JOURNAL OF BACTERIOLOGY, Feb. 1993, p. 973–981
0021-9193/93/040973-09$02.00/0

Vol. 175, No. 4

Leptospira Species Categorized by Arbitrarily Primed Polymerase Chain Reaction (PCR) and by Mapped Restriction Polymorphisms in PCR-Amplified rRNA Genes

DAVID RALPH,[1] MICHAEL McCLELLAND,[1]* JOHN WELSH,[1] GUY BARANTON,[2] AND PHILLIPE PEROLAT[3]

California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, California 92037,[1] and Unite de Bacteriologie Moleculaire et Medicale, Institut Pasteur, 75724 Paris Cedex 15,[2] and Laboratoire des Leptospires, Institut Pasteur de Noumea, Nouvelle Caledonie,[3] France

Received 22 June 1992/Accepted 1 December 1992

Reference strains from 48 selected serovars representing eight species of *Leptospira* were examined by two polymerase chain reaction (PCR)-based strategies. First, mapped restriction site polymorphisms (MRSP) were examined in PCR products from portions of *rrs* (16S rRNA gene) and *rrl* (23S rRNA gene). Twenty MRSP and 2 length polymorphisms were used to group reference strains into 16 MRSP profiles. Species assignments were consistent with those obtained by a second method, genomic fingerprinting with arbitrarily primed PCR, in which strains within a species were characterized by many shared arbitrarily primed PCR products. The results of both of these methods were in general agreement with those of previous studies that used DNA-DNA relatedness and confirmed the high level of divergence among the recognized species of *Leptospira*. However, *Leptospira meyeri* serovar ranarum and evansi strains were indistinguishable from some strains of *Leptospira interrogans* sensu stricto. Intervening sequences of about 485 to 740 bp were located near base 1230 in *rrl* of some strains.

A comparison of the many sequences available for *rrs* (16S rRNA gene) (11) indicates that restriction site polymorphisms can distinguish between taxa at the species, genus, and higher levels. It should also be possible to use other relatively conserved genes, such as *rrl* (23S rRNA gene), in the same way. Polymerase chain reaction (PCR) primers directed toward sequences that evolve slowly in rRNA genes can be used to amplify portions of these genes from widely divergent species (40). Previously, Vilgalys and Hester (39) used fingerprints of restriction sites in a 2-kb region of the 23S rRNA gene and part of the 5S rRNA gene to categorize strains and species of *Cryptococcus*. A strategy based on mapped restriction site polymorphisms (MRSP) in such PCR products should be useful for the rapid categorization of strains into subspecies, species, and genera. Potentially, MRSP data for a group of related species could also be used to construct a preliminary phylogeny (37).

Another potentially rapid and convenient strategy for classification involves arbitrarily primed PCR. This method generates a fingerprint of PCR products (41, 44, 45) that can be used to compare strains within a species or closely related species. For instance, in bacteria, arbitrarily primed PCR provides evidence of intra- and interspecific differences (41) and resolved strains of *Borrelia burgdorferi* into three phyletic groups (45), in agreement with DNA homology and multilocus enzyme electrophoresis data (3).

We applied these two PCR-based classification methods to the spirochete genus *Leptospira*, which originally was considered to contain two species, *Leptospira interrogans*, a species pathogenic for animals as well as humans, and *Leptospira biflexa*, a saprophytic species found in surface water and soil (7, 19). Traditionally, each *Leptospira* species was divided into taxa, called serovars, by a classification based on a pairwise comparison of the extent of cross-absorption of antisera raised against each serotype (8). Antigenically related serovars constitute serogroups; in total, *L. interrogans* included at least 202 serovars in 23 serogroups (21). The classification of strains into serovars is tedious, requiring the maintenance of a comprehensive collection of strains and their corresponding rabbit immune sera (19), and does not resolve strains into species. Recently, Yasuda et al. (47) further divided a collection of 40 pathogenic *L. interrogans* serovar reference strains into species by DNA-DNA hybridization with the hydroxyapatite method; they found that members of the pathogenic species *L. interrogans* sensu lato were very heterogeneous at the DNA level and separated them into at least six species, only one of which is now *L. interrogans* sensu stricto. Ramadass et al. (34) obtained similar conclusions by slot blot DNA hybridization and classified some serovars into a new species, *Leptospira kirschneri*. Restriction analysis by pulsed-field gel electrophoresis (16, 17) or by Southern blotting with recombinant DNA probes (23, 25, 29, 31, 38, 48) has been used to distinguish strains within species.

Only about one quarter of the known pathogenic *Leptospira* serovars have been studied by DNA relatedness. As it is difficult to obtain the quantities of leptospiral DNAs required for extensive DNA relatedness studies by conventional methods, we categorized species by using two PCR-based strategies that were both simple and reliable, used less DNA, and could be compared with the various methods used previously.

MATERIALS AND METHODS

Strains. Forty eight *Leptospira* strains from the French National Reference Center Collection are listed in Table 1.

Genomic DNAs. Total genomic DNAs were prepared as described by Perolat et al. (31).

Data base searches and alignments. 16S rRNA sequences available for members of the family *Leptospiraceae* included

* Corresponding author.

TABLE 1. *Leptospira* strains used in this study

Serovar	Strain	Species[a]	MRSP profile[b]	Lane[c]
icterohaemorrhagiae	RGA (type strain)	*L. interrogans*	inteA	1
copenhageni	M 20	*L. interrogans*	inteA	2
djasiman	Djasiman	*L. interrogans*	inteA	3
hebdomadis	Hebdomadis	*L. interrogans*	inteA	4
jalna	Jalna	*L. interrogans*	inteA	5
bratislava	Jez-bratislava	*L. interrogans*	inteA	8
gurungi	Gurung	*L. interrogans*	inteA	9
kremastos	Kremastos	*L. interrogans*	inteA	13
paidjan	Paidjan	*L. interrogans*	inteA (23S only)	16
bangkinang	Bangkinang I	*L. interrogans*	inteB	7
hardjo	Hardjoprajitno	*L. interrogans*	inteB	10
jonsis	Jones	*L. interrogans*	inteB	12
lai	Lai	*L. interrogans*	inteB	14
mooris	Moores	*L. interrogans*	inteB (23S only)	15
ranarum	ICF	*L. meyeri*	inteB	48
evansi	267-1348	*L. meyeri*	inteB	49
cynopteri	3522 C (type strain)	*L. kirschneri*	kirsA	17
grippotyphosa	Moskva V	*L. kirschneri*, *L. interrogans*	kirsA	19
vanderhoedeni	Kipod 179	*L. kirschneri*	kirsA	23
kamituga	Kamituga	*L. kirschneri*	kirsB	20
ndambari	Ndambari	*L. kirschneri*	kirsB	21
galtoni	LT 1014	*L. kirschneri*	kirsC	18
ratnapura	Wumalasena	*L. kirschneri*	kirsC	22
javanica	Veldrat Batavia 46 (type strain)	*L. borgpetersenii*	borgA	24
ballum	Mus 127	*L. borgpetersenii*	borgA	25
mini	Sari	*L. borgpetersenii*	borgA	26
sejroe	M 84	*L. borgpetersenii*	borgA	27
arborea	Arborea	*L. borgpetersenii*	borgA	29
jules	Jules	*L. borgpetersenii*	borgA	31
kenya	Njenga	*L. borgpetersenii*	borgA	32
tarassovi	Perepelicyn	*L. borgpetersenii*	borgB	28
saxkoebing	Mus 24	*L. interrogans*, *L. borgpetersenii*	borgC	6
balcanica	1627 Burgas	*L. borgpetersenii*	borgC	30
sorexjalna	Sorex Jalna	*L. borgpetersenii*	borgC	33
hardjo	Hardjo bovis/Sponselee	?	borgD	50
panama	CZ 214 (type strain)	*L. noguchii*	noguA	35
uncertain[d]	Uncertain[d]	*L. noguchii*	noguB[d]	36
louisiana	LSU 1945	*L. noguchii*	noguB	37
hermani	1342 K (type strain)	*L. santarosai*	santA	38
atlantae	LT 81	*L. santarosai*, *L. interrogans*	santA	39
bakeri	LT 79	*L. santarosai*	santA	40
babudieri	CI 40	*L. santarosai*	santA	42
canalzonae	CZ 188	*L. santarosai*	santA	43
navet	TRVL 109873	*L. santarosai*	santB	41
tabaquite	TRVL 3214	*L. santarosai*	santB	44
celledoni	Celledoni (type strain)	*L. weillii*	weilA	45
sarmin	Sarmin	*L. weillii*	weilB	46
patoc	Patoc I (type strain)	*L. biflexa*	biflexa	47

[a] Previous assignments of strains to species were as follows, with the strains represented as numbers listed in the righthand column of the table. For strains 1 to 5, 24 to 28, 35 to 41, and 45 to 48, the species were defined in reference 47. Our data conflict for serovars grippotyphosa and saxkoebing. For strains 7 to 9, 13, 15, 16, 18, 20, 21, 23, 29, 30, 33, 42, and 44, the species were defined in reference 20. For strains 10, 12, 14, 22, 31, 32, and 43, the species were suspected from rRNA gene restriction patterns (31). For strains 49 and 50, the species were defined elsewhere (30a). The results in reference 34 are consistent with our results for strains 1 to 6, 8, 10, 13, 17, 19, 26, 28, 30, 36 to 38, 40, and 45 but not strain 39. The results in reference 38 are consistent with our results for strain 50.

[b] Shown are the 16 different rRNA profiles of MRSP observed (see Table 2): inteA and inteB, interrogans A and B, respectively; kirsA, kirsB, and kirsC, kirschneri A, B, and C, respectively; borgA, borgB, borgC, and borgD, borgpetersenii A, B, C, and D, respectively; noguA and noguB, noguchii A and B, respectively; santA and santB, santarosai A and B, respectively; weilA and weilB, weillii A and B, respectively; and biflexa, biflexa.

[c] Listed are the lanes in which the PCR product from a strain appears in Fig. 1 and 2. Lane 50 is not present in Fig. 2.

[d] The status of this strain, which was initially believed to be fortbragg, is not yet settled.

L. interrogans serovars canicola (GenBank accession number X17547) and pomona (GenBank accession number M71241) and *Leptonema illini* (GenBank accession number M34118) (30). A 23S rRNA sequence was available for *L. interrogans* serovar canicola (GenBank accession number X14249) (12). By aligning these sequences with the eubacterial *rrs* and *rrl* genes available in GenBank version 69 by using the MacVector program (IBI-Kodak, New Haven, Conn.), we developed primers in highly conserved regions that showed at least a 90% match with almost all such genes.

PCR. Oligonucleotides manufactured by Genosys, Houston, Tex., were as follows: 16S-1507, CCAGATCTGAGC TCAAGGAGGTGATCCAGC; 16S-11, GGCTGCAGTC GACGTTTGATCCTGGCTCAG; 23S-1432, GGTGTCGAC TATGAACCTGCTTCCCATCGACTAC; and 23S-220, AAC CAGAATTCCGTCAGTAGCGGTGAGCGAA. The under-

lined sequences are restriction sites for cloning. At least 18 nucleotides at the 3′ end are homologous to the rRNA genes for each primer. The numbers indicate the first base showing homology with the rRNA gene sequence (see Results). The primers were located at bases 11 to 27 and at bases 1507 to 1492 in the *rrs* gene (GenBank accession number X17547) and at bases 220 to 239 and at bases 1466 to 1432 in the *rrl* gene (GenBank accession number X14249). The primer spanning bases 11 to 27 in the *rrs* gene was located at the same place as a conserved primer used by Weisburg et al. (40).

PCR was performed with approximately 1 ng of genomic DNA–500 nM primer–1.25 U of *Taq* polymerase (Perkin Elmer Cetus, Norwalk, Conn.)–0.2 mM each deoxynucleoside triphosphate–2.5 mCi of 3,000-Ci/mmol [^{32}P]dCTP in a volume of 50 μl of the recommended buffer (with 1.75 mM $MgCl_2$). Cycling was done as 30 cycles of 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C.

Restriction digestion. Restriction enzymes *Alu*I, *Ava*II, *Dde*I, *Fok*I, *Hha*I, *Hin*fI, *Hph*I, *Mnl*I, *Nla*IV, *Pal*I, *Rsa*I, *Sau*3A, *Sau*96I, *Scr*FI, and *Taq*I were purchased from Stratagene, La Jolla, Calif., *Aci*I, *Bfa*I, *Bsr*I, and *Mse*I were purchased from New England Biolabs, Beverly, Mass.

Four microliters of PCR product was cleaved without further processing by 1 to 5 U of each restriction enzyme in a reaction volume of 24 μl; the manufacturer's recommended buffer and incubation temperature were used. Each PCR amplification of 50 μl produced enough material for 12 restriction digestions.

Electrophoresis. Three microliters of each restriction digest was mixed with 1 μl of glycerol-dye loading buffer, and the mixture was electrophoresed on a 0.35-mm-thick 5% acrylamide native gel with 1× TBE (90 mM Tris-borate, 2 mM EDTA). After electrophoresis, the gel was dried on a vacuum dryer and autoradiographed for 12 to 72 h on Kodak X-Omat X-ray film.

Arbitrarily primed PCR. Fingerprinting was performed as described previously (41, 44) with minor modifications. Primers KG (5′-CACACGCACACGGAAGAA-3′) and KF (5′-CACGCACACGCACAGAGAG-3′) were purchased from Genosys. Fifty-microliter reaction mixtures were prepared with 5 or 1 ng of DNA–1× *Taq* polymerase buffer–4 mM $MgCl_2$–0.2 mM each deoxynucleoside triphosphate–1 μM single oligonucleotide primer–5 μCi of [^{32}P]dCTP–1.25 U of *Taq* polymerase (Perkin Elmer). These reactions were cycled 2 times through a low-stringency temperature profile (94°C for 5 min for denaturation, 40°C for 5 min for low-stringency annealing, and 72°C for 5 min for extension) and then 40 times through a high-stringency temperature profile (94°C for 1 min for denaturation, 60°C for 1 min for high-stringency annealing, and 72°C for 2 min for extension). Two microliters of each reaction mixture was combined with 10 μl of 80% formamide–dye, and this mixture was heated to 68°C for 15 min; 2.5 μl of each sample was loaded on a 4% acrylamide–50% urea sequencing gel and run in the Poker Face SE 1500 sequencer apparatus (Hoefer Scientific Instruments, San Francisco, Calif.). Electrophoresis was done at 1,500 V until the xylene cyanol tracking dye was approximately 10 cm from the bottom. The gel was exposed to Kodak X-Omat X-ray film.

Computer programs. The parsimony analysis package for the Macintosh computer was kindly supplied by D. L. Swofford (37). The PHYLIP package was developed by J. Felsenstein (9). The programs were run on a MacII ci.

RESULTS

PCR primers directed toward conserved regions of the 16S rRNA (*rrs*) and 23S rRNA (*rrl*) genes of eubacteria. *Leptospira rrs* and *rrl* gene sequences were aligned with the eubacterial *rrs* and *rrl* gene sequences available in GenBank version 69. We developed primers for highly conserved regions that showed at least a 90% match with almost all eubacterial genes. The primers allowed amplification of PCR products of bases 11 to 1507 in the *rrs* gene and bases 220 to 1466 in the *rrl* gene from all the *Leptospira* genomes listed in Table 1 and from genomes of a variety of other genera, such as *Borrelia* and *Pseudomonas* (data not shown). Most PCR amplifications contained only one prominent product of the expected size. An exception was a found in the PCR amplifications for all strains of *Leptospira borgpetersenii*. This product is labeled B in Fig. 1. The origin of the product is unknown.

MRSP. Restriction digestions were performed on the PCR-amplified portions of the rRNA genes by use of restriction endonucleases that are known to cleave frequently in these genes, and the resulting fragments were separated by electrophoresis. A total of 172 restriction sites representing about 700 bp were examined; 22 of these sites were polymorphic among *Leptospira* isolates (excluding the divergent *L. biflexa* group). The following enzymes yielded no polymorphisms in two strains from each of the seven pathogenic species examined and were not used against the set of 48 strains: *Aci*I, *Alu*I, *Ava*II, *Bfa*I, *Fok*I, *Mse*I, *Pal*I, *Sau*3A, *Sau*96I, and *Scr*FI.

Enzymes that detected polymorphisms were *Bsr*I, *Dde*I, *Hin*fI, *Hha*I, *Hph*I, *Mnl*I, *Nla*IV, *Rsa*I, and *Taq*I. As an example, the *Hin*fI digests of the PCR products from the *rrs* genes of all strains are presented in Fig. 1. Two of eight *Hin*fI sites were polymorphic. The map locations of the polymorphic restriction sites in the genes could be inferred from the known gene sequences for some strains (12, 40). The fragment labeled A is due to the loss of the site at base 974. When the site at base 974 is present, two fragments, of 302 and 445 bp, result. The fragments labeled C and D are derived from the 224-bp fragment when the site at base 198 is present.

A total of 20 MRSP and 2 length polymorphisms were detected in 47 strains of seven pathogenic *Leptospira* species. These polymorphisms occurred in 16 different patterns, or MRSP profiles (Table 2). The sequence divergence between MRSP profiles, estimated on the basis of divergence in the restriction sites, ranged from less than 0.5% base substitutions within species to about 2% base substitutions between the most divergent species (as previously defined [34, 47]).

The *rrs* and *rrl* genes occur in two copies in the *L. interrogans* genome (13) and, unusual among eubacteria, are apparently not closely linked (13, 14, 24, 35). Occasionally, a restriction site may differ between rRNA gene homologs in the same genome (e.g., reference 27 and references therein), but this situation was not observed and, in any case, should not interfere with the ability to group strains into species.

Arbitrarily primed PCR. Arbitrarily primed PCR fingerprints of leptospiral genomic DNA consisted of discrete and reproducible sets of products. An example of such a set of fingerprints is shown in Fig. 2. Strains within a species shared many arbitrarily primed PCR fingerprints, and the fingerprints differed markedly between strains from different species (as previously defined (34, 47]). Relationships observed with arbitrarily primed PCR data were entirely consistent with those observed with MRSP data.

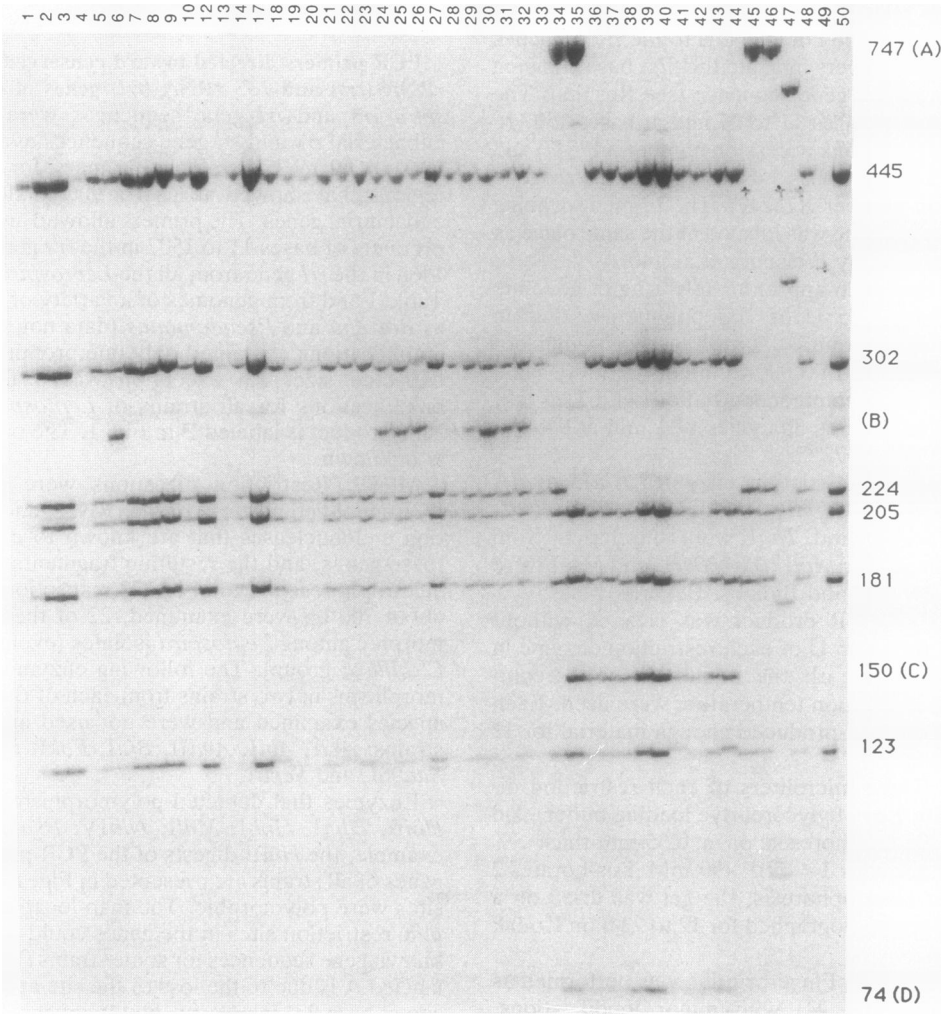

FIG. 1. *Hin*fI restriction digests of PCR-amplified portions of 16S genes. Shown is an autoradiograph of a native 5% acrylamide gel with TBE. The lane assignments are given in Table 1. The numbers on the right are nucleotides determined from the sequence of the *rrs* gene of *L. interrogans* (GenBank accession number X17547). *Hin*fI polymorphisms at bases 198 and 974 are responsible for the variations in restriction patterns. Fragments labeled A, B, C, and D are discussed in the text.

Although primers as short as 10 bases can be used for arbitrarily primed PCR (46), we used primers 18 or more bases long because they generally produce more polymorphic PCR products. These products are up to 1,500 bases long and are therefore easier to score reliably on a high-resolution acrylamide gel (41, 44, 45) than on an agarose gel (46).

Insertion element in the *rrl* gene. One striking feature of the restriction data for the *rrl* gene is the occurrence of insertions ranging from about 485 to 740 bp in some species. These various insertions are summarized as types b through f in Table 2. Preliminary DNA sequencing data indicate that these insertions replace a short stretch of sequence spanning base 1230 in the *rrl* gene (33).

Phylogenetic analysis. Restriction site polymorphisms can potentially be used to construct a phylogenetic relationship between taxa (9). If one can map restriction sites, parsimony analysis may use the data more efficiently than genetic distance analysis (6, 37).

Trees were constructed for the *Leptospira* data in Table 2 by a variety of methods, including genetic distance analysis with the additive or the ultrametric assumption (9) and parsimony analysis with a weighting in favor of site loss versus site gain (6, 37). A potential phylogenetic relationship between species was evident. An example of such a phylogenetic tree is shown in Fig. 3; for this experiment, parsimony analysis was performed with the PAUP phylogenetic package under the assumption that the loss of a site was favored three to one over the gain of a site. Most of the groupings seen in this hypothetical phylogenetic tree were also observed in a majority of trees generated by genetic distance analysis either with Fitch$^+$ or with Kitch$^+$ in the PHYLIP package of computer programs (9) or by the Felsenstein bootstrap method (10). The relative virtues of these methods and the different assumptions that they use are still a topic of active debate. The caveats attached to the phylogenetic hypothesis presented are discussed below.

DISCUSSION

We have used two simple PCR-based methods to classify a group of 48 *Leptospira* serovar reference strains. The two methods, (i) PCR of rRNA genes followed by restriction and

TABLE 2. Data matrix for polymorphisms detected by restriction digestion[a]

MRSP profile[b]	No. of strains examined	16S gene											23S gene									
		*Bsr*I 582	*Dde*I 80	*Dde*I 121	*Dde*I 216	*Hha*I 222	*Hha*I 640	*Hin*fI 198	*Hin*fI 974	*Nla*IV 165	*Taq*I 826	Length	*Bsr*I 606	*Hha*I 1110	*Hph*I 626	*Hph*I 697/788	*Mnl*I 259	*Mnl*I 279	*Mnl*I 377	*Mnl*I 543	*Rsa*I 499	*Rsa*I 548	Length
inteA	9	−	+	+	−	−	−	−	+	+	−	−	−	+	+	−	−	−	+	−	+	+	−
inteB[c]	7	−	+	+	−	−	−	−	+	+	−	−	−	+	+	−	+	+	+	−	+	+	−
kirsA	3	−	+	−	−	−	−	−	+	+	−	−	−	+	−	−	+	−	+	−	+	+	−
kirsB	2	−	+	−	−	−	−	−	+	+	−	−	−	+	+	−	+	−	−	−	+	+	−
kirsC	2	−	+	−	−	−	−	−	+	+	−	a	−	−	−	−	+	−	+	−	+	+	−
borgA	7	−	+	−	−	+	−	−	+	−	−	−	−	−	+	+	+	+	+	+	−	+	−
borgB	1	−	+	−	−	+	−	−	+	−	−	−	−	−	+	+	+	+	+	+	+	+	−
borgC	3	−	+	−	−	+	−	−	+	−	−	−	−	−	+	+	+	+	+	+	+	+	b
borgD	1	−	+	−	−	+	+	−	+	−	−	−	−	−	+	+	+	+	+	+	+	+	b
weilA	1	−	+	−	+	+	−	−	−	+	−	−	−	−	+	+	+	+	+	+	−	+	c
weilB	1	−	−	−	+	+	−	−	−	−	−	−	−	−	+	−	+	+	−	−	+	+	d
noguA	1	−	+	−	−	−	−	+	−	−	+	−	−	+	−	−	+	+	+	−	+	+	−
noguB	2	+	+	−	−	−	−	+	+	−	+	−	−	+	−	−	+	+	+	−	+	+	e
santA	5	−	+	−	−	−	−	+	+	−	+	−	+	−	−	−	+	+	+	−	+	−	f
santB	2	−	−	−	−	−	−	+	+	−	+	a	+	−	−	−	+	+	+	−	+	+	f
biflexa	1	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	+	+	+	?	−	−	−

[a] The position of each restriction polymorphism is given for the sequence having GenBank accession number X17547 or X14249. Polymorphisms that distinguish only *L. biflexa* are excluded. + and − indicate the presence and absence of a restriction site, respectively. The single length polymorphism between bases 27 and 133 in the *rrs* gene is scored as present (a) or absent (−). Length polymorphisms at about base 1230 in the *rrl* gene are scored as absent (−) or as lengths designated b through f. Restriction site polymorphisms within these insertions were not scored.

[b] The type strain in each species is designated as having profile A. See Table 1, footnote *b*, for explanations of designations.

[c] Seven for interrogans B and two for meyeri.

(ii) fingerprinting by arbitrarily primed PCR, provided taxonomic results that were in agreement with each other and generally in agreement with the species assignments based on DNA-DNA homology (34, 47).

The differences in MRSP and in arbitrarily primed PCR products for strains belonging to different species were usually considerable and presumably reflect the large amount of sequence divergence between species. More than one MRSP profile was obtained for most species, although the profiles for any single species generally differed by only one or two restriction sites. Thus, MRSP data alone will generally be insufficient to define a species. There are DNA-DNA hybridization data (34, 47) for almost all strains in each MRSP profile and arbitrarily primed PCR data for all strains (41, 43) to support the translation of MRSP profiles into species designations. Once species designations are established, as we have done here, it should be possible to use the MRSP profiles for the analysis of strains in the future, without recourse to serology or DNA-DNA hybridization. Similarly, each species was usually characterized by at least three prominent arbitrarily primed PCR fingerprinting products per primer that were not shared with the other

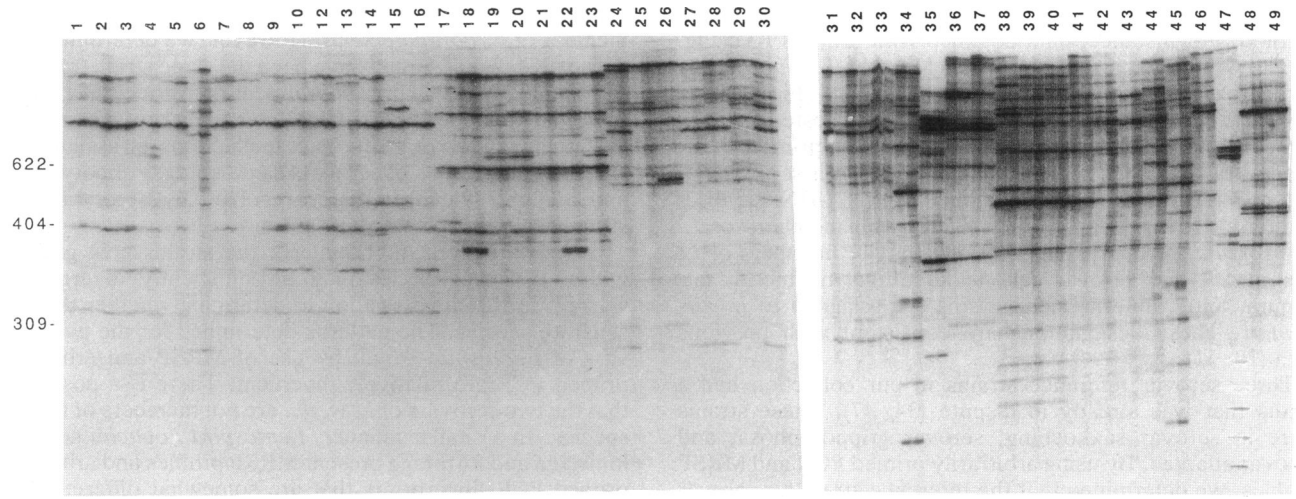

FIG. 2. Arbitrarily primed PCR fingerprints of *Leptospira* genomic DNAs. Shown is an autoradiograph of a denaturing 5% acrylamide gel with TBE. Arbitrarily primed PCR was performed with primer KG. The lane assignments are given in Table 1. The numbers on the left are nucleotides determined from pBR322 digested with *Hae*III.

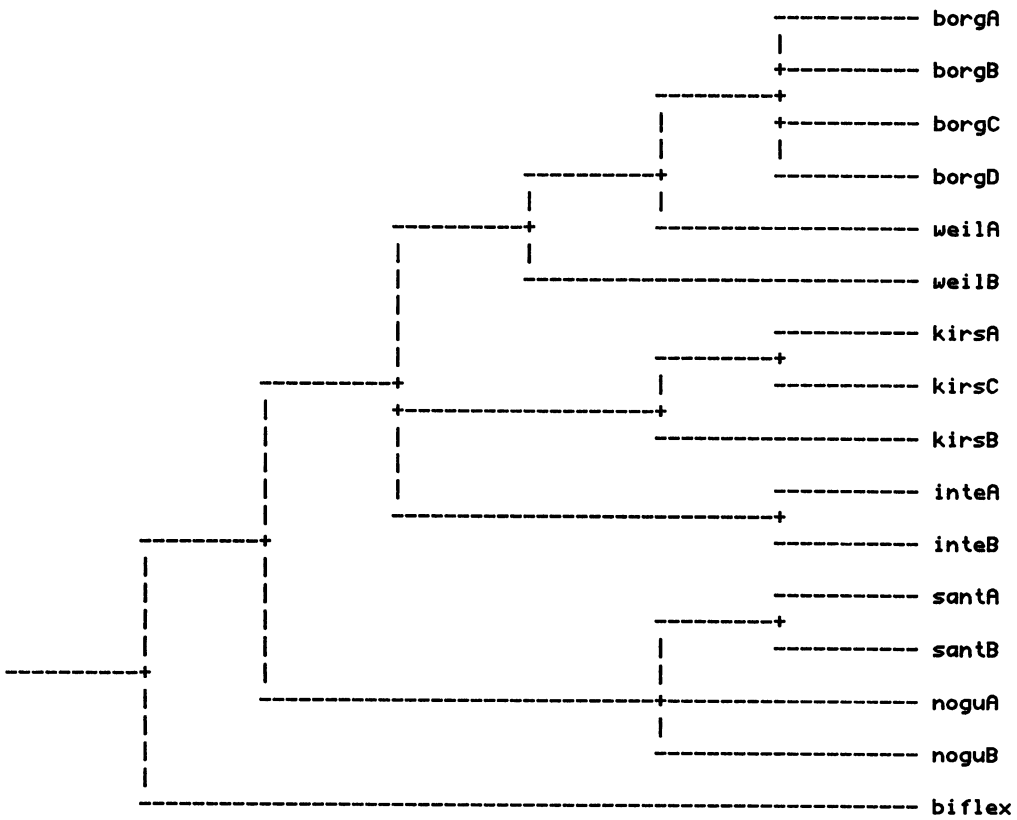

FIG. 3. Phylogenetic tree constructed by parsimony analysis. Shown is a cladogram constructed with the PAUP phylogenetic package of Swofford (37). All MRSP characters were weighted three to one against the reappearance of a restriction site (37). The insertion sequence in the *rrl* gene was set as a six-state unordered character type. A very small length polymorphism in the *rrs* gene was also set as unordered. Characters that did not show evidence of intraspecific variation were given a double weighting. *L. biflexa* was defined as the outgroup. The cladogram was constructed by use of a complete search by the Branch-Bound method (37). See Table 1, footnote *b*, for explanations of designations.

species. These shared PCR products allowed simple and rapid categorization of strains into the species *L. interrogans*, *L. kirschneri*, *L. borgpetersenii*, or *Leptospira santarosai*. However, arbitrarily primed PCR is probably more sensitive to sequence divergence than MRSP in rRNA genes. Thus, intraspecific comparisons can be made by this method. For instance, data obtained with primers KG (Fig. 2) and KF (data not shown) indicate that *L. interrogans* serovars icterohaemorrhagiae and copenhageni (Fig. 2, lanes 1 and 2) are closely related, as previously suggested (16, 31).

By use of MRSP profiles and arbitrarily primed PCR, species designations were obtained for several serovar reference strains not previously studied by DNA-DNA hybridization. Serovar jonsis and lai strains are members of *L. interrogans*, serovar canalzonae strains are members of *L. santarosai*, and serovar galtoni, kamituga, ndambari, ratnapura, and vanderhoedeni strains are assigned to *L. kirschneri*. These designations are consistent with previous data (20, 31).

Three serovar reference strains in our collection had a status that was already in dispute (34, 47). These strains were in serovar saxkoebing, serovar grippotyphosa, and serovar atlantae. By using arbitrarily primed PCR and MRSP profiles, we determined that the reference strain for serovar grippotyphosa is a strain of the species *L. kirschneri*. This result is consistent with the results of Ramadass et al. (34) and Kaufmann (20) but not consistent with the results of Yasuda et al. (47), who assigned this strain to *L. interrogans*. Conversely, our reference strain for serovar atlantae is a member of *L. santarosai*, as indicated by Yasuda et al. (47) but in contrast to the results of Ramadass et al. (34). The reference strain for serovar saxkoebing from the Pasteur collection and the collection of Ramadass et al. (34) is a member of *L. borgpetersenii*. However, we determined that an isolate of what should have been the same strain (received from David Miller, U.S. Department of Agriculture, Ames, Iowa) is a member of *L. interrogans* (data not shown). The disagreements among reference collections in assignments for reference strains of serovars saxkoebing, grippotyphosa, and atlantae will need to be resolved by serotyping the strains from each laboratory again.

Our data suggest that some of the strains now grouped within a single species may be sufficiently divergent to warrant further investigation to determine whether they are separate species. The patterns determined for the two serovars of *Leptospira weillii* by use of MRSP and arbitrarily primed PCR are relatively divergent. There is a possibility that the two serovars of *L. weillii* are not members of a single species. In a similar manner, *Leptospira noguchii* serovars louisiana and fortbragg present MRSP profiles and arbitrarily primed PCR fingerprints that are somewhat different from those generated by serovar panama (from the type strain). Since MRSP profile and arbitrarily primed PCR fingerprint data alone are generally insufficient to define a species, these

questions will remain until DNA hybridization or rRNA sequence data are available. At the other extreme, serovars evansi and ranarum (from reference strains of *Leptospira meyeri*) are very similar to each other (1, 2, 4, 21) and yield arbitrarily primed PCR fingerprints similar to those of *L. interrogans* and MRSP profiles identical to those of one group of *L. interrogans*. It is not clear why Yasuda et al. (47) determined the serovars they assigned to *L. meyeri* to be a species distinct from *L. interrogans*, whereas we and others (4, 36) did not.

Insertion sequence in the 23S rRNA gene. The presence of an insertion in the *rrl* gene was observed as a PCR length polymorphism (26, 42). An insertion ranging from about 485 to 740 bp occurs in the *rrl* gene in some strains of some species of *Leptospira*. The site of this insertion is approximately base 1230 in the *rrl* gene. The fact that only some strains of some species have this insertion suggests that the intervening sequence may be a sequence-specific mobile element.

The presence of an insertion in the *rrl* gene of some *Leptospira* strains is not unprecedented in eubacteria (5) but represents the first observation of this phenomenon outside the family *enterobacteriaceae*. Intervening sequences of about 90 bp have been found in the *rrl* gene of some strains of *Salmonella typhimurium* (5). One of the two places in the *S. typhimurium rrl* gene in which insertions were observed is very close to the site of insertion in the *Leptospira rrl* gene. The insertion sequences in *S. typhimurium* are cleaved from the mature 23S rRNAs but not spliced. Thus, the *rrl* gene is fragmented but apparently functional. Hsu et al. (18) showed that the 23S rRNA molecules of certain *Leptospira* strains are fragmented into 14S and 17S pieces. Our data include six of the strains that they examined, and our data for insertions in the *rrl* gene are completely concordant with the appearance of the fragmented 23S rRNA molecules, strongly suggesting that, as previously observed for *S. typhimurium*, the insertions are processed but not religated.

The sequences of the insertions in the *S. typhimurium* 23S rRNA gene show no evidence of the mechanism of insertion. The 720-bp insertion that we observed in some *Leptospira* strains is by far the largest insertion observed in an *rrl* gene in eubacteria to date and may retain evidence of the mechanism of insertion. We have cloned and are sequencing insertions. Preliminary data indicate that a conserved open reading frame occurs in a shared portion of these intervening sequences (33).

Hypothesized phylogeny. Restriction site polymorphisms can potentially be used to construct a phylogenetic relationship between taxa. Vilgalys and Hester (39) used the Fitch-Margoliash genetic distance method in the PHYLIP package (9). If one can map restriction sites, parsimony analysis may use the data more efficiently than genetic distance analysis (6, 37). However, rRNA genes undergo intense selection, so one can expect that some restriction sites will vary because of sequence changes over only a very limited repertoire of possibilities. In parsimony analysis, homoplasies due to the reappearance of a character can be expected to occur much more frequently during the evolution of highly selected DNA than during the drift of purely neutral DNA. Also, one must be cautious when comparing strains that are very closely related and that may undergo genetic exchange, since strains that undergo genetic exchange do not fit the assumption of a bifurcating tree, although the algorithms nevertheless build a tree. However, there is no evidence yet of genetic exchange in *Leptospira* or other spirochetes, and even if genetic exchange occurs, it is unlikely that there will be significant amounts of exchange between species.

With these considerations in mind, trees were constructed for the *Leptospira* data (Table 2) by a variety of methods, including genetic distance analysis with the additive or the ultrametric assumption (9) and parsimony analysis with a weighting in favor of site loss versus site gain (6, 37). An example of a phylogenetic tree constructed by parsimony analysis is shown in Fig. 3. This particular tree was constructed after weighting in favor of site loss and weighting in favor of characters that did not show evidence of intraspecific variation. A potential relationship between species was usually evident with a variety of different phylogenetic methods: (i) *L. noguchii* was grouped with *L. santarosai*, (ii) all three *L. kirschneri* profiles were grouped together, (iii) *L. interrogans* was grouped with *L. meyeri* and, more tentatively, *L. kirschneri*, and finally, (iv) *L. borgpetersenii* was grouped with *L. weillii*. However, the *L. weillii* profiles often seemed as closely related to *L. borgpetersenii* profiles as to each other, perhaps indicating, as did the arbitrarily primed PCR data, that the two *L. weillii* serovars are not members of the same species. The Felsenstein bootstrap method (10) yielded a phylogeny in which groupings i, ii, and iii were supported. The four groupings were observed in a majority of trees generated by genetic distance analysis either with Fitch$^+$ or with Kitch$^+$ in the PHYLIP package of computer programs (9). Even though these methods for constructing phylogenies all supported similar groupings, the nature of the characters and the small size of the data set make these groupings tentative.

The paucity of arbitrarily primed PCR characters shared between species limits phylogenetic analysis of genomic relationships among *Leptospira* species and reflects the high level of DNA divergence between *Leptospira* species: the highest level of DNA homology between species type strains is 54% (*L. borgpetersenii* versus *L. weillii* [48]), and almost all of the type strains show less than 40% DNA homology in pairwise comparisons (48). However, in principle, an intraspecific phylogenetic tree could be constructed on the basis of the shared derived characters observed within species, as was shown for the Lyme disease borreliae (45), although the possibility of extensive genetic exchange within species makes such phylogenies tentative.

Advantages of PCR-based methods as a way to group strains. MRSP profiles and arbitrarily primed PCR fingerprints generally supported previous species assignments and provided information for species designations of strains that had not been previously studied by DNA-based methods. MRSP profiles for rRNA genes and arbitrarily primed PCR fingerprints were consistent with the high genomic divergence of *Leptospira* species inferred from DNA-DNA hybridization data (34, 47).

Classification by use of MRSP in PCR products should be quite general. In the case of pathogenic *Leptospira* species, restriction digestion of the PCR product from *rrs* with *Dde*I, *Hha*I, *Nla*IV, and *Taq*I should categorize most new strains into the appropriate group. In addition, not all restriction site polymorphisms for these enzymes present in the population will have been detected by our sample, so additional new profiles may be found in a more extensive survey. Also, it is possible that the antigens responsible for serotypes are laterally transferred, in which case, strains with the same MRSP profiles or with similar arbitrarily primed PCR fingerprints may not be of the same serotype.

In experiments similar to those presented here, we have demonstrated that clinical isolates of *Pseudomonas aerugi-*

nosa can be distinguished into at least three profiles on the basis of restriction site polymorphisms and that these can be further divided by use of arbitrarily primed PCR (15). Similarly, members of the *B. burgdorferi* species cluster, which includes *B. burgdorferi* sensu stricto, *Borrelia garinii*, and *Borrelia* group VS461 (3, 45), can all be reliably distinguished by restriction site polymorphisms (32) and further divided by arbitrarily primed PCR (45).

RFLP analysis by Southern blotting, restriction analysis by pulsed-field gel electrophoresis (17), and DNA-DNA hybridization all have many virtues for categorizing strains, and DNA sequencing remains one of the most effective ways to obtain data for grouping strains and constructing phylogenies (e.g., 30). However, all of these methods are more laborious than scoring of restriction site polymorphisms in PCR products (32, 39) or arbitrarily primed PCR fingerprinting (41, 45). Furthermore, RFLP analysis and genomic fingerprinting by pulsed-field gel electrophoresis (17) do not allow for the unambiguous assignment of restriction site polymorphisms unless a restriction map can be produced, a limitation that reduces the usefulness of the observed differences in constructing phylogenetic relationships. DNA hybridization, in addition to being more laborious, yields a single measure of distance that is probably less reliable when highly divergent organisms are being compared (e.g., 22). Finally, PCR uses much less DNA, an issue for some organisms that grow slowly or are potentially pathogenic, and the use of MRSP in PCR-amplified rRNA genes should be applicable to species that cannot be cultured at all.

MRSP and arbitrarily primed PCR appear to be simple and fast tools for the study of the genomic diversity of *Leptospira* strains and for preliminary identification at the species level. These tools should be useful for the rapid preliminary classification of strains in many species of microorganisms.

ACKNOWLEDGMENTS

We thank D. Miller, U.S. Department of Agriculture, Ames, Iowa, for sending us strains. We thank Isabelle Saint Girons for critical reading of the manuscript, Rita Cheng for technical assistance, and Irène Lecuyer, Dominique Persan, Henri Thevenet, and Elizabeth Bellenger for strains and DNA preparation.

P.P. and G.B. were supported by the Institut Pasteur Foundation and Institut National de la Sante et de la Recherche Medicale (contract 910613). M.M. was supported by NIH grants HG00106 and HG00456, J.W. was supported by NIH grant AI32644, and David Ralph was supported by set-up funds at CIBR.

REFERENCES

1. **Alexander, A. D., L. B. Evans, M. F. Baker, H. J. Baker, D. Ellison, and M. Marriapan.** 1975. Pathogenic leptospiras isolated from Malaysian surface waters. Appl. Microbiol. **29:**30–33.
2. **Babudieri, B.** 1972. Systematics of a *Leptospira* strain isolated from frog. Experientia **28:**1252–1253.
3. **Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J.-C. Piffaretti, M. Assous, and P. A. D. Grimont.** 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. Int. J. Syst. Bacteriol. **42:**365–369.
4. **Brendle, J. J., M. Rogul, and A. D. Alexander.** 1974. DNA hybridization among selected leptospiral serotypes. Int. J. Syst. Bacteriol. **24:**205–214.
5. **Burgin, A. B., K. Parodos, D. J. Lane, and N. R. Pace.** 1990. The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. Cell **60:**405–414.
6. **Carr, S. M., A. J. Brothers, and A. C. Wilson.** 1987. Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of *Xenopus* frogs. Evolution **41:**176–188.
7. **Cole, J. R., Jr.** 1984. Spirochetes, p. 40–58. *In* G. R. Carter (ed.), Diagnostic procedures in veterinary bacteriology and mycology, 4th ed. Charles C Thomas, Publisher, Springfield, Ill.
8. **Dikken, H., and E. Kmety.** 1978. Serological typing methods of leptospires, p. 259–307. *In* T. Bergan and J. R. Norris (ed.), Methods in microbiology, vol. 11. Academic Press, Inc., New York.
9. **Felsenstein, J.** 1988. PHYLIP: phylogenetic inference package, version 3.0. University of Washington, Seattle.
10. **Felsenstein, J.** 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22:**521–565.
11. **Fox, G. E., E. Stackebrandt, R. B. Hepsell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209:**457–463.
12. **Fukunaga, M., I. Horie, and I. Mifuchi.** 1989. Nucleotide sequence of a 23S ribosomal RNA gene from *Leptospira interrogans* serovar *canicola* strain Moulton. Nucleic Acids Res. **17:**2123.
13. **Fukunaga, M., and I. Mifuchi.** 1992. Unique organization of *Leptospira interrogans* rRNA genes. J. Bacteriol. **171:**5763–5767.
14. **Fukunaga, M., and M. Sohnaka.** 1992. Tandem repeat of the 23S and 5S ribosomal RNA genes in *Borrelia burgdorferi*, the etiological agent of Lyme disease. Biochem. Biophys. Res. Commun. **183:**952–957.
15. **Heffernan, T., M. McClelland, and J. Welsh.** Unpublished data.
16. **Herrmann, J. L., C. Baril, E. Bellenger, P. Perolat, G. Baranton, and I. Saint Girons.** 1991. Conservation of the genome among isolates of *Leptospira interrogans*. J. Bacteriol. **173:**7582–7588.
17. **Herrmann, J. L., E. Bellenger, P. Perolat, G. Baranton, and I. Saint Girons.** 1992. Pulsed-field gel electrophoresis of *Not*I digests of leptospiral DNA: a new rapid method of serovar identification. J. Clin. Microbiol. **30:**1696–1702.
18. **Hsu, D., M.-J. Pan, Y. C. Zee, and R. B. LeFebre.** 1990. Unique ribosome structure of *Leptospira interrogans* is composed of four rRNA components. J. Bacteriol. **172:**3478–3480.
19. **Johnson, R. C., and S. Faine.** 1984. Family II. *Leptospiraceae*, p. 62–67. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. Williams & Wilkins, Baltimore.
20. **Kaufmann, A. F.** 1986. Unpublished data.
21. **Kmety, E., and H. Dikken.** 1988. Revised list of *Leptospira* serovars. University Press, Groningen, The Netherlands.
22. **Lawrence, J. G., H. Ochman, and D. L. Hartl.** 1991. Molecular and evolutionary relationships among enteric bacteria. J. Gen. Microbiol. **137:**1911–1921.
23. **Lefebvre, R. B.** 1987. DNA probe for detection of the *Leptospira interrogans* serovar *hardjo* genotype hardjo-bovis. J. Clin. Microbiol. **25:**2236–2238.
24. **Liesack, W., and E. Stackebrandt.** 1989. Evidence for unlinked *rrn* operons in the planctomycete *Pirellula marina*. J. Bacteriol. **171:**5025–5030.
25. **Marshall, R. B., B. E. Wilton, and A. J. Robinson.** 1981. Identification of *Leptospira* serovars by restriction endonuclease analysis. J. Med. Microbiol. **14:**163–166.
26. **McClelland, M., C. Petersen, and J. Welsh.** 1992. Length polymorphisms in tRNA intergenic spacers detected by using the polymerase chain reaction can distinguish streptococcal strains and species. J. Clin. Microbiol. **30:**1499–1504.
27. **Mylvganam, S., and P. P. Dennis.** 1992. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui*. Genetics **130:**399–410.
28. **Nielsen, J. N., C. H. Armstrong, and N. C. Nielsen.** 1989. Relationship among selected *Leptospira interrogans* serogroups as determined by nucleic acid hybridization. J. Clin. Microbiol. **27:**2724–2729.
29. **Pacciarini, M. L., M. L. Savio, S. Tagliabue, and C. Rossi.** 1992. Repetitive sequences cloned from *Leptospira interrogans* serovar *hardjo* genotype harjoprajitno and their application to sero-

var identification. J. Clin. Microbiol. **30:**1243–1249.
30. **Paster, B. J., F. E. Dewhirst, W. G. Weisburg, L. A. Tordoff, G. J. Fraser, R. B. Hespell, T. B. Staton, L. Zablen, L. Mandelco, and C. R. Woese.** 1991. Phylogenetic analysis of spirochetes. J. Bacteriol. **173:**6101–6109.
30a. **Perolat, P., and G. Baranton.** Unpublished data.
31. **Perolat, P., F. Grimont, B. Regnault, P. A. D. Grimont, E. Fournie, H. Thevenet, and G. Baranton.** 1990. rRNA gene restriction patterns of *Leptospira*: a molecular typing system. Res. Microbiol. **141:**159–171.
32. **Ralph, D., G. Baranton, D. Postic, and M. McClelland.** Unpublished data.
33. **Ralph, D., and M. McClelland.** Unpublished data.
34. **Ramadass, P., B. D. W. Jarvis, R. J. Corner, D. Penny, and R. B. Marshall.** 1992. Genetic characterization of pathogenic *Leptospira* species by DNA hybridization. Int. J. Syst. Bacteriol. **42:**215–219.
35. **Schwartz, J. J., A. Gazumyan, and I. Schwartz.** 1992. rRNA gene organization in the Lyme disease spirochete, *Borrelia burgdorferi*. J. Bacteriol. **174:**3757–3765.
36. **Stamm, L. V., and N. W. Charon.** 1988. Sensitivity of pathogenic and free-living *Leptospira* spp. to radiation and mitomycin C. Appl. Environ. Microbiol. **54:**728–733.
37. **Swofford, D. L.** 1991. PAUP: phylogenetic analysis using parsimony, version 3.0. Illinois Natural History Survey, Champaign.
38. **Van Eys, G. J. J. M., M. J. Gerritsen, H. Korver, G. J. Schoone, C. C. M. Kroon, and W. J. Terpstra.** 1991. Characterization of serovars of the genus *Leptospira* by DNA hybridization with *hardjobovis* and *icterohaemorrhagiae* recombinant probes with special attention to serogroup sejroe. J. Clin. Microbiol. **29:** 1042–1048.
39. **Vilgalys, R., and M. Hester.** 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. J. Bacteriol. **172:**4238–4246.
40. **Weisburg, W. G., S. M. Burns, D. A. Pelletier, and D. Lane.** 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. **173:**697–703.
41. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18:**7213–7218.
42. **Welsh, J., and M. McClelland.** 1992. PCR-amplified length polymorphisms in tRNA intergenic spacers for categorizing staphylococci. Mol. Microbiol. **6:**1673–1680.
43. **Welsh, J., P. Perolat, G. Baranton, and M. McClelland.** Unpublished data.
44. **Welsh, J., C. Petersen, and M. McClelland.** 1991. Polymorphisms generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping. Nucleic Acids Res. **19:**303–306.
45. **Welsh, J., C. Pretzman, D. Postic, I. Saint Girons, G. Baranton, and M. McClelland.** 1992. Genomic fingerprinting by arbitrarily primed polymerase chain reaction resolves *Borrelia burgdorferi* into three distinct phyletic groups. Int. J. Syst. Bacteriol. **42:**370–377.
46. **Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:**6531–6535.
47. **Yasuda, P. H., A. G. Steigerwalt, K. R. Sulzer, A. F. Kaufmann, F. Rogers, and D. J. Brenner.** 1987. Deoxyribonucleic acid relatedness between serogroups and serovars in the family *Leptospiraceae* with proposals for seven new *Leptospira* species. Int. J. Syst. Bacteriol. **37:**407–415.
48. **Zuerner, R. L., and C. A. Bolin.** 1991. Nucleic acid probe characterizes *Leptospira interrogans* serovars by restriction fragment length polymorphisms. Vet. Microbiol. **24:**355–366.