Assessment of Evolution of Pandemic *Vibrio parahaemolyticus* by Multilocus Sequence Typing

Nandini Roy Chowdhury,1,2 O. Colin Stine,2* J. Glenn Morris,2 and G. B. Nair1,3

**National Institute of Cholera and Enteric Diseases, Calcutta, India1; International Centre for Diarrhoeal Diseases, Bangladesh, Dhaka, Bangladesh2; and University of Maryland School of Medicine, Baltimore, Maryland2**

Received 8 April 2003/Returned for modification 25 August 2003/Accepted 21 November 2003

The genetic relatedness of 81 isolates of *Vibrio parahaemolyticus* was assessed by multilocus sequence typing. The strain with serotype O3:K6 emerged as a pandemic pathogen in 1996, with subsequent expansion to include strains having serotypes O1:KUT, O4:K68, and O1:K25. Sequence data from gyrB, recA, dnaE, and gnd revealed that 16 distinct serogroups isolated prior to the pandemic were highly variable and only isolates of serogroup O3:K6 shared two alleles with the pandemic strains. The pandemic strains regardless of serotype were clonal, with 51 of 54 isolates having the identical allelic profile (AP). Serotype alone did not adequately define a pandemic strain: among O1:KUT strains tested, seven strains with the identical pandemic AP carried previously described pandemic markers, while five nonpandemic strains had five distinct APs. Our sequence data provide strong molecular support for the clonal origin of pandemic *V. parahaemolyticus* O3:K6 and suggest that strains within such a clonal group may acquire previously identified serotypes.

*Vibrio parahaemolyticus* is a gram-negative bacterium that is often associated with self-limiting gastroenteritis following consumption of contaminated seafood or exposure to warm seawater (7). The pathogenicity of this organism has been correlated to the two well-characterized hemolysins—thermostable direct hemolysin (TDH) and the TDH-related hemolysin (26, 27). There was not a clear association between *V. parahaemolyticus*-mediated infection and serotype until 1996, when O3:K6 strains emerged as a major cause of illness. The O3:K6 strain appears to be a pandemic clone. It has been identified as a dominant serotype from clinical cases of diarrheaa reported from various Southeast Asian countries, including India, Japan, Thailand, Bangladesh, Taiwan, and Vietnam, as well as from the United States (1–7, 16, 22–24).

Molecular analyses of isolates of the recent O3:K6 strains from different countries revealed that these strains had probably evolved from a common ancestor. All the post-1995 O3:K6 strains were genetically identical by arbitrarily primed PCR, ribotyping, and pulsed-field gel electrophoresis (PFGE), but they were significantly different from the genetically variable pre-1995 O3:K6 strains (6, 16, 20). In addition, the newly emerged O3:K6 clone has diversified into three other serotypes—O1:KUT, O4:K68, and O1:K25—since its initial isolation. These strains are presumed to have emerged from the O3:K6 clone because they are genetically similar, based on group-specific (GS)-PCR for toxRS and ORF8 PCR for detection of the f237 filamentous phage and the molecular techniques of arbitrarily primed PCR, PFGE, and ribotyping (6, 16, 18).

An additional molecular technique, multilocus sequence typing (MLST), has not been applied to investigate the emergence and spread of this first recorded pandemic of *V. parahaemolyticus*. MLST has provided increased sensitivity compared to the other molecular typing methods for *Enterococcus* (17), *Salmonella* (14), and *Vibrio cholerae* (15). Additionally, MLST has provided insight into the role of recombination during bacterial evolution (10, 21).

Eighty-one isolates of *V. parahaemolyticus* collected between 1983 and 2000 were chosen. TCBS (thiosulfate, citrate, bile salts, and sucrose) agar medium followed by presumptive identification in Kapers multitest medium (1, 2, 3) was used for selection of *V. parahaemolyticus* strains. The strains were confirmed by the species-specific toxR assay (13). The commercially available *V. parahaemolyticus* antisera kit (Toshiba Kagaku Kogyo Co., Ltd., Tokyo, Japan) was employed for serological typing. Seven of the 81 selected isolates were O3:K6 from The Maldives, Singapore, and Thailand, collected between 1993 and 1995. Sixteen nonpandemic strains belonging to 11 different serotypes (O1:K38, O1:K1, O1:K56, O1:KUT, O2:K3, O3:K29, O4:K63, O4:K8, O5:KUT, O8:K21, and O8:K41) were isolated in Calcutta during 1994 and 1995, prior to the start of the pandemic. Fifty-eight isolates with pandemic serotypes collected from 1996 to 2000 were selected for sequence typing. Twenty-five had the O3:K6 antigenic combination, 14 had O4:K68, 11 had O1:KUT, and 8 had O1:K25. Thirty-four of these 58 were isolated from patients in Calcutta. There, the pandemic isolates constituted between 50 and 63% of the total number of *V. parahaemolyticus* isolates collected. Among the pandemic isolates, the non-O3:K6 isolates constituted 3% in 1996, 20% in 1997, 52% in 1998, 23% in 1999, and 73% in 2000. Representatives of each serotype were selected from each year that isolates were available. Serotype O3:K6 was represented by isolates from all years and O1:KUT and O4:K68 were represented in 1997, 1998, and 1999, while O1:K25 was represented from the only year it was identified, 2000. The other 24 of 58 isolates were from Japan, Korea, Nepal, Singapore, Thailand, the United States, and Vietnam and were provided by M. Nishibuchi. Of these, 2 were serotype O1:KUT, 13 were O3:K6, and 9 were O4:K68. Nonpandemic strains were...
The GS-PCR assay for toxRS (16) and the ORF8 PCR assay, which detects the presence of the f237 phage (18). By these assays, all of the strains isolated prior to 1996 and 4 strains of O1:KUT isolated between 1996 and 2000 were considered to be nonpandemic. A complete table of the isolates and alleles is available at http://medschool.umd.edu/departments/Epidemiology/vpta1b1.doc.

DNA was prepared from 5 μl of an overnight bacterial culture using Prepman Ultra (Applied Biosystems [ABI]). PCR was performed in a standard reaction using previously described PCR primers for gyrB, recA, dnaE, and gnd (11). The presence of amplified products was confirmed by electrophoresis in 1% agarose gel.

Sequencing was done using the ABI Prism Big Dye Terminator kit (ABI). The resulting fragments were separated and identified in an ABI Prism 3700 DNA analyzer. The trace files were read using Phred (8, 9) and Phrap (available with permission from P. Greene at http://www.washington.edu). Low-quality sequences at the ends were trimmed. The contigs from each individual isolate were aligned using Clustal X (12). The aligned sequences for each locus were analyzed by using the program Data Analysis in Molecular Biology and Evolution (25) and assigned into various allelic types (indicated numerically).

V. parahaemolyticus is clearly genetically diverse, with high variability noted in all four loci examined. There were 13 alleles for gyrB, 12 for recA, and 14 each for dnaE and gnd. These alleles were very closely related, since there were a total of 23, 29, 21, and 20 polymorphic nucleotides for gyrB, recA, dnaE, and gnd, respectively. The combination of the specific allele at each locus defined the allelic profile (AP) for each strain. There were 24 different APs among the 81 isolates.

The distribution of the APs among the strains was nonrandom. All of the O3:K6 isolates from 1983 to 1993 collected from diverse countries around the Indian Ocean were related. Four had the AP 1,2,2,1, and the other three had a single locus with an alternative allele, e.g., AP 1,0,2,1 where the recA allele is number 8 instead of allele number 2. These alleles differed by three nucleotides. Similarly, the APs of isolates collected in Calcutta in 1994 and 1995 with same serotype were related. The APs were identical, e.g., O5:KUT and O8:K41, or differed by one nucleotide. The AP of O8:K41 was 1,0,0,1, and was very similar to another AP (1,0,0,0). The AP of the four O1:KUT isolates was 1,1,1,1. These APs differ by three nucleotides.

The pandemic strain of V. parahaemolyticus arose from an O3:K6 progenitor and subsequently evolved by changing serotypes. Of the 12 serotypes examined, only those from the O3:K6 serotype, collected prior to the start of the pandemic, shared alleles at two of the four sequenced loci with the pandemic clone. A total of 51 of the 54 pandemic isolates were genetically identical, and the other 3 were very closely related and were likely clonal. This observed clonality was in agreement with arbitrarily primed PCR, ribotyping, and PFGE data obtained for pandemic strains that have been reported previously (5, 6).

Despite the genetic relatedness at the sequenced loci, the pandemic isolates were exceptional because they have four distinct serotypes (O3:K6, O4:K68, O1:KUT, and O1:K25). The acquisition of additional serotypes of the pandemic strain may be a selected response to host immunological pressure. Our results confirm the finding that multiple serotypes occur in a single genetic lineage (6, 16, 18).

Strains of a given serotype may or may not be genetically related. Two strains with the O5:KUT serotype were identical at all four loci, despite being collected in different years. The same was true for the two O8:K41 strains, and the two O8:K21 strains differed at a single locus. In contrast, the strains with serotype O1:KUT had one of six distinct APs—there were no shared alleles among the six, and all 24 possible alleles (six at each locus) were different. Thus, there are at least six distinct lineages within this single serotype, perhaps implying that the genes encoding the serotype have moved among lineages. The one collected before the pandemic had the AP 10,3,9,6. One of the APs (1,1,1,1) is the pandemic type. The other four O1:KUT isolates (VP105, VP166, VP230, and VP333), although they were collected during the pandemic, were distinct from the pandemic strains. They were negative for GS-PCR and had a unique AP. These data reinforce the suggestion that the serotypes are moving among lineages.

The pandemic isolates, as defined by the toxRS GS-PCR and the presence of ORF8 from f237, all have the AP 1,1,1,1. No other strains have three number 1 alleles in their AP. Although it is possible that, given a larger sample, variants will be found at the other loci among the pandemic strains, it is possible that they will remain invariant, as was seen among serotype A group III Neisseria (28) and Escherichia coli O157 (19). The very high correlation between the genetic type and the toxRS GS-PCR confirmed that toxRS and ORF8 of f237 are good markers for the pandemic strains of V. parahaemolyticus (18).

The genetic uniformity of the pandemic strains of V. parahaemolyticus is in stark contrast to the genetic diversity seen among strains of O139 V. cholerae (11). As previously reported, we found that >50% of V. cholerae O139 strains collected from a single hospital across an 8-year time span had novel APs. It is possible that the increased potential for genetic diversity in V. cholerae contributes in some way to its ability to cause recurrent pandemic disease in human populations; further studies of these species in other epidemic situations will be necessary to test this hypothesis.
Nucleotide sequence accession number. The allelic sequences have been assigned GenBank accession numbers AY423489 to AY423541.

This work was supported by an American Society of Microbiology Travel Award (N.R.C.) and by U.S. Public Health Service grant RO1 GM60791 (J.G.M.).

REFERENCES