Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species

Christel Cazalet, Sophie Jarraud, Yad Ghavi-Helm, et al.

Genome Res. 2008 18: 431-441
Access the most recent version at doi:10.1101/gr.7229808

References This article cites 45 articles, 25 of which can be accessed free at:
http://genome.cshlp.org/content/18/3/431.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

CRISPR and RNAi Genetic Screening.
Your new superpower.
LEARN MORE
CELLECTA

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © 2008, Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Letter

Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species

Christel Cazalet,[1,2] Sophie Jarraud,[3] Yad Ghavi-Helm,[1,2] Frank Kunst,[1,2] Philippe Glaser,[1,2] Jerome Etienne,[3] and Carmen Buchrieser[1,2,4,5]

[1]*Unité de Génomique des Microorganismes Pathogènes, Institut Pasteur, 75724 Paris Cedex 15, France;* [2]*CNRS URA 2171, 75724 Paris, France;* [3]*Centre National de Référence des Legionella, Laboratoire de Bactériologie INSERM U851, Faculté de Médecine, IFR 62, 69372 Lyon Cedex 08, France;* [4]*UP de Biologie des Bactéries Intracellulaires, Institut Pasteur, 75724 Paris Cedex 15, France*

Genomics can provide the basis for understanding the evolution of emerging, lethal human pathogens such as *Legionella pneumophila*, the causative agent of Legionnaires' disease. This bacterium replicates within amoebae and persists in the environment as a free-living microbe. Among the many *Legionella* species described, *L. pneumophila* is associated with 90% of human disease and within the 15 serogroups (Sg), *L. pneumophila* Sg1 causes over 84% of Legionnaires' disease worldwide. Why *L. pneumophila* Sg1 is so predominant is unknown. Here, we report the first comprehensive screen of the gene content of 217 *L. pneumophila* and 32 non-*L. pneumophila* strains isolated from humans and the environment using a *Legionella* DNA-array. Strikingly, we uncovered a high conservation of virulence- and eukaryotic-like genes, indicating strong environmental selection pressures for their preservation. No specific hybridization profile differentiated clinical and environmental strains or strains of different serogroups. Surprisingly, the gene cluster coding the determinants of the core and the O side-chain synthesis of the lipopolysaccaride (LPS cluster) determining Sg1 was present in diverse genomic backgrounds, strongly implicating the LPS of Sg1 itself as a principal cause of the high prevalence of Sg1 strains in human disease and suggesting that the LPS cluster can be transferred horizontally. Genomic analysis also revealed that *L. pneumophila* is a genetically diverse species, in part due to horizontal gene transfer of mobile genetic elements among *L. pneumophila* strains, but also between different *Legionella* species. However, the genomic background also plays a role in disease causation as demonstrated by the identification of a globally distributed epidemic strain exhibiting the genotype of the sequenced *L. pneumophila* strain Paris.

[Supplemental material is available online at www.genome.org.]

Although best known for its ability to cause severe pneumonia in persons with weakened immune defenses, *Legionella pneumophila* is one species of a large genus of bacteria that are ubiquitous in natural and drinking water supplies (Fields et al. 2002). The bacterium's survival and spread depend on the ability to replicate inside eukaryotic phagocytic cells like *Acanthamoeba castellanii* or other protozoa. Despite the presence of many different species of *Legionella* in aquatic reservoirs, the vast majority of human disease is caused by a single serogroup (Sg) of a single species, namely *L. pneumophila* Sg1, which is responsible for about 84% of all cases worldwide (Yu et al. 2002). Little is known about the distribution of *Legionella* species and serogroups in the environment. A recent study investigating 259 clinical and 3128 environmental strains isolated in France from 2001 to 2002 showed that *L. pneumophila* Sg1 accounts for 28.2% of environmental *Legionella* in contrast to 95.4% among clinical isolates (Doleans et al. 2004). Thus, the distribution of environmental *L. pneumophila* Sg1 differs significantly from the distribution of clinical *L. pneumophila* Sg1 strains. Furthermore, species like *Legionella anisa*, *Legionella dumofii*, or *Legionella feeleii* relatively frequently colonize water distribution systems but are rarely implicated in human disease (Muder and Yu 2002; Yu et al. 2002; Doleans et al. 2004). From these reports one can hypothesize that the high percentage of *L. pneumophila* Sg1 strains in human disease is not due to predominance in the environment, but rather connected with higher virulence of these strains for humans.

In previous work we have shown that the sequenced *L. pneumophila* strains possess highly diverse genomes (Cazalet et al. 2004; Cazalet and Buchrieser 2006). Genetic diversity was also suggested using different typing methods (Selander et al. 1985; Sabria et al. 2001; Aurell et al. 2003; Amemura-Maekawa et al. 2005), but attempts to correlate sequence types (ST) and epidemiological characteristics of strains were unsuccessful (Aurell et al. 2005). However, among *L. pneumophila* strains that cause human infections, significantly less diversity was found than among environmental isolates (Harrison et al. 2007) and a correlation between specific clones and the ecological niche was shown (Amemura-Maekawa et al. 2005).

In order to address these questions in more detail it is necessary to characterize the genetic variability of a large collection of strains. DNA-arrays are powerful tools to assess this inter- and intraspecies variability at the genome level. Examples are *Mycobacterium tuberculosis* (Behr et al. 1999; Brosch et al. 2002), *Vibrio cholerae* (Dziejman et al. 2005), *Staphylococcus aureus* (Fitzgerald et al. 2001; Trad et al. 2004), *Yersinia pestis* and *Yersinia pseudotuberculosis* (Hinchliffe et al. 2003), *Pseudomonas aeruginosa*

[5]**Corresponding author.**
E-mail cbuch@pasteur.fr; fax 33-1-45-68-87-86.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.7229808.

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

(Wolfgang et al. 2003), *Escherichia coli* (Dobrindt et al. 2003), or *Listeria monocytogenes* (Doumith et al. 2004). Knowledge gained from these surveys allowed scientists to develop typing tools, find vaccine candidates, and to understand the evolution and clonal expansion of particular virulent clones.

We have thus applied comparative genomic hybridization to 217 *L. pneumophila* and 32 *Legionella* strains belonging to 14 species with the aim to assess the genetic diversity and to search for correlations between functional traits, phylogeny, and epidemiological characteristics.

Results

Strain selection and macroarray conception

A well-characterized collection of 217 *L. pneumophila* and 32 *Legionella* strains belonging to 14 different species was selected based on epidemiological characteristics and geographical origin (Table 1; Supplemental Table S1). From the 217 *L. pneumophila* strains, 119 are clinical isolates, of which 37 were isolated during epidemics and 82 from sporadic cases. Eighty-nine are environmental isolates, of which 25 are an identified source of a clinical case or an epidemic, 64 are not disease related, and nine are monoclonal antibody (Mab)-type reference strains. A total of 151 strains are of Sg1, the remaining 66 belong to serogroups 2–15. The strains were isolated between 1980 and 2005 in Africa, Asia, Europe, and North America. All isolates were genomically characterized and compared by the use of a specifically designed focused macroarray. Based on the rationale that genes present in all three sequenced *L. pneumophila* strains (Cazalet et al. 2004; Chien et al. 2004) represent the core genome of this species, we selected 1027 probes representing the known flexible gene pool of *L. pneumophila*. In addition, 142 known and putative virulence genes of *L. pneumophila* were selected to study the correlation between the presence of virulence genes and epidemiological characteristics. These probes correspond to genes like those encoding types I, II, and IV secretion systems, certain effectors, virulence gene regulators, flagellar genes, eukaryotic-like genes, or the cluster required for lipopolysaccharide (LPS) synthesis (Luneberg et al. 2000) (Supplemental Table S2). Hybridization results were confirmed by hybridizing 10 reference strains to a multiple whole-genome microarray previously described (Brüggemann et al. 2006b) (see Supplemental information).

Comparative genomic analysis reflects the phylogeny of *Legionella*

Based on the hybridization data indicating presence and absence of genes, we used hierarchical clustering to build bifurcating trees to identify strains of similar gene content. As presented in Figure 1, the phylogeny deduced from these data revealed four main lineages (I, II, III, and IV) and distinguished *L. pneumophila* strains clearly from non-*pneumophila* strains. To avoid a possible bias due to loss or acquisition of large gene clusters en bloc, we weighted the results by considering each missing contiguous gene cluster as a single event. However, the same subgrouping was obtained. Lineage I contained 58 strains forming three subgroups (I.1–I.3), whereas lineage II is a small but distinct group of six strains, five of which were isolated during the epidemic in Lens (France, 2003/2004), the sixth was isolated in Germany. Lineage III was the most heterogeneous as it was subdivided into five groups (III.1–III.5) containing, respectively, 55, 11, 43, nine, and one strain. Finally, 34 strains, including the sequenced strain Philadelphia-1 compose lineage IV with two subgroups (IV.1 and IV.2) (Fig. 1).

To evaluate whether the phylogenies derived from the genomic profiles reflect phylogeny of the species as deduced from classical sequence typing, we determined the *rpoB* sequence of all *L. pneumophila* strains as previously described (Ko et al. 2006). *L. micdadei* ATCC33218 was used as an outgroup. The *rpoB*-derived phylogenetic tree (Supplemental Fig. S1) corresponded almost perfectly with the gene content-based tree (Fig. 1). All but one lineage I strain exhibited sequence type 1 (ST1). Consistent with its specific gene content, lineage II had a specific ST2. Like the DNA-array data, *rpoB* analysis showed that lineage III is constituted of isolates with broader genomic background, as 18 different STs were present among these strains. Finally, lineage IV is homogeneous, as all strains belonged to ST9. In conclusion, the

Table 1. General characteristics of the 249 *Legionella* strains studied

Species	Origin	Strains tested	Serogroup	Country of isolation[a]
L. pneumophila	Human-epidemic	37	1,2,4,10	Fr, Ch, Ge, Se, Nl
	Human-sporadic	82	1,2,3,5,6,8,10,12	Fr, Ge, At, Jp
	Environmental isolate of confirmed cases	25	1,4,10	Fr, Se, Ge, At
	Environmental	64	1–15	Fr, Ge, At, Jp, Sn
	Mab-type reference strains	9	1–15	
L. anisa	Human-sporadic, environmental, ATCC reference	3		
L. bozemanii	Human-sporadic	1		
L. dumoffii	Environmental, ATCC reference strain	2		
L. erythra	Environmental, ATCC reference strain	2		
L. gormanii	Human-sporadic, environmental, ATCC reference	3		
L. jordanis	Environmental, ATCC reference	2		
L. longbeachae	Human-sporadic	4	1	
	Environmental, ATCC reference strain	2	1	
L. micdadei	Human-sporadic, environmental, ATCC reference	3		
L. moravica	Environmental, ATCC reference strain	2		
L. parisiensis	Human-sporadic	1		
L. quinlivanii	Environmental, ATCC reference strain	2		
L. rubrilucens	Environmental, ATCC reference strain	2		
L. taurinensis	Environmental	1		
L. tusconensis	Human-sporadic	1		

[a](Fr) France; (Se) Sweden; (Ge) Germany; (Ch) Switzerland; (At) Austria; (Jp) Japan; (Sn) Senegal; (Nl) The Netherlands.

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

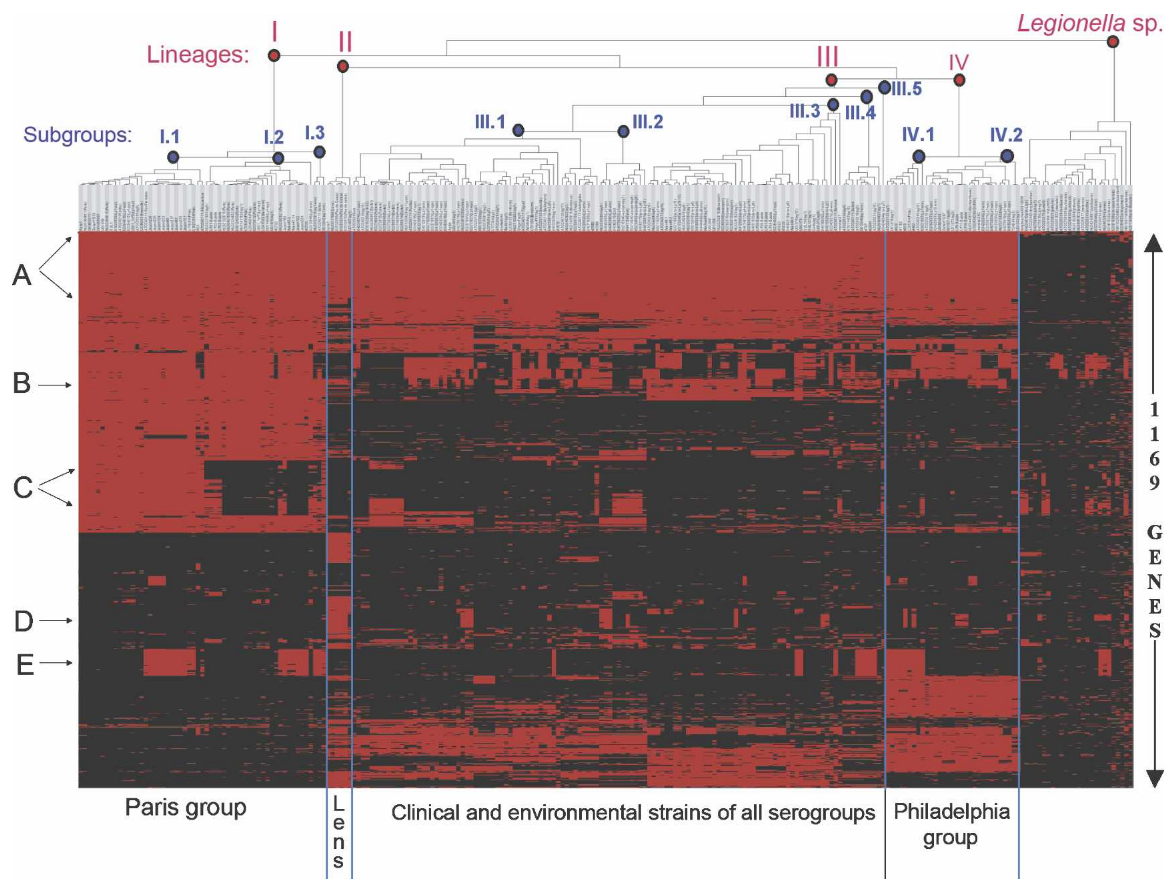

Figure 1. *Legionella* genetic diversity. Red and black denote the presence and absence of genes, respectively. Dendrogram shows estimates of genomic relationships as obtained by hierarchical cluster analysis. Blue lines indicate lineages I, II, III, and IV, and non-*pneumophila* strains. (*A*) Virulence genes; (*B*) Lvh region; (*C*) Plasmid of strain Paris (pLpp); (*D*) plasmid of strain Lens (pLpl); (*E*) 65-kb PAI described in strain Philadelphia-1.

rpoB-based sequence types match with the classification obtained from DNA-array analysis. Thus, our gene content-based results reflect the phylogeny of the species *L. pneumophila*.

Very similar results were obtained when DNA-array data were subjected to principal component analysis (PCA), a method of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. This method allows the number of variables in a multivariate data set to be reduced, whilst retaining as much as possible of the variation present in the data set (Härdle and Simar 2006). Based on the binary matrix indicating gene presence and absence within the species *L. pneumophila*, four lineages were defined by PCA that corresponded to lineage I–IV as defined by hierarchical clustering. Lineage I was the most distinct, indicating a specific gene content for these strains (Fig. 2).

Phylogenetic lineages do not correlate with geographical or epidemiological characteristics

Analysis of the distribution of the *L. pneumophila* strains among the different lineages with respect to geographical origin showed that no correlation with overall gene content data was found. Lineage I included isolates from all countries and continents investigated. Lineage III strains were isolated in six different countries and two continents, and lineage IV strains come from six countries. The only exception is lineage II that contains only French and one German strain.

Likewise, analysis of the distribution of the different serogroups showed that except for lineage II containing six Sg1 strains and lineage I containing 55 Sg1 and only three non-Sg1

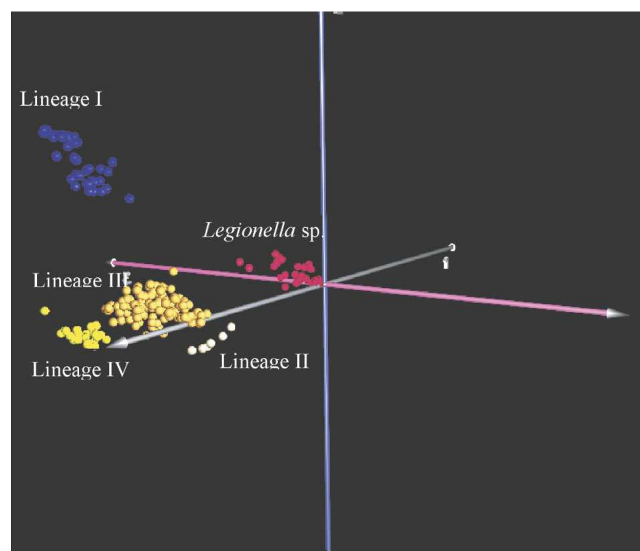

Figure 2. Principal component analysis based on gene content of 217 *L. pneumophila* and 32 *Legionella* sp. TMEV-4 software was used, based on a binary matrix indicating presence and absence of the investigated genes.

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

strains, isolates of different serogroups were found in all lineages. The same was true for clinical and environmental isolates, which were present in each lineage. Thus, clustering according to genomic profiles is independent of geographical origin, epidemiological characteristics, and largely independent of serogroups (Table 2; Supplemental Table S1).

Known virulence genes and eukaryotic-like genes are conserved throughout all lineages

Our data revealed that known and putative virulence genes of *L. pneumophila* were highly conserved throughout the species. As shown in Figure 1 and Supplemental Table S3, genes coding the Dot/Icm T4SS, T1SS, T2SS, and the Tat system were present in all *L. pneumophila* strains. In contrast, these genes are absent or highly divergent in all non-*L. pneumophila* species tested. Genes coding regulatory proteins implicated in virulence gene regulation like *letA/letS*, *csrA*, and *cpxR/cpxA* are also present in all strains. Exceptions are the *rtxA* and *ralF* genes. As depicted in Supplemental Figure S2A, the genetic locus coding RtxA, which contributes to entry, adherence, cytotoxicity, pore formation (Cirillo et al. 2001), and intracellular trafficking in amoebae (Cirillo et al. 2002) shows a variable organization. In strain Paris, the N-terminal part of this protein contains ~30 highly conserved tandem repeats of 549 bp fused to the gene *arpB* and a C-terminal part containing the typical glycine-rich nonapeptide motifs characteristic of Rtx toxins. Interestingly, strains Lens and Philadelphia each exhibit different N-terminal tandem repeat motifs (Supplemental Fig. S2A) (Cazalet et al. 2004; Cazalet and Buchrieser 2006). Among the 217 strains investigated, 60, eight, and 69 strains show an organization that resembles the ones of strains Paris, Lens, or Philadelphia, respectively. Eight new combinations of the N-terminal tandem repeats were identified (Supplemental Fig. S2B). Finally, 33 strains did not contain any of the four known repeats. Sequencing of the repeat-containing region identified two new motifs (Supplemental Fig. S2C). In addition, five of the six *L. pneumophila* Sg2 strains (ATCC33154, 98,010,759, HL02383010, L3739/04, LT77/04) lacked the gene *ralF,* coding for a Dot/Icm substrate implicated in the establishment of the replicative vacuole (Nagai et al. 2002). Sequence analysis revealed that in these five strains, *ralF* was deleted due to homologous recombination between an 11-bp direct repeat 18 bp upstream of the start codon and 3 bp downstream from the stop codon of *ralF*. This example shows that despite the high conservation of virulence genes in the species *L. pneumophila*, some strain-specific variations may influence the virulence of certain strains.

The identification of a high number of eukaryotic-like genes in the genome sequence of *L. pneumophila* was one of the most interesting findings of the genome analysis (Cazalet et al. 2004). Indeed, some of these proteins have meanwhile been shown to be implicated in host pathogen interactions (Newton et al. 2007; Sansom et al. 2007). As regards the eukaryotic-like genes (Cazalet et al. 2004; de Felipe et al. 2005; Brüggemann et al. 2006a) our analysis revealed that they are present only in *L. pneumophila* strains and are absent or far less conserved in non-*pneumophila* strains. As listed in Table 3, the two genes encoding homologs of apyrase (*lpp1033* and *lpp1880*), the gene encoding an homolog of the eukaryotic glucoamylase (*lpp0489*), the gene homologous to eukaryotic sphingosine-1-phosphate lyase (*lpp2128*), and a gene encoding a putative serine/threonine protein kinase (STPK) (*lpp0267*) are conserved in all *L. pneumophila* strains. The F-box proteins encoded by *lpp2082* and *lpp0233* that are present in the three sequenced genomes were present in 211 and 204 of the 217 strains, respectively. The high conservation of the eukaryotic-like proteins further strengthens the hypothesis that they are important for intracellular multiplication or persistence in the environment. However, variations in these proteins were also observed. For example, the U-box protein encoded by gene *lpp2887* in strain Paris that is absent from the other two sequenced strains is also less conserved among the tested strains, as it was present in only 155 of the 217 strains, or the two genes (*lpp1439* and *lpp2626*) coding additional STPKs are absent or divergent in the nine strains of subgroup III.4. Likewise, many of the ankyrin repeat proteins were absent in several strains (Table 3).

Horizontal gene transfer of mobile genetic elements among the different lineages and species

As previously shown, the *L. pneumophila* genomes contain at least three large plasticity zones. The region carrying the Lvh T4ASS (Cazalet et al. 2004; Chien et al. 2004), the 65-kb pathogenicity island (PAI) (Brassinga et al. 2003), and a 100-kb region containing several genes encoding efflux transporters for heavy metals and other toxic substances (Chien et al. 2004). As shown in Figure 1, these genetic elements were distributed heterogeneously among the investigated strains, indicating that they are mobile and can be exchanged by horizontal gene transfer. For example, the region carrying the Lvh T4SS was present in 147 of 217 *L. pneumophila* strains. Whereas Lvh was consistently present in lineage I, its distribution was heterogeneous in lineages II, III, and IV. This region exhibits a mosaic structure, as indicated by flanking sequences of the *lvh* genes that are different in each of the three sequenced strains (Cazalet and Buchrieser 2006). The same region also showed different rearrangements among the 147 strains (data not shown). All Lvh-positive strains contain a common phage integrase (*lpp0194-lpl0179*) and a phage excisionase (*lpp0193-lpl0178*) as well as a *traA*-like gene, suggesting that this region was acquired through various conjugative elements or phages, explaining their different flanking regions.

In contrast to previous reports arguing that the 65-kb PAI is specific for strain Philadelphia-1, we show here that it is present in 41 *L. pneumophila* strains belonging to all different lineages, and most interestingly, also in the three *L. anisa* strains tested (Fig. 1). Thus, horizontal gene transfer does not only occur among *L. pneumophila* strains, but also takes place among different *Legionella* species.

The plasmids of strains Paris (pLpp; 130kb) and Lens (pLpl, 60kb) also show a heterogeneous distribution, as they can be present completely, present in part, or be absent. pLpp for instance, is conserved in lineage I.1, but is present only in part or is missing completely in lineage I.2 strains. In the strains where pLpp is present only in part, a 85-kb region carrying the two-component system *lrpR-lskS* implicated in the virulence of *L. longbeachae* (Doyle and Heuzenroeder 2002) was present, whereas the 45-kb region (*plpp0001–plpp0045)* bordered by an integrase (*plpp0140*) and coding a homolog of CsrA and for plasmid-related functions like partition and replication proteins and the conjugation transfer proteins Tra, is missing. As integration and circularization of plasmids have been shown for *Legionella* (Luneberg et al. 2001; Doleans et al. 2006), it is tempting to assume that parts of pLpp may, under certain conditions, integrate into the chromosome.

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 2. Distribution of the 217 *L. pneumophila* strains among the four phylogenetic lineages and their epidemiological characteristics

	Number of strains per lineage				Number of strains per lineage						
	Lineage I				Lineage III					Lineage IV	
	I.1	I.2	I.3	Lineage II	III.1	III.2	III.3	III.4	III.5	IV.1	IV.2
Origin	29	26	3	6	55	11	43	9	1	12	22
Countries[a]	Fr, Ch, At, Ge, Jp, Sn, Nl	Fr, USA, At, Ge, Se, Jp	Fr	Fr, Ge	Fr, Ge, At, USA	Fr	Fr, Ge, USA, Se, Nl	Fr, USA	Fr	Fr, USA, At, Ge	Fr, Se Au, Ge, Sn
Human epidemic strains (%)	**4 (13.8)**	**4 (15.3)**	**0 (0)**	**3 (50)**	**10 (18.2)**	**2 (18.2)**	**5 (11.6)**	**0 (0)**	**0 (0)**	**5 (41.7)**	**4 (18.2)**
Human sporadic strains (%)	**9 (31)**	**10 (38.5)**	**0 (0)**	**1 (16.7)**	**26 (47.3)**	**3 (27.3)**	**12 (27.9)**	**4 (44.4)**	**0 (0)**	**6 (50)**	**11 (50)**
Environmental strains, disease related (%)	**4 (13.7)**	**4 (15.4)**	**0**	**2 (33.3)**	**5 (9)**	**2 (18.2)**	**6 (13.9)**	**0**	**0**	**0 (0)**	**2 (9)**
Environmental strains (%)	10 (34.5)	8 (30.8)	2 (66.7)	0 (0)	10 (18)	4 (36.4)	19 (44.2)	4 (44.4)	1 (100)	1 (8.3)	5 (22.7)
Mab-type reference strains	2	0	1 (33.3)	0	4	0	1	1	0	0	0
Sg1 strains (%)	**28 (96.5)**	**24 (92.3)**	**3 (100)**	**6 (100)**	**42 (76.3)**	**5 (45.4)**	**13 (30.2)**	**4 (44.4)**	**0 (0)**	**11 (91.7)**	**15 (68.2)**
Non-Sg1 strains (%)	1 (3.4)	2 (7.7)	0 (0)	0 (0)	13 (23.6)	6 (54.5)	30 (69.8)	5 (55.5)	1 (100)	1 (8.3)	7 (31.8)

Numbers in parentheses, percentages of strains/subgroup. Bold, clinical strains; italics, environmental strains.
[a](Fr) France; (Se) Sweden; (Ge) Germany; (Ch) Switzerland; (At) Austria; (Jp) Japan; (Sn) Senegal; (Nl) The Netherlands.

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 3. Occurrence of selected eukaryotic-like genes in 217 *L. pneumophila* strains

Gene	Annotation	Occurrence in 217 *L. pneumophila* strains tested
lpp1880	Putative apyrase	217
lpp0267	Putative serine threonine protein kinase	217
lpp2128	Putative sphingosine-1-phosphate lyase	217
lpp1033	Putative apyrase	217
lpp0489	Putative glucoamylase	217
lpp1683	Ankyrin repeat protein	217
lpp0126	Ankyrin repeat protein	217
lpp2641	Putative sphingomyelin phosphodiesterase	216
enhC	EnhC	216
lpp0547	Ankyrin repeat protein	215
lpp1933	Putative sphingosine kinase	213
ralF	RalF	211
lpp1439	Putative serine threonine protein kinase	210
lpp2082	F-box and ankyrin repeat protein	211
lpp0233	F-box protein	204
lpp2626	Putative serine threonine protein kinase	203
lpp0321	RNA-binding protein precursor	197
lpp2887	U-box protein	155
lpp1905	Ankyrin repeat protein	106
lpp2058	Ankyrin repeat protein	89
lpp2486	F-box protein	71
lpp2065	Ankyrin repeat protein	46

Predominance of *L. pneumophila* Sg1 in legionellosis seems to be linked to its specific lipopolysaccharide

When analyzing the position of *L. pneumophila* Sg1 strains on the phylogenetic tree, it is apparent that Sg1 strains of clinical and environmental origin were present in all lineages. Thus, disease-related Sg1 strains are not defined by a specific genetic profile, but have diverse genomic backgrounds. Nevertheless, we found a highly specific correlation between serogroup information and gene content of a 33-kb segment encoding proteins involved in LPS biosynthesis. As shown in Figure 3, the 33 kb absent from all non-*L. pneumophila* strains can be subdivided in block A (13 kb) and block B (20 kb). While almost all genes of block A (*lpp0814–lpp0826*) were present in all serogroups of *L. pneumophila*, the majority of genes of block B (*lpp0827–lpp0843*) were present only in Sg1 (the only exception was Sg1 strain HL04016006). This suggests that this part of the LPS cluster, determining the serogroup of *L. pneumophila*, can be transferred horizontally. In order to test this hypothesis in more detail, we hybridized the Sg9 reference strain ATCC35289, a rare non-Sg1 isolate that clustered in lineage I, with a multiple genome microarray described previously (Brüggemann et al. 2006b). Results revealed that this Sg9 strain and strain Paris (Sg1) had a nearly identical gene content. Of the 3077 genes of strain Paris, only 31 were missing from the Sg9 strain, among them the 16 genes constituting block B of the LPS gene cluster. This indicates that these genes are specific for Sg1 strains and supports the hypothesis of horizontal gene transfer of this part of the LPS cluster (Supplemental Table S4). Closer inspection of the DNA-array results of the 249 strains showed that three genes (*wzt*, *wzm*, *lpp0831*) are only present in Sg1 isolates. Thus, they represent interesting marker genes for rapid Sg1 identification. As LPS is not only the basis for the classification of serogroups, but is also a major immunodominant antigen of *L. pneumophila*, and membrane vesicles shed by virulent *L. pneumophila* containing LPS have been shown to be sufficient to inhibit phagosome-lysosome fusion (Fernandez-Moreira et al. 2006), our results suggest that LPS of Sg1 is implicated in the predominance of Sg1 strains in human disease (Yu et al. 2002; Doleans et al. 2004) compared with other serogroups of *L. pneumophila* and other *Legionella* species.

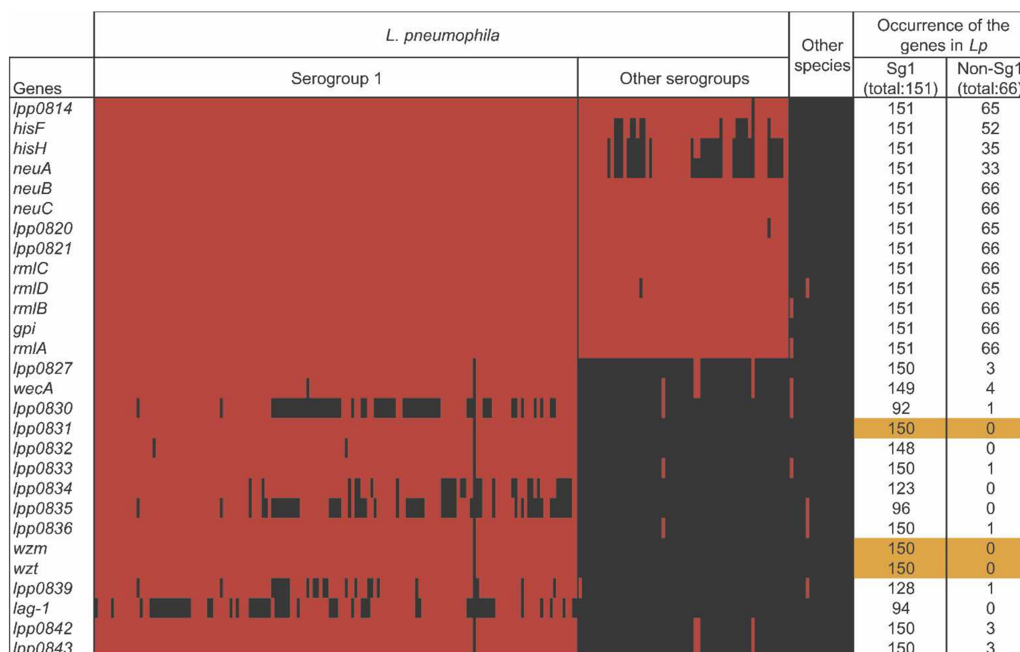

Figure 3. Distribution of the 30-kb gene cluster encoding LPS biosynthesis proteins in the 249 strains studied. Red and black denote the presence and absence of genes, respectively. Dendrogram was obtained by hierarchical cluster analysis based on the presence and absence of genes. The two *right* columns indicate the presence of the respective genes among the 151 tested *L. pneumophila* Sg1 and the 61 non-Sg1 strains. (Orange) Genes present in all and only Sg1 strains.

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

L. pneumophila Paris is a worldwide-distributed epidemic clone

Analysis of the strain distribution in each lineage revealed that lineage I contained the sequenced Paris isolate (Cazalet et al. 2004) and all other French isolates previously classified as "Paris" strains according to their PFGE profile (Aurell et al. 2003). Based on this finding and results from PCA that clearly separated lineage I strains from all other strains (Fig. 2), we analyzed this group of strains in more detail. As listed in Supplemental Table S1, lineage I contained isolates from many different countries in Europe (France, Sweden, Germany, Austria, Switzerland, the Netherlands), but also the United States, Japan, and Africa (Senegal). Furthermore, as presented in Figure 4, PFGE analysis of representative isolates of lineage I strains showed nearly indistinguishable profiles that were very different from those obtained for strains of the other three lineages. Thus, our DNA-array data together with the PFGE analysis identified for the first time that a globally distributed *L. pneumophila* clone exists that is implicated in outbreaks and sporadic cases of Legionnaires' disease. Clonality of the Paris strains is also emphasized through *rpoB* sequence typing, as all strains of lineage I are of ST1. Only minor differences in gene content among lineage I strains were observed. Furthermore, the variations are mainly present within the large mobile genetic elements and might thus be the result of few genetic recombination events. Thus, *L. pneumophila* strain Paris is a highly stable or recently emerged, worldwide distributed epidemic clone.

Similarly, lineage IV contained 34 strains sharing a genetic profile resembling that of the sequenced Philadelphia-1 isolate. These isolates were of ST9 but belonged to four different serogroups (Sg1, Sg3, Sg6, and Sg12). They are present in six countries and three continents (Supplemental Table S1). Thus, the group of Philadelphia-like strains seems to be also widely distributed, but it does not exhibit a highly specific gene content, as only 20 specific genes, mostly of unknown function, were identified (Supplemental Table S5).

In contrast, a search for specific genes of the "Paris" strains, which might be implicated in their apparent fitness in the environment and in human disease, identified 60 genes, 25 of which are clustered in three chromosomal regions (*lpp0068–lpp0083*, *lpp0291–lpp0297*, *lpp2111–lpp2125)* (Supplemental Table S6). Two of these regions show characteristics of genomic or pathogenicity islands. Region *lpp0068–lpp0083* is bordered by an Asn tRNA and has a highly variable GC content (29%–43%), suggesting that it was acquired through horizontal gene transfer. It carries three genes (*lpp0077-lpp0078-lpp0079*) coding proteins with similarity to the cation/multidrug efflux pump cluster *ceaABC*, which were recently shown to be induced intracellularly (Brüggemann et al. 2006b). Furthermore, a putative transcriptional regulator and *lpp0082,* predicted to encode a protein similar to serine/threonine protein kinases, are present. Region *lpp2111–lpp2125* is 11.3 kb long and is inserted in a Lys tRNA gene. It encodes an integrase, two phage-related genes, a putative SinR-like transcriptional regulator, and eight genes coding proteins with similarity to eukaryotic proteins having a low GC content of around 30%. Region *lpp0291–lpp0297* (GC content of 34%–42%) contains the genes *cyoABCD* encoding a specific *o* type quinol cytochrome oxidase and an homolog of proline/betaine transporter ProP. Five genes with similarity to eukaryotic genes (*lpp1824*, *lpp2139*, *lpp2168*, *lpp2485*, and *lpp2486*), *lpp0779* encoding an autotransporter and *lpp0197* and *lpp0198* encoding an adenine specific DNA methylase and a type-III restriction-modification enzyme

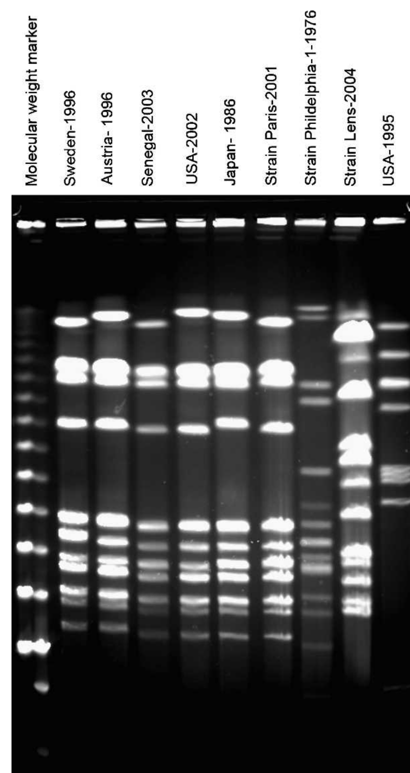

Figure 4. *Sfi*I DNA macrorestriction patterns of *L. pneumophila* clones Paris from geographically diverse origins. All strains clustered in lineage I. (Lane *1*) Molecular weight marker, kb. Strain designations (Supplemental Table S1): (lane *2*) LP22; (lane *3*) Wien 47–14; (lane *4*) 11559; (lane *5*) 2735; (lane *6*) NIIB228; (lane *7*) Paris; (lane *8*) Lens; (lane *9*) Philadelphia-1; (lane *10*) 1515.

are also highly specific for the Paris isolates. Thus, strain Paris seems to have acquired a specific set of genes.

Discussion

The genus *Legionella* is genetically highly diverse

In comparison with infectious agents that are known to cause human disease since ancient times (Zink et al. 2002), *L. pneumphila,* the causative agent of Legionnaires' disease is a recently recognized human pathogen whose transmission via aerosols is facilitated by modern warm-water distribution and air-conditioning systems as a consequence of altering the environment for human benefit (Muder et al. 1986). The majority of clinical cases are caused by *L. pneumophila* serogroup 1 (Sg1). As shown in our analysis, *L. pneumophila* strains were subdivided into four main genetic lineages (I–IV), but no specific correlation between the overall genotype and the potential ability to cause disease was found. Indeed, the genetic background of *L. pneumophila* was highly diverse. The main source of diversity seems to result from horizontal gene transfer of mobile genetic elements among *L. pneumophila*, but also between different *Legionella* species. Furthermore, plasmid excision and integration appear to constitute a particular mechanism of genome plasticity in *Legionella*. For example, the Lvh T4ASS can be present in an integrated and excised form (Cazalet et al. 2004; Chien et al. 2004; Doleans et al. 2006), similar to a 30-kb element specific for strain

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Olda associated with loss of virulence (Luneberg et al. 2001) and the two recently reported mobile elements carrying T4ASS in strain Corby (Glöckner et al. 2007). In addition, our results suggest that the 130-kb plasmid of strain Paris may integrate into the genome associated with loss of the genes necessary for plasmid replication. Importantly, most of these elements share a paralog of CsrA, a protein described as a repressor of transmission traits and an activator of replication leading to the hypothesis that the CsrA paralogs regulate the switch between integrated and circular forms in a growth phase-dependent manner.

However, despite genome plasticity, known virulence genes and also many recently identified eukaryotic-like genes are highly conserved throughout the species *L. pneumophila*, but are absent or highly divergent in non-*L. pneumophila* strains. This finding could mean that during their coevolution with protozoa, *L. pneumophila* strains have acquired a specific gene repertoire, in part from their hosts, that enables them now to more efficiently infect humans than other *Legionella* sp. Diversity in some of these genes like the ankyrin protein family or the additional STPKs might be involved in subtle differences, giving *L. pneumophila* advantages in the different environmental niches, but perhaps also the ability to persist in the human host. In particular, this hypothesis seems to hold true for the LPS of *L. pneumophila* Sg1 strains.

L. pneumophila serogroup 1 isolates carry a particular LPS gene cluster

In contrast to observations with respect to the overall gene content, where no clear correlation between genotype and serogroup of the isolate was detectable, closer inspection of particular gene clusters within the array data indicated that Sg1 strains harbor a particular set of genes involved in lipopolysaccharide synthesis (LPS cluster) that is absent from strains of other serogroups and species (Fig. 3). Interestingly, phylogenies derived from gene content and *rpoB* sequencing analysis showed that Sg1 strains are distributed in all different phylogenetic lineages (Fig. 1; Supplemental Fig. S1). This result suggests that the LPS cluster may be transferred horizontally among *L. pneumophila* strains similar to what is reported for other bacteria. Furthermore, we hypothesize that LPS of Sg1 per se may account for the high percentages of *L. pneumophila* Sg1 in human disease, as its predominance in legionellosis is independent of the genomic background but highly correlated with the LPS cluster determining the serogroup. Since LPS is highly antigenic, the variation at LPS loci is thought to be an advantage for evading the host immune system. For pathogens exposed to a strong immune response like *Haemophilus influenzae* or *Neisseria meningitidis*, the diversification of O-antigen-coding genes is probably due to the frequency-dependent selection imposed by the host immune system. However, this might not be true for the evolution and the maintenance of LPS encoding loci in *L. pneumophila*, as the human host appears to be a dead end for transmission and distribution of this pathogen. However, selection might be driven by the protozoan hosts, as different protozoa might recognize different forms of LPS with different efficiencies. Similar to what was proposed for the variability of the Fir–IcmQ protein pair (Feldman et al. 2005), variability of the LPS-encoding loci may be due to two forces: selection for protozoan cells that can avoid *Legionella* infection and selection for *Legionella* strains that can infect a broad host range of protozoa. Interestingly, it was recently suggested for *Salmonella* that diversifying selection of O-antigen is mediated by protozoa in the host intestinal environment (Wildschutte et al. 2004). Intestinal protozoa recognize antigenically diverse *Salmonella* with different efficiencies, and differences solely in the O-antigen are sufficient to allow for prey discrimination. It will be challenging in the future to understand the specificities of LPS of Sg1 in virulence for humans.

Given the importance of LPS, further investigation of the LPS cluster in strains of different serogroups should allow us to determine the genetic basis of serospecificity to explain virulence differences and to improve genotyping methods. As a first step, the selective genes present within block B of the LPS cluster of *L. pneumophila* Sg1 (*wzt*, *wzm*, *lpp0831*) identified here present promising markers for the development of rapid and accurate identification and subtyping tools that may be used in health institutions and for surveillance of hospital and other water supply systems.

Emergence of an epidemic *L. pneumophila* strain

Because of the highly different PFGE profiles previously obtained from different *L. pneumophila* clinical isolates of Sg1, it was thought that *Legionella* outbreaks and sporadic cases were mainly caused by different strains, for example, 16 outbreak strains investigated by PFGE each showed a unique genotype, and among 691 isolates 543 patterns were identified (Aurell et al. 2003). This hypothesis is further supported by the fact that the three *Legionella* outbreak strains whose genome sequence has been entirely determined (Paris, Lens, and Philadelphia-1) (Cazalet et al. 2004; Chien et al. 2004), show major differences in their overall gene content. However, when the macroarray hybridization data of different strains within the four lineages were compared, it was obvious that particularly the strains that clustered in lineage I together with strain Paris, although isolated from various outbreaks and geographical locations, showed high similarities. This observation prompted us to also determine their PFGE patterns, which were nearly identical (Fig. 4). Thus, strain Paris was identified here as a particular clone that is responsible for sporadic cases as well as for outbreaks reported since 1981 in different European countries (Sweden, Austria, Germany, Switzerland, The Netherlands), the USA, Africa (Senegal), and in Japan. Furthermore, our study showed that the clone, which is well adapted to cooling towers in Japan (Amemura-Maekawa et al. 2005) corresponds to strain Paris, and supposedly this clone is also present in Korea, as a recent study described Korean strains with the same *rpoB* ST and PFGE pattern (Ko et al. 2006). A search for specific genetic features allowing this clone its adaptation to the environment and its success in causing human disease identified three specific genomic regions of particular interest, as they contain genes encoding transporters, eukaryotic-like proteins, and a specific *o* type quinol cytochrome oxidase, which could allow better adaptation to changing oxygen tensions and a specific serine/threonine protein kinase. A putative type-III restriction-modification (RM) enzyme and its cognate DNA methylase are also specific and consistently present in all Paris strains. Similar comparative genomic hybridization studies on *Campylobacter jejuni* (Dorrell et al. 2001) and *V. cholerae* (Dziejman et al. 2005) identified strain specificities in RM systems. The presence of genes encoding type-I RM systems correlated with different gastric responses of the host to *Helicobacter pylori* (Bjorkholm et al. 2002), and it is thought to modulate the virulence potential of *Campylobacter* strains (Poly et al. 2005). The identification of a worldwide distributed *L. pneumophila* strain, resembling strain

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Paris, now opens exciting possibilities of research to find out whether the specific genes contribute to improved interaction with the host or to improved fitness in the environment.

Methods

Bacterial strains

Chromosomal DNA for array construction was isolated from *L. pneumophila* strains Paris CIP107629, Lens CIP108286, and Philadelphia-1 ATCC33152. Reference strains and French isolates were selected from the culture collection of the National Reference Center for *Legionella*, Lyon, France. The 10 strains from outbreaks in Sweden were kindly provided by S. Bernander (Karolinska Institute, Microbiology and Tumor Biology Center, Clinical Microbiology–KS, Stockholm, Sweden); the 12 Austrian isolates from G. Wewalka, (Austrian Agency for Health and Nutrition [AGES], Institute for Medical Microbiology and Hygiene, Vienna, Austria); the 29 strains isolated in Germany by C. Lück, (Institut for Medical Microbiology and Hygiene, National Reference Laboratory for *Legionella*, Technical University, Dresden, Germany); the 12 strains isolated in Japan by J. Amemura-Maekawa (Department of Bacteriology, National Institute of Infectious Diseases, Tokyo, Japan); the strains isolated in the USA by J.E. Stout (Special Pathogens Laboratory, Infectious Disease Section, VA Medical Center, Pittsburgh, VA, USA) and B. Fields (*Legionella* Section Chief, CDC, Atlanta, GA, USA), and the three strains isolated in Africa by A. Stevens (Institut Pasteur, Dakar, Senegal). The characteristics and origins of the 217 *L. pneumophila* strains (Sg1–Sg14) isolated from humans, and the environment, and the 32 *Legionella sp.* are described in Supplemental Table S1.

DNA-array design, construction, and hybridization.

Genes selected were those variable between the three completely sequenced *L. pneumophila* strains (Cazalet et al. 2004; Chien et al. 2004), all known or putative virulence factors of *L. pneumophila*, and 14 control genes. This represented 324, 242, and 227 PCR probes specific for genes uniquely present in strain Paris, Lens, or Philadelphia, respectively; 145 probes each present in only two of the three sequenced strains and 142 probes corresponding to known and putative virulence factors, 31 genes forming the LPS cluster of Sg1 (Luneberg et al. 2000), 30 genes of an unstable genetic element responsible for phase variation of LPS in strain Olda (Luneberg et al. 2001) and 14 control genes (Supplemental Tables S2A, S2B, S2C; *lpp, lpg, lpl* numbers indicate genes of *L. pneumophila* strains Paris, Philadelphia-1, and Lens. The respective genes and its annotations can be accessed at http://genolist.pasteur.fr/LegioList/). Internal PCR fragments for the genes conserved in strain Paris and Lens or conserved in strains Paris and Philadelphia were amplified with genomic DNA of strain Paris, and the ones conserved in strains Lens and Philadelphia were amplified with genomic DNA of strain Lens. Primers were designed using a modified version of Primer 3 software (CAAT-Box) (Frangeul et al. 2004) to amplify a specific fragment of 300–600 bp for each gene (melting temperatures 55–65°C) (Eurogentec). Amplification reactions were performed in a 100-µL reaction volume containing 6 ng of chromosomal DNA. The concentration and size of each PCR product was verified on agarose gels. Array preparation was done as described in Doumith et al. (2004). To control specificity, 70% of the PCR products were randomly chosen and sequenced. All corresponded to the expected PCR products. Hybridizations with chromosomal DNA from strains Paris, Lens, and Philadelphia were done to test the quality and the correct spotting of probes. For comparative hybridization, genomic DNA of each strain was extracted by the phenol/chloroform technique and was radiolabeled using a random priming DNA labeling kit (Roche) as described in Doumith et al. (2004). Probes and hybridization results are given in Supplemental Tables S2A, S2B, and S2C.

The microarrays used carry 70 bp of oligonucleotides specific for each gene of strains Paris, Lens, and Philadelphia. Design and construction are described in Brüggemann et al. (2006b). For competitive microarray hybridization of Sg9 reference strain ATCC35289, chromosomal DNA of strain Paris was used as reference. Then, 2 µg of genomic DNA was used as template for direct incorporation of alternate fluorescent analogs Cy5 and Cy3-dCTP by randomly primed polymerization reaction. Briefly, whole genomic DNA digested by RsaI and 9 µg of random hexamers (GIBCO BRL) in a reaction volume of 45 µL were denatured at 95°C for 10 min. Then, 5 µL of 10X Klenow buffer, 3 nmol of dye (Perkin Elmer), 1 µL of dNTP solution, and 50 U of Klenow (Biolabs) were added. The labeling reactions were incubated at 37°C for 3 h. Hybridization was performed following the manufacturers' recommendations (Corning) using 250 pmol of Cy3 and Cy5 labeled DNA. For each strain tested, a dye-swap experiment was carried out.

Data analysis

Macroarrays were scanned (Typhoon 9400; Amersham Biosciences); ArrayVision software (Imaging Research) was used for quantification of hybridization intensities and for normalization. For each spot, the hybridization intensity value was normalized by dividing it by the median of all significant intensity values on each filter. For ratio calculation, a reference array was built by combining the average normalized data of four replicate hybridizations of genomic DNA of strains Paris, Lens, and Philadelphia to the corresponding spots as described in Doumith et al. (2004). The cutoff ratio for the presence of a gene was defined by analyzing the results for Paris genes hybridized to Lens and Philadelphia genes and vice versa. The threshold for the presence of a gene was defined as a ratio of >0.3. This corresponds to a DNA similarity ≥80%, as verified by sequence comparisons. Data were converted into binary scores (ratio >0.3, present gene = 1; ratio <0.3, absent gene = 0) that were analyzed by hierarchical clustering (J-Express) (Dysvik and Jonassen 2001) and by intensive expert-based data mining with Excel spreadsheets. Probes and all hybridization data are given in Supplemental Table S2.

Microarrays were scanned on a GenePix 4000A scanner (Axon Instruments). Laser power and/or PMT were adjusted to balance the two channels. The resulting files were analyzed using Genepix Pro 4.0. For normalization and differential analysis, the R software (http://www.R-project.org) was used. A Loess normalization (Yang et al. 2002) was performed on a slide-by-slide basis (BioConductor package marray; http://www.bioconductor.org/packages/2.1/bioc/html/marray.html). Presence/absence of genes was determined using the GACK software (http://falkow.stanford.edu/whatwedo/software/software.html).

rpoB sequencing and analysis

For amplification of the 369-bp *rpoB* gene fragment, primers described in Ko et al. (2006) were used. Amplification was performed in 50 µL with 6 ng of chromosomal DNA. PCR products were verified on agarose gels and purified with exonuclease and Shrimp alkaline phosphatase for 15 min at 37°C and 15 min at 80°C. Sequencing reactions were performed twice (ABI PRISM BigDye Terminator cycle sequencing kit) on a 3700 Genetic Analyzer (Applied Biosystems), (GenBank/EMBL accession nos. AM931713–AM931928). The sequences were aligned using

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

ClustalW and the Neighbor joining method (MEGA; http://www.megasoftware.net/) was applied for inferring phylogeny.

Acknowledgments

We thank R. Brosch for critical reading of the manuscript and discussion of the results and J. Allignet for technical assistance. Financial support was received from the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS), the Agence française de sécurité sanitaire de l'environnement et du travail (AFSSET) Project no. ARCL-2005-001, the Israeli-French Research Networks Program in Bioinformatics, and the European contract LSHB-CT-2005-512061. C.C. is the holder of a fellowship jointly financed by the CNRS and VeoliaWater–Anjou Recherche.

References

Amemura-Maekawa, J., Kura, F., Chang, B., and Watanabe, H. 2005. *Legionella pneumophila* serogroup 1 isolates from cooling towers in Japan form a distinct genetic cluster. *Microbiol. Immunol.* **49:** 1027–1033.

Aurell, H., Etienne, J., Forey, F., Reyrolle, M., Girardo, P., Farge, P., Decludt, B., Vandenesch, F., and Jarraud, S. 2003. *Legionella pneumophila* serogroup 1 strain Paris: Endemic distribution throughout france. *J. Clin. Microbiol.* **41:** 3320–3322.

Aurell, H., Farge, P., Meugnier, H., Gouy, M., Forey, F., Lina, G., Vandenesch, F., Etienne, J., and Jarraud, S. 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.

Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., and Small, P.M. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284:** 1520–1523.

Bjorkholm, B.M., Guruge, J.L., Oh, J.D., Syder, A.J., Salama, N., Guillemin, K., Falkow, S., Nilsson, C., Falk, P.G., Engstrand, L., et al. 2002. Colonization of germ-free transgenic mice with genotyped *Helicobacter pylori* strains from a case-control study of gastric cancer reveals a correlation between host responses and HsdS components of type I restriction-modification systems. *J. Biol. Chem.* **277:** 34191–34197.

Brassinga, A.K., Hiltz, M.F., Sisson, G.R., Morash, M.G., Hill, N., Garduno, E., Edelstein, P.H., Garduno, R.A., and Hoffman, P.S. 2003. A 65-kilobase pathogenicity island is unique to Philadelphia-1 strains of *Legionella pneumophila*. *J. Bacteriol.* **185:** 4630–4637.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., et al. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci.* **99:** 3684–3689.

Brüggemann, H., Cazalet, C., and Buchrieser, C. 2006a. Adaptation of *Legionella pneumophila* to the host environment: Role of protein secretion, effectors and eukaryotic-like proteins. *Curr. Opin. Microbiol.* **9:** 86–94.

Brüggemann, H., Hagman, A., Jules, M., Sismeiro, O., Dillies, M., Gouyette, C., Kunst, F., Steinert, M., Heuner, K., Coppée, J., et al. 2006b. Virulence strategies for infecting phagocytes deduced from the in vivo transcriptional program of *Legionella pneumophila*. *Cell. Microbiol.* **8:** 1228–1240.

Cazalet, C. and Buchrieser, C. 2006. Genomics of *Legionella pneumophila*. In *Pathogenomics—Genome analysis of pathogenic microbes*, pp. 125–152. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Cazalet, C., Rusniok, C., Bruggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F., et al. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* **36:** 1165–1173.

Chien, M., Morozova, I., Shi, S., Sheng, H., Chen, J., Gomez, S.M., Asamani, G., Hill, K., Nuara, J., Feder, M., et al. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* **305:** 1966–1968.

Cirillo, S.L., Bermudez, L.E., El-Etr, S.H., Duhamel, G.E., and Cirillo, J.D. 2001. *Legionella pneumophila* entry gene *rtxA* is involved in virulence. *Infect. Immun.* **69:** 508–517.

Cirillo, S.L., Yan, L., Littman, M., Samrakandi, M.M., and Cirillo, J.D. 2002. Role of the *Legionella pneumophila rtxA* gene in amoebae. *Microbiol.* **148:** 1667–1677.

de Felipe, K.S., Pampou, S., Jovanovic, O.S., Pericone, C.D., Ye, S.F., Kalachikov, S., and Shuman, H.A. 2005. Evidence for acquisition of *Legionella* type IV secretion substrates via interdomain horizontal gene transfer. *J. Bacteriol.* **187:** 7716–7726.

Dobrindt, U., Blum-Oehler, G., Nagy, G., Schneider, G., Johann, A., Gottschalk, G., and Hacker, J. 2003. Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic *Escherichia coli* strain 536. *Infect. Immun.* **70:** 6365–6372.

Doleans, A., Aurell, H., Reyrolle, M., Lina, G., Freney, J., Vandenesch, F., Etienne, J., and Jarraud, S. 2004. Clinical and environmental distributions of *Legionella* strains in France are different. *J. Clin. Microbiol.* **42:** 458–460.

Doleans, A., Akermi, M., Ginevra, C., Cazalet, C., Kay, E., Schneider, D., Buchrieser, C., Atlan, D., Vandenesch, F., Etienne, J., et al. 2006. Growth phase-dependent mobility of the *lvh*-encoding region in *Legionella* strain Paris. *Microbiology* **152:** 3561–3568.

Dorrell, N., Mangan, J.A., Laing, K.G., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B.G., Parkhill, J., Stoker, N.G., Karlyshev, A.V., et al. 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* **11:** 1706–1715.

Doumith, M., Cazalet, C., Simoes, N., Frangeul, L., Jaquet, C., Kunst, F., Martin, P., Cossart, P., Glaser, P., and Buchrieser, C. 2004. New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics. *Infect. Immun.* **72:** 1072–1083.

Doyle, R.M. and Heuzenroeder, M.W. 2002. A mutation in an *ompR*-like gene on a *Legionella longbeachae* serogroup 1 plasmid attenuates virulence. *Int. J. Med. Microbiol.* **292:** 227–239.

Dysvik, B. and Jonassen, I. 2001. J-Express: Exploring gene expression data using Java. *Bioinformatics* **17:** 369–370.

Dziejman, M., Serruto, D., Tam, V.C., Sturtevant, D., Diraphat, P., Faruque, S.M., Rahman, M.H., Heidelberg, J.F., Decker, J., Li, L., et al. 2005. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proc. Natl. Acad. Sci.* **102:** 3465–3470.

Feldman, M., Zusman, T., Hagag, S., and Segal, G. 2005. Coevolution between nonhomologous but functionally similar proteins and their conserved partners in the *Legionella* pathogenesis system. *Proc. Natl. Acad. Sci.* **102:** 12206–12211.

Fernandez-Moreira, E., Helbig, J.H., and Swanson, M.S. 2006. Membrane vesicles shed by *Legionella pneumophila* inhibit fusion of phagosomes with lysosomes. *Infect. Immun.* **74:** 3285–3295.

Fields, B.S., Benson, R.F., and Besser, R.E. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin. Microbiol. Rev.* **15:** 506–526.

Fitzgerald, J.R., Sturdevant, D.E., Mackie, S.M., Gill, S.R., and Musser, J.M. 2001. Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc. Natl. Acad. Sci.* **98:** 8821–8826.

Frangeul, L., Glaser, P., Rusniok, C., Buchrieser, C., Duchaud, E., Dehoux, P., and Kunst, F. 2004. CAAT-Box, Contigs-Assembly and Annotation tool-box for genome sequencing projects. *Bioinformatics* **20:** 790–797.

Glöckner, G., Albert-Weissenberger, C., Weinmann, E., Jacobi, S., Schunder, E., Steinert, M., Hacker, J., and Heuner, K. 2007. Identification and characterization of a new conjugation type IVA secretion system (trb/tra) of *Legionella pneumophila* Corby localized on two mobile genomic islands. *Int. J. Med. Microbiol.* doi: 10.1016/j.ijmm.2007.07.012.

Härdle, W. and Simar, L. 2006. *Principal components analysis*. Springer Verlag, Berlin, Heidelberg, New York.

Harrison, T.G., Doshi, N., Fry, N.K., and Joseph, C.A. 2007. Comparison of clinical and environmental isolates of *Legionella pneumophila* obtained in the UK over 19 years. *Clin. Microbiol. Infect.* **13:** 78–85.

Hinchliffe, S.J., Isherwood, K.E., Stabler, R.A., Prentice, M.B., Rakin, A., Nichols, R.A., Oyston, P.C., Hinds, J., Titball, R.W., and Wren, B.W. 2003. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Res.* **13:** 2018–2029.

Ko, K.S., Miyamoto, H., Lee, H.K., Park, M.Y., Fukuda, K., Park, B.J., and Kook, Y.H. 2006. Genetic diversity of *Legionella pneumophila* inferred from *rpoB* and *dotA* sequences. *Clin. Microbiol. Infect.* **12:** 254–261.

Luneberg, E., Zetzmann, N., Alber, D., Knirel, Y.A., Kooistra, O., Zahringer, U., and Frosch, M. 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.

Luneberg, E., Mayer, B., Daryab, N., Kooistra, O., Zahringer, U., Rohde, M., Swanson, J., and Frosch, M. 2001. Chromosomal insertion and excision of a 30 kb unstable genetic element is responsible for phase

Downloaded from genome.cshlp.org on June 20, 2026 . Published by Cold Spring Harbor Laboratory Press

variation of lipopolysaccharide and other virulence determinants in *Legionella pneumophila*. *Mol. Microbiol.* **39:** 1259–1271.
Muder, R.R. and Yu, V.L. 2002. Infection due to *Legionella* species other than *L. pneumophila*. *Clin. Infect. Dis.* **35:** 990–998.
Muder, R.R., Yu, V.L., and Woo, A.H. 1986. Mode of transmission of *Legionella pneumophila*. A critical review. *Arch. Intern. Med.* **146:** 1607–1612.
Nagai, H., Kagan, J.C., Zhu, X., Kahn, R.A., and Roy, C.R. 2002. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295:** 679–682.
Newton, H.J., Sansom, F.M., Dao, J., McAlister, A.D., Sloan, J., Cianciotto, N.P., and Hartland, E.L. 2007. Sel1 repeat protein LpnE is a *Legionella pneumophila* virulence determinant that influences vacuolar trafficking. *Infect. Immun.* **75:** 5575–5585.
Poly, F., Threadgill, D., and Stintzi, A. 2005. Genomic diversity in *Campylobacter jejuni*: Identification of *C. jejuni* 81-176-specific genes. *J. Clin. Microbiol.* **43:** 2330–2338.
Sabria, M., Garcia-Nunez, M., Pedro-Botet, M.L., Sopena, N., Gimeno, J.M., Reynaga, E., Morera, J., and Rey-Joly, C. 2001. Presence and chromosomal subtyping of *Legionella* species in potable water systems in 20 hospitals of Catalonia, Spain. *Infect. Control Hosp. Epidemiol.* **22:** 673–676.
Sansom, F.M., Newton, H.J., Crikis, S., Cianciotto, N.P., Cowan, P.J., d'Apice, A.J., and Hartland, E.L. 2007. A bacterial ecto-triphosphate diphosphohydrolase similar to human CD39 is essential for intracellular multiplication of *Legionella pneumophila*. *Cell. Microbiol.* doi: 10.111/1.1462-5822.2007.00924.x.
Selander, R.K., McKinney, R.M., Whittam, T.S., Bibb, W.F., Brenner, D.J., Nolte, F.S., and Pattison, P.E. 1985. Genetic structure of populations of *Legionella pneumophila*. *J. Bacteriol.* **163:** 1021–1037.
Trad, S., Allignet, J., Frangeul, L., Davi, M., Vergassola, M., Couve, E., Morvan, A., Kechrid, A., Buchrieser, C., Glaser, P., et al. 2004. DNA macroarray for identification and typing of *Staphylococcus aureus* isolates. *J. Clin. Microbiol.* **42:** 2054–2064.
Wildschutte, H., Wolfe, D.M., Tamewitz, A., and Lawrence, J.G. 2004. Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *Proc. Natl. Acad. Sci.* **101:** 10644–10649.
Wolfgang, M.C., Kulasekara, B.R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C.G., and Lory, S. 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* **100:** 8484–8489.
Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. 2002. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30:** e15. doi: 10.1093/nar/30.4.e15.
Yu, V.L., Plouffe, J.F., Pastoris, M.C., Stout, J.E., Schousboe, M., Widmer, A., Summersgill, J., File, T., Heath, C.M., Paterson, D.L., et al. 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.
Zink, A.R., Reischl, U., Wolf, H., and Nerlich, A.G. 2002. Molecular analysis of ancient microbial infections. *FEMS Microbiol. Lett.* **213:** 141–147.

Received October 4, 2007; accepted in revised form December 7, 2007.