

# Pandemic Spread of Cholera: Genetic Diversity and Relationships within the Seventh Pandemic Clone of *Vibrio cholerae* Determined by Amplified Fragment Length Polymorphism

Ruiting Lan and Peter R. Reeves\*

Department of Microbiology, University of Sydney, NSW 2006, Australia

Received 11 May 2001/Returned for modification 9 August 2001/Accepted 15 October 2001

**The seventh cholera pandemic started in 1961 and continues today. A collection of 45 seventh pandemic isolates of *V. cholerae* sampled over a 33-year period were analyzed by amplified fragment length polymorphism (AFLP) fingerprinting. All but four pairs and one set of three isolates were distinguished. AFLP revealed far more variation than ribotyping, which was until now the most useful method of revealing variation within the pandemic clone. Unfortunately, the ribotype variation observed is mainly due to recombination between the multiple copies of the *rrn* genes (R. Lan and P. R. Reeves, *Microbiology* 144:1213–1221, 1998), which makes changes susceptible to repeat occurrences and reversion. This AFLP study shows that particularly for the common ribotypes G and H, such events have indeed occurred. AFLP grouped most of the 45 isolates into two clusters. Cluster I consists mainly of strains from the 1960s and 1970s, while cluster II contains mainly strains from the 1980s and 1990s, revealing a temporal pattern of change in the clone. This is best seen in the relationships of the strains from Africa, which correlate with the epidemiology of epidemics on that continent. The data confirm independent introductions to Africa during the 1970s outbreak and reveal several other African introductions. In the 1991 cholera upsurge, isolates from the Southern and Eastern African epidemic focus are markedly different from those from the West African epidemic focus. An isolate from 1987 in Algeria was identical to the West epidemic isolates, suggesting that the strain was present in Africa at least 3 years before causing large outbreaks. These observations have major implications for our understanding of cholera epidemiology.**

Cholera is still a very significant disease in developing countries with an estimated incidence of more than five million cases per year (41). Seven pandemics of cholera have been recorded since 1817, the ongoing seventh pandemic having started in 1961 in Asia. There have been two major upsurges of this pandemic: one in the 1970s spread to Africa, and the other in 1991 took cholera to South America. Both continents had been free from pandemic cholera for over a century.

In this paper, we look at the relationships among the seventh pandemic isolates of *Vibrio cholerae*, the etiological agent for cholera, from different countries and dates. Multilocus enzyme electrophoresis (MLEE) is not useful for this purpose, because all except the South American variant belong to a single electrophoretic type (31, 46). Sequencing of housekeeping genes, which was used to resolve relationships between pandemic clones and to environmental strains, is also not suitable for studying variation within the seventh pandemic clone, because no variation was found in any of several housekeeping genes sequenced (4, 20).

Other molecular methods, including pulse-field gel electrophoresis (PFGE) and ribotyping, have been employed to study the diversity and epidemiology of the seventh pandemic (6, 35, 46). Ribotyping revealed much variation within the clone and provided some interesting insight into its evolution (19). How-

ever, subsequent analysis of the molecular basis of ribotype variation revealed that it was generated by recombination between *rrn* operons (27). The localized nature of such recombination makes ribotyping less valuable for evolutionary studies, because any such change is readily reversed, making the true relationships of isolates difficult to reveal. PFGE is expensive and technically difficult to standardize. Furthermore, the variation detected by PFGE is often a result of genome instability (genome rearrangement, etc.) (29, 42). Dissimilarity in PFGE patterns is not a good measure of evolutionary divergence. An alternative method is necessary.

Amplified fragment length polymorphism (AFLP) is a novel PCR-based DNA fingerprinting method (44) revealing variation around the whole genome by selectively amplifying a subset of restriction fragments for comparison. The AFLP procedure involves digestion of DNA with *EcoRI* and *MseI* or another combination of enzymes with 6- and 4-base recognition sites. Double digestion with *EcoRI* and *MseI* will produce around 2,000 *EcoRI-MseI* fragments, 16,000 *MseI-MseI* fragments, and very few *EcoRI-EcoRI* fragments for a genome of 4,000 kb. This is followed by ligation of adapters to both ends of the fragments and amplification with primers designed so as to amplify *EcoRI-MseI* fragments. This preamplification step is followed by a second “selective” amplification with primers that include 1 or 2 bases in addition to the segment based on the restriction site. Each additional base reduces the number of effective substrates fourfold. With one selective base on each primer, the number of fragments is reduced 16-fold,

\* Corresponding author. Mailing address: Department of Microbiology, University of Sydney, NSW 2006, Australia. Phone: 61-2 9351 6045. Fax: 61-2 9351 4571. E-mail: reeves@angis.usyd.edu.au.

which can be resolved to 1 base by using polyacrylamide sequencing gels. The ability to resolve to 1 bp enables precise comparison of bands across samples.

Many studies have shown the value of AFLP in typing of microorganisms (1, 2, 11, 13, 14, 21, 23–25, 30, 34, 37). Recently AFLP was applied to the study of genetic diversity of clinical and environmental isolates of *V. cholerae* (15, 16), but again there was no variation observed within the seventh pandemic. In this study, we used AFLP with a range of primer pairs and found sufficient variation to draw conclusions about the genetic relationships of the isolates, revealing, for example, independent introductions of epidemic cholera into Africa in both the 1970s and the 1990s. The AFLP data in this study give new insights into the spread of cholera and open up the possibility of greatly enhancing our understanding of its epidemiology.

#### MATERIALS AND METHODS

**Strains.** The strains used (Table 1) were a collection of isolates from 1961 to 1993 from many countries. Chromosomal DNA was prepared as described previously (27).

**AFLP.** The AFLP technique (44) was performed essentially as described by Lan and Reeves (28), which allowed the use of hot start PCR. One hundred nanograms of DNA was used for digestion and ligation of adapters, which was done simultaneously. The restriction enzyme combination used was *EcoRI* and *MseI*. Adapters and primers were designed as described in reference 28 (Table 2). For AFLP, the primers can have at the 3' end of the base of the restriction site one or more additional (selective) bases. Nonselective primers (with no additional selective base) amplify all *EcoRI-MseI* fragments, whereas each additional (selective) base reduces the number of fragments amplified to one-fourth.

The ligation mix was denatured and diluted 1:10, and 1  $\mu$ l was used for preamplification with nonselective primers. Preamplification was done with PCR mix containing 0.2  $\mu$ M each deoxynucleoside triphosphate (dNTP), bovine serum albumin (4  $\mu$ g/ $\mu$ l), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2 mM MgCl<sub>2</sub>, and 0.5 U of AmpliTaq Gold (PE Applied Biosystems, Melbourne, Australia) plus 0.3  $\mu$ M preamplification primers. The cycling profile was 20 cycles of denaturation at 94°C for 15 s, except for 10 min for the first cycle; annealing at 56°C for 30 s; and extension at 72°C for 1 min. For final (selective) amplification, 1  $\mu$ l of diluted PCR product (1:10) from preamplification was mixed with 19  $\mu$ l of PCR mix as described above for preamplification, but containing 0.3  $\mu$ M unlabeled selective primer and 0.1  $\mu$ M <sup>33</sup>P-labeled selective primer. One-base selection was used for each of the *EcoRI* and *MseI* primers. The basic cycling parameters were denaturation at 94°C for 15 s, except for 10 min for the first cycle, annealing at the temperatures specified below for 30 s, and extension at 72°C for 1 min. The first 10 cycles were touchdown cycles from 66 to 56°C, decreasing by 1°C/cycle, followed by 20 cycles at 56°C. AFLP products were run on standard 6% polyacrylamide sequencing gels and visualized by exposure to Kodak BioMax-MR film overnight.

**Analysis of fingerprinting patterns.** Autoradiographs were analyzed manually and scored for the presence or absence of a band. Bands with a weak signal or blurred appearance were excluded, as well as bands from the two ends of the gel. Samples for each primer pair were run on the same gel; manual scoring of the presence or absence of polymorphic bands is fast and accurate. It was considerably more difficult to achieve current scoring by using commercial software. The band patterns from the manual scoring were converted to distance between isolates according to the Dice coefficient of similarity (33). The Dice coefficient was used to construct phylogenetic trees by using programs in the PHYLIP package (version 3.5, written by J. Felsenstein, Department of Genetics, University of Washington, Seattle). Both the unweighted pair group method with arithmetic mean (UPGMA) (38) and the neighbor-joining method (36) were used to obtain an indication of validity of the relationships by alternative methods. UPGMA has been widely used for clustering based on genetic distance. The major difference between the two algorithms is that UPGMA assumes equal rates of evolution along all branches. The Simpson index of diversity (*D*) was calculated by the formula  $D = 1 - \{\sum [n_j(n_j - 1)]\} / [N(N - 1)]$ , where *N* is the number of strains tested and *n<sub>j</sub>* is the number of strains belonging to the *j*th type (12, 39). An in-house program, MLEECOMP (available upon request), was used to calculate the Dice coefficient and Simpson index of diversity. Bootstrap sam-

pling (9) was done with MLEECOMP. In this analysis, the 966 fragments were sampled repeatedly to give a new set of the same size (966 fragments) in which by chance some of the original fragments will be omitted and others will be present more than once. This derived set of fragments was then used to generate a new tree that was compared with the original tree at each node. The above resampling and tree generation process was repeated 1,000 times. Each node divides the strains into two groups, and the proportion of the 1,000 samples that gives the same division as in the original set is the bootstrap value for that node. A value of 100% means that each of the 1,000 resampled data set gave the same distribution of strains at that node.

#### RESULTS

**AFLP analysis.** All 16 *EcoRI-MseI* +1/+1 primer combinations were used for the 45 isolates, and a total of 996 fragments were scored with an average of 62 fragments per primer pair (Table 3). One hundred twenty-two fragments (13.3%) are polymorphic. Fifty are phylogenetically informative (i.e., both presence and absence are observed in two or more isolates) and hence affect the branching pattern of a phylogenetic tree, but more than half (72 bands) are singularly polymorphic, being either present or absent in one isolate only. Primer pair 14 detected the largest number of polymorphic bands (14 bands), while primer pair 8 detected the least (2 bands). Given that the average *EcoRI-MseI* AFLP fragment is 256 bp long (based on analysis of genome data of strain N16961 [10], but coinciding with the expected frequency of *MseI* fragments), we have surveyed approximately 6.3% (254,976 nucleotides [996 total fragments  $\times$  256 bp]) of the genome for *EcoRI-MseI* internal (length) variation. Also, given that there are 9 nucleotides associated with each *EcoRI-MseI* fragment (3 bases for *EcoRI* site, because most *EcoRI* sites are present in two fragments; 4 for *MseI* site, because each site is rarely present in two fragments; and 2 for +1 selection at each), we have screened approximately 0.22% (8,964 [996  $\times$  9] nucleotides) for point mutations.

**Genetic relationships.** We included a sixth pandemic isolate, a U.S. Gulf isolate, and two pre-seventh pandemic isolates in this study (data not shown). The UPGMA tree (Fig. 1) shows that relationships between the sixth pandemic, U.S. Gulf, and seventh pandemic isolates are similar to those revealed by sequence variation of several housekeeping genes (4, 20). However, sequence data did not reveal any variation within or between the seventh pandemic and prepandemic isolates, whereas in this study, the two pre-seventh pandemic isolates were clearly placed outside of, but closely related to, the seventh pandemic isolates. The seventh pandemic clone seems not to be a direct descendant of the two prepandemic strains, suggesting that there were several related strains, one of which gave rise to the seventh pandemic by rapid expansion.

The UPGMA tree for the seventh pandemic isolates is shown in Fig. 2. The two pre-seventh pandemic isolates were used as an outgroup. There are two main clusters. Cluster I strains are mainly from the 1960s and 1970s, while cluster II strains are from the 1980s and 1990s. Within cluster I, there are three subclusters, marked Ia, Ib, and Ic in Fig 2. Ia contains primarily strains from Asia in the 1960s, Ib contains primarily isolates from Africa in the 1970s, and Ic contains those from Asia (4) and Africa (1) from the 1980s. Cluster II isolates are from Asia from the late 1970s to 1990s and Africa in the 1990s. The neighbor-joining tree yielded comparable results for the major clusters (data not shown).

TABLE 1. Strains used in this study

Strain	Yr isolated	Country of isolation	Source <sup>a</sup>	Original laboratory identification
6th Pandemic M642	1943	India	NCTC	6585
7th Pandemic M793	1961	Indonesia	University of Maryland	E9120
M803	1961	Hong Kong	Institut Pasteur	HK1
M804	1962	India	Institut Pasteur	930030
M805	1963	Cambodia	Institut Pasteur	930059
M806	1964	India	Institut Pasteur	CRC1106
M807	1966	Vietnam	Institut Pasteur	601
M686	1968	Thailand	AFRIMS	SP-EV-29-1
M808	1969	Vietnam	Institut Pasteur	1536
M647	1970	Bangladesh	CCUG	13119
M809	1970	Sierra Leone	Institut Pasteur	930037
M810	1970	Ethiopia	Institut Pasteur	930038
M811	1971	Burma	Institut Pasteur	930029
M812	1971	Chad	Institut Pasteur	930046
M813	1972	Senegal	Institut Pasteur	9292
M814	1972	Morocco	Institut Pasteur	113
M815	1973	Philippines	Institut Pasteur	430035
M816	1974	Senegal	Institut Pasteur	B998C
M817	1974	Chad	Institut Pasteur	99
M818	1975	Comoros Islands	Institut Pasteur	102
M819	1975	Germany <sup>b</sup>	Institut Pasteur	232
M650	1976	India	National Institute of Cholera	762/76
M795	1976	Bangladesh	University of Maryland	30167
M820	1978	Malaysia	Institut Pasteur	EB 251/1MR
M714	1979	Thailand	AFRIMS	96A/CO
M646	1979	Bangladesh	CCUG	9193
M652	1981	India	National Institute of Cholera	1200/81
M723	1982	Thailand	AFRIMS	WR-32
M821	1982	France <sup>b</sup>	Institut Pasteur	Assous M
M822	1983	Vietnam	Institut Pasteur	359
M823	1984	Algeria	Institut Pasteur	Marquez
M740	1985	Thailand	AFRIMS	D-145
M797	1986	Hong Kong	University of Hong Kong	V31
M824	1987	Algeria	Institut Pasteur	Mekki
M825	1988	Zaire	Institut Pasteur	Zaire1
M764	1989	Thailand	AFRIMS	FX-41-3
M799	1989	Hong Kong	University of Hong Kong	In21
M826	1990	Malawi	Institut Pasteur	Bakala Malenge
M827	1990	Guinea	Institut Pasteur	Guinea1
M791	1991	Thailand	AFRIMS	CX-043-0
M654	1991	India	National Institute of Cholera	413/91
M828	1991	Morocco	Institut Pasteur	Akretche
M663	1992	Indonesia (Bali)	IMVS	2100
M829	1992	Malawi	Institut Pasteur	F. Francisco
M662	1993	Indonesia (Bali)	State Health Laboratory, Perth	7340
Pre-7th pandemic M802	1937	Indonesia (Sulawesi)	Institut Pasteur	66-2
M640	1954	Egypt	NCTC	9420
U.S. Gulf M796	1978	Louisiana	University of Maryland	4808
Latin America M830	1993	French Guiana	Institut Pasteur	Modesto

<sup>a</sup> NCTC, National Collection of Type Cultures, London, England; CCUG, Culture Collection of the University of Goteborg, Goteborg, Sweden; Institut Pasteur, Paris, France; University of Maryland, Center for Vaccine Development, University of Maryland, Baltimore; AFRIMS, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand; National Institute of Cholera, National Institute of Cholera and Enteric Diseases, Calcutta, India; University of Hong Kong, Hong Kong, China; IMVS, Institute of Medical and Veterinary Sciences, Adelaide, Australia; State Health Laboratory, Perth, Australia; ICDDR, International Center for Diarrhoeal Disease Research, Dhaka, Bangladesh.

<sup>b</sup> Probably isolated from travellers from Asia or Africa.

M793 was isolated in 1961 in Indonesia, where the pandemic started, and was placed outermost of the seventh pandemic isolates. M793 lacks VPI (18). However, deletion of the VPI region of DNA, which would appear as the absence of AFLP

bands in M793 only, should not affect placement in the tree but only branch length.

How robust are the branches? The robustness of the branching order is commonly assessed by bootstrap analysis. A value

TABLE 2. Adapters and primers used

Oligonucleotide name	Sequence (5'→3')	Selective base	Enzyme end
<b>Adapters<sup>a</sup></b>			
924	ACTGCGTACC	N/A	<i>EcoRI</i>
925	AATTGGTACGCACTCAGTGAGGTTTCATTACCATCC	N/A	<i>EcoRI</i>
930	TACTCAGGACTC	N/A	<i>MseI</i>
931	CCTGATTGCTACAACCTGAACGATGAGTCCTGAG	N/A	<i>MseI</i>
<b>Preamplification primers</b>			
926	GGATGGTAATGAACCTCACT	None	<i>EcoRI</i>
932	CCTGATTGCTACAACCTGAAC	None	<i>MseI</i>
<b>Selective primers</b>			
927	GACTGCGTACCAATTCA	A	<i>EcoRI</i>
1269	GACTGCGTACCAATTCT	T	<i>EcoRI</i>
1270	GACTGCGTACCAATTCG	G	<i>EcoRI</i>
1271	GACTGCGTACCAATTCC	C	<i>EcoRI</i>
933	GATGAGTCCTGAGTAAC	C	<i>MseI</i>
1272	GATGAGTCCTGAGTAAA	A	<i>MseI</i>
1273	GATGAGTCCTGAGTAAT	T	<i>MseI</i>
1274	GATGAGTCCTGAGTAAG	G	<i>MseI</i>

<sup>a</sup> See reference 28 for details on adapter preparations.

of 50% or more is usually indicative of support for a branch. However, only two branches met this standard: one for the placement of prepandemic isolates and the other for placement of the earliest seventh pandemic isolate, M793. However, bootstrap analysis is a very stringent test. In a data set in which there is a large number of strains, a small proportion of atypical strains can lead to greatly reduced bootstrap values. This will occur if, for example, some strains are frequently partitioned differently from the original during resampling at a given node due to recombination, or parallel or reverse mutation, even if the majority of strains give a consistent distribu-

tion. The good correlation of major branches with locality indicates that the groupings are phylogenetically meaningful. We therefore looked at band distribution (Table 4) for evidence of support of major branches with low bootstrap values. For the division of clusters I and II, there is good but not absolute support by band patterns 2, 3, 7, 8, 12, 16, and 20. Within cluster I, patterns 10, 11, and 18 distinguish Ia from Ib and Ic, and patterns 22 and 23 distinguish Ib and Ic. Patterns 13, 14, and 15 also gave support to cluster Ic.

**DISCUSSION**

**Resolving power of AFLP.** The discriminatory power of a typing technique is measured by the Simpson index of diversity (*D*) (12, 39). *D* should be greater than 0.95 for a typing system to be used as single typing method with the generally accepted probability of 5% of type I errors (39). The *D* values for each primer pair were calculated (Table 3). The lowest is 0.189, while the highest is 0.837. This shows that the power of AFLP fingerprinting will depend on the choice of primer pairs. It is highly discriminatory for a single primer pair to reach 0.837. With the combination of two primer pairs, the highest *D* value (primer pairs 11 and 6) reaches 0.944. When all primer pairs are considered, the *D* value is 0.993. AFLP has outstanding resolving power for the very closely related seventh pandemic isolates.

TABLE 3. Numbers and types of polymorphisms observed in the 45 seventh pandemic isolates

Pair no.	Primers <sup>a</sup>	No. of fragments				<i>D</i>
		Total <sup>b</sup>	Constant <sup>c</sup>	Polymorphic	Informative	
1	A/C	74	64	10	4	0.680
2	T/C	68	61	7	4	0.375
3	G/C	66	60	6	1	0.609
4	C/C	49	43	6	3	0.660
5	A/A	77	73	4	1	0.522
6	T/A	56	46	10	3	0.802
7	G/A	58	53	5	4	0.730
8	C/A	57	55	2	2	0.226
9	A/T	84	76	8	2	0.603
10	T/T	55	47	8	3	0.725
11	G/T	56	47	9	3	0.837
12	C/T	73	62	11	6	0.559
13	A/G	66	57	9	5	0.591
14	T/G	48	34	14	6	0.663
15	G/G	67	58	9	2	0.649
16	C/G	42	38	4	1	0.189
Total		996	874	122	50	0.993

<sup>a</sup> *EcoRI-MseI* +1/+1, denoting *EcoRI* and *MseI* primers, each with one selective base. The letters (such as A/C, for example) indicate selective bases (A and C, respectively, for example) for *EcoRI* and *MseI*.

<sup>b</sup> Bands scored excluding some small and large fragments at the ends of the gel.

<sup>c</sup> Constant fragments refer to fragments present in all isolates. (See text for definition of polymorphic and informative fragments.)

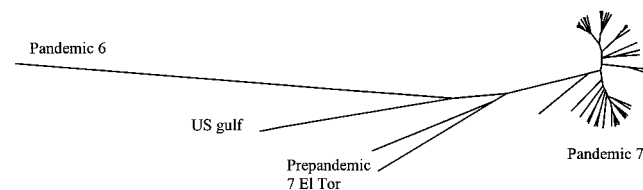


FIG. 1. Relationships among the three toxigenic clones, U.S. gulf isolate, sixth pandemic isolate, and isolates of the seventh pandemic clone constructed by using the UPGMA algorithm based on the Dice coefficient obtained after pairwise comparison of AFLP variation.



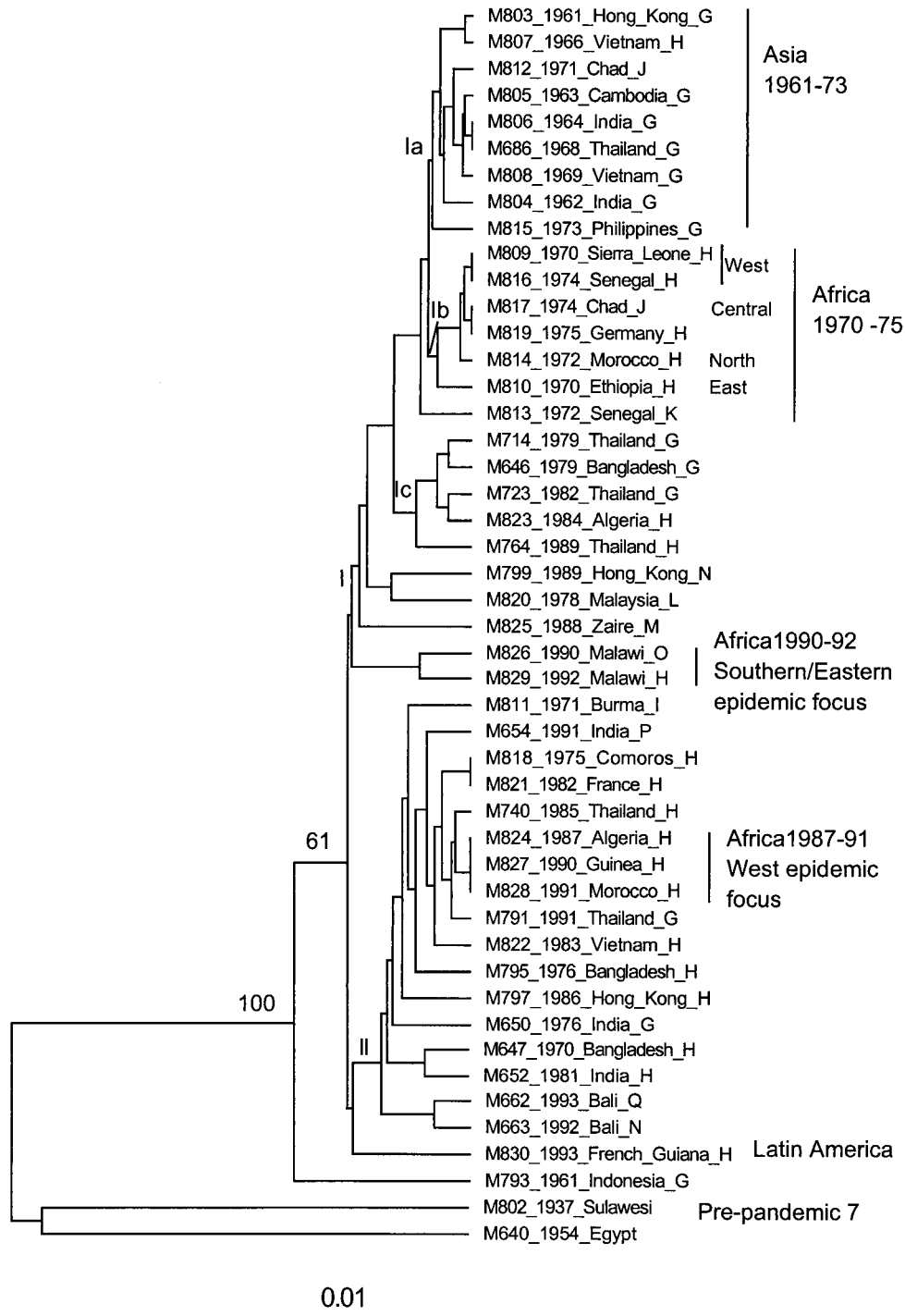


FIG. 2. Dendrogram of the 45 isolates of the seventh pandemic clone constructed by using the UPGMA algorithm based on the Dice coefficient obtained after pairwise comparison of AFLP variation. Bootstrap values are percentages of 1,000 replications and are indicated at the nodes if greater than 50%. The strain information given is name, year and place of isolation, and ribotype. Ribotype data were from our previous study (19).

Ribotyping has been the most useful method for revealing variation within the seventh pandemic clone (6–8, 19, 22, 35, 45). The strains used in this study had been typed by ribotyping previously (19), allowing a direct comparison of the discriminatory power of the two techniques. The *D* values for ribotyping are 0.714 for the *Bgl*I digest, 0.664 for the *Sa*I digest, and

0.715 for the data combined. Four of the 16 AFLP primer pairs have a higher *D* value than ribotyping. Therefore, AFLP, even with a single primer pair, easily surpasses the power of ribotyping.

It is known that different enzyme and/or primer combinations produce AFLP patterns of different complexity (13). A



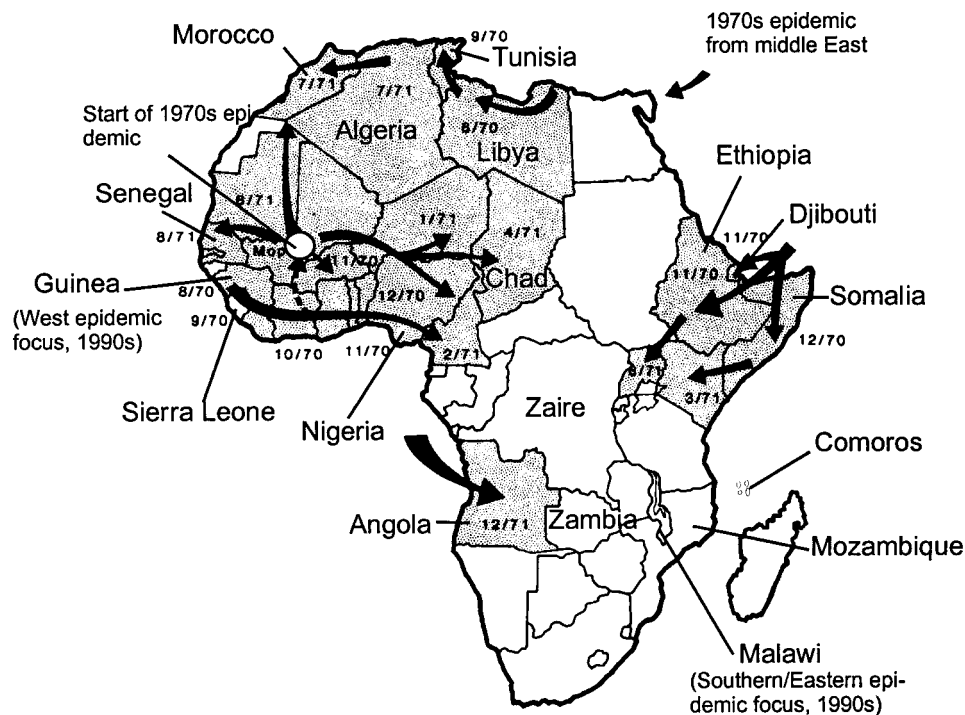


FIG. 3. Epidemics of cholera in Africa. The relevant country or region is annotated on the map to assist interpretation of strain relationships and epidemiological data. Adapted from reference 47 with permission from the World Health Organization (the map in reference 47 shows the route of transmission and date of emergence in the 1970s).

recent study by Jiang et al. (16) showed that AFLP distinguished closely related toxigenic *V. cholerae* strains, but failed to differentiate the 26 seventh pandemic isolates by using a single pair of *ApaI-TaqI* primers of +1 selection. We used *EcoRI* and *MseI* and obtained very useful subdivision by use of all primer pairs, also of +1 selection. The differences between the two studies may lie in the number of primer pairs used and the use of different enzymes.

**Pandemic spread and AFLP relationships.** The seventh pandemic started in Indonesia in 1961 and spread through Asia and the Middle East (3). There was a lull in the late 1960s, but it spread to Africa in 1970 during a major surge in the pandemic. The African outbreak started in the West (Guinea, Sierra Leone, Liberia, Nigeria, and other countries) and spread inland along rivers and trade routes. At the same time, there were outbreaks of cholera in North Africa (Libya, Tunisia, Algeria, and Morocco) and East Africa (Djibouti, Ethiopia, and Somalia), which were thought to have originated in the Middle East (40, 47) (Fig. 3). Although the majority of African isolates are in subcluster Ib, the relationship among the strains supports the proposal that there were at least two introductions of cholera in the 1970s. Isolates from West Africa (Sierra Leone [M809] and Senegal [M816]) are identical, but different from a strain (M810) from East Africa (Ethiopia) in 5 of the 50 informative bands (Table 4). The isolate M814 from North Africa (Morocco) was expected to be closer to M810, but is actually closer to the West African isolates, with only two differences in informative bands. This suggests that the North Africa outbreak was from a third source. However, there is no epidemiological evidence for the North and East epidemics originating from different sources. More isolates

would be helpful, because we have used only a single isolate from each of the regions.

The two strains from Chad are interesting. Chad was affected in 1971 by cholera from the West (40, 47). M817, isolated in 1974 and presumably a descendant from the 1970 outbreak, is in the African subcluster Ib, as expected. However, M812, isolated in 1971, belongs to subcluster Ia and is clearly of Asian origin. We have no explanation for this anomaly, but it is probable that M812 represents yet another separate introduction of cholera into Africa.

In the 1991 resurgence of cholera in Africa, there were two main epidemic foci: in South and East Africa (Zambia, Mozambique, Malawi, and Angola) and in West Africa (40, 47). AFLP data from isolates from the 1990s show close correlation with epidemiological data (Fig. 3). The two Malawi isolates (M826 and M829) are within cluster I, but distantly related to subcluster Ib, representing strains from the 1991 Southern and Eastern epidemic focus. Strains M827 and M828 in cluster II, from Guinea and Morocco respectively, are clearly from the 1990s West epidemic focus and not remnants of the outbreak in the 1970s based on comparison of M828 with the 1970s Morocco isolate (M814), which is in subcluster Ib of African isolates from the 1970s. The strain causing outbreaks in the West is clearly from Asia and may have been introduced before 1991. An isolate (M824) identical to the West epidemic focus isolates M814 and M828 was isolated from Algeria in 1987. Thus, the strain causing the 1991 outbreak in the West appears to have been present in the continent at least 3 years before that.

Cholera remained at very low levels in Africa from 1972 to 1977 and only came back to high levels in 1991 (40). Interest-

ingly, during this period, strains that do not fit into the groups or clusters representing the major outbreaks were also isolated, suggesting yet more introductions of cholera. M823, a 1984 Algerian isolate, is in the subcluster Ic of generally Asian isolates, and M825, isolated in 1988 from Zaire, stands on its own. M825 was previously observed to be very different (19, 27). A 1975 isolate from Comoros off the coast of Africa, which was the only newly affected country that year (3), is not related to isolates from the start of the African epidemic, but to cluster II of Asian isolates.

The AFLP data presented above give a very detailed picture of the spread of cholera in Africa. Asia is regarded as the source of the seventh pandemic spread to Africa in the 1970s, and the AFLP relationships confirm that. The surprising finding is that the diversity in Africa parallels that found in Asia, suggesting that there were many introductions of cholera after the 1970s outbreak, indicating a continuous flow of cholera from regions of endemicity. Furthermore, the strain causing an outbreak could have been present in the local environment many years before. For example, the 1987 Algeria strain is identical in AFLP fingerprint to strains from the West epidemic focus of the 1991 upsurge of cholera. This is a very significant finding for the monitoring of cholera.

There are two other possible explanations to be considered for the observed pattern of variation in *V. cholerae* from Africa. One is that the variation arose independently in Africa. Jiang et al. (16) recently used AFLP to study genetic diversity of clinical and environmental isolates and showed that some clinical O1 isolates from Mexico are very different from the seventh pandemic clone isolates, suggesting an independent origin of these isolates. However, all African strains clearly fall within the seventh pandemic in the tree; they are unlikely to have originated in situ. We must also consider the possibility that the strains all evolved from the 1970s outbreak strain. The fact that 1991 West African epidemic isolates cluster with the 1980s Asian isolates argues strongly against this hypothesis.

The relationships among the strains isolated in the 1960s in Asia are less useful for inferring details of spread of the disease, perhaps because less variation had developed in the first few years of the pandemic. The pandemic traveled from one country to another in a well-documented fashion: first to Hong Kong and the Philippines in 1961, to Cambodia and Thailand in 1963, and then to India and Vietnam in 1964. (For a full list of the countries affected each year, see the review by Barua [3].) That the M807 isolated in 1966 in Vietnam is very similar to M803 isolated in 1961 from Hong Kong is consistent with the spread. However, M805 isolated from Cambodia in 1963 is more distantly related to M803 and M807.

The seventh pandemic spread to Latin America in 1991. One isolate (M830) from the 1991 outbreak was included in this study. Latin American isolates differ from other seventh pandemic isolates with a new MLEE allele of leucine aminopeptidase (46). Although variation within Latin American isolates has been reported (5), all are generally very similar and differ from other seventh pandemic isolates (45). M830 represents the Latin American epidemic in our study and seems to be a divergent member of cluster II, generally comprising 1990s isolates. The origin of the 1991 Latin American epidemic is unclear, with speculation that it was imported from an area in which cholera is endemic but not represented in the strain

collections usually used (41). Clearly AFLP analysis has demonstrated the power that, with more isolates, should make it possible to trace the source of this variant of the seventh pandemic clone.

**Comparison of ribotype and AFLP variation.** We previously identified 11 *Bgl*I ribotypes in the same 45 strains (19) used in this study. Ribotypes G and H are most frequent, with 13 and 21 isolates, respectively (Fig. 2). The difference between the two ribotypes is the presence of a tandem operon in ribotype G and its absence in ribotype H (27). All but one 1960s isolate are ribotype G. The parsimonious assumption is that the progenitor of the seventh pandemic was of ribotype G. The earliest ribotype H isolate is M807, isolated in 1966. The distribution of ribotype H isolates among ribotype G isolates (clusters Ib, Ic, and II) indicates that ribotype H has arisen from ribotype G several times. Most ribotype G isolates are in clusters Ia and Ic. However, two ribotype G isolates in cluster II are grouped with ribotype H isolates, suggesting that this may be a reversion of ribotype H to ribotype G by regenerating the tandem operon deleted in the ribotype H form.

The two ribotype N isolates, M663 and M799, also show independent derivations. M663 is very similar to M662 (ribotype Q) in cluster II, while M799 is far away in cluster I. It is also interesting to note that M817 and M812, both from Chad, have identical ribotypes (J), but different AFLP patterns.

We have previously shown that ribotype variation is mainly due to changes within the *rm* operons that result from recombination between the operons (27). This comparison between AFLP and ribotype relationships confirms our previous suggestion that ribotyping should not be used for evolutionary studies, because changes of this nature could easily revert (26).

AFLP has very high discriminatory power, as shown in this and many other studies (17, 32, 43). It is well suited for epidemiological investigations of a homogenous clone, such as the seventh pandemic clone. However, it becomes impractical for AFLP to be used for routine typing if the proportion of polymorphic fragments is so low that multiple primer pairs have to be used to differentiate isolates. As discussed below, a major advantage of AFLP is that polymorphic fragments found to be useful can be isolated and used to develop a PCR-based typing scheme using only the most useful AFLP fragments.

**Concluding comments.** Evolutionary and epidemiological analyses of a clone are often hampered by the lack of methods to find variation, because few mutations will have accumulated if it is of recent origin. AFLP, which scans genomewide for variation, offers almost unlimited power by use of combinations of enzymes and selective primer extensions. We applied AFLP to 45 isolates of the *V. cholerae* seventh pandemic clone sampled over a 33-year period. Using all 16 pairs of *Eco*RI-*Mse*I selective primers with a 1-base extension, AFLP was able to distinguish all but 4 pairs (M805 and M806, M809 and M816, M817 and M819, and M818 and M821) and one set of 3 (M824, M827, and M828) of the 45 isolates (Fig. 1 and Table 4). This demonstrates the sensitivity and value of AFLP for finding variation in a clone that arose in 1961 and has developed relatively little diversity.

The AFLP analysis has provided by far the best data yet on the evolution of the seventh pandemic clone, allowing us to track outbreaks in detail. This study confirmed the epidemio-



logical evidence for two independent introductions of cholera into Africa in the 1970s. More significantly the two epidemic foci (Southern/Eastern and Western) in the 1991 upsurge of cholera in Africa were shown to be due to yet another two separate introductions. Several other introductions of *V. cholerae* during the 21-year period (1970 to 1991) were also indicated. Our data have further shown that a strain identical in AFLP to the strain causing the 1991 epidemic in West Africa had been isolated 3 years before, indicating that an introduced strain could be present in the continent some years before causing large outbreaks. These findings have significant implications for cholera epidemiology.

AFLP offers great potential for further study of the seventh pandemic clone. The set of strains we used are broadly representative of the seventh pandemic, but the details of the relationships will be much more informative if AFLP analysis is extended to a much larger set of strains. There is also great potential to make use of the AFLP variation observed. It is possible to clone and sequence the bands, which will give insight into the nature of the polymorphic markers. Recent completion of the genome sequence of a seventh pandemic isolate (10) will facilitate such an analysis, giving us a better understanding of the evolution of the seventh pandemic clone. It would then be possible to use the knowledge gained to design PCR primers that allow amplification of only isolates that have a particular AFLP band. This would allow a single multiplex PCR assay to replace the multiple AFLP gels used in this study, with the obvious potential to develop PCR-based typing for epidemiological analysis.

#### ACKNOWLEDGMENT

This research is supported by a grant from the National Health and Medical Research Council of Australia.

#### REFERENCES

- Aarts, H. J., L. A. van Lith, and J. Keijer. 1998. High-resolution genotyping of *Salmonella* strains by AFLP-fingerprinting. *Lett. Appl. Microbiol.* **26**:131–135.
- Arnold, C., L. Metherell, G. Willshaw, A. Maggs, and J. Stanley. 1999. Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high-resolution typing method with phylogenetic significance. *J. Clin. Microbiol.* **37**:1274–1279.
- Barua, D. 1992. History of cholera, p. 1–36. In D. Barua and W. B. Greenough III (ed.), *Cholera*. Plenum Press, New York, N.Y.
- Byun, R., L. D. H. Elbourne, R. Lan, and P. R. Reeves. 1999. Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infect. Immun.* **67**:1116–1124.
- Dalsgaard, A., M. N. Skov, O. Serichantalergs, P. Echeverria, R. Meza, and D. N. Taylor. 1997. Molecular evolution of *Vibrio cholerae* O1 strains isolated in Lima, Peru, from 1991 to 1995. *J. Clin. Microbiol.* **35**:1151–1156.
- Evens, G. M., D. N. Cameron, J. G. Wells, K. D. Greene, T. Popovic, S. Giono-Cerezo, I. K. Wachsmuth, and R. V. Tauxe. 1995. The emerging diversity of the electrophoretic types of *Vibrio cholerae* in the Western hemisphere. *J. Infect. Dis.* **172**:173–179.
- Faruque, S. M., A. R. M. A. Alim, S. K. Roy, F. Khan, G. B. Nair, R. B. Sack, and M. J. Albert. 1994. Molecular analysis of rRNA and cholera toxin genes carried by the new epidemic strain of toxigenic *Vibrio cholerae* O139 synonym Bengal. *J. Clin. Microbiol.* **32**:1050–1053.
- Faruque, S. M., S. K. Roy, A. R. M. A. Alim, A. K. Siddique, and M. J. Albert. 1995. Molecular epidemiology of toxigenic *Vibrio cholerae* in Bangladesh studied by numerical analysis of rRNA gene restriction patterns. *J. Clin. Microbiol.* **33**:2833–2838.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using bootstrap. *Evolution* **39**:783–791.
- 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 et al. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477–483.
- Hookey, J. V., V. Edwards, S. Patel, J. F. Richardson, and B. D. Cookson. 1999. Use of an amplified-fragment length polymorphism technique to fingerprint and differentiate isolates of *Helicobacter pylori*. *J. Microbiol. Methods* **37**:2580–2585.
- Hunter, P. R. 1990. Reproducibility and indices of discriminatory power of microbial typing methods. *J. Clin. Microbiol.* **28**:1903–1905.
- Janssen, P., R. Coopman, G. Huys, J. Swings, M. Bleeker, P. Vos, M. Zabeau, and K. Kersters. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology* **142**:1881–1893.
- Janssen, P., and L. Dijkshoorn. 1996. High resolution DNA fingerprinting of *Acinetobacter* outbreak strains. *FEMS Microbiol.* **142**:191–194.
- Jiang, S. C., V. Louis, N. Choopun, A. Sharma, A. Huq, and R. R. Colwell. 2000. Genetic diversity of *Vibrio cholerae* in Chesapeake Bay determined by amplified fragment length polymorphism fingerprinting. *Appl. Environ. Microbiol.* **66**:140–147.
- Jiang, S. C., M. Matte, G. Matte, A. Huq, and R. R. Colwell. 2000. Genetic diversity of clinical and environmental isolates of *Vibrio cholerae* determined by amplified fragment length polymorphism fingerprinting. *Appl. Environ. Microbiol.* **66**:148–153.
- Jonas, D., H. G. Meyer, P. Matthes, D. Hartung, B. Jahn, F. D. Daschner, and B. Jansen. 2000. Comparative evaluation of three different genotyping methods for investigation of nosocomial outbreaks of Legionnaires' disease in hospitals. *J. Clin. Microbiol.* **38**:2284–2291.
- Karaolis, D. K. R., J. A. Johnson, C. C. Bailey, E. C. Boedeker, J. B. Kaper, and P. R. Reeves. 1998. A *Vibrio cholerae* pathogenicity island (VPI) associated with epidemic and pandemic strains. *Proc. Natl. Acad. Sci. USA* **95**:3134–3139.
- Karaolis, D. K. R., R. Lan, and P. R. Reeves. 1994. Molecular evolution of the seventh pandemic clone of *Vibrio cholerae* and its relationship to other pandemic and epidemic *V. cholerae* isolates. *J. Bacteriol.* **176**:6199–6206.
- Karaolis, D. K. R., R. Lan, and P. R. Reeves. 1995. The sixth and seventh cholera pandemics are due to independent clones separately derived from environmental, nontoxicogenic, non-O1 *Vibrio cholerae*. *J. Bacteriol.* **177**:3191–3198.
- Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P. Jackson. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. *J. Bacteriol.* **179**:818–824.
- Koblavi, S., F. Grimont, and P. A. D. Grimont. 1990. Clonal diversity of *Vibrio cholerae* O1 evidenced by rRNA gene restriction patterns. *Res. Microbiol.* **141**:645–657.
- Koeleman, J. G. M., J. Stoof, D. J. Biesmans, P. H. M. Savelkoul, and C. M. J. E. Vandenbroucke-Grauls. 1998. Comparison of amplified ribosomal DNA restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. *J. Clin. Microbiol.* **36**:2522–2529.
- Kokotovic, B., N. F. Friis, J. S. Jensen, and P. Ahrens. 1999. Amplified-fragment length polymorphism fingerprinting of *Mycoplasma* species. *J. Clin. Microbiol.* **37**:3300–3307.
- Kühn, I., M. J. Albert, M. Ansaruzzaman, N. A. Bhuiyan, S. A. Alabi, M. S. Islam, P. K. B. Neogi, G. Huys, P. Janssen, K. Kersters, and R. Möllby. 1997. Characterization of *Aeromonas* spp. isolated from humans with diarrhea, from healthy controls, and from surface water in Bangladesh. *J. Clin. Microbiol.* **35**:369–373.
- Lan, R., and P. R. Reeves. 1998. Molecular basis of ribotype variation in the seventh pandemic clone and its O139 variant of *Vibrio cholerae*. *Mem. Inst. Oswaldo Cruz* **93**:595–599.
- Lan, R., and P. R. Reeves. 1998. Recombination between rRNA operons created most of the ribotype variation observed in the seventh pandemic clone of *Vibrio cholerae*. *Microbiology* **144**:1213–1221.
- Lan, R., and P. R. Reeves. 2000. Unique adaptor design for AFLP fingerprinting. *BioTechniques* **29**:745–746, 748, 750.
- Liu, S.-L., and K. E. Sanderson. 1996. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc. Natl. Acad. Sci. USA* **93**:10303–10308.
- Lyoda, S., A. Wada, J. Weller, S. J. A. Flood, E. Schreiber, B. Tucker, and H. Watanabe. 1999. Evaluation of AFLP, a high-resolution DNA fingerprinting method, as a tool for molecular subtyping of enterohemorrhagic *Escherichia coli* O157:H7 isolates. *Microbiol. Immunol.* **43**:803–806.
- Momen, H., and C. A. Salles. 1985. Enzyme markers for *Vibrio cholerae*: identification of classical, El Tor, and environmental strains. *Trans. Soc. Trop. Med. Hyg.* **79**:773–776.
- Nair, S., E. Schreiber, K. L. Thong, T. Pang, and M. Altwegg. 2000. Genotypic characterization of *Salmonella typhi* by amplified fragment length polymorphism fingerprinting provides increased discrimination as compared to pulsed-field gel electrophoresis and ribotyping. *J. Microbiol. Methods* **41**:35–43.
- Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**:5269–5273.
- Picardeau, M., G. Prod'Hom, L. Raskine, M. P. LePennec, and V. Vincent.

1997. Genotypic characterization of five subspecies of *Mycobacterium kansasii*. *J. Clin. Microbiol.* **35**:25–32.
35. Popovic, T., C. A. Bopp, Ø. Ølsvik, and K. Wachsmuth. 1993. Epidemiologic application of a standardized ribotype scheme for *Vibrio cholerae* O1. *J. Clin. Microbiol.* **31**:2474–2482.
36. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
37. Sloos, J. H., P. Janssen, C. P. van Boven, and L. Dijkshoorn. 1998. AFLP typing of *Staphylococcus epidermidis* in multiple sequential blood cultures. *Res. Microbiol.* **149**:221–228.
38. Sneath, P. H. A., and R. R. Sokal. 1973. Numerical taxonomy. W. H. Freeman and Company, San Francisco, Calif.
39. Struelens, M. J. 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin. Microbiol. Infect.* **2**:2–11.
40. Swerdlow, D. L., and M. Isaäcson. 1994. Epidemiology of cholera in Africa, p. 297–307. In I. K. Wachsmuth, P. A. Blake, and Ø. Ølsvik (ed.), *Vibrio cholerae* and cholera: molecular to global perspectives. ASM Press, Washington, D.C.
41. Tauxe, R., L. Seminario, R. Tapia, and M. Libel. 1994. The Latin American epidemic, p. 321–344. In I. K. Wachsmuth, P. A. Blake, and Ø. Ølsvik (ed.), *Vibrio cholerae* and cholera: molecular to global perspectives. ASM Press, Washington, D.C.
42. Thal, L. A., J. Silverman, S. Donabedian, and M. J. Zervos. 1997. The effect of Tn916 insertions on contour-clamped homogeneous electrophoresis patterns of *Enterococcus faecalis*. *J. Clin. Microbiol.* **35**:969–972.
43. van Eldere, J., P. Janssen, A. Hoefnagels-Schuermans, S. van Lierde, and W. E. Peetermans. 1999. Amplified-fragment length polymorphism analysis versus macro-restriction fragment analysis for molecular typing of *Streptococcus pneumoniae* isolates. *J. Clin. Microbiol.* **37**:2053–2057.
44. Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**:4407–4414.
45. Wachsmuth, I. K., G. M. Evins, P. I. Fields, Ø. Ølsvik, T. Popovic, C. A. Bopp, J. G. Wells, C. Carrillo, and P. A. Blake. 1993. The molecular epidemiology of cholera in Latin America. *J. Infect. Dis.* **167**:621–626.
46. Wachsmuth, K., Ø. Ølsvik, G. M. Evins, and T. Popovic. 1994. Molecular epidemiology of cholera, p. 357–370. In I. K. Wachsmuth, P. A. Blake, and Ø. Ølsvik (ed.), *Vibrio cholerae* and cholera: molecular to global perspectives. ASM Press, Washington, D.C.
47. World Health Organization. 1991. Cholera in Africa. *Wkly. Epidemiol. Rec.* **66**:305–311.