

Phylogenetic Evidence for Horizontal Transfer of an Intervening Sequence between Species in a Spirochete Genus

DAVID RALPH AND MICHAEL McCLELLAND*

California Institute of Biological Research, La Jolla, California 92037

Received 31 March 1994/Accepted 25 July 1994

The 23S rRNA genes (*rrl* genes) of some strains of certain species of the spirochete genus *Leptospira* carry an intervening sequence (IVS) of 485 to 759 bases flanked by terminal inverted repeat and encoding an open reading frame for a putative protein of over 120 amino acids. The structure and the sporadic distribution of the IVS suggest that it might be a mobile element that can be horizontally transferred within or between species. Phylogenetic hypotheses based on the sequences for six IVS open reading frames from various species were compared with hypotheses constructed by using DNA sequences from the 16S rRNA gene (*rrs*), which is not closely linked to *rrl* in this genus. The predicted phylogenies for the IVS and *rrs* differed in a major respect: one strain that claded with *L. weilii* in the tree based on the *rrs* data claded with *L. noguchi* in the tree based on the IVS data. Neither set of data supported a tree in which this strain was constrained to be in the same clade as was supported by the other set of data. This result indicates a probable horizontal transfer of the IVS from a recent ancestor of *L. noguchi* to a recent ancestor of one of the *L. weilii* strains. This observation is the first indication of horizontal transfer of elements encoded on the chromosomes of spirochetes.

An intervening sequence (IVS) with an extensive open reading frame (ORF) is observed in the 23S rRNA genes (*rrl* genes) of some but not all species of *Leptospira* and in some but not all strains in those species (22, 23). This sporadic distribution of the IVS suggests the possibility of horizontal transfer. This possibility can be examined by comparing phylogenetic data from the IVS with phylogenetic data collected from elsewhere in the genome. If the phylogenetic hypotheses constructed for the IVS and for a gene located elsewhere in the genome are concordant, then lateral transfer is not supported. However, if the phylogenetic trees for the IVS and another gene are different, then this difference is circumstantial evidence for horizontal transfer.

Horizontal transfer between species has been observed in a large number of organisms and is now recognized as an important process in evolution (see, e.g., reference 27). Most examples of horizontal transfer of chromosomal genes have been in naturally transformable species (reviewed in reference 17). Some of the best examples are the transfer of antibiotic resistance genes on plasmids or transposons in bacteria (14) and the movement of retroelements (29). Other good examples include group I and II introns that seem to have moved between species by horizontal transfer, on the basis of remarkable similarities in sequence in divergent organisms (2, 4, 10, 15, 16, 18, 19, 31).

Spirochetes may exchange genetic information relatively rarely. Sequence evidence from a spirochete, *Borrelia burgdorferi*, indicates clonality of the chromosome (5). No case of natural horizontal transfer of an element on a chromosome has been confirmed in these organisms. However, transfer of plasmids between species has been suspected (5), and it has been speculated that the unusual ends of the linear *Borrelia* plasmids have a common ancestry with certain eukaryotic

viruses (12). Gene transfer in the laboratory has proved difficult although not impossible (24a, 28).

The IVS in *rrl* of *Leptospira* species showed all the hallmarks of a horizontally transferable element (22, 23). We investigated the phylogeny of this element because it might show the first evidence of natural horizontal transfer between spirochete chromosomes.

MATERIALS AND METHODS

Genomic DNAs. Genomic DNAs from the strains listed in Fig. 1 were prepared by Phillippe Perolat, using a previously described method (21). The same strains were also acquired from David Miller (U.S. Department of Agriculture, Ames, Iowa).

PCR. PCR was performed on approximately 1 ng of genomic DNAs, using 500 nM primers, 1.25 U of *Taq* polymerase (Perkin-Elmer Cetus, Norwalk, Conn.), and 0.2 mM each deoxynucleoside triphosphate along with 2.5 μ Ci of 3,000-Ci/mmol [α - 32 P]dCTP in a volume of 50 μ l of the recommended buffer plus 1.75 mM MgCl₂. The temperature profile was 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C for 35 cycles. Ten microliters of each PCR mixture was loaded on a 1% agarose gel (FMC, Rockland, Maine), and DNA products were separated by electrophoresis. All DNAs gave successful amplifications.

The IVS in *rrl* was amplified by PCR using the primers 23S-1150 and 23S-1432R (listed below) directed against conserved *rrl* sequences flanking the insertion. These primer sequences were located at bases 1147 to 1171 and bases 1466 to 1432, respectively, in the sequence under GenBank accession number X14249 (7). The *rrs* genes were amplified by PCR using the primers 16S-20 and 16S-1500 (listed below), which were located at bases 10 to 27 and bases 1492 to 1507, respectively, in the sequence under GenBank accession number X17547 (8).

Oligodeoxyribonucleotide primers. All primers were manufactured by Genosys, Woodlands, Tex. The primers are named for the gene and then the location within the gene. Orientation

* Corresponding author. Mailing address: California Institute of Biological Research, 11099 N. Torrey Pines Rd., La Jolla, CA 92037. Phone: (619) 535-5486. Fax: (619) 535-5472.

NB36	M	V	S	F	K	K	I	E	D	10	D	V	F	K	R	A	Y	D	L	S	20
WA52	M	V	G	F	K	K	S	I	E	D	F	D	V	F	N	R	A	Y	D	L	S
WA45	M	P	G	V	K	K	S	Y	E	D	F	D	V	F	N	R	A	Y	D	L	S
WB46	M	A	G	L	K	R	N	I	E	D	F	D	V	F	N	R	A	Y	D	L	S
SA39	M	A	G	L	K	R	N	I	E	D	F	D	V	F	N	R	A	Y	D	L	S
BD50	M	D	K	I	R	K	Y	E	D	L	F	E	V	F	R	K	S	Y	L	V	S
Node	M	A	G	L	K	S	Y	E	D	L	F	E	V	F	R	K	S	Y	L	V	S
30																					
NB36	L	K	I	H	K	F	S	L	S	L	P	K	F	E	Q	S	D	L	A	N	40
WA52	L	R	I	H	K	F	S	L	S	L	P	R	F	E	Q	S	D	L	A	N	40
WA45	L	K	V	H	H	F	S	L	S	L	P	R	F	E	Q	S	D	L	A	N	40
WB46	L	K	I	H	K	L	S	L	E	M	P	K	F	E	Q	L	D	L	A	N	40
SA39	L	K	I	H	K	L	S	L	E	F	P	K	F	E	Q	S	D	L	A	N	40
BD50	L	E	I	H	K	K	T	L	E	F	P	K	F	E	Q	Y	G	L	A	S	40
Node	L	K	I	H	K	F	S	L	E	L	P	K	F	E	Q	S	D	L	A	N	40
50																					
NB36	Q	L	R	R	S	S	K	S	V	C	S	N	F	A	E	G	F	A	K	Q	60
WA52	Q	L	R	R	S	S	K	S	I	C	S	N	F	A	E	G	F	A	K	Q	60
WA45	Q	L	R	R	S	S	K	S	I	C	S	N	F	A	E	G	F	A	K	Q	60
WB46	Q	L	R	R	S	S	K	S	V	C	S	N	F	A	E	G	F	A	K	Q	60
SA39	Q	L	R	R	S	S	K	S	V	C	S	N	F	A	E	G	F	A	K	Q	60
BD50	Q	L	R	R	S	S	K	S	I	C	S	N	F	A	E	G	F	A	K	Q	60
Node	R	L	R	R	S	S	K	S	I	C	S	N	F	A	E	G	F	A	K	Q	60
70																					
NB36	N	F	S	V	K	E	F	R	K	Y	L	I	I	A	L	G	S	S	D	E	80
WA52	N	F	S	A	K	E	F	R	K	Y	L	I	I	A	L	G	S	S	D	E	80
WA45	N	F	S	I	R	E	F	R	K	Y	L	I	I	A	I	G	S	S	N	E	80
WB46	N	F	S	K	R	E	F	R	K	Y	L	I	I	A	I	G	S	S	N	E	80
SA39	S	F	S	A	K	E	F	R	K	Y	L	I	I	G	L	G	S	S	D	E	80
BD50	S	Q	S	K	Q	E	F	R	R	F	L	S	M	T	I	G	S	S	D	E	80
Node	S	F	S	K	R	E	F	R	K	Y	L	I	I	A	I	G	S	S	N	E	80
90																					
NB36	S	M	L	W	L	R	Y	C	K	D	L	G	Y	L	S	E	S	E	S	D	100
WA52	S	M	L	W	L	R	Y	C	K	D	L	G	Y	L	S	E	S	E	S	D	100
WA45	S	I	V	W	L	K	Y	C	K	E	D	L	G	Y	L	S	E	T	E	F	A
WB46	S	A	L	W	L	R	Y	C	K	D	L	G	Y	L	S	E	T	E	D	Q	Y
SA39	S	I	L	W	L	R	Y	C	K	D	L	G	Y	L	S	E	K	V	W	L	
BD50	T	K	V	W	L	N	Y	A	C	D	L	G	Y	L	S	E	S	L	Y	T	
Node	S	I	L	W	L	K	Y	C	K	D	L	G	Y	L	S	E	T	E	Y	S	
110																					
NB36	E	W	T	E	E	Y	L	Q	V	S	K	M	L	A	R	L	A	K	L	Q	120
WA52	E	W	T	K	E	Y	L	Q	I	S	K	M	L	A	R	L	A	K	L	Q	120
WA45	E	W	S	D	E	Y	E	Q	I	A	R	M	L	S	K	L	A	Q	N	Q	120
WB46	E	W	S	D	E	C	L	Q	I	A	K	I	L	S	K	L	A	Q	I	Q	120
SA39	E	M	T	N	E	Y	L	E	I	A	S	K	X	L	A	V	L	A	K	K	120
BD50	S	W	K	G	T	Y	V	D	I	A	K	M	L	S	G	L	Y	A	K	S	120
Node	E	W	S	D	E	Y	E	Q	I	A	K	M	L	S	K	L	A	K	I	Q	120
130																					
NB36	D	S	K	F	L	R	Q	T	V	N	R	Q	Y	*							
WA52	D	S	K	S	L	R	*														
WA45	Q	*																			
WB46	A	V	S	E	*																
SA39	S	N	*																		
BD50	E	*																			
Node	E																				

FIG. 1. Amino acid sequences of ORFs in the IVS. The conceptual amino acid sequences of the ORFs are shown. Strains examined were as follows: WA52, *L. weillii* serovar worfoldi strain Worfold (GenBank accession number U12675); WA45, *L. weillii* serovar celledoni strain Celledoni (U12674); WB46, *L. weillii* serovar sarmin strain Sarmin (L14266); NB36, *L. noguchi* fortbragg (serovar and strain uncertain) (L14264); SA39, *L. santarosai* serovar atlantae strain LT 81 (L14265); and BD50, *L. borgpetersenii* serovar hardjo, strain Hardjo bovis/Sponselee (L14263). The abbreviations WA52, WA45, WB46, NB36, SA39, and BD50 refer to the species, the MRSP classification, and the strain number in reference 23. The sequence labeled Node is the node connecting the strain carrying the IVS with the greatest genetic distance, BD50, to the other strains in Fig. 2A. This node represents the best current prediction of the ancestral protein sequence.

in the antisense strand is indicated with an R. The underlined sequences are restriction sites used for cloning.

- 23S-1150: 5'-GCAGGAATTCTTTAAAGAGTGC GTAATAGCTCAC
EcoRI
- 23S-1432R: 5'-GGTGTCGACTATGAACCTGCTCCCATCGACTAC
SalI
- LEP23S-1295: 5'-GGAGGAGTAAAGAAATGAAGA
- LEP23S-2567R: 5'-GCCGACATCGAGGTGCCAAA
GGCTGCAGTCGACGTTTGATCCTGGCTCAG
PstI SalI
- 16S-1500: CCAGATCTGAGCTCAAGGAGGTGATCCAGC
BglIII SacI
- LEP16S-326: 5'-TGCTGCCTCCCGTAGG
- LEP16S-696: 5'-CTTCGCCACTGGTGTTT
- LEP16S-910R: 5'-TAAACCACATGCTCCACC

LEP16S-1199R: 5'-CTTTGTACCGGCCATTGTAG
IVSa: 5'-GAACAGTTGGATTGGCGAACCAA

Cloning. The IVS PCR product was cleaved with *EcoRI* and *SalI*, and the *rrs* product was cleaved with *SacI* and *SalI*; this was followed by ligation with T4 DNA ligase. The recognition sites for these enzymes were engineered into the oligonucleotides used in these PCRs and permitted directional insertion of the PCR products into the pBSK and pBKS polylinkers (Stratagene Inc., La Jolla, Calif.).

Sequencing. Single-stranded DNAs for sequencing were rescued by using the helper phage VCSM13 (Stratagene). DNA sequence determination was performed by using the Sequenase reagent kit (U.S. Biochemicals), various oligonucleotide primers, and α -³⁵S-dATP (Amersham). The -40 primer (U.S. Biochemicals) and T7 primer (Stratagene) anneal to sequences flanking the insert in the Bluescript vector. The primers LEP16S-326 and LEP16S-696 (listed above) anneal to the sense strand of the 16S rRNA gene at nucleotides 326 and 696, respectively. LEP16S-910R and LEP16S-1199R anneal to the antisense strand at nucleotides 910 and 1199, respectively. In addition, primer IVSa was designed to hybridize to a sequence within the conserved ORFs of the IVSs and was used for sequencing the 3' ends of the sense strand of the ORFs and the 3' untranslated region. The resulting reaction products were separated by electrophoresis on 5% polyacrylamide-7 M urea gels. The gels were fixed in 10% glacial acetic acid-12% methanol, dried, and autoradiographed with XAR-5 film (Kodak). In all cases, at least two independent clones from each ligation were sequenced.

Mapped restriction site polymorphisms (MRSPs). Restriction enzymes *DdeI*, *HinfI*, *MnlI*, and *TaqI* were purchased from Stratagene. *Bsp* 50I, *BsrI*, and *BbvI* were purchased from New England Biolabs, Beverly, Mass.

PCR of *rml* was performed with the primers LEP23S-1295 and LEP23S-2567R. PCR products (4 μ l) were cleaved without further processing with 1 to 5 U of each restriction enzyme in a reaction volume of 24 μ l, using the manufacturers' recommended buffer and incubation temperature. Each 50- μ l PCR amplification produced enough material for 12 restriction digests.

Electrophoresis. Three microliters of each restriction digestion mixture was mixed with 1 μ l of Ficoll-dye loading buffer and electrophoresed on a 0.35-mm-thick 5% acrylamide native gel in 1 \times 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.

Computer programs. The parsimony analysis package for the Macintosh was kindly supplied by D. L. Swofford (26). The program was run on a MacII ci.

Nucleotide sequence accession numbers. The GenBank accession numbers for the U12669 sequences obtained here are through U12677.

RESULTS AND DISCUSSION

The sporadic occurrence of an IVS in some strains of some species of *Leptospira* might be explained by the occurrence of the IVS in a common ancestor and loss in some subsequent lines. However, the IVS could be a transposable element that is sometimes moved between strains and sometimes lost. To examine the possibility that the IVS has moved between strains, we compared the sequences of the IVS ORFs between species and within a species, selecting those IVSs that seemed to vary within species on the basis of MRSPs (22, 23). Such variation could be due to divergence or could reflect lateral

transfer. In order to distinguish between these possibilities, we compared the phylogenetics of the IVSs with the phylogenetics of another gene, *rrs*, located many kilobases distant in the same genomes (9). Discordance between these phylogenies would be evidence for lateral transfer between strains.

Phylogenetic tree for the IVS. We had previously sequenced four IVSs found in the *rml* genes of certain strains of *Leptospira* (22, 23). Two more sequences from *L. weilli* serovar *worsfoldi* strain *Worfold* and *L. weilli* serovar *celledoni* strain *Celledoni* were obtained by PCR between conserved portions of *rml* flanking the IVS. These IVSs were chosen for sequencing because MRSP data indicated that they seemed to be from strains within the *L. weilli* group but the IVSs seemed to be different from each other and from the closest relative we had already sequenced (reference 23 and data not shown). The alignment of the DNA and amino acid sequences from the six ORFs is presented in Fig. 1.

Phylogenetic hypotheses were constructed for these data. We used parsimony analysis because this may be the most effective strategy for constructing and assessing the confidence of a cladogram (26). First, a most-parsimonious tree for the ORF in the IVS DNA sequences was constructed by using the PAUP program (26). Only data from the DNA sequence in the ORF were used because this was the only part of the IVS that could be confidently aligned between species. This tree assumed that transition mutations would be twice as common as transversions. One method to ascertain the confidence of such a hypothesis is to perform repeated tree constructions from random samples of the data and build a consensus tree (6). This strategy allows an estimate to be made regarding the proportion of data supporting each node. When such a "bootstrap" tree was built (Fig. 2A), each branch was supported and the topology was the same as that of the most-parsimonious tree, lending considerable support to the phylogeny predicted. The major discrepancy between this phylogeny and the previously assigned species designations was the clading of the IVS from *L. weilli* WA52 with the IVS of *L. noguchi* NB36.

Phylogenetic hypothesis for the rRNA genes. To confirm species assignments, the *rrs* genes were investigated. PCR using primers directed toward highly conserved regions near the 5' and 3' ends of *rrs* and *rml* generated products that could be used to perform restriction digestions. In a previous study 20 MRSPs had been collected from the *rrs* and *rml* genes by this strategy (23). An additional 10 MRSPs were collected, and all are presented in Table 1.

The complete set of 30 MRSPs was analyzed by parsimony, using the assumption that during DNA sequence divergence the disappearance of a restriction site was more likely than the reappearance of the site. In evolutionarily neutral DNA the ratio of loss to reappearance of a four-base restriction site is about 4:1 (3). The exact ratio of gain to loss in a real sequence is difficult to assess and almost certainly varies widely between sites in the highly selected rRNA genes. Nevertheless, most sites are probably very rarely regained once they have been lost. We investigated a variety of gain-to-loss ratios, and there was little effect on the tree topology in the range from 5:1 to 2:1. Earlier evidence from DNA-DNA hybridization (24, 30) and arbitrarily primed PCR (23) allowed us to constrain the tree to make each species monophyletic except *L. weilli*, for which we had evidence of significant intraspecific divergence (23). A bootstrap tree from 100 replications optimized by heuristic TBR (Tree bisection-reconnection) was constructed on the basis of the MRSP data with a gain-to-loss weighting of 4:1 (data not shown). There were clearly discrepancies in the positions of WA52, WA45, and NB36 between the cladograms in Fig. 2A and the cladogram constructed by using the MRSP

data (not shown) that might indicate a horizontal transfer event. However, use of the MRSP data alone in a bootstrap consensus revealed that there was not enough information to support the discrepancy between the trees. For this reason more phylogenetically informative characters were sought.

Sequences of 16S rRNA genes. Sequencing of *rrs* was performed to obtain more phylogenetic data from this gene. A region spanning 1,465 bp was amplified by using the same primers as used for generating the MRSP data. The resulting products were cloned and sequenced (see Materials and Methods). A phylogenetic tree was constructed for these data and is presented in Fig. 2B.

A pairwise comparison of the genetic similarities between strains from the IVS-ORF and *rrs* sequence data is presented in Table 2. The *rrs* sequences are identical for *L. weilli* WA45 and WA52, yet WA52 carries an IVS that is much more similar to one found in *L. noguchi* NB36 than to the IVS found in WA45, consistent with the data in Fig. 2A and B.

Interestingly, WA45 and WA52 also differs in their *rrs* sequences compared with the other *L. weilli* strain, WB46, by more than they do compared with *L. santarosai*. Strains WA45 and WA46 also differ substantially in their arbitrarily primed PCR fingerprints (23). These strains had been grouped previously on the basis of serology, while strain WA52 had remained unclassifiable on this basis (20a). While the definition of species is somewhat arbitrary, particularly in organisms such as spirochetes that may behave largely clonally (5), it is clear that *L. weilli* consists of at least two distinct clades that may differ sufficiently to warrant designation as different species.

Phylogenetic trees of the *Leptospira* 16S rRNA genes were also constructed after adding the data of Fukunaga et al. (7) and Paster et al. (20) and the recent extensive data of Hookey et al. (13). The topologies of the taxa in Fig. 2A did not change. While we are very confident of our data, we found that some published sequence determinations were different from ours. These discrepancies did not change the topologies or the conclusions drawn.

It could be argued that it is not acceptable to use the combined *rrs* and *rml* data to construct a cladogram because *rml* is so intimately linked to the IVS that the *rml* gene might be transferred along with the IVS. Below we explain the evidence indicating that this is unlikely. Thus, the MRSP data from *rml* were combined with *rrs* sequence data in order to produce the most data available to date for relating the species. The resulting cladogram, shown in Fig. 3, is concordant with that presented in Fig. 2A. The apparent differences are due to the cladogram in Fig. 3 being rooted by using *L. biflexa* as an outgroup. Discrepancies between this phylogeny and previous published phylogenies (13, 23) may be attributed to an increase in the number of informative characters used here.

Difference between the two observed tree topologies. The phylogenetic hypothesis for the IVS in Fig. 2A differs from that for the *rrs* in Fig. 2B and 3. In particular, the location of strain WA52 differs in these trees. WA52 clades with NB36 in the tree constructed by using the IVS data and with WA45 in the tree constructed by using the *rrs* data. New phylogenetic hypotheses were generated in which the tree topologies were constrained to force WA52 to clade with *L. weilli* WA45 when the IVS data were used (Fig. 2C). Similarly, WA52 was constrained to clade with *L. noguchi* NB36 in a tree constructed by using *rrs* sequence data (Fig. 2D). These constrained trees are considerably less parsimonious than the trees constructed without forcing the topology. Tree lengths of 617 versus 646 steps were obtained for the IVS data, and lengths of 45 versus 63 steps were obtained for the *rrs* sequence data. Of the 105 tree topologies that are possible for six taxa, the

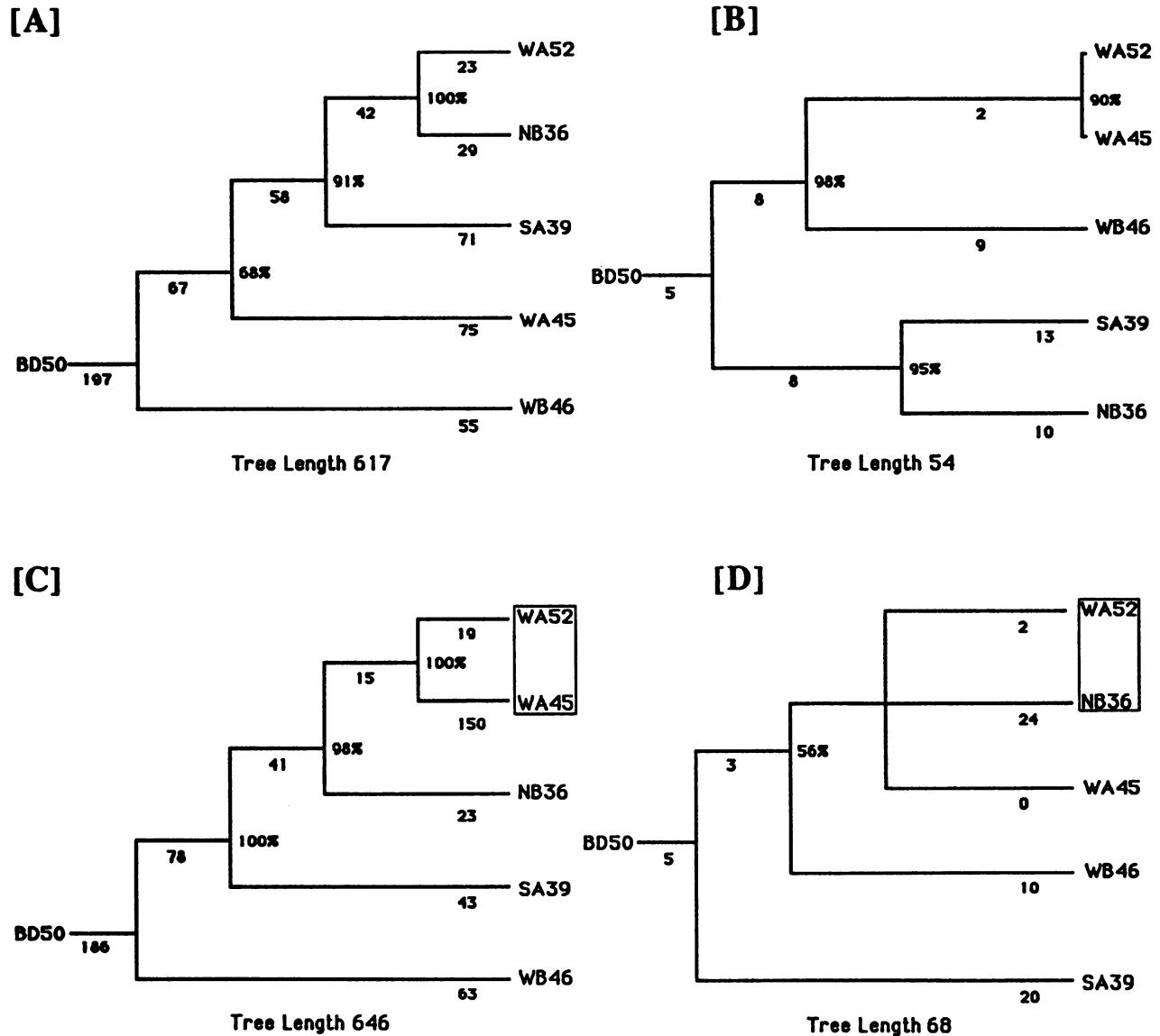


FIG. 2. Phylogenetic hypotheses for the IVS and for *rrs*. The strains used are listed in Fig. 1. Each unrooted tree represents the consensus generated after 100 bootstrap replications of the Branch and Bound algorithm (26). BD50 is shown on the left of each cladogram to assist in comparing topologies. The percentage of bootstrap trees consistent with each clade is indicated at each node. The branch lengths between nodes are indicated below each branch. Transversions versus transitions were weighted 2:1. (A) Phylogenetic hypothesis for the ORF in the IVS. (B) Phylogenetic hypothesis for the 16S rRNA gene (*rrs*) based on sequence data. (C) Tree constructed by using the IVS-ORF sequence data with the constraint that WA45 and WA52 clade together. (D) Tree constructed by using the *rrs* sequence data with the constraint that WA52 and NB36 clade together.

artificially constrained trees constructed by using the IVS and *rrs* sequence data were longer than 15 and 43, respectively.

The possibility of confusion among strains was eliminated by analysis of two further sets of the six strains, one repeat shipment from P. Perolat (New Caledonia, France) and a shipment from a separately maintained collection held by David Miller (U.S. Department of Agriculture, Ames, Iowa). MRSP data were acquired for *rrs*, *rfl*, and the IVS for each of these independently acquired strains (23). All of the strains gave restriction data entirely consistent with the strain designations and sequences we report here.

There is little previous evidence for lateral transfer of genes in spirochetes. Nevertheless, the data presented here are most

simply explained by the horizontal transfer of the IVS from an ancestor of *L. noguchii* NB36 to an ancestor of *L. weilli* WA52. The alternative explanation that is now much less likely is that the sporadic distribution of the IVS is due solely to the loss in some lines of the IVS insertion that was present in the ancestor of all these strains.

The possibility that *rrs* was transferred from an ancestor of WA45 to an *L. noguchii* ancestor of WA52 is not supported by the following data. First, arbitrarily primed PCR data show that the whole genomes of WA45 and WA52 are very similar (23; data not shown). Second, there are five MRSPs in *rfl* that distinguish NB36 from WA45/WA52 (Table 1). Thus, any gene conversion event from WA45 to WA46 that involved not only

TABLE 1. Data matrix for polymorphisms detected by restriction digestion^a

MRSP profiles ^b	No. of strains examined	<i>rrs</i>														<i>rml</i>														IVS						
		<i>Bsr</i> I 582	<i>Bbv</i> I 609	<i>Dde</i> I 80	<i>Dde</i> I 121	<i>Dde</i> I 216	<i>Hha</i> I 222	<i>Hha</i> I 640	<i>Hinf</i> I 198	<i>Hinf</i> I 974	<i>Nla</i> IV 165	<i>Taq</i> I 120	<i>Bsp</i> 501 1584	<i>Bsr</i> I 606	<i>Bsr</i> I 1600	<i>Dde</i> I 1702	<i>Hha</i> I 1110	<i>Hinf</i> I 1517	<i>Hph</i> I 626	<i>Hph</i> I 697/788	<i>Mni</i> II 259	<i>Mni</i> II 279	<i>Mni</i> II 377	<i>Mni</i> II 543	<i>Mni</i> II 1411	<i>Mni</i> II 2179	<i>Mni</i> II 2328	<i>Rsa</i> I 499	<i>Rsa</i> I 548		<i>Taq</i> I 1803	<i>Taq</i> I 1615				
<i>L. interrogans</i> A	9	-	+	+	+	-	-	-	+	+	-	+	-	-	+	+	+	+	+	-	-	-	+	-	+	+	+	+	+	+	+	+	+	+	-	
<i>L. interrogans</i> B and <i>L. meyeri</i>	7 ^c	-	+	+	+	-	-	-	-	+	+	-	+	-	-	+	+	+	+	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	-	
<i>L. kirschneri</i> A	3 ^d	-	+	+	-	-	-	-	-	+	+	-	-	-	+	+	+	+	-	-	+	-	+	-	+	+	-	+	+	+	+	+	+	+	-	
<i>L. kirschneri</i> B	2	-	+	+	-	-	-	-	-	+	+	-	-	-	+	+	+	+	-	-	+	-	+	-	+	+	-	+	+	+	+	+	+	+	-	
<i>L. kirschneri</i> C	2	-	+	-	-	-	-	-	-	+	+	-	-	-	+	+	-	+	-	-	+	-	+	-	+	+	-	+	+	+	+	+	+	+	-	
<i>L. borgpetersenii</i> A	8	-	+	+	-	-	+	-	-	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	
<i>L. borgpetersenii</i> B	1	-	+	+	-	-	+	-	-	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	
<i>L. borgpetersenii</i> C	3	-	+	+	-	-	+	-	-	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
<i>L. borgpetersenii</i> D	1	-	+	+	-	-	+	+	-	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
<i>L. weillii</i> A	1	-	+	+	-	+	+	-	-	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
<i>L. weillii</i> B	1	-	+	+	-	+	+	-	-	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
<i>L. noguchii</i> A	1	-	+	+	-	-	-	-	-	+	-	-	+	-	-	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	
<i>L. noguchii</i> B	2	+	+	+	-	-	-	-	+	+	-	+	-	-	+	+	+	+	-	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	-	+
<i>L. santarosai</i> A	5	-	+	+	-	-	-	-	-	+	+	-	+	+	+	+	+	+	-	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
<i>L. santarosai</i> B	2	-	+	-	-	-	-	-	-	+	+	-	+	+	-	-	+	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	+
<i>L. biflexa</i>	1	-	?	-	-	-	-	-	-	-	-	-	?	-	?	?	+	?	+	-	+	+	+	?	?	?	?	?	-	-	-	?	?	-	-	

^a The position of each restriction polymorphism is given for GenBank accession numbers X17547 (*rrs*) and X14249 (*rml*). Polymorphisms that distinguish only the outgroup *L. biflexa* are excluded. + and - indicate the presence and absence of a restriction site, respectively. The length polymorphisms, caused by the replacement of bases 1224 to 1245 in the *rml* gene by the IVS, are scored absent (-) or present (+). Restriction site polymorphisms within these insertions were not scored.

^b The type strain in each species is designated profile A. The strains assigned to each MRSP profile are listed in Table 1 of reference 23.

^c Five *L. interrogans* B strains and two *L. meyeri* strains.

^d Serotype *kambale* strain Kambale has been assigned to *L. kirschneri*.

rrs but also *rml* would need to avoid conversion of the IVS embedded in *rml*. One of the MRSPs is 150 bases from the site of IVS insertion; the most conservative interpretation of this is that the IVS is transferred alone or with very little flanking *rml* sequence.

The indication of horizontal transfer between species of spirochetes that is provided by phylogenetic data leads to some

interesting possibilities. The IVS is transcribed in the sense strand of the 23S rRNA, and the IVS is released after transcription (22), so the ORF should be available for translation. It can be speculated that the putative protein encoded

TABLE 2. Pairwise genetic similarities for the IVS-ORF and *rrs* DNA sequences^a

Strain	Sequence	Pairwise similarity				
		NB36	SA39	WA52	WA45	WB46
SA39	IVS	0.47				
	<i>rrs</i>	0.35				
WA52	IVS	0.76	0.50			
	<i>rrs</i>	0.11	0.22			
WA45	IVS	0.29	0.28	0.34		
	<i>rrs</i>	0.18	0.31	1.00		
WB46	IVS	0.27	0.28	0.33	0.29	
	<i>rrs</i>	0.00	0.03	0.59	0.56	
BD50	IVS	0.03	0.02	0.00	0.04	0.09
	<i>rrs</i>	0.39	0.32	0.52	0.66	0.35

^a Strain designations are in Fig. 1. The number of differences in DNA sequence between each pair of strains was calculated for the IVS-ORF and *rrs* sequence data. The largest pairwise distances were 266 and 21 for the IVS and *rrs* data, respectively. These genetic distances were rescaled for each data set to cover a range from 1 to 0, with 1.0 indicating a perfect match and 0.0 indicating the largest pairwise distance observed. Rescaling allowed the similarities between strains measured for the IVS and *rrs* data to be directly compared and discrepancies between them to be noted more easily. The top number in every pair is the rescaled IVS genetic similarity, and the bottom number is the rescaled *rrs* similarity. The two greatest discrepancies between the pairwise genetic similarities for the IVS and *rrs* data are shown in boldface. These discrepancies involve strains WA52, WA45, and NB36 and are consistent with the indication of lateral transfer from parsimony analysis (Fig. 2).

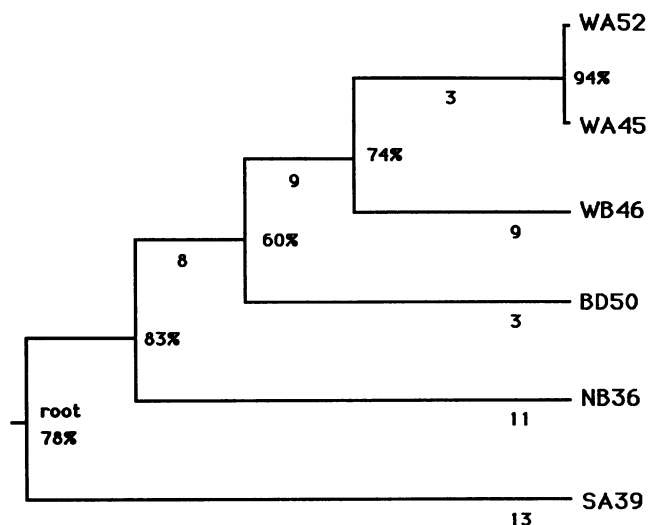


FIG. 3. Phylogenetic hypothesis based on combined data for *rml* and *rrs*. This phylogenetic hypothesis is a rooted tree obtained by using 100 bootstrap replications of the Branch and Bound algorithm (26) on the MRSP data for *rml* in Table 1 and the sequence data for *rrs*. *L. biflexa* was used as an outgroup. Transversions versus transitions were weighted 2:1. Gain versus loss of a restriction site was weighted 4:1. Data from the IVS were excluded. The branch lengths between nodes are indicated below each branch.

by the conserved ORF is involved in some step in the transfer process. However, other eubacterial 23S rRNA genes (*rriI*) that carry small insertions of about 90 to 112 bases that do not carry ORFs have been observed (1, 11, 25). Therefore, the ORF could have some other function. We have recently expressed a protein from this ORF in *Escherichia coli* (22a). It will be interesting to investigate this protein further and determine its mode of action. Ultimately, it may be possible to genetically engineer the IVS and use this sequence as an avenue to insert and express genes in the *Leptospira* genome. With suitable manipulation, the gene might even be transferable to other genera of spirochetes.

ACKNOWLEDGMENTS

We particularly thank Guy Baranton and P. Perolat for supplying strains and for their enthusiastic collaboration, without which this work would not have been possible. We thank John Welsh and N. Yue for critical reading of the manuscript.

D.R. was supported by CIBR set-up funds. This work was supported in part by NIH grants AI34829, NS33377, and HG00456 to M.M. and AI32644 to John Welsh.

REFERENCES

- 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.
- Burke, J. M., C. Breitenberger, J. E. Heckman, B. Dujon, and U. L. RajBhandary. 1984. Cytochrome *b* gene of *Neurospora crassa* mitochondria: partial sequence and location of introns at sites different from those in *Saccharomyces cerevisiae* and *Aspergillus nidulans*. *J. Biol. Chem.* **259**:504–511.
- 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.
- Dujon, B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene* **82**:91–114.
- Dykhuizen, D. E., D. S. Polin, J. J. Dunn, B. Wilske, V. Preac-Mursic, R. J. Dattwyler, and B. J. Luft. 1993. *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc. Natl. Acad. Sci. USA* **90**:10163–10167.
- Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**:521–565.
- 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.
- Fukunaga, M., I. Horie, N. Okuzako, and I. Mifuchi. 1990. Nucleotide sequence of a 16S rRNA gene for *Leptospira interrogans* serovar canicola strain Moulton. *Nucleic Acids Res.* **18**:366.
- Fukunaga, M., and I. Mifuchi. 1989. Unique organization of *Leptospira interrogans* rRNA genes. *J. Bacteriol.* **171**:5763–5767.
- Gilbert, H. J., G. P. Hazlewood, J. I. Laurie, C. G. Orpin, and G. P. Xue. 1992. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. *Mol. Microbiol.* **6**:2065–2072.
- Haraszthy, V. L., G. J. Sunday, L. A. Bobek, T. S. Motley, H. Preus, and J. J. Zambon. 1992. Identification and analysis of the gap region in the 23S ribosomal RNA from *Actinobacillus actinomycetemcomitans*. *J. Dent. Res.* **71**:1561–1568.
- Hinnebusch, J., and A. G. Barbour. 1991. Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. *J. Bacteriol.* **173**:7233–7239.
- Hookey, J. V., J. Bryden, and L. Gatehouse. 1993. The use of 16S rDNA sequence analysis to investigate the phylogeny of Leptospiraceae and related spirochaetes. *J. Gen. Microbiol.* **139**:2585–2590.
- Krawiec, S., and M. Riley. 1990. Organization of the bacterial chromosome. *Microbiol. Rev.* **54**:502–539.
- Lambowitz, A. M. 1989. Infectious introns. *Cell* **56**:323–326.
- Lang, B. F. 1984. The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*: highly homologous introns are inserted at the same position in the otherwise less conserved *coxI* genes in *S. pombe* and *Aspergillus nidulans*. *EMBO J.* **3**:2129–2136.
- Maynard Smith J., C. G. Dowson, and B. G. Spratt. 1991. Localized sex in bacteria. *Nature (London)* **349**:29–31.
- Michel, F., and D. J. Cummings. 1985. Analysis of class I introns in a mitochondrial plasmid associated with senescence of *Podospora anserina* reveals extraordinary resemblance to the *Tetrahymena* ribosomal intron. *Curr. Genet.* **10**:69–79.
- Michel, F., and B. Dujon. 1986. Genetic exchange between bacteriophage T4 and filamentous fungi? *Cell* **46**:323.
- 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.
- Perolat, P. Personal communication.
- 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.
- Ralph, D., and M. McClelland. 1993. A novel class of intervening sequences with conserved open reading frames found in eubacterial 23S rRNA genes. *Proc. Natl. Acad. Sci. USA* **90**:6864–6868.
- Ralph, D., and M. McClelland. Unpublished data.
- Ralph, D., M. McClelland, J. Welsh, G. Baranton, and P. Perolat. 1993. *Leptospira* species categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes. *J. Bacteriol.* **175**:973–981.
- 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.
- Samuels, S. Personal communication.
- Skurnik, M., and P. Toivanen. 1991. Intervening sequences (IVSs) in the 23S ribosomal RNA genes of pathogenic *Yersinia enterocolitica* strains. The IVSs in *Y. enterocolitica* and *Salmonella typhimurium* have a common origin. *Mol. Microbiol.* **5**:585–593.
- Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1. Illinois Natural History Survey, Champaign.
- Syvanen, M. J. 1987. Molecular clocks and evolutionary relationships: possible distortions due to horizontal gene flow. *J. Mol. Evol.* **26**:16–23.
- Ter Huurne, A. A. H. M., M. van Houten, S. Muir, J. G. Kusters, B. A. M. van der Zeijst, and W. Gastra. 1992. Inactivation of a *Serpula (Treponema) hyodysenteriae* hemolysin gene by homologous recombination: importance of this hemolysin in pathogenesis in mice. *FEMS Microbiol.* **92**:109–114.
- Xiong, Y., and T. H. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**:3353–3362.
- 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.
- Zimmer, M., F. Welsch, G. Oraler, and K. Wolf. 1987. Distribution of mitochondrial introns in the species *Schizosaccharomyces pombe* and the origin of the group II intron in the gene encoding apocytochrome *b*. *Curr. Genet.* **12**:329–336.