

# Clinical and Environmental Isolates of *Legionella pneumophila* Serogroup 1 Cannot Be Distinguished by Sequence Analysis of Two Surface Protein Genes and Three Housekeeping Genes

Helena Aurell,<sup>1</sup> Pierre Farge,<sup>1</sup> H  l  ne Meugnier,<sup>1</sup> Manolo Gouy,<sup>2</sup> Fran  oise Forey,<sup>1</sup>  
G  rard Lina,<sup>1</sup> Fran  ois Vandenesch,<sup>1</sup> Jerome Etienne,<sup>1</sup>  
and Sophie Jarraud<sup>1\*</sup>

Centre National de R  f  rence des *Legionella*, INSERM E-0230, Laboratoire de Bact  riologie, Facult   de M  decine Laennec IFR 62, Lyon,<sup>1</sup> and Laboratoire de Biom  trie et Biologie Evolutive, UMR CNRS 5558, Universit   Claude Bernard, Villeurbanne,<sup>2</sup> France

Received 10 March 2004/Accepted 26 August 2004

**We used gene sequencing to determine whether clinical (sporadic, epidemic, and endemic) and environmental isolates of *Legionella pneumophila* serogroup (sg) 1 belong to specific lineages. A total of 178 clinical and environmental *L. pneumophila* sg 1 isolates, defined by pulsed-field gel electrophoresis and epidemiological data as sporadic, epidemic, or endemic, were analyzed for polymorphisms in five gene fragments. The fragments belonged to three housekeeping genes (coding for aconitase [*acn*], aspartate- $\beta$ -semialdehyde dehydrogenase [*asd*], and RNA polymerase  $\beta$  subunit [*rpoB*]) and two surface protein genes (coding for the macrophage infectivity potentiator [*mip*] and the major outer membrane protein [*mompS*]). The phylogenetic tree inferred from sequence polymorphisms of the five genes identified two large clusters, one consisting of 133 poorly differentiated strains and containing two smaller clusters (10 and 2 strains) unrelated to each other and the other consisting of 42 strains. Clinical and environmental isolates could not be distinguished on this basis, and no link between genetic background and epidemiological type was found, suggesting that other factors are responsible for differences in pathogenicity.**

*Legionella*, an environmental bacterium, is a common cause of hospital- and community-acquired pneumonia. At least 47 species of *Legionella* have been described, of which 21 have been associated with human infection (31). There are more environmental species (5) than pathogenic species. One species, *Legionella pneumophila*, is responsible for about 90% of cases of legionellosis, and serogroup (sg) 1 accounts for about 84% of cases (52).

Intraspecific-level analytical methods are necessary to discriminate among clinical *L. pneumophila* sg 1 isolates and to identify environmental sources (26, 34, 41, 47). These methods include pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism, arbitrarily primed PCR, multilocus variable-number repeat analysis, and sequence-based typing (11, 17, 33). PFGE typing of clinical *L. pneumophila* isolates reveals a large number of clones. In combination with epidemiological data, PFGE has identified several epidemiological groups of clinical strains (1): (i) sporadic strains are defined as isolates that have a unique PFGE profile and are epidemiologically unrelated, (ii) epidemic strains share identical PFGE profiles and are responsible for several cases of legionellosis at the same location over a limited period of time, and (iii) endemic isolates share identical PFGE pulsotypes but

cause several epidemiologically unrelated cases (1, 25). Some *L. pneumophila* sg 1 strains that colonize water sources do not appear to cause legionellosis in exposed patients (34, 35), while others, such as epidemic strains and the endemic Paris strain, are more frequently associated with clinical cases.

MLST and sequence-based typing have been used to study long-term evolutionary relationships among strains of various bacterial genera (6, 12), notably separating virulent and less virulent clones of *Streptococcus pneumoniae* sg 6 (39). Sequence polymorphism analysis has also been applied to *Legionella* isolates. Ko et al. divided 79 Korean isolates, mostly of environmental origin, into six clonal populations by analyzing the *rpoB* and *dotA* genes (20), while Bumbaugh et al. separated a collection of 17 clinical and environmental isolates into two clusters by analyzing the *dotA* and *mip* genes (2).

Here, we used gene sequencing to study the relationship between clinical and environmental *L. pneumophila* sg 1 isolates and to determine whether sporadic, epidemic, and endemic isolates belong to specific lineages.

## MATERIALS AND METHODS

**Bacterial isolates.** Legionellosis is a notifiable disease in France. The French National Reference Centre for *Legionella* (FNRLC) collects all clinical *Legionella* isolates and types them by means of conventional PFGE with the SfiI restriction enzyme (25). PFGE banding patterns differing by less than two bands define a specific profile. These PFGE profiles, together with clinical data, are used to identify the sporadic, epidemic, or endemic nature of the isolates (1).

For this study, we selected 134 clinical *L. pneumophila* sg 1 isolates (sporadic, epidemic, or endemic) collected throughout France between 1998 and 2002, together with 15 clinical *L. pneumophila* sg 1 isolates from the European Work-

\* Corresponding author. Mailing address: Centre National de R  f  rence des *Legionella*, INSERM E-0230, Laboratoire de Bact  riologie, Facult   de M  decine Laennec IFR 62, 7 rue Guillaume Paradin, 69372 Lyon cedex 08, France. Phone: 33 4 78 77 86 57. Fax: 33 4 78 77 86 58. E-mail: sophie.jarraud@univ-lyon1.fr.

TABLE 1. Primers used for PCR amplification of *L. pneumophila* sg 1 gene fragments

Primer	Target	Oligonucleotide sequence (5'-3')
acn 1	<i>acn</i>	CAC CTC TAA TCC AAG TGT ACT AAT GG
acn 2	<i>acn</i>	TGT TTG AAT GGC TTG CCA ATG AGC
rpoB 1	<i>rpoB</i>	TGG ATT TGA GGT CCG TGA CG
rpoB 2	<i>rpoB</i>	CTC CAC CTC TCT TCG CAA C
asd 1	<i>asd</i>	CCC TAA TTG CTC TAC CAT TCA GAT G
asd 2	<i>asd</i>	CGA ATG TTA TCT GCG ACT ATC CAC
mip 1	<i>mip</i>	GCT GCA ACC GAT GCC AC
mip 2	<i>mip</i>	CAT ATG CAA GAC CTG AGG GAA C
mompS 1	<i>mompS</i>	CAA CTA CAA CAA CAA GTG GGA TGC
mompS 2	<i>mompS</i>	TGG ATA AAT TAT CCA GCC GGA CTT C

ing Group on Legionella Infections (EWGLI) collection (10) and 28 environmental isolates collected in 2003 from hospital water systems in France (details on the isolates are available at <http://www.lyon.inserm.fr/strain-legio/table.pdf>). The reference *L. pneumophila* sg 1 Philadelphia strain (ATCC 33152) was used as a reference for sequence analysis. A clinical *L. pneumophila* sg 7 isolate was used as an outgroup external control to root the phylogenetic tree.

**French clinical isolates.** Fifty-four sporadic isolates were collected throughout France. Seventeen epidemic isolates were recovered during five recent outbreaks. They consisted of (i) one strain (Montparnasse) from an outbreak in Paris in 1999, (ii) seven isolates from an outbreak in Rennes in 2000 (two strains, designated Rennes a and Rennes b, differed by one PFGE band), (iii) four isolates (strain Lyon) from a 2001 outbreak in Lyon, (iv) three isolates from a nosocomial outbreak in Sarlat in 2002, and (v) two isolates from a nosocomial outbreak in Besançon in 2002. Forty-nine endemic isolates designated Paris (CIP 107-629-T) (1) were isolated from various parts of France. Twelve endemic isolates, designated Axa a and Axa b, had PFGE patterns differing by only one band and were isolated in the Paris area and two other French towns. Two endemic isolates with identical PFGE profiles were isolated in two cases of nosocomial legionellosis that occurred 4 years apart in the same Bordeaux hospital.

**EWGLI isolates.** Fifteen epidemiologically unrelated European clinical *L. pneumophila* sg 1 isolates were obtained from the reference EWGLI collection (10). The PFGE pattern of seven isolates differed from that of the endemic Paris strain by a single band.

**French environmental isolates.** Twenty-eight environmental isolates were collected from hospital water distribution systems that had never been linked to cases of human infection. Six isolates had the Paris PFGE profile. The remaining PFGE profiles were absent from the FNRLC database of clinical PFGE profiles.

**Gene selection.** As the aims of this study were both to identify virulent lineages and to investigate evolutionary relationships among isolates, we analyzed both selectively neutral housekeeping genes and virulence-related genes. The three housekeeping genes encoded aconitase (*acn*), aspartate- $\beta$ -semialdehyde dehydrogenase (*asd*), and the RNA polymerase  $\beta$  subunit (*rpoB*); the two surface protein genes coded for macrophage infectivity potentiator (*mip*) (3) and outer membrane surface protein (*mompS*) (24). Comparison with sequence data from the *Legionella* genome project (<http://genome3.cpmc.columbia.edu/~legion/>) showed that the five genes were located in different parts of the chromosome. The DNA sequences of the genes and the deduced amino acid sequences of their protein products were aligned to identify the most variable regions by using the GenBank database (<http://www.ncbi.nlm.nih.gov/GenBank/>). Primers were manually designed from the sequences of highly conserved regions that flanked the most variable regions (Table 1). Each primer pair amplified an internal fragment of the gene (about 450 bp) and permitted accurate sequencing of fragments of the same size on both strands.

**Amplification and nucleotide sequencing.** DNA was retrieved by standard phenol-chloroform extraction (38). PCR was performed under standard conditions with 32 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s. An annealing temperature of 62°C was required for the *asd* gene fragment. PCR products were purified with a PCR 96 Cleanup kit (Millipore, Bedford, Mass.) and sequenced from both ends with a MegaBACE/1000 automated sequencer (Amersham Pharmacia Biotech, Amersham, England) and a DYEamic ET Dye Terminator kit (Amersham Pharmacia Biotech).

**Nucleotide sequence variations and phylogenetic analysis.** For each *Legionella* isolate, the sequences of the five loci were determined, and each locus was assigned an allelic number. Alleles of the reference *L. pneumophila* Philadelphia

strain (ATCC 33152) were numbered "1." Each single-nucleotide variation was designated by a new allelic number. The alleles at all five loci defined the allelic profile, which corresponded to the sequence type (ST). The STs were assigned arbitrary numbers in order of description. For each polymorphic site, synonymous and nonsynonymous base pair substitutions were identified, and the  $d_s/d_N$  ratio of synonymous to nonsynonymous base pair substitutions was calculated by using the SNAP program (available at <http://www.hiv.lanl.gov/content/hiv-db/SNAP>) based on the method of Nei and Gojobori (22, 30). The gene sequences were aligned with the multiple-alignment program ClustalX (49), and amino acid sequences were deduced with Gene Jockey software (Biosoft, Cambridge, England). A phylogenetic tree was computed with ClustalX by using the neighbor-joining method applied to Kimura's two-parameter distances between pairs of DNA sequences. The tree was drawn with the NJplot program, and bootstrap confidence levels were determined by randomly resampling the sequence data 1,000 times (32). The index of association ( $I_A$ ) was calculated as described previously by Smith et al. (45). The five individual gene data sets of the 178 *L. pneumophila* sg 1 isolates were compared statistically for incongruence by using the incongruence length differences (ILD) test (8). The ILD test was performed by using the PAUP\* package with 100 random bipartitions of the data for each test (48).

**Discrimination index.** Simpson's index of diversity (IOD), defined as the probability with which two isolates chosen at random will be of a different type, was used to estimate the discriminatory capacity of the typing method (16, 44). The following unrelated strains were included: the reference *L. pneumophila* sg 1 Philadelphia strain ( $n = 1$ ), all EWGLI isolates, all sporadic strains ( $n = 54$ ), one randomly chosen epidemic strain per outbreak ( $n = 7$ ), all epidemiologically unrelated endemic Paris strains ( $n = 25$ ), Axa strains (7 Axa a and 3 Axa b strains), and Bordeaux strains, defined as one isolate per town or, when several isolates were available from a given town, one community-acquired isolate and all epidemiologically unrelated hospital-acquired isolates. Confidence intervals for discriminatory indices were calculated as previously described (13).

**Nucleotide sequence accession numbers.** The sequences of all alleles of *acn* (17), *rpoB* (13), *asd* (15), *mip* (15), and *mompS* (17) have been deposited in GenBank (National Center for Biotechnology Information, Bethesda, Md.) under the accession numbers given in Fig. 1.

## RESULTS

**Sequence diversity.** Nucleotide sequences were obtained for all five gene fragments for all *Legionella* isolates. The size of the sequenced fragments was between 402 and 485 bp, and the number of alleles ranged from 14 to 17 (Table 2). The number of polymorphic sites ranged from 14 to 35, and the minimal and maximal numbers were observed for the *mip* and *mompS* genes, respectively. There were no preferential regions of polymorphic sites on the gene fragments, except on the *mip* gene (polymorphic sites started at nucleotide number 112). The polymorphisms for the outgroup (ST20) and for three environmental isolates (ST60, ST61, and ST46) were too numerous to be shown in Fig. 1, except for *mip* gene, which contained seven identical polymorphism sites for the four STs. The percentage of polymorphism sites for specific alleles of these four STs ranged from 5 to 10% for the four other genes (except for allele 10 of *asd* and allele 17 of *acn*, which contained less than 2% polymorphism). The  $d_s/d_N$  values were  $>1$  for four genes, indicating that the number of synonymous substitutions (selectively neutral) was larger than the number of nonsynonymous substitutions (subject to positive selection) (Table 2). The  $d_s/d_N$  ratio could not be calculated for the *mip* gene because the proportion of synonymous or nonsynonymous substitutions was  $\geq 0.75$  in some alleles, and the Jukes-Cantor transformation could not be applied (Table 2) (<http://www.hiv.lanl.gov/content/hiv-db/SNAP>).

**Comparison of PFGE and gene sequence analysis.** Sequence-based strain identification was generally in keeping with PFGE-based typing results. A total of 69 STs were ob-

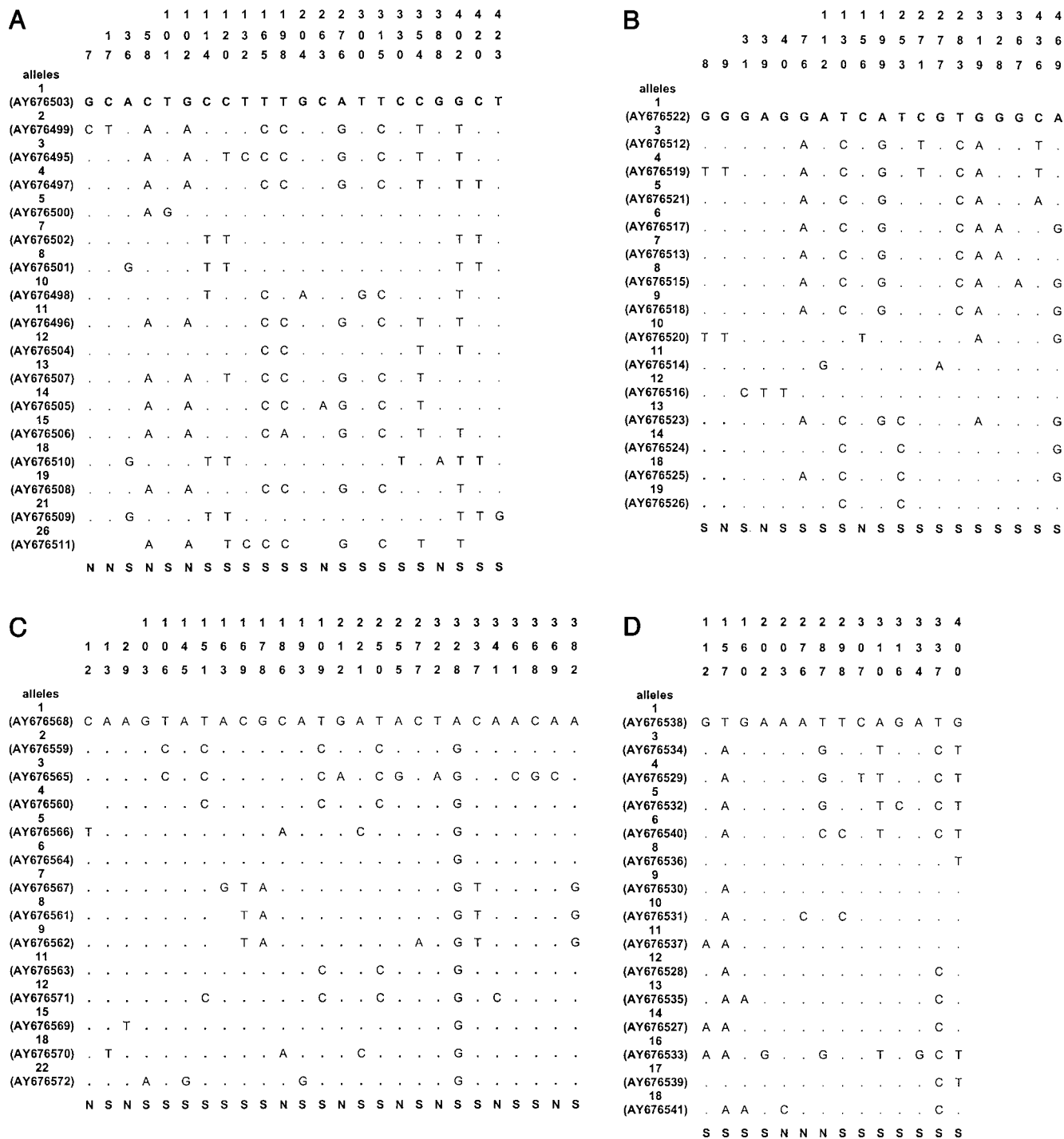


FIG. 1. Polymorphic sites in the five gene fragments. The nucleotides present at each variable site are shown for strain ATCC 33152 (allele 1). For the other alleles, only sites that differ among the 175 clinical and environmental *L. pneumophila* sg 1 isolates are shown (ST20, ST46, ST60, and ST61 are not included); sites that are the same as those in allele 1 are shown by periods. The sites are numbered above in vertical format based on the nucleotide numbering of *L. pneumophila* strain ATCC 33152. The polymorphisms that are synonymous (S) and nonsynonymous (N) are shown below the panels. GenBank accession numbers of alleles are in parentheses. A, *acn*; B, *asd*; C, *tpoB*; D, *mip*; E, *mompS*.

tained for the 179 strains (93 PFGE profiles) (<http://www.lyon.inserm.fr/strain-legio/table.pdf>). The 49 clinical Paris isolates and the 7 EWGLI Paris isolates were ST2. Similarly, almost all the epidemic isolates had a single ST during each outbreak, in

keeping with PFGE typing results. However, the seven isolates from the Rennes outbreak had four different STs (ST24, ST28, ST32, and ST33), while the PFGE patterns of these isolates differed by a single band (data not shown). Besançon clinical



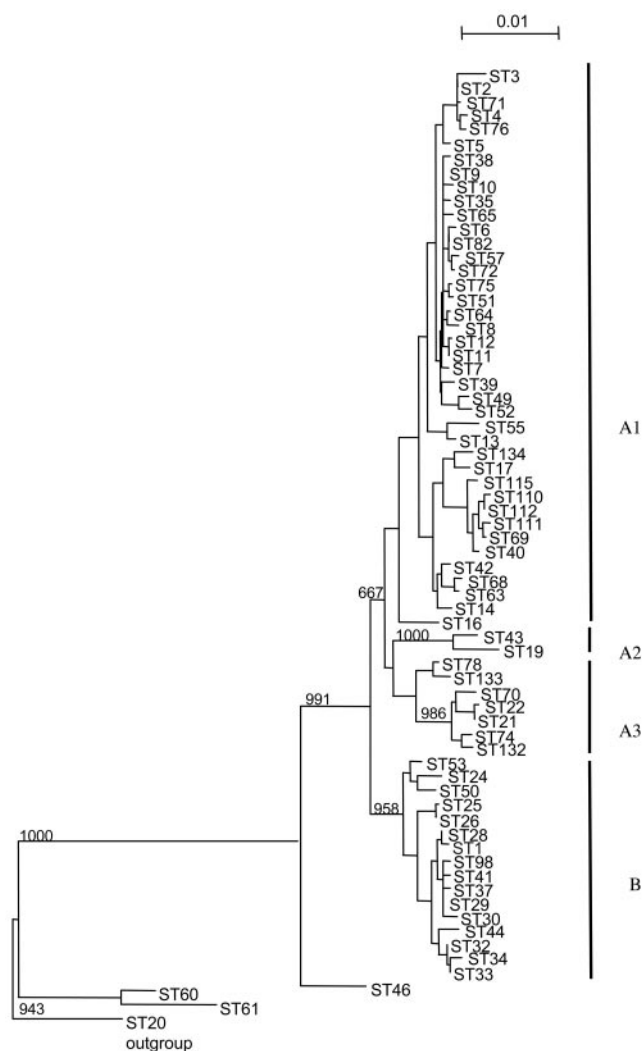


FIG. 2. Dendrogram of genetic relationships among 178 *L. pneumophila* sg1 isolates based on five gene fragments. The tree was constructed by the neighbor-joining method with 1,000 replications using *L. pneumophila* sg 7 (ST20) as an outgroup. Linkage distance is indicated by a scale bar on the top. Clusters are indicated by thick vertical bars (A1, A2, A3, and B).

**Discrimination of clinical *L. pneumophila* isolates.** The IOD was calculated as the number of isolates per ST for all five genes, using the isolates defined in Materials and Methods. The overall IOD for the clinical isolates was 90.9% (95% confidence interval, 86.4 to 95.5%). The IOD of PFGE typing for the same isolates was 91.6% (95% confidence interval, 87.0 to 96.3%). When only one isolate with the Paris PFGE profile was included, the IOD was 97.9% (95% confidence interval, 96.5 to 99.4%) for gene sequence analysis and 99.2% (95% confidence interval, 98.3 to 100%) for PFGE.

**Phylogenetic analysis.** The phylogenetic tree constructed from the sequences of the five gene fragments identified two major clusters (A and B [Fig. 2]); 150 clinical isolates and 25 environmental isolates were contained in the two clusters. A large subcluster (cluster A1) was formed by 120 strains which were poorly differentiated and whose order of divergence was

not significantly resolved. This subcluster comprised 34 French sporadic strains, all 49 French clinical ST2 isolates, the 7 European EWGLI strains with ST2, and 4 EWGLI strains with other STs. This subcluster also included nine isolates originating from the three outbreaks in Lyon, Sarlat, and Besançon. Eleven environmental isolates with ST2 and six environmental isolates with other STs were also included in cluster A1. Two smaller subclusters (A2 and A3) were closely related to cluster A1. Subcluster A3 (11 isolates) comprised eight sporadic isolates, two isolates from the Bordeaux outbreak, and one EWGLI isolate. Subcluster A2 comprised only two sporadic clinical isolates. Cluster B corresponded to a subset of 42 isolates equally distributed among the different epidemiological types as follows: 11 sporadic strains, the 12 Axa a and Axa b endemic strains, 8 epidemic isolates from the Rennes and Montparnasse outbreaks, 2 EWGLI strains, and the reference Philadelphia strain, ATCC 33152. In addition, eight environmental isolates of ST1, ST32, ST33, and ST34 were included in cluster B. Three environmental strains and the *L. pneumophila* sg 7 outgroup strain segregated separately from clusters A and B.

ILD tests were conducted to determine whether *asd*, *acn*, *rpoB*, *mip*, and *mompS* data sets were coalescent together or not (8). The results of the ILD tests for 10 combined data sets were significantly incongruent ( $P < 0.01$ ).

The level of linkage disequilibrium between alleles was assessed in terms of the  $I_A$  (45). An  $I_A$  of 2.66 ( $\pm 0.101$ ) was obtained when all 179 isolates were included. When one isolate from each of the 69 STs was included, the  $I_A$  was 1.14 ( $\pm 0.184$ ). The  $I_A$  for cluster A (A1, A2, and A3) and cluster B were calculated separately. The  $I_A$  was significantly greater than zero for cluster A ( $3.10 \pm 0.112$ ), and an  $I_A$  of 0.01 ( $\pm 0.211$ ) was obtained for cluster B. These results suggested that the population structure was clonal, except for cluster B, which seemed to show a sexual population structure.

## DISCUSSION

Legionnaires' disease is generally sporadic, but large outbreaks (4, 9) and endemicity (1, 26, 29) are regularly reported. It is unclear whether the particular epidemiological characteristics of clinical cases are due to strain variability in terms of environmental adaptation, virulence factors, or other mechanisms. The sporadic, epidemic, or endemic nature of *Legionella* isolates is determined on the basis of their PFGE patterns and epidemiological data (1). Environmental isolates with typing profiles that have never been associated with human infection have also been described previously (34, 35, 50). It is unclear whether clinical strains are more pathogenic than environmental strains. In this study, we sought a relationship between the genetic background and clinical or environmental origin and also investigated the possible relationship between genetic background and the epidemiological type of clinical isolates by analyzing the sequences of five genes (*rpoB*, *asd*, *acn*, *mompS*, and *mip*) in a large collection of isolates previously characterized by PFGE and epidemiological investigations. The results of gene sequence analysis were generally in keeping with PFGE results: isolates with identical PFGE profiles generally had the same ST, and isolates with unique PFGE patterns were generally confined to unique STs. However, several exceptions

were observed: strains with identical STs sometimes had different PFGE patterns, and vice versa. The discriminatory power of sequence polymorphisms in the *rpoB*, *asd*, *acn*, *mompS*, and *mip* genes (IOD, 97.9%) was close to that of PFGE (IOD, 99.4%) when calculated by using a single Paris strain (1). However, the IOD of gene sequence polymorphisms fell to 90.9% and that of PFGE fell to 91.6% when we included 25 French and 7 European isolates of the endemic Paris clone that had the same ST (ST2). We chose to include a large number of Paris isolates in the study because this strain predominates in France, accounting for 12% of cases of legionellosis (1), and PFGE typing does not allow for discrimination. A gene sequence-based typing method was recently developed by Gaia et al. (11) using the partial sequences of three genes under probable selective pressure (*mompS*, *proA*, and *flaA*) and was used to type *L. pneumophila* sg 1 isolates. When combined, these three genes gave an IOD of 92.4%, which is sufficiently correct for interpretation (11, 16). We obtained a similar IOD when a large number of endemic Paris PFGE profile isolates were included, confirming that our choice of genes permitted satisfactory discrimination. When only one Paris isolate was included, the IOD increased to 97.8%, indicating the major impact of endemic strains on epidemiological studies. Indeed, our gene sequence analysis identified several new endemic strains such as ST9 (six strains with different PFGE profiles).

In contrast to the objectives of Gaia et al., who sought to develop a typing method, our objectives were to analyze potential relationships between strains of different origins and to identify virulence-linked lineages. The extent of recombination events in the population is important when interpreting a potential genetic relationship between bacterial strains (45, 46). We therefore assessed the level of recombination in our *L. pneumophila* sg 1 population by using the  $I_A$  (45). The  $I_A$  results should be treated with caution, as only five loci were analyzed, but the population appeared to be clonal, as the  $I_A$  value was significantly higher than zero, indicating that recombination events were rare in the clinical and environmental strain population. Evidence of clonality persisted when a single isolate from each ST was included in the analysis, allowing us to rule out an epidemic population structure. These data confirm those of previous studies (2, 20, 43, 45). Surprisingly, when the  $I_A$  was calculated separately for each cluster (A and B), cluster A was clonal, while cluster B had a sexual structure. Cluster B comprised various clinical and environmental isolates of different geographical origins. Interestingly, this cluster included epidemic Rennes a and b isolates with two closely related PFGE profiles as well as several epidemic Axa a and b isolates, also with two closely related PFGE profiles.

The construction of the phylogenetic trees was justified by the  $I_A$  values significantly greater than zero (45). The trees derived from each of the five genes were not fully congruent with the global tree (data not shown), and the ILD test revealed significant differences in the phylogeny of the five genes, indicating that recombination events might have occurred. All of this suggest that some recombination occurred during the evolution of the compared strains (as assessed by the ILD test, which is a very sensitive test) but that a majority of strains had a clonal evolution (reflected by the positive  $I_A$  test), except for cluster B, whose features indicate a high level of recombina-

tion. Recent sequence-based studies focusing on limited numbers of loci (*dot/icm*, *rpoB*, and *mip*) suggest that *L. pneumophila* has an intermediate level of clonality but that horizontal transfer events may occur (2, 15, 18–20).

We then investigated whether the clinical or environmental origin of the isolates was linked to a specific genetic background. The phylogenetic tree inferred from sequence polymorphisms of the five genes identified two groups: a large cluster (A) of 133 poorly differentiated strains, with two smaller clusters (A2 and A3) that were unrelated to each other, and a second large cluster (B) containing 42 strains. The environmental isolates were found in both large groups but not in the small subclusters A2 and A3. Thus, gene sequence analysis failed to segregate either clinical or environmental strains. These results confirm the results of the Bumbaugh et al. study of the *dotA* and *mip* genes in a smaller collection of isolates, in which the two observed clusters were not linked to the geographical location or source (2).

The link between epidemic isolates and a specific genetic background has been discussed for several bacterial species (7, 28). For example, despite the broad diversity of natural populations of *L. monocytogenes*, only two clones, identified by their genetic backgrounds, have been responsible for most major outbreaks in recent decades in Europe and North America (40). PFGE analysis grouped clinical and environmental strains of *Salmonella* spp. into distinct clusters, and DNA sequence analysis of the *manB* gene also separated clinical and environmental strains (23).

We examined whether the epidemiological type of *L. pneumophila* sg 1 strains was linked to a specific genetic background. The epidemic, endemic, or sporadic nature of the isolates could not be distinguished on the basis of polymorphisms in the five genes. These findings suggest that discrete markers not restricted to a subset of closely related *L. pneumophila* sg 1 strains might be capable of explaining the pathogenic nature of epidemic and endemic strains. Such markers may correspond to virulence-associated factors or to elements providing increased environmental adaptation. Putative genetic candidates for distinguishing virulent *L. pneumophila* clones from nonvirulent clones, such as pathogenicity islands (e.g., the *lvh* region encoding a type IV secretion system) (42) and an unstable 30-kb genetic element (27) which is excised from the chromosome in nonvirulent strains, have been described in the literature. Adaptation to the intracellular environment is a key determinant of virulence acquisition by *Legionella* (14). Thus, genes involved in the initial steps of cell invasion, such as the *mompS* and *mip* genes, were good candidates to explain better adaptation to the human host. The 25-kDa Momp protein of *L. pneumophila* is a membrane protein involved in host-cell recognition, which plays a major role in the virulence of the organism for the chicken embryo (24). The *mip* gene encodes a homodimeric cell surface protein required for early intracellular survival both in protozoa and in human macrophages and monocytes (3, 21, 37, 51). Analysis of the DNA sequences showed that the allelic diversity was similar and that the  $d_S/d_N$  ratio was  $>1$  for the three housekeeping genes (*rpoB*, *asd*, and *acn*) and *mompS*, suggesting that it is involved in some housekeeping function. The *mip* gene displayed a similar total number of substitutions, although these were preferentially situated in a particular part of the fragment

coding for the C-terminal FK506-binding protein-homologous domain (21), whose exact role in vivo remains to be elucidated. However, it is unclear whether this gene is involved directly in pathogenesis or is also involved in biosynthetic or metabolic processes, since it is present in all strains regardless of their virulence (36).

In conclusion, we sequenced five genes in a panel of *L. pneumophila* sg 1 clinical and environmental isolates in order to determine whether sporadic, epidemic, and endemic isolates of *L. pneumophila* sg 1 and environmental isolates belong to specific lineages. We found that clinical and environmental isolates could not be distinguished on this basis. Likewise, no association between genetic background and epidemiological type was found, suggesting that other factors might be responsible for differences in pathogenicity.

#### ACKNOWLEDGMENT

We thank David Young for editing the manuscript.

#### REFERENCES

- Aurell, H., J. Etienne, F. Forey, M. Reyrolle, P. Girardo, P. Farge, B. Decludt, C. Campese, F. Vandenesch, and S. Jarraud. 2003. *Legionella pneumophila* serogroup 1 strain Paris: distribution throughout France. *J. Clin. Microbiol.* **41**:3320–3322.
- Bumbaugh, A. C., E. A. McGraw, K. L. Page, R. K. Selander, and T. S. Whittam. 2002. Sequence polymorphism of *dotA* and *mip* alleles mediating invasion and intracellular replication of *Legionella pneumophila*. *Curr. Microbiol.* **44**:314–322.
- Cianciotto, N. P., J. M. Bangsberg, B. I. Eisenstein, and N. C. Engleberg. 1990. Identification of *mip*-like genes in the genus *Legionella*. *Infect. Immun.* **58**:2912–2918.
- Den Boer, J. W., E. P. Yzerman, J. Schellekens, K. D. Lettinga, H. C. Boshuizen, J. E. Van Steenbergen, A. Bosman, S. Van den Hof, H. A. Van Vliet, M. F. Peeters, R. J. Van Ketel, P. Speelman, J. L. Kool, and M. A. Conyn-Van Spaendonck. 2002. A large outbreak of Legionnaires' disease at a flower show, The Netherlands, 1999. *Emerg. Infect. Dis.* **8**:37–43.
- Doleans, A., H. Aurell, M. Reyrolle, G. Lina, J. Freney, F. Vandenesch, J. Etienne, and S. Jarraud. 2004. Clinical and environmental distributions of *Legionella* strains in France are different. *J. Clin. Microbiol.* **42**:458–460.
- Enright, M. C., and B. G. Spratt. 1999. Multilocus sequence typing. *Trends Microbiol.* **7**:482–487.
- Enright, M. C., and B. G. Spratt. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* **144**:3049–3060.
- Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1997. Testing significance of incongruence. *Cladistics* **10**:315–319.
- Fernandez, J. A., P. Lopez, D. Orozco, and J. Merino. 2002. Clinical study of an outbreak of Legionnaire's disease in Alcoy, southeastern Spain. *Eur. J. Clin. Microbiol. Infect. Dis.* **21**:729–735.
- Fry, N. K., S. Alexiou-Daniel, J. M. Bangsberg, S. Bernander, M. C. Pastoris, J. Etienne, B. Forsblom, V. Gaia, J. H. Helbig, D. Lindsay, P. C. Lück, C. Pelaz, S. A. Uldum, and T. G. Harrison. 1999. A multicenter evaluation of genotypic methods for the epidemiologic typing of *Legionella pneumophila* serogroup 1: results of a pan-European study. *Clin. Microbiol. Infect.* **5**:462–477.
- Gaia, V., N. K. Fry, T. G. Harrison, and R. Peduzzi. 2003. Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. *J. Clin. Microbiol.* **41**:2932–2939.
- Grundmann, H., S. Hori, M. C. Enright, C. Webster, A. Tami, E. J. Feil, and T. Pitt. 2002. Determining the genetic structure of the natural population of *Staphylococcus aureus*: a comparison of multilocus sequence typing with pulsed-field gel electrophoresis, randomly amplified polymorphic DNA analysis, and phage typing. *J. Clin. Microbiol.* **40**:4544–4546.
- Grundmann, H., S. Hori, and G. Tanner. 2001. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J. Clin. Microbiol.* **39**:4190–4192.
- Harb, O. S., L.-Y. Gao, and Y. A. Kwaik. 2000. From protozoa to mammalian cells: a new paradigm in the life cycle of intracellular bacterial pathogens. *Environ. Microbiol.* **2**:251–265.
- Heuner, K., M. Steinert, R. Marre, and J. Hacker. 2002. Genomic structure and evolution of *Legionella* species. *Curr. Top. Microbiol. Immunol.* **264**:61–78.
- Hunter, P. R., and M. A. Gaston. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**:2465–2466.
- 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.
- Ko, K. S., S. K. Hong, H. K. Lee, M. Y. Park, and Y. H. Kook. 2003. Mosaic structure of pathogenicity islands in *Legionella pneumophila*. *J. Mol. Evol.* **57**:63–72.
- Ko, K. S., S. K. Hong, H. K. Lee, M. Y. Park, and Y. H. Kook. 2003. Molecular evolution of the *dotA* gene in *Legionella pneumophila*. *J. Bacteriol.* **185**:6269–6277.
- Ko, K. S., H. K. Lee, M. Y. Park, M. S. Park, K. H. Lee, S. Y. Woo, Y. J. Yun, and Y. H. Kook. 2002. Population genetic structure of *Legionella pneumophila* inferred from RNA polymerase gene (*rpoB*) and *DotA* gene (*dotA*) sequences. *J. Bacteriol.* **184**:2123–2130.
- Kohler, R., J. Fanghanel, B. König, E. Luneberg, M. Frosch, J. U. Rahfeld, R. Hilgenfeld, G. Fischer, J. Hacker, and M. Steinert. 2003. Biochemical and functional analyses of the Mip protein: influence of the N-terminal half and of peptidylprolyl isomerase activity on the virulence of *Legionella pneumophila*. *Infect. Immun.* **71**:4389–4397.
- Korber, B. 2000. Signature and sequence variation analysis., p. 55–72. In A. G. Rodrigo and G. H. Learn (ed.), *Computational analysis of HIV molecular sequences*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Kotetishvili, M., O. C. Stine, A. Kreger, J. G. Morris, Jr., and A. Sulakvelidze. 2002. Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. *J. Clin. Microbiol.* **40**:1626–1635.
- Krinos, C., A. S. High, and F. G. Rodgers. 1999. Role of the 25 kDa major outer membrane protein of *Legionella pneumophila* in attachment to U-937 cells and its potential as a virulence factor for chick embryos. *J. Appl. Microbiol.* **86**:237–244.
- Lawrence, C., M. Reyrolle, S. Dubrou, F. Forey, B. Decludt, C. Goulvestre, P. Matsiota-Bernard, J. Etienne, and C. Nauciel. 1999. Single clonal origin of a high proportion of *Legionella pneumophila* serogroup 1 isolates from patients and the environment in the area of Paris, France, over a 10-year period. *J. Clin. Microbiol.* **37**:2652–2655.
- Lück, P. C., J. H. Helbig, V. Günter, M. Assman, R. Blau, H. Koch, and M. Klepp. 1994. Epidemiologic investigation by macrorestriction analysis and by using monoclonal antibodies of nosocomial pneumonia caused by *Legionella pneumophila* serogroup 10. *J. Clin. Microbiol.* **32**:2692–2697.
- Luneberg, E., N. Zetzmann, D. Alber, Y. A. Knirel, O. Kooistra, U. Zahringer, and M. Frosch. 2000. Cloning and functional characterization of a 30 kb gene locus required for lipopolysaccharide biosynthesis in *Legionella pneumophila*. *Int. J. Med. Microbiol.* **290**:37–49.
- Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
- Montanaro-Punzengruber, J. C., L. Hicks, W. Meyer, and G. L. Gilbert. 1999. Australian isolates of *Legionella longbeachae* are not a clonal population. *J. Clin. Microbiol.* **37**:3249–3254.
- Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
- Park, M. Y., K. S. Ko, H. K. Lee, M. S. Park, and Y. H. Kook. 2003. *Legionella busanensis* sp. nov., isolated from cooling tower water in Korea. *Int. J. Syst. Evol. Microbiol.* **53**:77–80.
- Perriere, G., and M. Gouy. 1996. WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**:364–369.
- Pourcel, C., Y. Vidgop, F. Ramière, G. Vergnaud, and C. Tram. 2003. Characterization of a tandem repeat polymorphism in *Legionella pneumophila* and its use for genotyping. *J. Clin. Microbiol.* **41**:1819–1826.
- Pruckler, J. M., L. A. Mermel, R. F. Benson, C. Giorgio, P. K. Cassidy, R. F. Breiman, C. G. Whitney, and B. S. Fields. 1995. Comparison of *Legionella pneumophila* isolates by arbitrarily primed PCR and pulsed-field gel electrophoresis: analysis from seven epidemic investigations. *J. Clin. Microbiol.* **33**:2872–2875.
- Rangel-Frausto, M. S., P. Rhomberg, R. J. Hollis, M. A. Pfaller, R. P. Wenzel, C. M. Helms, and L. A. Herwaldt. 1999. Persistence of *Legionella pneumophila* in a hospital's water system: a 13-year survey. *Infect. Control Hosp. Epidemiol.* **20**:793–797.
- Ratcliff, R. M., J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 1998. Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. *J. Clin. Microbiol.* **36**:1560–1567.
- Riboldi-Tunnicliffe, A., B. König, S. Jessen, M. S. Weiss, J. Rahfeld, J. Hacker, G. Fischer, and R. Hilgenfeld. 2001. Crystal structure of Mip, a prolyl isomerase from *Legionella pneumophila*. *Nat. Struct. Biol.* **8**:779–783.
- Riffard, S., F. Lo Presti, F. Vandenesch, F. Forey, M. Reyrolle, and J.

- Etienne.** 1998. Comparative analysis of infrequent-restriction-site PCR and pulsed-field gel electrophoresis for epidemiological typing of *Legionella pneumophila* serogroup 1 strains. *J. Clin. Microbiol.* **36**:161–167.
39. **Robinson, D. A., D. E. Briles, M. J. Crain, and S. K. Hollingshead.** 2002. Evolution and virulence of serogroup 6 pneumococci on a global scale. *J. Bacteriol.* **184**:6367–6375.
40. **Salcedo, C., L. Arreaza, B. Alcalá, L. De La Fuente, and J. A. Vazquez.** 2003. Development of a multilocus sequence typing method for analysis of *Listeria monocytogenes* clones. *J. Clin. Microbiol.* **41**:757–762.
41. **Schoonmaker, D., T. Heimberger, and G. Birkhead.** 1992. Comparison of ribotyping and restriction enzyme analysis using pulsed-field gel electrophoresis for distinguishing *Legionella pneumophila* isolates obtained during a nosocomial outbreak. *J. Clin. Microbiol.* **30**:1491–1498.
42. **Segal, G., J. J. Russo, and H. A. Shuman.** 1999. Relationships between a new type IV secretion system and the *icm/dot* virulence system of *Legionella pneumophila*. *Mol. Microbiol.* **34**:799–809.
43. **Selander, R. K., R. M. McKinney, T. S. Whittam, W. F. Bibb, D. J. Brenner, F. S. Nolte, and P. E. Pattison.** 1985. Genetic structure of populations of *Legionella pneumophila*. *J. Bacteriol.* **163**:1021–1037.
44. **Simpson, E. H.** 1949. Measurement of diversity. *Nature* **163**:688.
45. **Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**:4384–4388.
46. **Spratt, B. G.** 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Curr. Opin. Microbiol.* **2**:312–316.
47. **Struelens, M. J.** 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin. Microbiol. Infect.* **2**:2–11.
48. **Swofford, D. L.** 1999. PAUP\*: phylogenetic analysis using parsimony (\*and other methods), version 4. Sinauer Associates, Sunderland, Mass.
49. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
50. **Visca, P., P. Goldoni, P. C. Lück, J. H. Helbig, L. Cattani, G. Giltri, S. Bramati, and M. C. Pastoris.** 1999. Multiple types of *Legionella pneumophila* serogroup 6 in a hospital heated-water system associated with sporadic infections. *J. Clin. Microbiol.* **37**:2189–2196.
51. **Wintermeyer, E., B. Ludwig, M. Steinert, B. Schmidt, G. Fischer, and J. Hacker.** 1995. Influence of site specifically altered Mip proteins on intracellular survival of *Legionella pneumophila* in eukaryotic cells. *Infect. Immun.* **63**:4576–4583.
52. **Yu, V. L., J. F. Plouffe, M. C. Pastoris, J. E. Stout, M. Schousboe, A. Widmer, J. Summersgill, T. File, C. M. Heath, D. L. Paterson, and A. Cheresky.** 2002. Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *J. Infect. Dis.* **186**:127–128.