pair numbers and targeting regions are described in Fig.3B.

[b)] About 1.3 kbp smaller than the product size of O4:K68 strains.

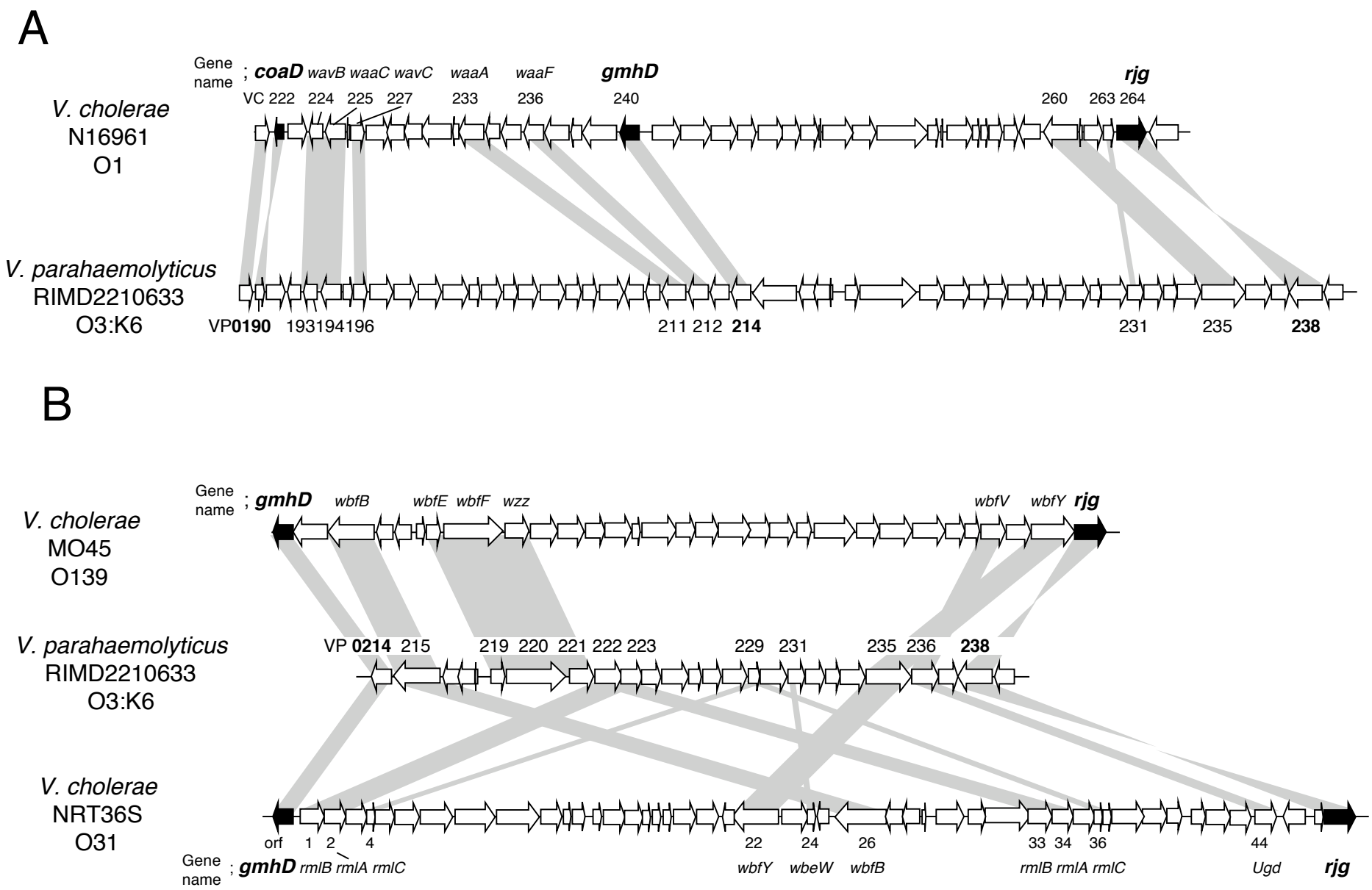
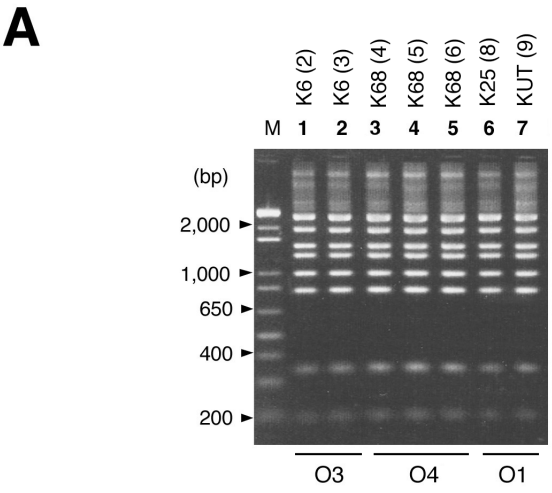
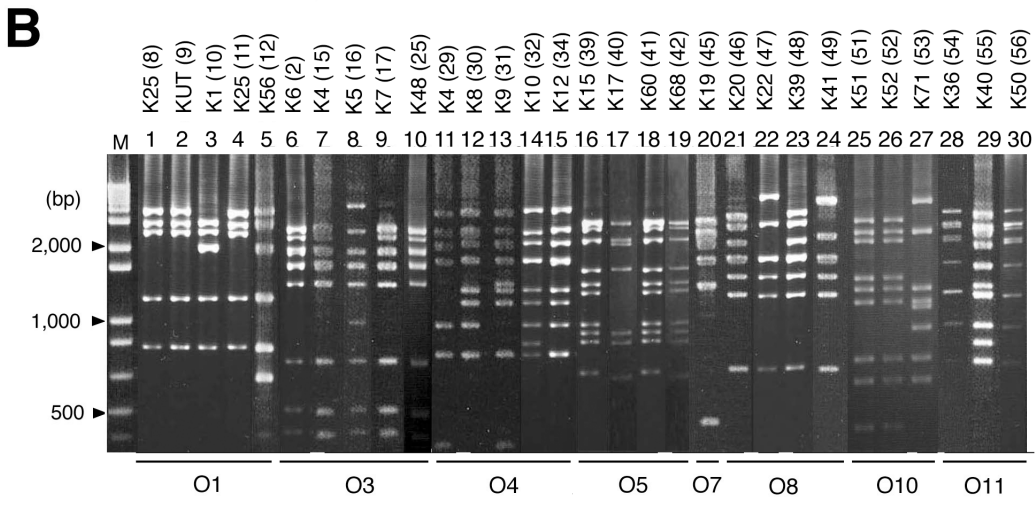

A
Gene name
; coaD wavB waaC wavC waaA waaF gmhD rjg
V. cholerae
N16961
O1
VC 222 224 225 227 233 236 240 260 263 264
V. parahaemolyticus
RIMD2210633
O3:K6
VP0190 193194196 211 212 214 231 235 238
B
Gene name
; gmhD wbfB wbfE wbfF wzz wbfV wbfY rjg
V. cholerae
MO45
O139
V. parahaemolyticus
RIMD2210633
O3:K6
VP 0214 215 219 220 221 222 223 229 231 235 236 238
V. cholerae
NRT36S
O31
orf 1 2 4 22 24 26 33 34 36 44
Gene name
; gmhD rmlB rmlA rmlC wbfY wbeW wbfB rmlB rmlA rmlC Ugd rjg

Fig. 1


A
K6 (2)
K6 (3)
K68 (4)
K68 (5)
K68 (6)
K25 (8)
KUT (9)
M 1 2 3 4 5 6 7
(bp)
2,000
1,000
650
400
200
O3
O4
O1


B
K25 (8) KUT (9) K1 (10) K25 (11) K56 (12) K6 (2) K4 (15) K5 (16) K7 (17) K48 (25) K4 (29) K8 (30) K9 (31) K10 (32) K12 (34) K15 (39) K17 (40) K60 (41) K68 (42) K19 (45) K20 (46) K22 (47) K39 (48) K41 (49) K51 (51) K52 (52) K71 (53) K36 (54) K40 (55) K50 (56)
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
(bp)
2,000
1,000
500
O1
O3
O4
O5
O7
O8
O10
O11

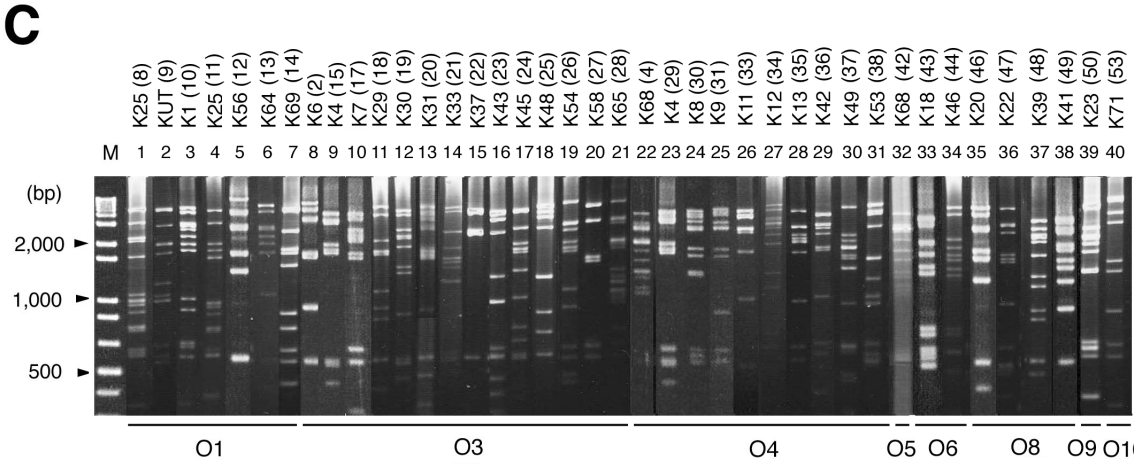

C
K25 (8) KUT (9) K1 (10) K25 (11) K56 (12) K64 (13) K69 (14) K6 (2) K4 (15) K7 (17) K29 (18) K30 (19) K31 (20) K33 (21) K37 (22) K43 (23) K45 (24) K48 (25) K54 (26) K58 (27) K65 (28) K68 (4) K4 (29) K8 (30) K9 (31) K11 (33) K12 (34) K13 (35) K42 (36) K49 (37) K53 (38) K68 (42) K18 (43) K46 (44) K20 (46) K22 (47) K39 (48) K41 (49) K23 (50) K71 (53)
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
(bp)
2,000
1,000
500
O1
O3
O4
O5
O6
O8
O9

Fig.2

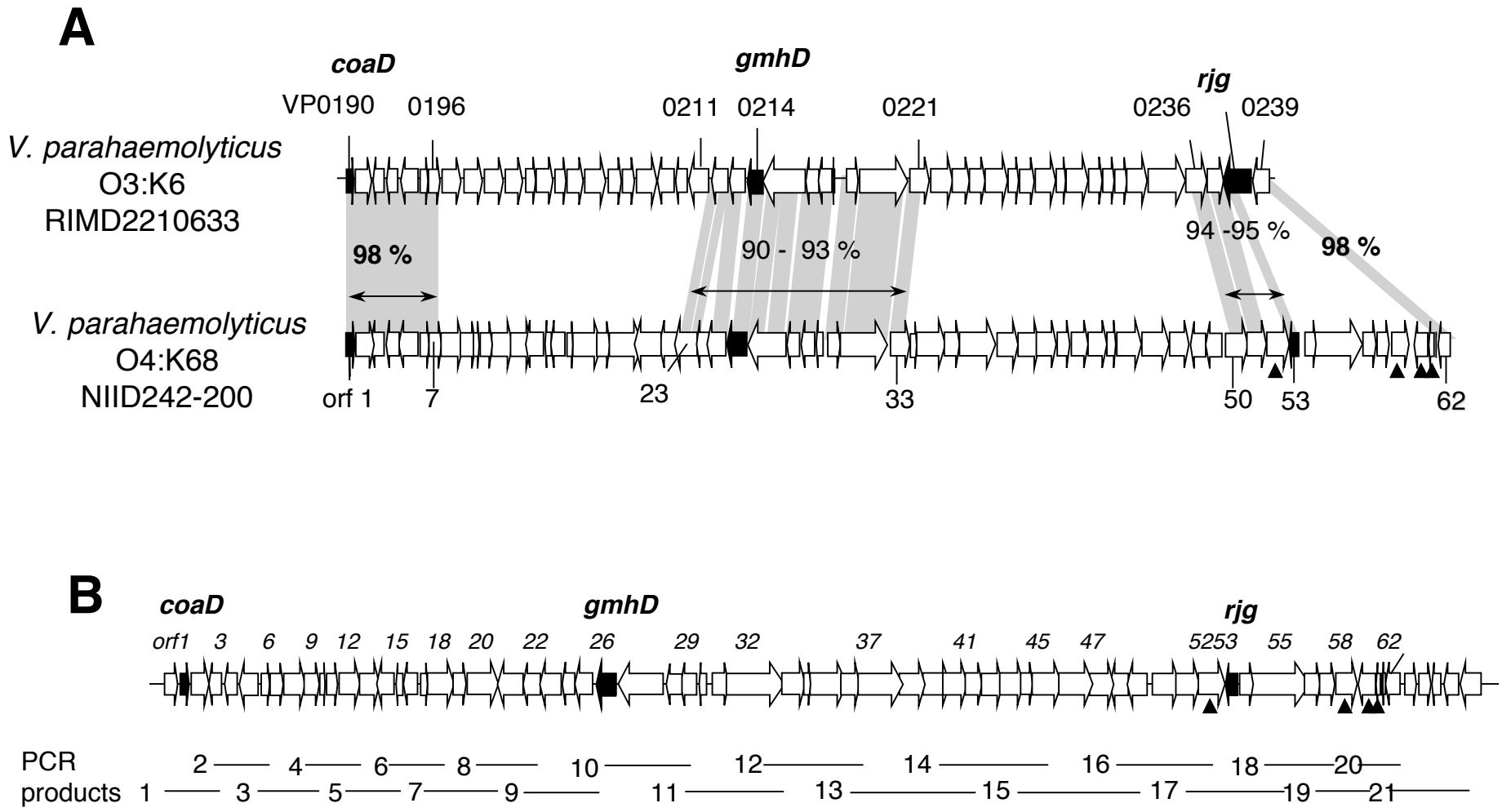

A
coaD
gmhD
rjg
VP0190
0196
0211
0214
0221
0236
0239
V. parahaemolyticus
O3:K6
RIMD2210633
98 %
90 - 93 %
94 -95 %
98 %
V. parahaemolyticus
O4:K68
NIID242-200
orf 1
7
23
33
50
53
62
B
coaD
gmhD
rjg
orf1
3
6
9
12
15
18
20
22
26
29
32
37
41
45
47
52
53
55
58
62
PCR
products
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

Fig.3

A

```
RIMD2210633(pandemic O3:K6)    GTGTGGATTATTGATTTTGATAAGTGCCGCAAACAAGAGCACGGCGACTGGAAAAAGCAA  207038
NIID242-200(pandemic O4:K68)   ************************************************************  600
AP18000(pandemic O1:K25)       ************************************************************  600
AN16000(pandemic O1:KUT)       ************************************************************  600

RIMD2210633(pandemic O3:K6)    AACTTAGAAAGGCTTCTTCGCTCATTCAAAAAAGAACTGCTCAAACGCCAAATACATTGG  207098
NIID242-200(pandemic O4:K68)   ***C*********************************************CAT****  660
AP18000(pandemic O1:K25)       ************************************************CAT****  660
AN16000(pandemic O1:KUT)       ***C*********************************************CAT****  660

RIMD2210633(pandemic O3:K6)    AAGGAACGTGACTTCGCTGTGCTTACTGAAGCGTTATCGTGCTTAGACATAAAATAATTC  207158
NIID242-200(pandemic O4:K68)   C*A*CGTCC**T**T*A*CATG*ACAAA**AGT*ATA*TGAAAATACGTA**GTG*GC*T  720
AP18000(pandemic O1:K25)       C*A*CGTCC**T**T*A*CATG*ACAAA**AGT*ATA*TGAAAATA*GTA**GTG*GC*T  720
AN16000(pandemic O1:KUT)       C*A*CGTCC**T**T*A*CATG*ACAAA**AGTCATA*TGAAAATA*GTA**GTG*GC*T  720
```

B

```
RIMD2210633(pandemic O3:K6)    TCTGGTTGTGTTATTGCTTACAAATTTAAAGGTCAGCGTTTTGACTGCGGCAGTGTAGAA  251671
AN16000(pandemic O1:KUT)       ***********T*G**A**************************T*****T**C******   60

RIMD2210633(pandemic O3:K6)    GGTTACATCGAGGCGACAAACTACTGCTACGAGAACCTATACAAAAAAGCGTCTAGTGCT  251731
AN16000(pandemic O1:KUT)       **C*****T**A**A*****T*****************T**T**G****AC**CT****G  120

RIMD2210633(pandemic O3:K6)    GAGCTTGCTAAAACACCAACCGTGAAGGTTTCTGAACCAGCATA-GTTGTTCAGGCTCGT  251790
AN16000(pandemic O1:KUT)       *****C**G****AG**********AA**G*C**G**AT*G**A****----------  169

RIMD2210633(pandemic O3:K6)    TTAGGCTCCTCGGAATGAC-------------------------------------------  253493
NIID242-200(pandemic O4:K68)   ******************GGGATTCTCATAGAATATGTCTTTTGTTTGGTTGTGTTGGA   60
AP18000(pandemic O1:K25)       ******************GGGATTCTCATAGAATATGTCTTTTGTTTGGTTGTGTTGGA   60
AN16000(pandemic O1:KUT)       ------------------------------------------------------------

RIMD2210633(pandemic O3:K6)    ----------------------------------------------------AAAACTC  253500
NIID242-200(pandemic O4:K68)   TATGCAGAGGGATAAGGTTTCTAATGTCACATAAGCTTTTGGGAGTGACAGAATTC*ACG  120
AP18000(pandemic O1:K25)       TATGCAGAGGGATAAGGTTTCTAATGTCACATAAGCTTTTGGGAGTGACAGAATTC*ACG  120
AN16000(pandemic O1:KUT)       -------------------------------------GAACTTTCTATTAGGT*C*C  189

RIMD2210633(pandemic O3:K6)    ATTAGATCGTAGGCAAACAAAAAGGGATGCATGATGCATCCCTTTGTTATTTCTGCT-TA  253559
NIID242-200(pandemic O4:K68)   GG*C*T**TC***************A****G**********************C***-**  179
AP18000(pandemic O1:K25)       GG*C*T**TC***************A****G**********************C***-**  179
AN16000(pandemic O1:KUT)       CGAT*GCACA**A*******************************************A**C*C  249

RIMD2210633(pandemic O3:K6)    AT-AAAAGCGAAATTAAGCTTTAGCTGCTGCCGCTGCTTTAGCGATTGCTGCGAAGCTCT  253618
NIID242-200(pandemic O4:K68)   **A*********************************************************  238
AP18000(pandemic O1:K25)       **A*********************************************************  238
AN16000(pandemic O1:KUT)       *GTT***A***G**********T***TT*********************************  309
```

C

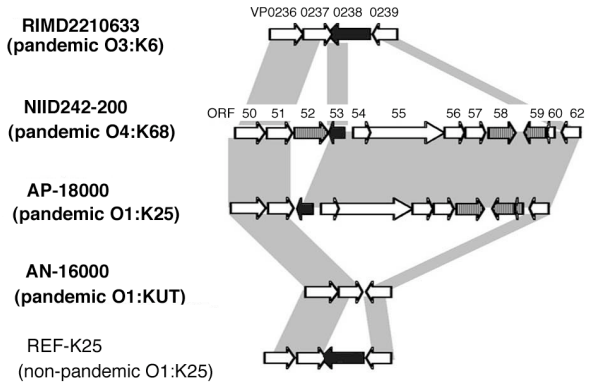


Fig. 4