

## Insertion Sequences as Highly Resolutive Genomic Markers for Sequence Type 1 *Legionella pneumophila* Paris<sup>∇†</sup>

Mike Vergnes,<sup>1,2</sup> Christophe Ginevra,<sup>3,4</sup> Elisabeth Kay,<sup>1,2</sup> Philippe Normand,<sup>5</sup> Jean Thioulouse,<sup>6</sup> Sophie Jarraud,<sup>3,4,7</sup> Max Maurin,<sup>1,2,8</sup> and Dominique Schneider<sup>1,2\*</sup>

Laboratoire Adaptation et Pathogénie des Microorganismes, Université Joseph Fourier Grenoble 1, BP 170, F-38042 Grenoble cedex 9, France<sup>1</sup>; CNRS UMR5163, F-38042 Grenoble cedex 9, France<sup>2</sup>; Centre National de Référence des Légionelles, Université Lyon 1, Lyon, France<sup>3</sup>; INSERM U851, IFR128, Faculté de Médecine Lyon Est, 7 rue Guillaume Paradin, F-69372 Lyon cedex 08, France<sup>4</sup>; Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS UMR5557, Laboratoire d'Ecologie Microbienne, F-69622 Villeurbanne, France<sup>5</sup>; Université de Lyon, F-69000, Lyon, Université de Lyon 1, CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France<sup>6</sup>; Hospices Civils de Lyon, Laboratoire de Bactériologie, Centre de Biologie et Pathologie Est, 59 Boulevard Pinel, Bron, France<sup>7</sup>; and CHU Grenoble, Université Joseph Fourier, BP 217, 38043 Grenoble, France<sup>8</sup>

Received 21 June 2010/Returned for modification 2 August 2010/Accepted 20 October 2010

The causative agent of legionellosis, *Legionella pneumophila*, colonizes all natural and human-made water networks, thus constituting the source of contaminated aerosols responsible for airborne human infections. Efficient control of infections, especially during epidemics, necessitates the fastest and most resolutive identification possible of the bacterial source for subsequent disinfection of reservoirs. We thus compared recognized typing approaches for *Legionella* with a method based on characterization of insertion sequence (IS) content. A total of 86 clinical or environmental isolates of *L. pneumophila*, including 84 Paris isolates, sampled from 25 clinical investigations in France between 2001 and 2007, were obtained from the *Legionella* National Reference Center. All strains were typed by monoclonal antibody subgrouping, sequence-based typing, pulsed-field gel electrophoresis, and restriction fragment length polymorphism based on the presence or absence of IS elements. We identified six different types of IS elements in *L. pneumophila* Paris and used them as genomic markers in hybridization experiments. One IS type, ISLpn11, revealed a high discriminatory power. Simpson's index of discrimination, calculated from the distribution of IS elements, was higher than that obtained with the other typing methods used for *L. pneumophila* Paris. Moreover, specific ISLpn11 copies were found only in strains isolated from particular cities. In more than half of the cases, each clinical isolate had an ISLpn11 profile that was recovered in at least one environmental isolate from the same geographical location, suggesting that our method could identify the infection source. Phylogenetic analysis suggests a clonal expansion for the *L. pneumophila* Paris strain.

*Legionella* species are facultative intracellular Gram-negative bacteria causing legionellosis, which may develop into a severe pneumonia called Legionnaires' disease but also a much more benign flu-like illness known as Pontiac fever. *Legionella* is widely distributed in the environment and colonizes all natural and human-made water networks, such as those of hospitals, thermal baths, hotels, etc., providing an abundant and prevalent reservoir for human infections (16, 41, 62). In these environments, *Legionella* can multiply inside protozoan cells (51). Human infection occurs through inhalation of aerosolized water contaminated with *Legionella* cells. Although the genus *Legionella* comprises more than 50 species and 64 serogroups, *Legionella pneumophila* serogroup 1 is the most common pathogenic species, accounting for more than 90% of legionellosis cases (43, 63) despite representing only 28% of environmental isolates (13). To offer efficient control of human

infections, fast and accurate identification of the isolates involved and their water reservoirs is necessary. This is particularly important in the context of a legionellosis outbreak in which rapid treatment of contaminated water sources will allow effective prevention of the occurrence of new infections. Highly resolutive typing will also further optimize current surveillance measures of water systems for prevention of the disease. Over the last decades and besides culture on selective media (54), several methods have been used for *Legionella* typing. Before giving an overview of these methods, it is important to emphasize that no clear differentiation of clinical versus environmental isolates is presently possible for *L. pneumophila* strains (2, 4).

Classification of *Legionella* spp. into serogroups and monoclonal subgroups has been performed by monoclonal antibody (MAb) typing using an international panel of MAbs (27, 32). Many genotyping methods have been developed, including pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), restriction endonuclease analysis, or arbitrarily primed PCR (21). However, these methods have limitations in both resolution level and interlaboratory typing reproducibility. Therefore, a multilocus sequence typing approach, called sequence-based typing (SBT), applicable either to bacterial strains or directly to clinical samples, was developed by members of the European Working Group for

\* Corresponding author. Mailing address: Laboratoire Adaptation et Pathogénie des Microorganismes, CNRS UMR5163, Université Joseph Fourier, Institut Jean Roget, Campus Santé, Domaine de la Merci, BP170, 38042 Grenoble Cedex 9, France. Phone: 33 (0)4 76 63 74 90. Fax: 33 (0)4 76 63 74 97. E-mail: dominique.schneider@ujf-grenoble.fr.

† Supplemental material for this article may be found at <http://jcm.asm.org/>.

∇ Published ahead of print on 27 October 2010.

*Legionella* Infections (EWGLI). This method is based on the allelic profiling of clinical and environmental isolates based on nucleotide sequences of protein-coding loci (22, 23). Seven-locus typing is now proposed as a standard epidemiological method allowing the classification of isolates into sequence types (ST) (46). A further improvement using nested PCR-based SBT was recently applied to clinical samples (24). Recently, the availability of DNA arrays for *Legionella* permitted the use of differential genomic hybridization, but this did not allow the inference of the genetic origin of the infection (4). Moreover, DNA arrays may be difficult to set up in laboratories.

Several studies combined these different methodologies to infer the distribution of *Legionella* isolates, especially from *L. pneumophila* serogroup 1, using collections of isolates sampled all over the world—in Japan (1), Canada (49, 58), the United States (35), South Korea (37), and Europe (22, 23, 26). A consensus trend emerged, with all data revealing both a high level of genetic diversity in *Legionella* and the emergence of specific groups of clones that showed high similarities and were responsible for both sporadic legionellosis cases and worldwide outbreaks (1, 4, 26, 34, 35, 37, 49, 58). In particular, isolates from ST1 (which includes the widely distributed Paris strain), ST47, and ST222 have been detected in sporadic and outbreak cases. Among clinical isolates sampled in France over the 2001-2007 period, 6.4% to 11.3% were found to belong to the Paris strain, the proportion in 2009 being 10.8% among a total of 213 studied isolates (data from the *Legionella* National Reference Center, Lyon, France). Expansion of these particular clones may be related to specific fitness and/or virulence traits. Although combinations of MAb and molecular typing methods are used in epidemiological investigations to compare human and environmental *L. pneumophila* strains, the source of infection remains often difficult, or even impossible, to identify, emphasizing the need for additional more-resolutive approaches. We decided in this study to use insertion sequence (IS) mobile genetic elements as genomic markers for typing *L. pneumophila* serogroup 1 Paris strains and to investigate both their discriminatory power and ability to identify the environmental source of legionellosis outbreaks. Such highly discriminant genetic tools may also be applied to the identification of clinically relevant bacterial subtypes and the investigation of genetic determinants of bacterial fitness and virulence.

IS elements are small DNA sequences (500 to 2,500 bp) that encode generally no function other than those involved in their mobility (39). They are widely distributed in virtually all bacterial genomes and show a huge diversity in element types and copy numbers. They are often considered selfish parasitic DNA elements, but their activity may also confer to the cells that harbor them increased fitness in certain circumstances (9, 44). They contribute significantly to spontaneous mutagenesis in bacteria and can have profound effects on gene expression through their transposing activity (see reference 39 for a review). Moreover, DNA rearrangements can also occur through recombination between IS copies. For all these reasons, IS elements have been used extensively in molecular epidemiological studies, and in some cases, these genotyping tools have been standardized by the international community. Moreover, IS elements have even been reported to play a role in virulence (31). Genotyping has been performed using specific IS elements for pneumonic pathogens other than *Legionella*, for

example *Streptococcus pneumoniae* (38, 50, 56), *Staphylococcus aureus* (30, 64), *Staphylococcus epidermidis* (12), and *Coxiella burnetii* (60). More generally, restriction fragment length polymorphism using IS elements (RFLP-IS) has been performed for typing natural isolates in numerous bacterial human pathogens, including *Yersinia* (59), *Mycobacterium* (61), *Salmonella* (20), and *Leptospira* (65). Their discrimination power was also used in experimental evolution experiments to differentiate clones derived from the same common ancestral strain and to measure genetic diversity (44).

In the work reported here, we identified the IS elements in the genome sequence of the *L. pneumophila* Paris strain (5) and subsequently used them as genomic markers in RFLP-IS experiments to characterize 86 clinical or environmental isolates of *L. pneumophila*, including 84 Paris isolates, in combination with MAb, PFGE, and SBT typing methods. We found that a single IS element, IS*Lpn11*, had by itself a significantly higher discrimination power than other methods. Moreover, in more than half of the cases, a genetic link could be detected between clinical and environmental isolates sampled at identical geographical locations during independent epidemiological investigations. Phylogenetic analyses using the obtained RFLP-IS profiles revealed, as previously suggested, a clonal expansion of the Paris strain. However, between-group analysis revealed some correlation between geographical locations and specific IS copies.

## MATERIALS AND METHODS

**Bacterial strains and growth conditions.** The bacterial strains used in this study were provided by the *Legionella* National Reference Center (Lyon, France) and are listed in Tables 1 and 2. All bacterial strains were plated on *N*-(2-acetamido)-2-aminoethanesulfonic acid (ACES)-buffered charcoal yeast extract (BCYE) agar plates and incubated at 37°C for 2 or 3 days for PFGE and MAb subtyping or genomic DNA extraction, respectively.

**Genomic DNA extraction.** Between 10<sup>9</sup> and 10<sup>10</sup> bacterial cells of each strain were scraped from petri dishes and resuspended in 500 µl of lysis solution (50 mM Tris-HCl, 20 mM EDTA, 100 mM glucose, and 4 mg/ml lysozyme). After incubation on ice for 10 min, 100 µl of 10% (wt/vol) SDS was added, and the mixture further incubated at 37°C for 10 min. Proteins and cell debris were removed by adding 800 µl of acetate-phenol (50/50 [vol/vol]), vortexing for 10 min at room temperature, and centrifuging at 10,000 × *g* for 10 min at 4°C. The supernatant was collected, and genomic DNA precipitated by the addition of a 1/10 volume of 3 M sodium acetate, pH 5.2, and 3 volumes of cold absolute ethanol. The DNA pellet was then washed with cold 70% (vol/vol) ethanol to remove salts, dried, and resuspended in 200 µl of sterile water.

**Southern blot hybridization.** Genomic DNA was digested by BamHI and PvuII, since the IS elements studied all lacked these restriction sites, and the resulting fragments were separated by electrophoresis in 0.8% agarose gels in 1× Tris-acetate-EDTA (TAE) buffer. The DNA fragments were then transferred onto a nylon membrane (Schleicher and Schuell Bioscience, Dassel, Germany). Hybridizations were performed at high stringency (68°C) using internal fragments of the IS elements as probes. These internal sequences all lacked BamHI and PvuII restriction sites, so that each hybridizing fragment represented a unique copy of the relevant IS element. The IS internal fragments were generated by PCR using primer pairs listed in Table 3 and cold-labeled using the DIG system (Roche, Meylan, France). Filters for Southern blot hybridizations were probed with each IS element successively, and each hybridized probe was stripped prior to reprobing. An example of Southern hybridization is given in Fig. S1 in the supplemental material. Each of the 86 strains was scored for the presence (score of 1) or absence (score of 0) of each fragment carrying a particular IS element, representing a genotype-specific IS fingerprint. This scoring yielded a matrix of 86 clones and 35 fragments over the six IS elements that were used in subsequent analyses. (For the most discriminating element, IS*Lpn11* [see Results], a total of 15 IS-containing fragments was detected.) The RFLP-IS scoring data are given in Table S1 in the supplemental material for the 76 *L. pneumophila* Paris isolates (Table 2).

**PFGE subtyping.** *L. pneumophila* cells were resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8) and mixed with an equal volume of 1% SeaKem

TABLE 1. Strains of *L. pneumophila* isolated from nonrelated cases, with their PFGE, SBT, MAb, and RFLP-IS profiles

City	Isolate no.	PFGE <sup>a</sup>	SBT <sup>b</sup>	MAb <sup>c</sup>	RFLP-IS <sup>d</sup>
Paris Clamart	HL 0213 2010	Paris	ST1	Olda	000001011001101
Paris	HL 0051 4008	Paris	ST1	Olda	000011110001101
Annecy	HL 0228 2021	Paris	ST1	Philadelphia	000011011001101
Compiègne	HL 0215 1004	Paris	ST1	Philadelphia	000011011001101
Meaux	HL 0230 4016	Paris	ST1	Philadelphia	000011011011101
Paris	HL 0036 4001	Paris	ST1	Philadelphia	000011011001101
Paris	HL 0337 3012	Paris	ST1	Philadelphia	000010011001101
Paris	HL 0102 3035	Paris	ST1	Philadelphia	000011011001101
Paris	HL 0501 3042	Unique	ST79	Olda	001001010001001
Poitiers	HL 0323 1005	Unique	NA	NA	000001011001111

<sup>a</sup> PFGE, pulsed field gel electrophoresis. The PFGE profile is given for each isolate, with the Paris profile being similar to the one of the reference strain CIP107629 (5), while unique profiles differ by at least one fragment from the ones present in the library of the *Legionella* National Reference Center.

<sup>b</sup> SBT, sequence-based typing. The ST number is given for each isolate, with ST1 and ST79 corresponding to the 1-4-3-1-1-1-1 and 1-4-3-4-1-1-15 allelic codes, respectively. NA, not applicable, since the *L. pneumophila* HL 0323 1005 isolate belongs to serogroup 8. Its ST number has not been assigned, as the *neuA* gene could not be sequenced, which is typical for isolates of *L. pneumophila* serogroups other than 1.

<sup>c</sup> MAb, monoclonal antibody typing. Isolates from the group Olda react with MAb8/4 but not with MAb3/1 and MAb26/1, while isolates from the group Philadelphia react with MAb8/4 and with MAb3/1 but not with MAb3 (27). NA, not applicable since MAb subtyping is specific to *L. pneumophila* serogroup 1 and the *L. pneumophila* HL 0323 1005 isolate belongs to serogroup 8. Although MAb subtyping may not be reliable since the expression of the lipopolysaccharide types that react with the monoclonal antibodies is prone to phase variation and may also depend on the growth stage (28), it was shown recently that the MAb3/1 epitope is expressed during all growth phases (48). Moreover, MAb typing is widely used according to strictly standardized protocols to take these potential questions into consideration.

<sup>d</sup> RFLP-IS, restriction fragment length polymorphism using IS elements as probes. Here, only the RFLP-IS hybridization scoring for the most discriminating *ISLpn11* is given, with 0 and 1 corresponding to the absence and presence, respectively, of each of the 15 hybridizing IS-containing fragments.

gold agarose (Tebu, Le Perray En Yvelines, France) for casting plugs. Plugs were then treated with proteinase K (50 µg/ml) in TE buffer for 24 h at 55°C. DNA was then digested with 20 units of the SfiI restriction enzyme (Roche, Meylan, France) for 16 h at 50°C. Fragments of DNA were separated by electrophoresis in 0.8% agarose gels and run in 0.5× Tris-borate-EDTA buffer (pH 8.3) in a contour-clamped homogeneous field apparatus (CHEF DRII system; Bio-Rad, Ivry-sur-Seine, France) with a constant voltage of 150 V. Runs were carried out with constant pulse times (25 s) for 11 h at 10°C followed by increasing pulse times (35 to 60 s) for 11 h at 10°C. PFGE patterns were analyzed using the GelComparII software (Applied Maths, Kortrijk, Belgium).

**MAB subgrouping.** These experiments were based on the international MAB subgrouping scheme (27). The *L. pneumophila* cells were resuspended in saline buffer and analyzed by indirect immunofluorescence using the Dresden panel monoclonal antibodies. The bacterial suspensions (10 µl) were allowed to dry at room temperature on each well of Teflon-coated microscope slides. After 1-h incubation of the slides on a heater plate, monoclonal antibodies were added and incubated in a humid chamber for 45 min at 37°C. The slides were then washed for 5 min in 10 mM phosphate-buffered saline (PBS), pH 7.2, and air dried. Fluorescein-labeled goat anti-mouse immunoglobulin was then added and incubated in a humid chamber for 30 min at 37°C. After a PBS wash, slides were mounted and examined with an epifluorescence microscope (Axioskop; Zeiss).

**SBT.** All isolates were processed for SBT using the EWGLI standard scheme. Seven genes, namely, *asd*, *flaA*, *mip*, *mompS*, *pilE*, *proA*, and *neuA* (22, 46), were PCR amplified using the same oligonucleotide primers and conditions described previously (22, 46) (<http://www.ewgli.org>). They yielded 245- to 648-bp PCR products that were then sequenced (GATC Biotech, Konstanz, Germany). The resulting sequences for each gene locus were trimmed and compared to previously assigned allele numbers in the online database at <http://www.ewgli.org>. Based on a predetermined order (*flaA*, *pilE*, *asd*, *mip*, *mompS*, *proA*, and *neuA*), the combination of seven alleles at each of the loci was defined as a seven-digit allelic profile (e.g., 1-4-3-1-1-1-1), and each combination type was attributed an allelic profile number or ST (e.g., ST1) using the EWGLI sequence quality tool ([http://www.hpa-bioinformatics.org.uk/legionella/legionella\\_sbt/php/sbt\\_homepage.php](http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php)). No new variant was found in this study. The SBT methodology applied to *Legionella* is distinct from the more typical MLST scheme since it includes genes under diversifying pressure, whereas MLST is based exclusively on housekeeping genes.

**Simpson's diversity index.** Discriminating power was assessed by computing the diversity index of Simpson (29) according to the following equation:

$$D = 1 - \frac{\sum_{i=1}^S n_i(n_i - 1)}{N(N - 1)},$$

where  $S$  is the number of different profiles,  $N$  is the total number of sampled profiles, and  $n_i$  is the number of isolates with the same profile.

**Phylogenetic analyses.** Identical profiles were eliminated from the IS-derived matrix comprising 86 strains and 15 IS fragments (containing the discriminating

*ISLpn11* element) and resolved as 17 operational taxonomic units (OTUs). The matrix was then formatted for treatment using the Phylip software (19).

A parsimony approach was used with the Wagner method (15, 33), where both 0 to 1 and 1 to 0 changes are allowed, with the assumption that 0 is the ancestral state. It is also assumed that different characters evolve and that different lineages evolve independently. The different trees yielded by the approach were condensed using consensus, the consensus tree program written by Margush and McMorris (40). Another parsimony approach using the Dollo method (17) is possible. However, it assumes that the direction of mutation is known, with 0 to 1 being much more probable than the reverse, while the Wagner method does not assume such a preferential mutational direction. Moreover, the Wagner method is more largely used in the literature (3, 53, 55) for comparable data sets. We thus also applied the Dollo method, which may further confirm the robustness of branches in the phylogenetic trees.

The IS elements were treated as 1,000-nucleotide loci using RestDist, written initially for restriction site analysis (42). A distance matrix (substitutions/site) was generated and treated by neighbor-joining according to Saitou and Nei (52). A bootstrap analysis was then performed according to Felsenstein (18).

**BGA.** Analyses of the differences between cities in the distribution of IS elements in *Legionella* strains were performed by between-group analysis (BGA) (11, 14), which is a multivariate analysis technique derived from principal component analysis (PCA) (45). The aim of PCA is to summarize a data table by searching orthogonal axes on which the projection of items (rows of the table) has the highest possible variance. This characteristic ensures that the associated graphs, or principal component (PC) maps, will best represent the initial data set. PCs have the property of having the highest possible correlation with the original variables (columns of the table).

From a theoretical point of view, BGA is the particular case of PCA, with respect to instrumental variables (PCAIV) (36, 45), in which the instrumental variable table is reduced to only one dummy variable. This variable defines groups of rows in the data table, and BGA is simply the PCA of the table of group means (here, the groups correspond to the cities where *Legionella* strains were collected). The group means table has a number of rows equal to the number of groups, and the same columns as the original table. The objective of BGA is to separate the groups, and it can be considered a robust alternative to discriminate analysis. A Monte Carlo (permutation) test was used to check the significance of the differences between groups. Computations and graphical displays were carried out with the R software (47), using both the ade4 (6) and ade4TkGUI (57) packages. This software is available at <http://pbil.univ-lyon1.fr/ade4/>.

## RESULTS

**IS families in *L. pneumophila* Paris.** We analyzed the available genome sequence of *L. pneumophila* Paris (5) for the

TABLE 2. Isolates of *L. pneumophila* Paris, as determined by PFGE typing, sampled from patients and from putative corresponding contaminating sources of 25 epidemiological investigations

Epidemiological investigation no.	Yr of isolation	Isolate	Origin	City <sup>a</sup>
1	2001	HL 0128 1038	Clinical	Paris, 15th
		2001 I no. 1	Environmental	Paris, 15th
		2001 I no. 2	Environmental	Paris, 15th
2	2002	HL 0128 4004	Environmental	Paris, 15th
		HL 0230 4014	Clinical	Meaux
		HL 0230 4015	Clinical	Meaux
		HL 0230 4016	Clinical	Meaux
		HL 0230 4019	Environmental	Meaux
3	2002	HL 0230 4021	Environmental	Meaux
		HL 0233 3024	Clinical	Evry
		HL 0234 2010	Environmental	Evry
4	2003	HL 0234 2011	Environmental	Evry
		HL 0311 1002	Clinical	Nice
5	2003	HL 0311 1005	Environmental	Nice
		HL 0311 1006	Environmental	Nice
		HL 0312 2043	Clinical	Clermont Ferrand
6	2003	HL 0312 2044	Environmental	St. Etienne
		HL 0315 5019	Clinical	Nice
7	2003	HL 0317 3023	Clinical	Nice
		HL 0317 3011	Environmental	Nice
		HL 0317 3008	Environmental	Nice
		HL 0332 1022	Clinical	Rouen
8	2004	HL 0333 2001	Environmental	Rouen
		HL 0333 2002	Environmental	Rouen
		HL 0416 3014	Clinical	Brest
9	2004	HL 0416 3015	Environmental	Brest
		HL 0416 3016	Environmental	Brest
		HL 0425 3033	Clinical	Eaubonne
10	2004	HL 0427 4016	Clinical	Eaubonne
		HL 0427 4017	Environmental	Eaubonne
		HL 0427 4018	Environmental	Eaubonne
11	2004	HL 0428 3017	Clinical	Strasbourg
		HL 0430 4020	Environmental	Strasbourg
12	2005	HL 0430 4040	Clinical	Nice
		HL 0430 4043	Environmental	Nice
		HL 0430 4044	Environmental	Nice
13	2005	HL 0504 5030	Clinical	Marseille
		HL 0504 5031	Environmental	Marseille
14	2005	HL 0525 4042	Clinical	Nice
		HL 0526 3026	Environmental	Nice
15	2005	HL 0530 1024	Clinical	Nantes
		HL 0531 4007	Environmental	Poitiers
		HL 0531 4008	Environmental	Poitiers
		HL 0650 5002	Environmental	Poitiers
16	2005	HL 0533 5004	Clinical	Amiens
		HL 0532 3007	Environmental	Tours
		HL 0532 3008	Environmental	Tours
17	2005	HL 0534 5006	Clinical	Beauvais
		HL 0534 5007	Environmental	Beauvais
		HL 0534 5008	Environmental	Beauvais
18	2006	HL 0534 3013	Clinical	Grenoble
		HL 0540 1036	Environmental	Grenoble
19	2006	HL 0631 2017	Clinical	Argenteuil
		HL 0631 5002	Environmental	Eaubonne
		HL 0631 5003	Environmental	Eaubonne
20	2006	HL 0641 5016	Clinical	Montpellier
		HL 0644 4026	Environmental	Fréjus
		HL 0644 1005	Environmental	Fréjus
21	2007	HL 0644 2011	Clinical	Nice
		HL 0644 2012	Clinical	Nice
		HL 0644 5004	Environmental	Nice
21	2007	HL 0701 3011	Clinical	Rueil Malmaison

Continued on following page



TABLE 2—Continued

Epidemiological investigation no.	Yr of isolation	Isolate	Origin	City <sup>a</sup>
22	2006	HL 0701 3003	Environmental	Rueil Malmaison
		HL 0701 3004	Environmental	Rueil Malmaison
		HL 0651 3023	Clinical	Anncy
		HL 0703 5017	Environmental	Anncy
23	2007	HL 0703 5018	Environmental	Anncy
		HL 0712 5004	Clinical	Aulnay
		HL 0714 2027	Environmental	Aulnay
24	2007	LG 0713 5006	Clinical	Paris, 15th
		LG 0713 5007	Environmental	Paris, 15th
		LG 0713 5008	Environmental	Paris, 15th
		LG 0725 3019	Environmental	Paris, 15th
		LG 0725 3022	Environmental	Paris, 15th
25	2007	LG 0725 4014	Clinical	Argenteuil
		LG 0725 5011	Environmental	Eaubonne

<sup>a</sup> For clinical isolates, the city where the patients were hospitalized is given. 15th, fifteenth arrondissement.

presence of different IS element families, based on four IS features: the presence of terminal inverted repeats (IRs), the presence of small duplications directly outside the IRs representing the target sites of the IS elements, the presence of one or two open reading frames (ORF), and/or the homology of the translated products with known transposases (39). Identities with known IS families are expected to yield information about the type of IS detected. We identified six different types of potentially functional IS elements in *L. pneumophila* Paris, which are all present in several copies in the genome (Table 4). The 6 IS elements belonged to 5 different IS families. Comparison with the genomes of three other *L. pneumophila* strains, Lens (5), Philadelphia (7), and Corby (25), revealed a differential distribution of these IS elements, reflecting the high plasticity of the *Legionella* genome. Two IS elements, *ISLpn1* and *ISLpn10*, were specifically found only in the Paris strain, whereas *ISLpn9* and *ISLpn12* were present in only one of the other *L. pneumophila* strains, Philadelphia and Corby, respectively. Two IS elements were detected in two of the other strains: *ISLpn8* in Philadelphia and Lens and *ISLpn11* in Philadelphia and Corby. The copy number of the different IS elements was also found to be variable across the four genomes (Table 4).

**Validation of RFLP-IS as a typing method of *L. pneumophila* isolates.** As a first step to evaluate RFLP-IS typing, we used 8 isolates of *L. pneumophila*, including two classified as sero-

group 1, PFGE Paris, ST1, and MAb subgroup Olda and six as Philadelphia (Table 1). Using the six different IS elements that were identified in *L. pneumophila* Paris as probes, we obtained five different hybridization profiles, and this discrimination threshold was obtained with a unique IS element, *ISLpn11*, the other probes adding no further discriminating power. The RFLP-IS therefore yielded a very high Simpson's diversity index *D* value of 79%, compared to SBT or PFGE, which could not discriminate these isolates, and to MAb typing, with a Simpson's index value of only 43%. These experiments also revealed *ISLpn11* as a strongly discriminating molecular tool that alone was better than the combination of the three other typing methods.

To assess whether the RFLP-IS method could be extended to non-Paris *L. pneumophila* isolates, two isolates were investigated, including one from serogroup 1 but with a unique PFGE profile, ST79, and MAb Olda, and one from serogroup 8. As expected for a discriminating tool, two distinct and unique RFLP-IS profiles were obtained (Table 1).

As a second step, we therefore extended these studies to a set of specifically selected *L. pneumophila* Paris isolates presenting the same PFGE typing profile. We performed RFLP-IS typing, and the data were treated in three ways: (i) phylogenetic reconstruction analysis, (ii) between-group analyses to measure differences between cities in the distribution of IS copies in *Legionella* strains, and (iii) comparative estimation

TABLE 3. Primers used to generate the IS-derived probes

Primer	Sequence	IS
PLP71	5'-GAATTCGTTGAAGTAGCTCACTGATTCATTG-3'	<i>ISLpn9</i>
PLP72	5'-TAAAATTGCTAATCTTCACAACAGTTTTC-3'	
PLP73	5'-ATTAATTGAAAAGAAGGCAGTTTTATAGTGG-3'	<i>ISLpn10</i>
PLP74	5'-GCGCATCCAAATGAGCGCTCCGGCCAAATG-3'	
PLP75	5'-GAGCAGTTACTTTTTAAAGTGTACATTAAC-3'	<i>ISLpn1</i>
PLP76	5'-ACTCAGCTCAAAGCGCCACATTCAATAAATG-3'	
PLP77	5'-ATGCCCAAGAAAAGATTGTATCCTAAATTTAC-3'	<i>ISLpn11</i>
PLP78	5'-TCAAGAACACAAAACCTTAACCTTAACCTAT-3'	
PLP81	5'-TAGTTGTATTCTGATTAGAGTTCTCCAAG-3'	<i>ISLpn8</i>
PLP82	5'-CCCCATATTTAGTGGCATCCTAAATTAGAG-3'	
PLP85	5'-TTTGCCGACTGAAAATTGATCCCCCTACTATAG-3'	<i>ISLpn12</i>
PLP86	5'-AAAAAATTAATCAGTTATATTATTTGCTC-3'	

TABLE 4. Characteristics of potentially functional IS elements detected in *L. pneumophila* Paris<sup>c</sup>

Name	Family	Subgroup	Gene for ORF1	Gene for ORF2	End	IR (bp)	DR (bp)	Length (bp)	Copy no.			
									Paris	Philadelphia	Lens	Corby
<i>ISLpn1</i>	IS4	IS10	<i>lpp0125</i>	No	CT	16	0	1,355	2	0	0	0
<i>ISLpn8</i>	IS3	IS407	<i>lpp0199</i>	<i>lpp0200</i>	TG	32	0	1,199	2	2	1	0
<i>ISLpn9</i>	IS4	IS10	<i>lpp0038</i>	<i>lpp0039</i>	CT	60	9	1,628	7 (+2) <sup>a</sup>	1	0	0
<i>ISLpn10</i>	IS5	IS427	<i>lpp0083</i>	No	GG	22	2	896	2	0	0	0
<i>ISLpn11<sup>b</sup></i>	ISL3		<i>lpp2037</i>	No	GG	20	8	1,326	6	0 (+1) <sup>a</sup>	0	1
<i>ISLpn12</i>	IS21		<i>lpp1563</i>	<i>lpp1564</i>	TG	41	7	1,996	2	0	0	1

<sup>a</sup> In some cases, truncated copies of some IS elements are present, and their numbers are indicated in parentheses.

<sup>b</sup> While one of the six *ISLpn11* copies in *L. pneumophila* Paris presented no duplication of the target site, the five others led to 8-bp duplications of the insertion target, with no obvious similarity between them (data not shown). Moreover, the insertion sites are widely distributed on the Paris chromosome, even if two insertions occurred close to tRNA-encoding genes, tRNA<sup>Lys</sup> and tRNA<sup>Met</sup>. In the Philadelphia strain, the unique *ISLpn11* copy is also localized close to the tRNA<sup>Met</sup>-encoding gene.

<sup>c</sup> For each IS element, the following features are given: family and subgroup, name of the genes encoding ORF1 and ORF2, last two nucleotides (End), length of IRs, length of the direct repeats (DR) generated upon insertion of the IS, total length, and copy number in the genome of the four strains Paris, Lens (5), Philadelphia (7), and Corby (25).

of the discriminating power of RFLP-IS versus SBT and MAB typing methods as well as its ability to relate clinical isolates to the environmental source clone.

**Phylogenetic analyses.** We performed RFLP-IS typing on a set of 76 *L. pneumophila* Paris isolates sampled during 25 epidemiological investigations that occurred between 2001 and 2007, including 30 isolates from patients and 46 from putative contaminating roots (Table 2). Again, *ISLpn11* was the only IS element that discriminated the isolates, as all the other IS elements gave identical profiles for all isolates. The RFLP-*ISLpn11* scoring for this set of 76 isolates is given in Table S1 in the supplemental material, together with an example of Southern hybridization in Fig. S1 in the supplemental material. Combining the data obtained for this set of 76 isolates (Table S1) with the above set of 10 clinical isolates (Table 1), a total of 15 different copies of *ISLpn11*, with a mean copy number of  $6.87 \pm 0.81$ , was found. We found 17 different hybridization profiles. Ten isolates were found to have unique profiles, while the most represented one was found 32 times (Fig. 1 and Table 1; see also Table S1 in the supplemental material). This particular profile will be called Ref7 hereinafter; it belongs to the

reference *L. pneumophila* Paris strain initially isolated from a clinical case (5). One IS copy was present in all isolates tested, while three IS copies were present in a single isolate.

The Wagner tree (Fig. 1), including all 86 isolates, was found to group reliably OTUs 76, 55, 114, and 32 and OTUs 132, 86, and 29 at consensus levels of 0.84 and 1.00, respectively. Several OTUs with a single representative (OTUs 91, 52, 61, 33, and 35) were grouped at a consensus level of 0.51. The neighbor-joining tree was found to group OTUs 132, 86, and 29, as for the Wagner parsimony (data not shown). Although the Wagner method is most likely to better reflect our data set (see Materials and Methods), we also applied the Dollo parsimony method to see if some branches could also be recovered with a less suitable method, thus reinforcing the robustness of these relationships. It was indeed the case for OTUs 91 and 52, for OTUs 132, 86, and 29, and for others, if we tolerate one missing OTU: OTUs 91, 52, 61, 33, and 35 and OTUs 78, 32, 98, 76, and 55.

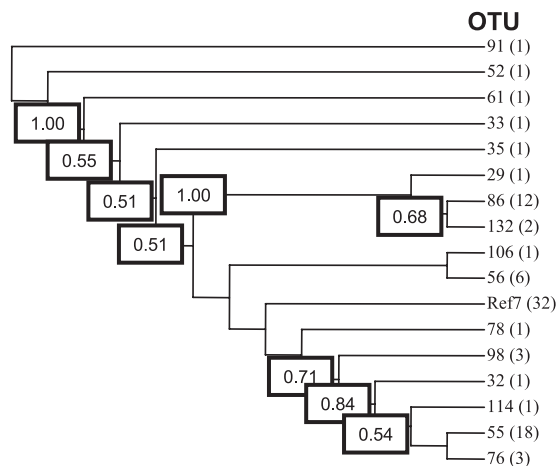


FIG. 1. Wagner parsimony phylogenetic tree. By using RFLP-IS, 17 different OTUs were found, and the number of isolates showing the corresponding profile is indicated in parentheses. Ref7 is the OTU of the reference *L. pneumophila* strain Paris as well as the most represented profile. Boxes indicate consensus levels above 50%.

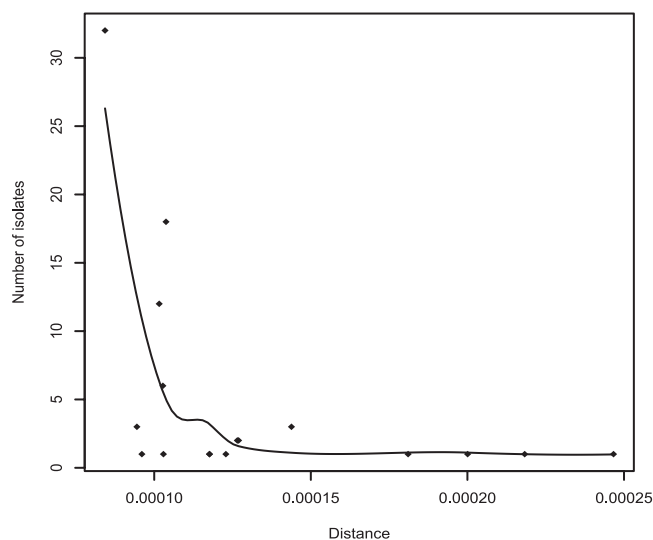


FIG. 2. Relationship between distance and the number of isolates within each OTU. For each of the 17 OTUs, the average distance is calculated to each other and is expressed as substitutions per site. Diamonds represent data points, and the curve is a locally weighted regression (Loess model [8]) of the number of isolates against the average distance between OTUs.

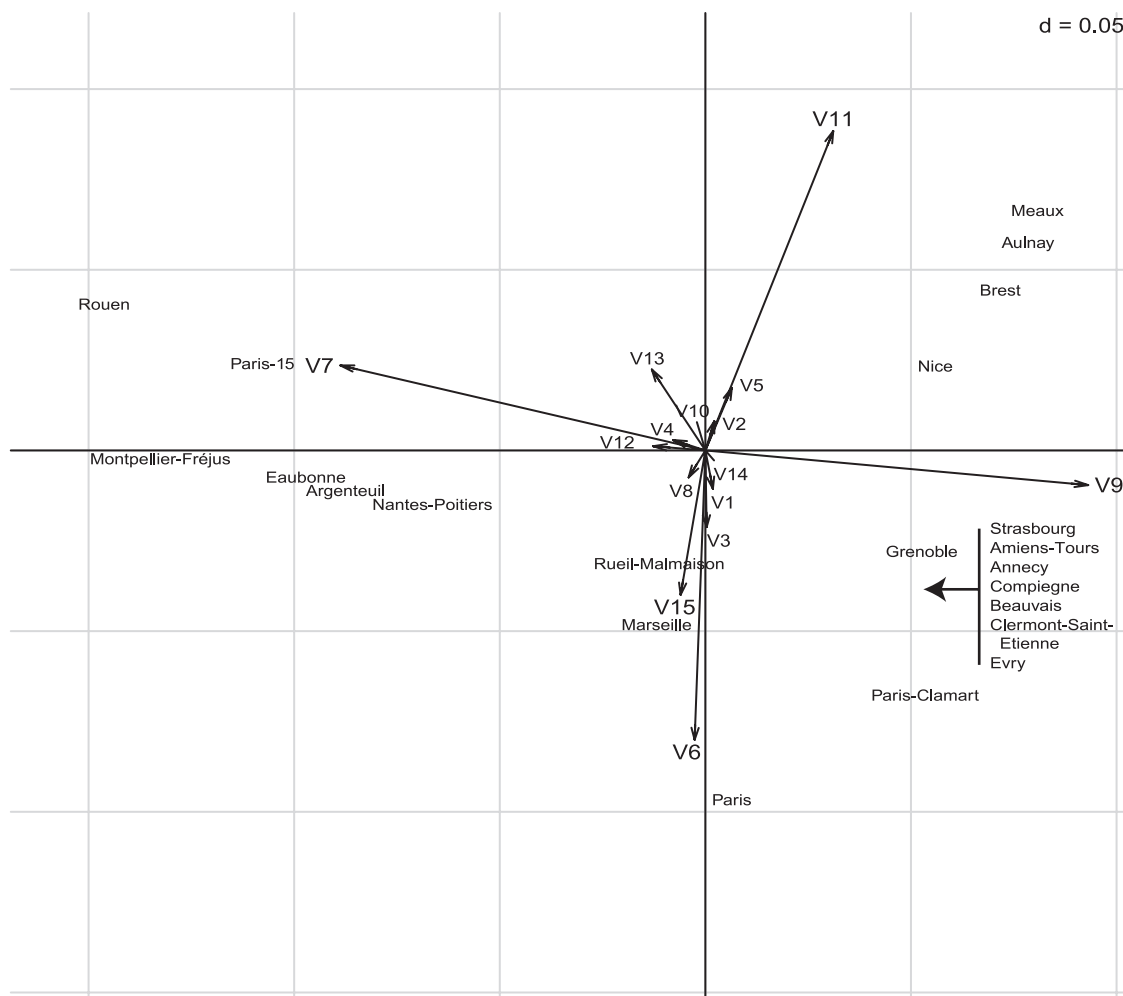


FIG. 3. Between-group analysis of the presence of IS elements in 86 *Legionella* isolates sampled from 26 cities. The 86 isolates shown in Tables 1 and 2 were included in the analysis. Differences between cities were highly statistically significant ( $P \approx 0.001$ ). The scale is given in the upper right corner of the graph as the value  $d$ , which is the size of the background grid.

The matrix distance allowed us to compute the average distance of each OTU to the others (Fig. 2). The OTU with the smallest distance was the reference strain with the profile Ref7, with 0.08 substitution per site, while the OTU with the longest distance was OTU 78, with 0.25. An inverse relationship between distance and number of isolates per OTU was thus found, suggesting the emergence of better-adapted isolates linked to the IS transposition-induced mutation rate.

**BGA.** Next, we analyzed our data according to the city of origin of the different isolates. All 86 isolates originated from 26 different cities (Tables 1 and 2). BGA was performed to investigate whether the IS distribution was affected by the city of origin of the different *L. pneumophila* isolates (Fig. 3). The raw data table comprised 86 rows corresponding to the 86 *Legionella* isolates analyzed here and 15 columns corresponding to the 15 *ISLpn11* copies labeled V1 to V15. Values in the table correspond to the presence (value of 1) or absence (value of 0) of each IS copy in each isolate. Since the IS copy number varies among isolates, these values were first transformed to row percentages, and the BGA was then computed to yield the mean values for the 26 cities. The Monte Carlo test of the

BGA shows that the difference between cities was statistically highly significant ( $P \approx 0.001$ ). The BGA was displayed under the form of a biplot with both IS copies and cities on the same graph, where long arrows correspond to IS copies that discriminated cities and shorter arrows correspond to common IS copies (Fig. 3). Each of four specific IS copies, V6, V7, V9, and V11, was characteristic of each of four groups of cities. The *Legionella* isolates coming from cities that are located in the same direction as one of these IS copies tended to possess that particular IS copy. It must be noted that the directions, symbolized by arrows, are important in both ways: for example, the position of Brest city on the graph is due to both the presence of the IS copy V9 and the absence of the IS copy V6, but not to the presence of the IS copy V11 in the isolates collected in Brest.

**Discriminatory power of RFLP-IS and identification of the source of infection.** We used here the data of RFLP-IS obtained with the 76 *L. pneumophila* Paris isolates (Table 2) to check whether the environmental isolates sampled during epidemiological investigations might be related to the pathogenic isolates sampled from the patients. For comparative purposes,

TABLE 5. Differentiation of 76 *L. pneumophila* Paris isolates sampled from patients and potential contaminating roots by SBT, MAb typing, and RFLP-IS<sup>a</sup>

Parameter	1 method			2 methods			3 methods (RFLP-IS + SBT + MAb)
	SBT	MAb	RFLP-IS	SBT + MAb	RFLP-IS + SBT	RFLP-IS + MAb	
Different profiles (no.)	4	6	14	8	17	23	25
Unique profiles (no.)	1	0	7	1	8	10	11
Related cases (%)	24 (96)	17 (68)	18 (72)	17 (68)	18 (72)	14 (56)	14 (56)
Simpson's index ( <i>D</i> ) (%)	14	53	80	58	82	88	90

<sup>a</sup> For each typing method and each combination of methods, the number of different and unique profiles is given together with the number of related cases which correspond to epidemiological investigations with one or more environmental isolates presenting the same profile as the corresponding isolate sampled from patients. Simpson's index (*D*) was calculated for each method or combination of methods.

each isolate was also tested by SBT and MAb typing. All data are given in the supplemental material in Table S1. First, we confirmed that the discriminatory power (*D*) of the RFLP-IS was far better (80%) compared to each of the SBT and MAb methods (14 and 53%, respectively). Moreover, when two typing techniques were simultaneously used, the best combinations always included the RFLP-IS typing, with *D* increasing to 82 and 88% when RFLP-IS was combined with SBT and MAb typing, respectively, while only 58% was obtained with both SBT and MAb typing (Table 5). Combining all three typing methods did not significantly improve *D*, which increased to 90%.

Second, we compared the different typing methods by investigating the same set of isolates for the level of relationship between environmental and clinical isolates sampled at the same geographical locations during the different epidemiological investigations (Table 5). At some extreme, the strong discriminating power of RFLP-IS may lead to the absence of detection of the isolate that may constitute the source of the infection in the environmental settings. However, while more than two-thirds of the cases allowed the recovery of the profile of the isolates sampled from a patient in environmental isolates, this was also true in about half of the cases when either RFLP-IS and MAb typing or all three typing methods were combined. Therefore, RFLP-IS *pn11* increased the confidence of detecting the source of infection by refining the molecular identification and differentiation ability between closely related isolates.

## DISCUSSION

Based on the available genome sequence of *L. pneumophila* Paris, which is one of the major pathogenic strains of this bacterial species, we identified the potentially functional IS elements. We used them as hybridization probes in RFLP experiments to evaluate their value as epidemiological tools, compared to other typing methods.

Discriminating genotyping methods are of high importance to precisely evaluate the genetic diversity in *L. pneumophila* and more generally in pathogenic bacteria, since it has important consequences for a precise understanding of epidemiology and persistence in the environment. Here, we used two sets of strains, including 10 independent clinical isolates and 76 *L. pneumophila* Paris isolates, with both environmental and clinical isolates sampled during 25 epidemiological investigations that occurred between 2001 and 2007. For this particular set of isolates, we showed that RFLP-IS had the highest discrimina-

tory power compared to MAb typing, PFGE, and SBT, which was described as a very efficient tool to generate epidemiological and highly discriminatory data in the closely related *L. pneumophila* serogroup 1 strains (46). This was even better when both RFLP-IS and MAb typing were combined. A very high level of genetic diversity was detected even for the closely related Paris isolates, confirming the genetic plasticity of *Legionella* (5). However, one major profile, Ref7, was repeatedly observed, and our phylogenetic analyses provided further support for a clonal population structure for *L. pneumophila* (10), with a dominant epidemic clone that showed diversification profiles.

In almost all cases, a common profile was detected by SBT between environmental and clinical isolates sampled from each geographical location (Table 5). Using the more discriminatory RFLP-IS tool, related profiles were observed in about three quarters of the cases. Combining RFLP-IS and MAb typing still allowed the detection of related profiles in more than half of the cases. It has to be emphasized that we used a small number of environmental and clinical isolates sampled from the same geographical locations. Yet, even under these conditions, we detected related profiles for more than half of the cases, and increasing the number of analyzed samples should allow an even better coverage. Therefore, the RFLP-IS tool not only allowed a high discrimination but also improved the confidence for detection of the source of infection compared to SBT and MAb typing approaches, which should result in a better control of legionellosis infections. This is a huge advantage of the technique in the context of epidemic situations. Comparison of clinical and environmental isolates using the RFLP-IS tool may offer for the first time the opportunity to rapidly and accurately identify the waterborne source of infection and to prevent extension of outbreaks by effective decontamination of the bacterial reservoir. By defining subgroups of *L. pneumophila* Paris isolates with potentially longer persistence in the environment and/or higher infectious traits for humans, the RFLP-IS tool may also reinforce current surveillance procedures of human-made water distribution systems recommended for legionellosis prevention purposes. It should, however, be noted that we used a relatively small number of isolates that belong mostly to the Paris ST1 group. Since the RFLP-IS method generated convincing results with this set of isolates, its broader use as an epidemiological tool can now be tested with other related isolates, including, for example, the Philadelphia 1 group.

The high discriminatory power that we detected was associ-



ated with the diversity generated by a single IS element, *ISLpn11*, suggesting that this mobile element is actively moving in the Paris isolates, thereby contributing largely to their polymorphism and genomic plasticity. Bacterial IS elements are known to transpose at a high frequency, depending on both environmental conditions and host factors (39). Moreover, identical copies of IS elements may serve as substrates for homologous recombination, which may lead to chromosomal rearrangements, like inversions, deletions, or duplications. Therefore, all the changes that we detected in the RFLP profiles using *ISLpn11* may be related to either transposition events or rearrangements associated with recombination between homologous copies of *ISLpn11*. The *ISLpn11* element therefore constitutes a very efficient genomic marker for epidemiological purposes in *L. pneumophila*. In some extreme cases, we could even detect *ISLpn11* copies that were specific to some cities. It would be of interest to understand the molecular biology and physiology linked to the transposition of this element. In particular, one could investigate whether its transposition is dependent upon differential environmental conditions, including the contact with eukaryotic host cells, and second, one could quantify the fitness of the strains and the *ISLpn11* link with particular genes up- or downregulated following IS transposition. To the best of our knowledge, nothing is currently known about the stability of *ISLpn11* in *L. pneumophila* and about its specificity of integration. We detected no obvious preferences either for target sites or for chromosomal localization in the distribution of *ISLpn11* in the sequenced genomes (Table 4), suggesting that it may transpose into various sites.

RFLP-IS was revealed to be a very reliable, simple, and easy-to-reproduce genotyping method, since data were obtained for all analyzed isolates. Moreover, the possibility to express the data as matrices with binary codes allows easy standardization and transfer of results. The availability of a wide range of epidemiological tools is of primary importance to investigate the extent and nature of genetic diversity of bacterial pathogens. This is particularly important for addressing epidemiological questions, identifying the source of infection episodes, and understanding the evolutionary processes driving their distribution and phenotypic properties.

#### ACKNOWLEDGMENTS

This work was supported by a grant from the French Centre National de la Recherche Scientifique, a Bonus Qualité Recherche grant from the University Joseph Fourier, and a grant from the Agence Française de Sécurité Sanitaire de l'Environnement et du Travail (AFSSET; grant EST-2007-1) (to D.S.). M.V. thanks the French Ministry of Education and Research for a Ph.D. fellowship and the University Joseph Fourier for ATER funding.

We thank J. H. Helbig for providing antibodies for MAb subgrouping. We thank all our colleagues for sending *Legionella* isolates to the National Reference Center.

#### REFERENCES

- Amemura-Maekawa, J., F. Kura, B. Chang, and H. Watanabe. 2005. *Legionella pneumophila* serogroup 1 isolates from cooling towers in Japan form a distinct genetic cluster. *Microbiol. Immunol.* **49**:1027–1033.
- Aurell, H., P. Farge, H. Meugnier, M. Gouy, F. Forey, G. Lina, F. Vandenesch, J. Etienne, and S. Jarraud. 2005. Clinical and environmental isolates of *Legionella pneumophila* serogroup 1 cannot be distinguished by sequence analysis of two surface protein genes and three housekeeping genes. *Appl. Environ. Microbiol.* **71**:282–289.
- Bukhari, Y. M., K. Koivu, and P. M. A. Tigerstedt. 1999. Phylogenetic analysis of *Acacia* (*Mimosaceae*) as revealed from chloroplast RFLP data. *Theor. Appl. Genet.* **98**:291–298.
- Cazalet, C., S. Jarraud, Y. Ghavi-Helm, F. Kunst, P. Glaser, J. Etienne, and C. Buchrieser. 2008. Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. *Genome Res.* **18**:431–441.
- Cazalet, C., C. Rusniok, H. Bruggemann, N. Zidane, A. Magnier, L. Ma, M. Tichit, S. Jarraud, C. Bouchier, F. Vandenesch, F. Kunst, J. Etienne, P. Glaser, and C. Buchrieser. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* **36**:1165–1173.
- Chessel, D., A.-B. Dufour, and J. Thioulouse. 2004. The ade4 package - I: one-table methods. *R News* **4**:5–10.
- Chien, M., I. Morozova, S. Shi, H. Sheng, J. Chen, S. M. Gomez, G. Asamani, K. Hill, J. Nuara, M. Feder, et al. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* **305**:1966–1968.
- Cleveland, W. S., E. Grosse, and W. M. Shyu. 1992. Local regression models, p. 309–376. In J. M. Chambers and T. J. Hastie (ed.), *Statistical models in S*. Wadsworth and Brooks/Cole, Boca Raton, FL.
- Cooper, V. S., D. Schneider, M. Blot, and R. E. Lenski. 2001. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J. Bacteriol.* **183**:2834–2841.
- Coscollá, M., M. J. Gosalbes, V. Catalan, and F. Gonzalez-Candelas. 2006. Genetic variability in environmental isolates of *Legionella pneumophila* from Comunidad Valenciana (Spain). *Environ. Microbiol.* **8**:1056–1063.
- Culhane, A. C., G. Perrière, E. C. Considine, T. G. Cotter, and D. G. Higgins. 2002. Between-group analysis of microarray data. *Bioinformatics* **18**:1600–1608.
- Deplano, A., M. Vanechoutte, G. Verschraegen, and M. J. Struelens. 1997. Typing of *Staphylococcus aureus* and *Staphylococcus epidermidis* strains by PCR analysis of inter-IS256 spacer length polymorphisms. *J. Clin. Microbiol.* **35**:2580–2587.
- 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.
- Dolédec, S., and D. Chessel. 1987. Rythmes saisonniers et composantes stationnelles en milieu aquatique. I. Description d'un plan d'observations complet par projection de variables. *Acta Oecol.* **8**:403–426.
- Eck, R. V., and M. O. Dayhoff. 1966. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, MD.
- Edagawa, A., A. Kimura, H. Doi, H. Tanaka, K. Tomioka, K. Sakabe, C. Nakajima, and Y. Suzuki. 2008. Detection of culturable and nonculturable *Legionella* species from hot water systems of public buildings in Japan. *J. Appl. Microbiol.* **105**:2104–2114.
- Farris, J. S. 1977. Phylogenetic analysis under Dollo's law. *Syst. Zool.* **26**:77–88.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* **5**:164–166.
- Fica, A. E., H. W. Horowitz, H. Lior, and F. C. Cabello. 1994. Demonstration of persistence of *Salmonella typhimurium* in an AIDS patient by molecular methods. *J. Clin. Microbiol.* **32**:2327–2330.
- Fry, N. K., S. Alexiou-Daniel, J. M. Bangsberg, S. Bernander, M. Castellani 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, B. Afshar, P. C. Lück, H. Meugnier, J. Etienne, R. Peduzzi, and T. G. Harrison. 2005. Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. *J. Clin. Microbiol.* **43**:2047–2052.
- 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.
- Ginevra, C., M. Lopez, F. Forey, M. Reyrolle, H. Meugnier, F. Vandenesch, J. Etienne, S. Jarraud, and M. Molmeret. 2009. Evaluation of a nested-PCR-derived sequence-based typing method applied directly to respiratory samples from patients with Legionnaires' disease. *J. Clin. Microbiol.* **47**:981–987.
- Glöckner, G., C. Albert-Weissenberger, E. Weinmann, S. Jacobi, E. Schunder, M. Steinert, J. Hacker, and K. Heuner. 2008. Identification and characterization of a new conjugation/type IVA secretion system (*trb/trt*) of *Legionella pneumophila* Corby localized on two mobile genomic islands. *Int. J. Med. Microbiol.* **298**:411–428.
- Harrison, T. G., B. Afshar, N. Doshi, N. K. Fry, and J. V. Lee. 2009. Distribution of *Legionella pneumophila* serogroups, monoclonal antibody subgroups and DNA sequence types in recent clinical and environmental isolates from England and Wales (2000–2008). *Eur. J. Clin. Microbiol. Infect. Dis.* **28**:781–791.
- Helbig, J. H., S. Bernander, M. Castellani Pastoris, J. Etienne, V. Gaia, S.

- Lauwers, D. Lindsay, P. C. Lück, T. Marques, S. Mentula, M. F. Peeters, C. Pelaz, M. Struelens, S. A. Uldum, G. Wewalka, and T. G. Harrison. 2002. Pan-European study on culture-proven Legionnaires' disease: distribution of *Legionella pneumophila* serogroups and monoclonal subgroups. *Eur. J. Clin. Microbiol. Infect. Dis.* **21**:710–716.
28. Helbig, J. H., E. Fernandez-Moreira, E. Jacobs, P. C. Lück, and M. Witt. 2006. Lipopolysaccharide architecture of *Legionella pneumophila* grown in broth and host cells, p. 261–264. In N. P. Cianciotto (ed.), *Legionella*: state of the art 30 years after its recognition. ASM Press, Washington, DC.
29. 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.
30. Huygens, F., G. R. Nimmo, J. Schooneveldt, W. J. Munckhof, and P. M. Giffard. 2002. Genotyping of methicillin-resistant *Staphylococcus aureus* by assaying for the presence of variable elements associated with *mecA*. *J. Clin. Microbiol.* **40**:3093–3097.
31. Ichikawa, K., T. Yagi, M. Moriyama, T. Inagaki, T. Nakagawa, K. Uchiya, T. Nikai, and K. Ogawa. 2009. Characterization of *Mycobacterium avium* clinical isolates in Japan using subspecies-specific insertion sequences, and identification of a new insertion sequence, IS*Mav6*. *J. Med. Microbiol.* **58**:945–950.
32. Joly, J. R., R. M. McKinney, J. O. Tobin, W. F. Bibb, I. D. Watkins, and D. Ramsay. 1986. Development of a standardized subgrouping scheme for *Legionella pneumophila* serogroup 1 using monoclonal antibodies. *J. Clin. Microbiol.* **23**:768–771.
33. Kluge, A. G., and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**:1–32.
34. Ko, K. S., H. Miyamoto, H. K. Lee, M. Y. Park, K. Fukuda, B. J. Park, and Y. H. Kook. 2006. Genetic diversity of *Legionella pneumophila* inferred from *rrpB* and *dotA* sequences. *Clin. Microbiol. Infect.* **12**:254–261.
35. Kozak, N. A., R. F. Benson, E. Brown, N. T. Alexander, T. H. Taylor, Jr., B. G. Shelton, and B. S. Fields. 2009. Distribution of *lag-1* alleles and sequence-based types among *Legionella pneumophila* serogroup 1 clinical and environmental isolates in the United States. *J. Clin. Microbiol.* **47**:2525–2535.
36. Lebreton, J. D., R. Sabatier, G. Banco, and A. M. Bacou. 1991. Principal component and correspondence analyses with respect to instrumental variables: an overview of their role in studies of structure-activity and species-environment relationships, p. 85–114. In J. Devillers, and W. Karcher (ed.), *Applied multivariate analysis in SAR and environmental studies*, Kluwer Academic Publishers, Dordrecht, Netherlands.
37. Lee, H. K., J. I. Shim, H. E. Kim, J. Y. Yu, and Y. H. Kang. 2010. Distribution of *Legionella* species from environmental water sources of public facilities and genetic diversity of *L. pneumophila* serogroup 1 in South Korea. *Appl. Environ. Microbiol.* **76**:6547–6554.
38. Luna, V. A., D. B. Jernigan, A. Tice, J. D. Kellner, and M. C. Roberts. 2000. A novel multiresistant *Streptococcus pneumoniae* serogroup 19 clone from Washington State identified by pulsed-field gel electrophoresis and restriction fragment length patterns. *J. Clin. Microbiol.* **38**:1575–1580.
39. Mahillon, J., and M. Chandler. 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725–774.
40. Margush, T., and F. R. McMorris. 1981. Consensus n-trees. *Bull. Math. Biol.* **43**:239–244.
41. Mouchtouri, V., E. Velonakis, and C. Hadjichristodoulou. 2007. Thermal disinfection of hotels, hospitals, and athletic venues hot water distribution systems contaminated by *Legionella* species. *Am. J. Infect. Control* **35**:623–627.
42. Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **76**:5269–5273.
43. Newton, H. J., D. K. Y. Ang, I. R. van Driel, and E. L. Hartland. 2010. Molecular pathogenesis of infections caused by *Legionella pneumophila*. *Clin. Microbiol. Rev.* **23**:274–298.
44. Papadopoulos, D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **96**:3807–3812.
45. Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya* **26**:329–359.
46. Ratzow, S., V. Gaia, J. H. Helbig, N. K. Fry, and P. C. Lück. 2007. Addition of *neuA*, the gene encoding *N*-acetylneuraminyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. *J. Clin. Microbiol.* **45**:1965–1968.
47. R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
48. Reichardt, K., E. Jacobs, I. Röske, and J. H. Helbig. 2010. *Legionella pneumophila* carrying the virulence-associated lipopolysaccharide epitope possesses two functionally different LPS components. *Microbiology* **156**:2953–2961.
49. Reimer, A. R., S. Au, S. Schindle, and K. A. Bernard. 2010. *Legionella pneumophila* monoclonal antibody subgroups and DNA sequence types isolated in Canada between 1981 and 2009: laboratory component of national surveillance. *Eur. J. Clin. Microbiol. Infect. Dis.* **29**:191–205.
50. Robinson, D. A., S. K. Hollingshead, J. M. Musser, A. J. Parkinson, D. E. Briles, and M. J. Crain. 1998. The IS*1167* insertion sequence is a phylogenetically informative marker among isolates of serotype 6B *Streptococcus pneumoniae*. *J. Mol. Evol.* **47**:222–229.
51. Rowbotham, T. J. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J. Clin. Pathol.* **33**:1179–1183.
52. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
53. Schilling, E. E. 1997. Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction site data. *Theor. Appl. Genet.* **94**:925–933.
54. Stout, J. E., and V. L. Yu. 1997. Legionellosis. *N. Engl. J. Med.* **337**:682–687.
55. Suzuki, H., Y. Kawamoto, O. Takenaka, I. Munechika, H. Hori, and S. Sakurai. 1994. Phylogenetic relationships among *Homo sapiens* and related species based on restriction site variations in rDNA spacers. *Biochem. Genet.* **32**:257–269.
56. Tamura, G. S., M. Herndon, J. Przekwas, C. E. Rubens, P. Ferrieri, and S. L. Hillier. 2000. Analysis of restriction fragment length polymorphisms of the insertion sequence IS*1381* in group B streptococci. *J. Infect. Dis.* **181**:364–368.
57. Thioulouse, J., and S. Dray. 2007. Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. *J. Stat. Softw.* **22**:1–14.
58. Tijet, N., P. Tang, M. Romilowich, C. Duncan, V. Ng, D. N. Fisman, F. Jamieson, D. E. Low, and C. Guyard. 2010. New endemic *Legionella pneumophila* serogroup 1 clones, Ontario, Canada. *Emerg. Infect. Dis.* **16**:447–454.
59. Torrea, G., V. Chenal-Francois, A. Leclercq, and E. Carniel. 2006. Efficient tracing of global isolates of *Yersinia pestis* by restriction fragment length polymorphism analysis using three insertion sequences as probes. *J. Clin. Microbiol.* **44**:2084–2092.
60. Vaidya, V. M., S. V. S. Malik, S. Kaur, S. Kumar, and S. B. Barbuddhe. 2008. Comparison of PCR, immunofluorescence assay, and pathogen isolation for diagnosis of Q fever in humans with spontaneous abortions. *J. Clin. Microbiol.* **46**:2038–2044.
61. Warren, R. M., T. C. Victor, E. M. Streicher, M. Richardson, G. D. van der Spuy, R. Johnson, V. N. Chihota, C. Loch, P. Supply, and P. D. van Helden. 2004. Clonal expansion of a globally disseminated lineage of *Mycobacterium tuberculosis* with low IS*6110* copy numbers. *J. Clin. Microbiol.* **42**:5774–5782.
62. Wellinghausen, N., C. Frost, and R. Marre. 2001. Detection of legionellae in hospital water samples by quantitative real-time LightCycler PCR. *Appl. Environ. Microbiol.* **67**:3985–3993.
63. Yáñez, M. A., C. Carrasco-Serrano, V. M. Barberá, and V. Catalán. 2005. Quantitative detection of *Legionella pneumophila* in water samples by immunomagnetic purification and real-time PCR amplification of the *dotA* gene. *Appl. Environ. Microbiol.* **71**:3433–3441.
64. Yoshida, T., N. Kondo, Y. A. Hanifah, and K. Hiramoto. 1997. Combined use of ribotyping, PFGE typing and IS*431* typing in the discrimination of nosocomial strains of methicillin-resistant *Staphylococcus aureus*. *Microbiol. Immunol.* **41**:687–695.
65. Zuerner, R. L., and C. A. Bolin. 1997. Differentiation of *Leptospira interrogans* isolates by IS*1500* hybridization and PCR assays. *J. Clin. Microbiol.* **35**:2612–2617.