

Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity

Christel Cazalet¹, Christophe Rusniok¹, Holger Brüggemann¹, Nora Zidane², Arnaud Magnier², Laurence Ma², Magalie Tichit², Sophie Jarraud³, Christiane Bouchier², François Vandenesch³, Frank Kunst¹, Jérôme Etienne³, Philippe Glaser¹ & Carmen Buchrieser¹

Legionella pneumophila, the causative agent of Legionnaires' disease, replicates as an intracellular parasite of amoebae and persists in the environment as a free-living microbe. Here we have analyzed the complete genome sequences of *L. pneumophila* Paris (3,503,610 bp, 3,077 genes), an endemic strain that is predominant in France, and Lens (3,345,687 bp, 2,932 genes), an epidemic strain responsible for a major outbreak of disease in France. The *L. pneumophila* genomes show marked plasticity, with three different plasmids and with about 13% of the sequence differing between the two strains. Only strain Paris contains a type V secretion system, and its Lvh type IV secretion system is encoded by a 36-kb region that is either carried on a multicopy plasmid or integrated into the chromosome. Genetic mobility may enhance the versatility of *L. pneumophila*. Numerous genes encode eukaryotic-like proteins or motifs that are predicted to modulate host cell functions to the pathogen's advantage. The genome thus reflects the history and lifestyle of *L. pneumophila*, a human pathogen of macrophages that coevolved with fresh-water amoebae.

L. pneumophila is the causative agent of legionellosis, an atypical pneumonia that can be fatal if not treated promptly¹. This Gram-negative facultative intracellular pathogen can adapt to both the aquatic environment and the intracellular milieu of phagocytic cells of a human host². When inhaled in contaminated aerosols, *L. pneumophila* can reach the alveoli of the lung, where it is engulfed by macrophages. In contrast to most bacteria, which are destroyed there, *L. pneumophila* can multiply inside its phagosome and eventually kill the macrophage, resulting in legionellosis³.

L. pneumophila and other *Legionella* species are inhabitants of natural aquatic biotopes and of man-made water systems, such as air-conditioning cooling towers³. *Legionella* has been detected by culture in 40%, and by PCR in up to 80%, of freshwater environments, where it is known to survive and to replicate intracellularly in free-living protozoa, often within aquatic biofilms⁴. Its ability to exploit the basic cellular mechanisms of numerous protozoal eukaryotic hosts also enables *Legionella* to infect human cells⁵. Indeed, the capability of *L. pneumophila* to multiply intracellularly in amoeba has been shown to contribute to disease, although little is known about the mechanisms that govern host-microbe interactions. By contrast, the biphasic life cycle of *L. pneumophila*, its change from parasitic replicative cells to extracellular transmissible forms, and the complex regulatory network that governs this change are understood in part⁶.

The genus *Legionella* comprises 48 species, but over 90% of clinical cases of legionellosis are caused by *L. pneumophila* and, even more strikingly, up to 84% are caused by *L. pneumophila* serogroup 1 (ref. 7). To provide insight into the genetic characteristics of *L. pneumophila*, and to identify properties that have been selected in niches that are specific for the pathogenicity and the life cycle of *L. pneumophila*, we have determined the complete genome sequences of two clinical isolates of serogroup 1, strains Paris and Lens. Strain Paris is the only endemic strain known so far, accounting for 12.7% of cases of legionellosis in France and 33% of those occurring in the Paris area⁸. It is associated with hospital- and community-acquired forms of the disease that occur as outbreaks or sporadic cases. From November 2003 to January 2004, strain Lens caused an outbreak of 86 cases resulting in 17 deaths in northern France, suggesting that it is particularly successful in causing disease in humans. Comparative genomic assessment of an endemic and an epidemic isolate provides the basis for understanding strain specificities and might provide clues to the particular adaptability and stability of *L. pneumophila* strain Paris.

RESULTS

General features

L. pneumophila strain Paris and strain Lens each contain one circular chromosome of about 3,503,610 bp and 3,345,687 bp, respectively,

¹Laboratoire de Génétique des Microorganismes Pathogènes and CNRS URA 2171, Institut Pasteur, 28, rue du Dr. Roux, 75724 Paris Cedex 15, France. ²Plate-forme Génomique, Pasteur Génopole Ile de France, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris, France. ³Centre National de Référence des Legionella, Laboratoire de Bactériologie INSERM E-0230, Faculté de Médecine, IFR 62, 69372 Lyon Cedex 08, France. Correspondence should be addressed to C.B. (cbuch@pasteur.fr).

Table 1 General features of the two *Legionella pneumophila* genomes

	<i>L. pneumophila</i> Paris	<i>L. pneumophila</i> Lens
Chromosome size	3,503,610 bp	3,345,87 bp
G+C content	38.3% (37.4%)	38.4% (38.4%)
G+C content of protein-coding genes	39.1% (37.9%)	39.4% (39.1%)
Total number of protein-coding genes	3,076 (141)	2,931 (57)
Average length (no. of codons) of protein-coding genes	331	333
Number of rRNA operons (16S, 23S, 5S)	3	3
Number of tRNA genes	43	43
Percentage coding	87.9% (92%)	88% (83.7%)
Plasmid	1 (131.9 kb)	1 (59.8 kb)
Number of strain-specific genes	428 (125)	280 (44)
Number of orthologous genes	2,664	2,664

Values in parentheses refer to the plasmids.

with an average G+C content of 38% (Table 1 and Fig. 1). *L. pneumophila* Paris contains one plasmid of 131,885 bp and strain Lens contains a plasmid of 59,832 bp.

We identified 3,077 genes in the chromosome of *L. pneumophila* Paris and 2,932 in that of *L. pneumophila* Lens (Supplementary Table 1 online). No function could be predicted for 42.1% (1,354) of the *L. pneumophila* Paris and 44.1% (1,320) of the *L. pneumophila* Lens genes, a proportion similar to that found in most other sequenced bacterial genomes. A high proportion of the predicted genes (21% in strain Paris and 20.4% in strain Lens) are unique to the genus *Legionella* and thus might encode *Legionella*-specific functions.

Exploitation and modulation of host cell functions

An intriguing issue is how *Legionella* subverts host functions to enter, survive, replicate and evade amoebae or alveolar macrophages. *L. pneumophila* encodes an abundance of eukaryotic-like proteins in its genome. Indeed, 30 genes encode proteins with high similarity to eukaryotic proteins (Table 2), and 32 genes encode proteins with eukaryotic domains that are implicated in protein-protein interactions (Table 3). We highlight here proteins that are predicted either to divert eukaryotic regulatory pathways or to be secreted into eukaryotic cells, making them strong candidates for directing *Legionella* invasion, its trafficking in the host cell, and its modulation or evasion of host cell functions.

Tetratricopeptide repeats (TPRs) are degenerate repeated motifs of 34 amino acids that are present in tandem arrays of 3–16 motifs and form scaffolds to mediate protein-protein interactions. TPR proteins

contribute to cell cycle control, transcription repression, the stress response, protein kinase inhibition, mitochondrial and peroxisomal protein transport, and neurogenesis⁹. Sel-1 repeats represent a sub-family of TPR sequences. In *L. pneumophila*, five proteins containing Sel-1 domains were identified. Two of those (EnhC, LidL) have been previously implicated in interactions with host cells or in the early signaling events that regulate *L. pneumophila* trafficking decisions in macrophages^{10,11}. Therefore, the three newly identified proteins are also probably involved in host-pathogen interactions (Table 3).

After internalization, *L. pneumophila* manipulates the host endosomal-lysosomal degradation pathway to survive and to replicate in a vacuole derived from the endoplasmic reticulum (ER). One *L. pneumophila* protein, RalF, which is thought to contribute to ER recruitment, contains a eukaryotic Sec 7 domain. RalF, a substrate of the type IV secretion system, is required for association of the regulatory protein ARF with the *L. pneumophila* phagosomes¹². Both *L. pneumophila* strains encode three eukaryotic-like serine/threonine protein kinases (STPKs; Table 3). Multiple sequence comparisons of kinase domains from *L. pneumophila* Paris and other prokaryotic and eukaryotic STPKs revealed that Lpp2626 and Lpp1439 of *L. pneumophila* cluster in the group of eukaryotic STPKs, close to STPKs from *Entamoeba histolytica* (Fig. 2). *Mycobacterium tuberculosis*, which like *L. pneumophila* blocks phagosome lysosome fusion, produces 11 such eukaryotic-like STPKs¹³. In particular, PknG STPK of *M. tuberculosis* inhibits phagosome-lysosome fusion and promotes intracellular survival¹⁴. The STPK domain of the Lpp0267 and Lp10262 (of *L. pneumophila* strains Paris and Lens, respectively) is closely related to PknG and to the *Y. pseudotuberculosis* STPK YpkA (Fig. 2), an enzyme that is translocated into eukaryotic cells, where it counteracts host defense by interfering with eukaryotic signal transduction pathways¹⁵. This finding suggests that the *L. pneumophila* STPKs may also modulate eukaryotic signal transduction mechanisms and can modify host cell trafficking pathways.

Twenty proteins contain ankyrin domains (Table 3), tandem repeats of around 33 amino acids that represent one of the most common modular protein-protein interaction motifs of eukaryotes. So far, the only prokaryotic genomes known to encode large families of ankyrin domain proteins are those of *Coxiella burnetii* and *Wolbachia pipientis*, which encode 13 and 23 members,

Figure 1 Circular genome map of *L. pneumophila* strain Paris and specific genes of *L. pneumophila* strain Lens. From the outside, the first circle indicates strain Paris genes on the + and – strands, respectively. Red bar indicates an inversion in strain Lens. Green indicates strain Paris genes, black indicates rRNA operons, and red indicates known virulence genes as follows: 1, *lvh-lvr* type IV secretion system (*lvrABC*, *lvhB2B3B4B5*, *lvrD*, *lvhB6B8B9B10B11D4*, *lvrE*); 2, *dot-icm* type IV secretion system (*icmTSRQOMLKEGCDJBF*); 3, *mip*; 4, *lspA*; 5, *lspDE*; 6, *htrA*; 7, *lspFGHIJK*; 8, *enhABC*; 9, *dot-icm* type IV secretion system (*icmVWX* and *dotABCD*); 10, *momp*. The second circle shows genes specific for strain Lens with respect to strain Paris. The third circle shows the G/C bias (G+C/G–C) of strain Paris. The fourth circle shows the G+C content of strain Paris: light yellow, <32.5% G+C; yellow, between 32.5 and 44.1% G+C; dark yellow, >44.1% G+C. The scale (in Mb) is indicated on the outside, with the origin of replication at position 0.

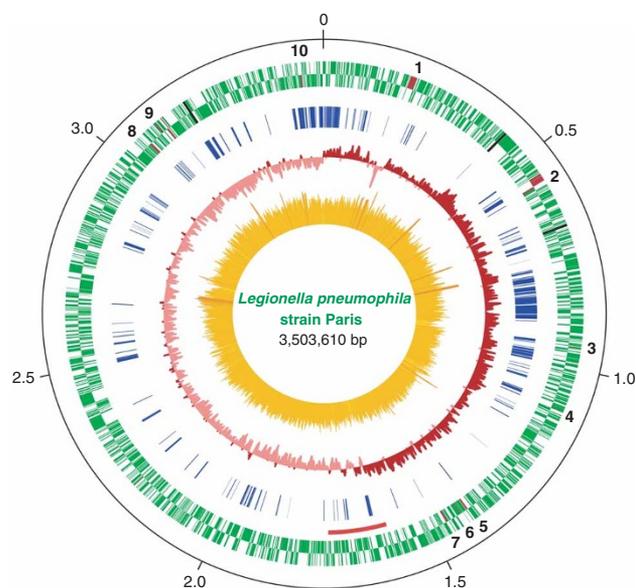


Table 2 *L. pneumophila* proteins with the highest similarity score to eukaryotic proteins

<i>L.p.</i> Paris ^a	Predicted product	<i>L.p.</i> Lens ^a	G-C content	Percentage of identity ^b
<i>lpp1647</i>	purC	<i>lpl1640</i>	38%	61%/100% (AAR06292.1 <i>Nicotiana tabacum</i>)
<i>lpp0702</i>	exoA exodeoxyribonuclease III	<i>lpl0684</i>	39%	58%/100% (EAA20230.1 <i>Plasmodium yoelii yoelii</i>)
<i>lpp0321</i>	RNA-binding protein precursor	—	34%	45%/50% (AAL07519 <i>Solanum tuberosum</i>)
<i>lpp1157</i>	Pyruvate decarboxylase	<i>lpl1162</i>	39%	50%/100% (AAB16855.1 <i>Arabidopsis thaliana</i>)
<i>lpp1522</i>	Thiamine biosynthesis protein NMT-1	<i>lpl1461</i>	38%	49%/100% (AAC64375.1 <i>Botryotinia fuckeliana</i>)
<i>lpp2832</i>	nuoE NADH dehydrogenase I chain E	<i>lpl2701</i>	38%	49%/82% (BAA25988.1 <i>Homo sapiens</i>)
<i>lpp0050</i>	Hypersensitive-induced response protein	—	36%	48%/100% (AAN17462.1 <i>Hordeum vulgare subsp. Vulgare</i>)
<i>lpp0634</i>	Hypothetical protein	<i>lpl0618</i>	39%	48%/100% (XP_306643.1 <i>Anopheles gambiae</i>)
<i>lpp0965</i>	Protease	<i>lpl0935</i>	39%	45%/100% (NP_189431.2 <i>A. thaliana</i>)
<i>lpp2748</i>	Phytanoyl-coA dioxygenase	<i>lpl2621</i>	36%	44%/100% (XP_372144.1 <i>H. sapiens</i>)
<i>lpp2128</i>	Sphingosine 1-phosphate lyase	<i>lpl2102</i>	41%	36%/100% (NP_775139.1 <i>Rattus norvegicus</i>) ^c
<i>lpp0489</i>	Glucoamylase	<i>lpl0465</i>	39%	32%/100% (P42042 <i>Arxula adenivorans</i>)
<i>lpp0955</i>	Cytokinin oxidase	<i>lpl0925</i>	39%	32%/100% (NP_484368.1 <i>Nostoc</i> sp.)
<i>lpp0578</i>	Phytanoyl coA dioxygenase	<i>lpl0554</i>	36%	31%/100% (EAA70100.1 <i>Gibberella zeae</i>)
<i>lpp0379</i>	Hypothetical protein	<i>lpl0354</i>	39%	31%/100% (CAD21525.1 <i>Taenia solium</i>)
<i>lpp1033</i>	Ectonucleoside triphosphate diphosphohydrolase (apyrase)	<i>lpl1000</i>	40%	25%/100% (Q9MYU4 <i>Sus scrofa</i>) contains a nucleoside phosphatase signature
<i>lpp2923</i>	6-Pyruvoyl-tetrahydropterin synthase	<i>lpl2777</i>	34%	26%/100% (NP_703938.1 <i>Plasmodium falciparum</i>)
<i>lpp3071</i>	Zn ²⁺ metalloproteinase	<i>lpl2927</i>	38%	38%/100% (AAF56122.1 <i>Drosophila melanogaster</i>)
<i>lpp2134</i>	SAM-dependent methyltransferase	<i>lpl2109</i>	35%	24%/100% (BAC98835.1 <i>Bombyx mori</i>)
<i>lpp1880</i>	Ectonucleoside triphosphate diphosphohydrolase (apyrase)	<i>lpl1869</i>	39%	26%/100% (CAE70887.1 <i>Caenorhabditis briggsae</i>)
<i>lpp2747</i>	SAM-dependent methyltransferase	<i>lpl2620</i>	35%	33%/56% (EAA20288.1 <i>Plasmodium yoelii yoelii</i>)
<i>lpp2468</i>	Cytochrome <i>P450</i>	<i>lpl2326</i>	39%	20%/75% (NP_487786.1 <i>Nostoc</i> sp.)
<i>lpp1824</i>	Nuclear membrane binding protein	—	34%	19%/40% (NP_082559.1 <i>Mus musculus</i>)
<i>lpp1665</i>	Uracyl DNA glycosylase	<i>lpl1659</i>	36%	27%/80% (EAA36774.1 <i>Giardia lamblia</i>)
<i>lpp1959</i>	Chromosome condensation 1-like	<i>lpl1953</i>	41%	Pattern of chromosome condensation regulator conserved
<i>lpp0358</i>	Hypothetical protein	<i>lpl334</i>	38%	37%/53% (EAA20288.1 <i>Plasmodium yoelii yoelii</i>)
<i>lpp1127</i>	Ca ²⁺ -transporting ATPase	<i>lpl1131</i>	37%	22%/34% (AAB81284.1 <i>Paramecium tetraurelia</i>)
<i>lpp1167</i>	Uridine kinase	<i>lpl1173</i>	33%	35%/65% (AAM09314.2 <i>Dictyostelium discoideum</i>)
<i>lpp2626</i>	STPK domain	<i>lpl2481</i>	32%	Conserved domain ^c
<i>lpp1439</i>	STPK	<i>lpl1545</i>	36%	Conserved domain ^c

^a*lpp* indicates predicted coding sequences of *L. pneumophila* Paris; *lpl* indicates predicted coding sequences of *L. pneumophila* Lens. ^bPercentage identity over percentage length of the most similar protein. Protein accession numbers and organism are given in parentheses. ^cProteins are also mentioned in **Table 3** because of their conserved eukaryotic domains.

respectively^{16,17}. Similar to *L. pneumophila*, *C. burnetii* is an intracellular pathogen that is highly adapted for life in the eukaryotic phagolysosome, whereas *W. pipientis* is a parasitic 'endosymbiont' that lives inside the reproductive cells of various arthropods. Thus, ankyrin domains may be involved in a common microbial mechanism to manipulate the host cell physiology.

A possible function of the ankyrin repeat-containing proteins (hereafter called ankyrin proteins) of *L. pneumophila* is to modify interactions with the host cytoskeleton, because many eukaryotic ankyrin proteins are thought to function as linkers between membrane proteins and the cytoskeleton¹⁸ and are important for targeting proteins to the plasma membrane or to the ER. Ankyrin domains are also components of transcriptional regulators, which suggests that they can influence host cell gene expression, as has been proposed for *Ehrlichia phagocytophila*¹⁹. Indeed, one of the *L. pneumophila* ankyrin proteins (*Lpp3991/Lpl0559*) also contains a eukaryotic SET domain, which is known to bind host chromatin and to influence host cell gene expression²⁰. In contrast to the ankyrin proteins of *W. pipientis*, none of the *L. pneumophila* ankyrin proteins contains a signal peptide¹⁷. Instead, some of them might be secreted via the type IV secretion system, a pathway that is independent of typical targeting signals.

The last stages in the intracellular life cycle of *L. pneumophila* involve the killing of and escape from the host cell, a mechanism that

is not understood. A class of proteins that may affect the control of host cell division (*Lpp2082/Lpl2072*, *Lpp2486*, *Lpp0233/Lpl10234*) contains eukaryotic F-box domains, which constitute sites of protein-protein interactions. Eukaryotic F-box domains are typically associated with other interaction domains²¹. Consistent with this, two of the F-box proteins identified in *L. pneumophila* are associated with an ankyrin repeat or a coiled-coil motif (**Table 3**). F-box proteins assembled into SCF ubiquitin-ligase complexes determine which substrates will be targeted for ubiquitination and subsequent proteolysis by the proteasome. Because the targeted substrates include promoters and inhibitors of the cell cycle, as well as signal transduction components²², F-box proteins can regulate cell division and differentiation. To our knowledge, the only other prokaryotic F-box protein that has been described is *Agrobacterium tumefaciens* VirF, a protein that is thought to interact with host proteins through its F-box domain to target them for proteolysis²³. Another motif implicated in eukaryotic ubiquitination is the U-box motif. Protein *Lpp2887*, present in strain Paris but not in strain Lens, contains such a motif. To our knowledge, it is the first one that has been identified in a prokaryotic organism.

Additional eukaryotic-like proteins identified in the genomes of both *L. pneumophila* strains are sphingosine 1-phosphate lyase (*Lpp2128/Lpl2102*) and two secreted apyrases (*Lpp1033*, *Lpp1880/*

Table 3 *L. pneumophila* proteins encoding domains preferentially found in eukaryotic proteins

<i>L.p.</i> Paris ^a	<i>L.p.</i> Lens ^a	Motif identified ^b	G-C content	Putative function	
<i>lpp2692</i> (<i>enhC</i>)	<i>lpl2564</i> (<i>enhC</i>)	21 Sel-1 domains	39%	Invasion and trafficking into host cells	
<i>lpp1174</i> (<i>lidL</i> ; <i>enhC</i> paralog)	<i>lpl1180</i> (<i>lidL</i> ; <i>enhC</i> paralog)	6 Sel-1 domains	38%		
<i>lpp1310</i> (<i>enhC</i> paralog)	<i>lpl1307</i> (<i>enhC</i> paralog)	4 Sel-1 domains	41%		
<i>lpp2174</i> (<i>enhC</i> paralog)	<i>lpl1303</i> (<i>enhC</i> paralog)	3 Sel-1 domains	40%		
–	<i>lpl1059</i> (<i>enhC</i> paralog)	7 Sel-1 domains	45%		
<i>ralF</i> (<i>lpp1932</i>)	<i>ralF</i> (<i>lpl1919</i>)	Sec 7 domain	34%		
<i>lpp0267</i>	<i>lpl0262</i>	STPK domain	38%	Modulation of host cell functions	
<i>lpp2626</i>	<i>lpl2481</i>	STPK domain	32%		
<i>lpp1439</i>	<i>lpl1545</i>	STPK domain	36%		
<i>lpp2065</i>	<i>lpl2055</i>	Ankyrin repeat	37%		
<i>lpp0037</i>	<i>lpl0038</i>	Ankyrin repeat	38%		
<i>plpp0098</i>	–	Ankyrin repeat	37%		
<i>lpp2058</i>	<i>lpl2048</i>	Ankyrin repeat	38%		
<i>lpp0750</i>	<i>lpl0732</i>	Ankyrin repeat	35%		
<i>lpp2061</i>	<i>lpl2051</i>	Ankyrin repeat	39%		
<i>lpp2270</i>	<i>lpl2242</i>	Ankyrin repeat	34%		
<i>lpp0503</i>	<i>lpl0479</i>	Ankyrin repeat	38%		
<i>lpp1905</i>	–	Ankyrin repeat	35%		
<i>lpp1683</i>	<i>lpl1682</i>	Ankyrin repeat + SET domain	33%		
<i>lpp2248</i>	<i>lpl2219</i>	Ankyrin repeat	39%		
<i>lpp0202</i>	–	Ankyrin repeat	38%		
<i>lpp0469</i>	<i>lpl0445</i>	Ankyrin repeat	38%		
<i>lpp2517</i>	<i>lpl2370</i>	Ankyrin repeat	36%		
<i>lpp1100</i>	–	Ankyrin repeat	48%		
<i>lpp0126</i>	<i>lpl0111</i>	Ankyrin repeat	39%		
<i>lpp0356</i>	–	Ankyrin repeat	38%		
<i>lpp2522</i>	<i>lpl2375</i>	Ankyrin repeat	39%		
<i>lpp0547</i>	<i>lpl0523</i>	Ankyrin repeat	40%		
–	<i>lpl1681</i>	Ankyrin repeat	34%		
–	<i>lpl2344</i>	Ankyrin repeat	35%		
–	<i>lpl2058</i>	Ankyrin repeat	40%		
<i>lpp2082</i>	<i>lpl2072</i>	F-box domain + ankyrin repeat	36%		Control of host cell division, evasion
<i>lpp2486</i>	–	F-box domain + coiled coil	34%		
<i>lpp0233</i>	<i>lpl0234</i>	F-box domain	39%		
<i>lpp2887</i>	–	2 U-box domains	35%		
<i>lpp2128</i> (sphingosine 1-phosphate lyase)	<i>lpp2102</i> (sphingosine 1-phosphate lyase)		41%		

^a*lpp* indicates predicted coding sequences of *L. pneumophila* Paris; *lpl* indicates predicted coding sequences of *L. pneumophila* Lens. ^bNumbers indicate the number of domains identified in a protein.

Lpl1000, Lpl1869), suggesting that *L. pneumophila* modulates the host cell cycle to its advantage. The widely expressed enzyme sphingosine 1-phosphate lyase catalyzes the essentially irreversible cleavage of the signaling molecule sphingosine 1-phosphate, a product of sphingomyelin degradation that regulates cell proliferation and cell death in eukaryotes. Indeed, overexpression of sphingosine 1-phosphate lyase can induce apoptosis in eukaryotes, identifying this enzyme as a dual modulator of sphingosine 1-phosphate and ceramide metabolism, as well as a regulator of cell fate decisions²⁴.

In addition, sphingosine 1-phosphate has a central role in the development of the amoeba *Dictyostelium discoideum*, because disruption of its gene results in aberrant actin distribution, an abnormal morphogenetic phenotype and an increase in viability during the stationary phase²⁵. The two *L. pneumophila* apyrases are the only apyrases so far identified in prokaryotes. The apyrase protein family comprises enzymes capable of cleaving nucleotide tri- (NTPs) and diphosphates (NDPs) in a Ca²⁺- or Mg²⁺-dependent manner. Apyrase

has been isolated in the autophagy vacuole²⁶, suggesting that the two *L. pneumophila* proteins influence the fate of the *L. pneumophila* phagosome by decreasing the concentration of NTPs and NDPs during cell parasitism.

We also identified 246 proteins (7.6%) in strain Paris and 231 proteins (7.7%) in strain Lens with predicted coiled-coil domains (Supplementary Table 2 online), many of which also show weak similarities to eukaryotic proteins. Coiled-coil domains mediate protein-protein interactions either for protein multimerization or for macromolecular recognition. Thus, coiled-coil domains may target proteins to their appropriate localization in the eukaryotic host. Notably, all nine substrates (SidA-SidH, SdeC) of the type IV secretion system²⁷ and LidA, LepA and LepB contain coiled-coil domains.

To affect the eukaryotic cell, *L. pneumophila* must translocate these eukaryotic-like proteins to the host cytoplasm. These proteins are therefore candidate substrates for the type IV secretion system, as are VirF in *A. tumefaciens*²⁸ and RalF in *L. pneumophila*¹².



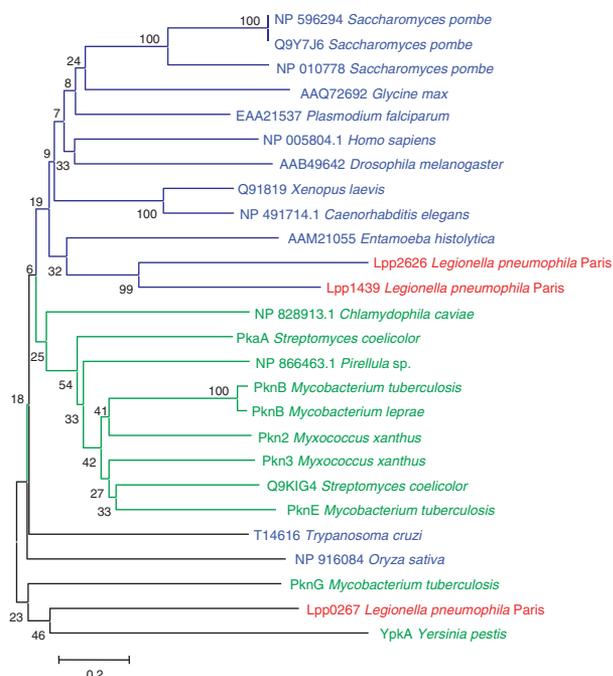


Figure 2 Phylogenetic tree of kinase domains from *Legionella pneumophila* strain Paris and other prokaryotic and eukaryotic kinases constructed by the program MEGA. The calculation was done by using Poisson correction as the distance method and neighbor joining as the tree construction method. The value of 0.2 on the scale bar indicates two amino acid substitutions per ten sites. The Nrprot accession numbers, gene names and organism names are shown. Numbers indicate bootstrap values.

Secretion systems

Central to the pathogenesis of *L. pneumophila* are the *dot* and *icm* loci, which together direct assembly of a type IV secretion apparatus. Although both strains contain the complete *dot-icm* loci, their sequences show variations. Indeed, a previous comparison of the sequence of the *dot-icm* genes of 18 different *L. pneumophila* strains identified considerable sequence variation and placed the strains into seven different phylogenetic groups; however, no correlation with virulence was apparent²⁹.

A previously unknown putative virulence factor of *L. pneumophila*, restricted to strain Paris, is a predicted autotransporter protein. Lpp0779 contains several hallmarks of type V secretion systems, including an N-terminal leader peptide for secretion across the inner membrane and a dedicated C-terminal domain that forms a pore in the outer membrane through which the passenger domain passes to the cell surface³⁰. The passenger domain is composed of hemagglutinin repeats—which are known to be involved in cell-cell aggregation—that are highly similar to those of the *Escherichia coli* autotransporters AIDA-I and Ag43, two proteins that are implicated in virulence. The bacterial surface protein AIDA-I mediates adherence to mammalian cells³¹, whereas Ag43 not only confers a low level of adhesion to certain mammalian cells, but also promotes the autoaggregation that is important for biofilm formation³².

Similarly, the *L. pneumophila* autotransporter may be involved in adhesion to the host cell and in biofilm formation. In contrast to AIDA-I and Ag43, the *L. pneumophila* autotransporter does not possess an RGD motif, which is implicated in binding to human integrins, and thus may have another interaction domain. The autotransporter was probably acquired by horizontal gene transfer, as is suggested by its

many upstream remnants of insertion sequences and its GC content of 41%, which exceeds the genome average of 38% (**Supplementary Fig. 1** online). Studies on the distribution of this gene in clinical and environmental *L. pneumophila* strains, together with research into its function, should provide insight to its importance.

In addition, both *L. pneumophila* strains contain a twin arginine translocation (Tat) secretion pathway (TatAB, TatC) and complete type I and II secretion systems. The type II system encoded by the *lspA* and *lspD* to *lspJ* genes is required for the secretion of several enzymes such as lipase A and B (Lpp0533, Lpp1159/Lpl0509, Lpl1164), acid phosphatase Map and SurE (Lpp1120, Lpp1245/Lpl1124, Lpl1245), lysophospholipase A (Lpp2291/Lpl2264) and the phospholipase PlA (Lpp1568/Lpl422), proteins that are all present in both strains.

Metabolism

The metabolic pathways used by *L. pneumophila* to multiply in eukaryotic cells are not known. The bacterium seems to prefer proteinaceous substrates, because a large set of oligopeptide and amino acid uptake and degradation systems are encoded in the genome. In particular, in addition to the *Pseudomonas aeruginosa* elastase homolog ProA (Lpp0532/Lpl0508), three secreted paralogous metalloproteases and 46 additional peptidases are present (**Supplementary Table 3** online).

By contrast, systems for sugar uptake are rare, although the complete Embden-Meyerhof and Entner-Doudoroff pathways are present. In neither *L. pneumophila* strain was a phosphotransferase-like uptake system identified. Some of the 55 ABC-type transport systems might be involved in sugar uptake, however, because the bacterium does possess a few systems for degrading complex sugars such as trehalase, polysaccharide deacetylase, eukaryotic-like glucoamylase (Lpp0489/Lpl0465), β -hexosaminidase and chitinases (Lpp1117/Lpl1121). Both strains encode proteins that are highly homologous to glycerol phosphate ABC transporters (Lpp1696, Lpp1695, Lpp1694/Lpl1695, Lpl1694, Lpl1693) and a hexose phosphate transporter (Lpp2623/Lpl2474), which may be important during intracellular growth. We also identified several enzymes that are probably involved in myo-inositol use and that may interfere with the host cell signaling mediated by this intracellular messenger.

L. pneumophila is predicted to encode an extensive aerobic respiratory chain consisting of NADH dehydrogenase, cytochrome-dependent succinate dehydrogenase, ubiquinol-cytochrome *c* reductase and four terminal oxidases (*aa*₃ cytochrome, two *bd* type quinol cytochromes, *o* type quinol cytochrome oxidase), which guarantee its capacity to cope with changing oxygen tension. The *o* type quinol cytochrome oxidase is absent in strain Lens. Systems involved in anaerobic respiration are apparently absent in all strains. Both strains encode a F₀F₁-type ATP synthase typical of γ -proteobacteria; in addition, strain Paris encodes a second ATP synthase similar to uncharacterized systems of archaea and marine bacteria.

L. pneumophila also encodes at least four Na⁺/H⁺ antiporters (Lpp1464, Lpp2448, Lpp0868, Lpp0667/Lpl1519, Lpl2304, Lpl0839, Lpl0651), which probably modulate the H⁺ and Na⁺ gradients across the cytoplasmic membrane. Thus, a Na⁺ motive force may be used for cellular activities. In this respect, the presence of a component of Na⁺-type polar flagellar motors, MotY, as well as two significantly different *motA-motB* gene clusters leads to the prediction that motility might be powered by Na⁺ as well as H⁺ motive forces. A particular feature of *L. pneumophila* is its differentiation to a mature intracellular form that accumulates inclusions of poly-hydroxybutyrate as a carbon and energy reserve. In agreement with this, strain Paris encodes four and strain Lens encodes three paralogous polyhydroxybutyrate synthases

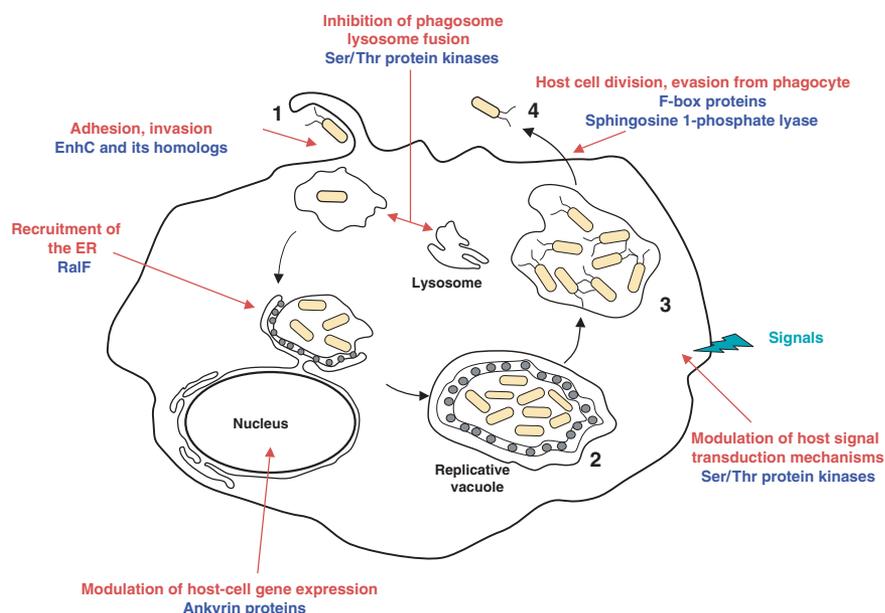


Figure 5 Steps in the intracellular growth of *L. pneumophila* in macrophages. Four different steps are shown: 1, *L. pneumophila* adhesion to and invasion of the host cell; 2, recruitment of organelles and their conversion into a rough endoplasmic reticulum-like compartment (note that the phagosome does not fuse with lysosomes); 3, intracellular replication of non-flagellated *L. pneumophila* inside a phagosome; and 4, release of flagellated *L. pneumophila*. Red indicates steps that are important in the infectious cycle of *L. pneumophila*. Blue indicates newly identified proteins that could interfere at these steps in the cycle.

Except for the above-mentioned genes, this 65-kb pathogenicity island is absent from strains Paris and Lens; however, the corresponding chromosomal location in strain Paris is the insertion site of an integrative plasmid discussed below. Thus, these two regions might be hot spots for genomic rearrangements. Genomic variation is also evident from the array of mobile elements, which includes 10 integrases, 58 insertion sequences (34 complete and 24 truncated ones; **Supplementary Table 7** online) and phage-related proteins. In addition, the genomes contain a large set of repeated sequences organized as inverted repeats, which are reminiscent of enterobacterial repeated intergenic consensus (ERIC) sequences. The *Legionella* or 'LeRIC' sequences fall into seven classes that are present in many copies (80, 18, 18, 25, 9, 9 and 6 in strain Paris) (**Supplementary Fig. 3** online).

L. pneumophila strains Paris and Lens contain *lvh*, a region that encodes a second type IV secretion system previously characterized in strain Philadelphia³⁷. Notably, the *lvh* region of *L. pneumophila* Paris is encoded in a 36-kb region that is either integrated in the chromosome or excised as a multicopy plasmid (data not shown). This pattern is similar to that described for the 30-kb unstable element of strain Oolda, which is possibly derived from phage and is involved in phase variation³⁸. The *lvh* region has a G+C content (43%) that differs from that of the rest of the chromosome (38%), and it contains some phage-related genes, suggesting a possible phage origin. The exact mode of excision and integration is not, however, understood. An appealing hypothesis is that the integration and excision of particular regions of the chromosome constitute a specific mechanism of *L. pneumophila* to increase versatility.

The second plasmid of strain Paris (132 kb) includes known virulence factors, mobile genetic elements and antibiotic resistance genes. The two-component regulator system *lrpR-lskS* present on this plasmid has been found on a *Legionella longbeachae* plasmid implicated in virulence of this species³⁹. The high conservation (93–98% protein identity) of the six genes sequenced on the 135-kb *L. longbeachae* plasmid might indicate recent horizontal transfer between *L. pneumophila* and *L. longbeachae*. *L. pneumophila* Lens contains a 60-kb plasmid that encodes several proteins homologous to the transfer region of the F plasmid of *E. coli*. All three plasmids of strains Paris and Lens encode a paralog of CsrA—a repressor of transmission traits and activator of replication⁴⁰.

Although the role of plasmids in *L. pneumophila* virulence remains to be determined, correlation between strains containing a plasmid and virulence in a mouse model has been described⁴¹. In addition, *L. pneumophila* strains with plasmids seem to persist longer in the environment than do strains lacking plasmids⁴². The identification in both clinical isolates of plasmids encoding putative virulence factors is another indication of the importance of plasmids for pathogenicity of *Legionella*.

The *L. pneumophila* genomes also show plasticity at the gene level. The *enh* loci, implicated in the entry into host cells⁴³, are present in strains Paris and Lens. One of the proteins encoded by these loci is RtxA, which contributes to entry, adherence, cytotoxicity, pore formation⁴³ and intracellular trafficking in amoebae⁴⁴. Unlike in strain AA100 (ref. 10), in strain Paris, *rtxA* is fused to *arpB* and a second gene with roughly 30 highly conserved tandem repeats of 549 bp. A similar structure is encoded by strain Lens; however, the latter strain contains two motifs in the repeated region, both of which differ from that of strain Paris although the number of repetitions seems to be the same (**Fig. 4**). It is possible that variations in the number and sequence of the repeats contribute to *L. pneumophila* versatility and to virulence.

Consistent with the plasticity apparent in their genomes, the *L. pneumophila* strains encode type IV pili—organelles that are required for natural competence for DNA transformation⁴⁵. The organization of the genes coding type IV pili is similar to that in *P. aeruginosa*, where they are crucial for bacterial adherence to and colonization of mucosal surfaces and for twitching motility. Another mechanism in *L. pneumophila* that contributes to genome plasticity is conjugative transfer mediated by the type IV secretion of plasmids⁴⁶ and chromosomal DNA⁴⁷.

DISCUSSION

Analysis and comparison of the genome sequences of the Paris and Lens strains of clinical *L. pneumophila* identify this bacterium as a highly versatile organism that shows extensive genome plasticity and diversity. The excision and integration of plasmids or genes might represent one mechanism that *L. pneumophila* exploits to adapt to different environments. Its large cohort of eukaryotic-like proteins is predicted to manipulate the host cell to the advantage of the pathogen (**Fig. 5**). Proteins of putative eukaryotic origin have been also

identified in other intracellular pathogens, including *Coxiella*, *Wolliella*, *Agrobacterium*, *Mycobacterium* and *Ehrlichia*, but currently *L. pneumophila* ranks as the prokaryote with the widest variety of eukaryotic-like proteins. Presumably, pathogenic *L. pneumophila* acquired DNA by horizontal transfer from its host or by convergent evolution during its coevolution with free-living amoebae. These proteins may also contribute to the infection of human macrophages.

On the basis of these genome sequences, future comparative and functional studies will enable us to define survival tactics of intracellular parasites and to identify the special attributes of endemic and epidemic *L. pneumophila*. To combat the increasing threat of *L. pneumophila* strains that are resistant to the chemicals used to decontaminate public water systems, including those of hospitals, the genome sequences could be used to identify targets for new biocides that are active against *L. pneumophila*.

METHODS

DNA preparation and sequencing. *L. pneumophila* strains Paris and Lens were grown on BCYE agar at 37 °C for 3 d and chromosomal DNA was isolated by standard protocols. Cloning, sequencing and assembly were done as described⁴⁸. For both genomes, two libraries (inserts of 1–2 and 2–3 kb) were generated by random mechanical shearing of genomic DNA, followed by cloning the fragments into pDNA-2.1 (Invitrogen). A scaffold was obtained by end-sequencing clones from a BAC library constructed as described⁴⁹ using pIndigoBac (Epicentre) as a vector. For *L. pneumophila* strain Paris, we constructed a medium-sized insert library (8–10 kb) in the low-copy number vector pSYX34. Plasmid DNA purification was done with either a Montage Plasmid Miniprep96 kit (Millipore) or a TempliPhi DNA sequencing template amplification kit (Amersham Biosciences). Sequencing reactions were done with an ABI PRISM BigDye Terminator cycle sequencing ready reactions kit and a 3700 or a 3730 XI Genetic Analyzer (Applied Biosystems). We obtained, assembled and finished 47,200 sequences for *L. pneumophila* strain Paris and 47,231 sequences for strain Lens from four libraries each as described⁴⁸.

Annotation and analysis. Definition of coding sequences and annotation were done as described⁴⁸ by using CAAT-box software⁵⁰. All predicted coding sequences were examined visually. Function predictions were based on BLASTp similarity searches and on the analysis of motifs using the PFAM, Prosite and SMART databases. We identified orthologous genes by reciprocal best-match BLAST and FASTA comparisons. To identify coiled-coil domains, we used the publicly available software PairCoil and Coilsan. Pseudogenes had one or more mutations that would prevent complete translation. Repetitive DNA sequences were identified by BLASTN comparisons of the intergenic regions and the complete genome. To predict the folding of single-strand DNA molecules, we used MFOLD software.

URLs. The sequence and the annotation of both *L. pneumophila* genomes are accessible at the LegioList Web Server (<http://genolist.pasteur.fr/LegioList>). For annotation and analysis we used PairCoil (<http://paircoil.lcs.mit.edu/cgi-bin/paircoil>) and Coilsan (<http://www.biology.wustl.edu/gcg/coilsan.html>). MFOLD server, <http://www.bioinfo.rpi.edu/applications/mfold/old/dna/>.

Accession numbers. GenBank: *L. pneumophila* strain Paris, CR628336; *L. pneumophila* strain Lens, CR628337; *L. pneumophila* Paris plasmid, CR628338; *L. pneumophila* Lens plasmid, CR628339. GenBank protein: strain AA100 *rtxA* locus, AAD41583.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank S.T. Cole for support and critical reading of the manuscript and A.S. Le Peuple for support and discussions. This work received financial support from the Institut Pasteur, the Centre National de la Recherche and the Consortium National de Recherche en Génomique du Réseau National des Genopoles, France. Public access to the partial genome sequence of *L. pneumophila* strain Philadelphia, sequenced at the Columbia Genome Center, New York, is

acknowledged. C.C. holds a fellowship jointly financed by the Centre National de la Recherche, France, and VeoliaWater–Anjou Recherche; H.B. holds a fellowship of the German Academy of Natural Scientists Leopoldina funded by the German Ministry of Education and Research.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 28 June; accepted 9 September 2004

Published online at <http://www.nature.com/naturegenetics/>

- McDade, J.E. *et al.* Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N. Engl. J. Med.* **297**, 1197–1203 (1977).
- Fields, B.S. The social life of Legionellae. in *Legionella* (eds. Marre R. *et al.*) 135–142 (ASM, Washington DC, 2002).
- Steinert, M., Hentschel, U. & Hacker, J. *Legionella pneumophila*: an aquatic microbe goes astray. *FEMS Microbiol. Rev.* **26**, 149–162 (2002).
- Fields, B.S., Benson, R.F. & Besser, R.E. Legionella and Legionnaires' disease: 25 years of investigation. *Clin. Microbiol. Rev.* **15**, 506–526 (2002).
- Atlas, R.M. *Legionella*: from environmental habitats to disease pathology, detection and control. *Environ. Microbiol.* **1**, 283–293 (1999).
- Molofsky, A.B. & Swanson, M.S. Differentiate to thrive: lessons from the *Legionella pneumophila* life cycle. *Mol. Microbiol.* **53**, 29–40 (2004).
- Yu, V.L. *et al.* 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 (2002).
- Aurell, H. *et al.* *Legionella pneumophila* serogroup 1 strain Paris: endemic distribution throughout France. *J. Clin. Microbiol.* **41**, 3320–3322 (2003).
- Goebel, M. & Yanagida, M. The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem. Sci.* **16**, 173–177 (1991).
- Cirillo, S.L.G., Lum, J. & Cirillo, J.D. Identification of a novel loci involved in entry by *Legionella pneumophila*. *Microbiology* **146**, 1345–1359 (2000).
- Conover, G.M., Derre, I., Vogel, J.P. & Isberg, R.R. The *Legionella pneumophila* LidA protein: a translocated substrate of the Dot/Icm system associated with maintenance of bacterial integrity. *Mol. Microbiol.* **48**, 305–321 (2003).
- Nagai, H., Kagan, J.C., Zhu, X., Kahn, R.A. & Roy, C.R. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295**, 679–682 (2002).
- Cole, S.T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
- Walburger, A. *et al.* Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* **304**, 1800–1804 (2004).
- Barz, C., Abahji, T.N., Trulzsch, K. & Heesemann, J. The *Yersinia* Ser/Thr protein kinase YpkA/YopO directly interacts with the small GTPases RhoA and Rac-1. *FEBS Lett.* **482**, 139–143 (2000).
- Seshadri, R. *et al.* Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. USA* **100**, 5455–5460 (2003).
- Wu, M. *et al.* Phylogenomics of the reproductive parasite *Wolbachia pipiensis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, E69 (2004).
- Hryniewicz-Jankowska, A., Czogalla, A., Bok, E. & Sikorski, A.F. Ankyrins, multifunctional proteins involved in many cellular pathways. *Folia Histochem. Cytobiol.* **40**, 239–249 (2002).
- Caturegli, P. *et al.* *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect. Immun.* **68**, 5277–5283 (2000).
- Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci.* **54**, 80–93 (1998).
- Kipreos, E.T. & Pagano, M. The F-box protein family. *Genome Biol.* **1** REVIEWS3002 (2000).
- Craig, K.L. & Tyers, M. The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog. Biophys. Mol. Biol.* **72**, 299–328 (1999).
- Schrammeijer, B. *et al.* Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. *Curr. Biol.* **11**, 258–262 (2001).
- Reiss, U. *et al.* Sphingosine-phosphate lyase enhances stress-induced ceramide generation and apoptosis. *J. Biol. Chem.* **279**, 1281–1290 (2004).
- Li, G., Foote, C., Alexander, S. & Alexander, H. Sphingosine-1-phosphate lyase has a central role in the development of *Dictyostelium discoideum*. *Development* **128**, 3473–3483 (2001).
- Biederbeck, A., Rose, S. & Elsasser, H.P. A human intracellular apyrase-like protein, LALP70, localizes to lysosomal/autophagic vacuoles. *J. Cell. Sci.* **112**, 2473–2484 (1999).
- Luo, Z.Q. & Isberg, R.R. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc. Natl. Acad. Sci. USA* **101**, 841–846 (2004).
- Vergunst, A.C. *et al.* VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science* **290**, 979–982 (2000).
- Morozova, I. *et al.* Comparative sequence analysis of the *icm/dot* genes in *Legionella*. *Plasmid* **51**, 127–147 (2004).
- Desvaux, M., Parham, N.J. & Henderson, I.R. The autotransporter secretion system. *Res. Microbiol.* **155**, 53–60 (2004).

31. Benz, I. & Schmidt, M.A. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. *Mol. Microbiol.* **6**, 1539–1546 (1992).
32. Klemm, P., Hjerrild, L., Gjermansen, M. & Schembri, M.A. Structure-function analysis of the self-recognizing antigen 43 autotransporter protein from *Escherichia coli*. *Mol. Microbiol.* **51**, 283–296 (2004).
33. Smith, R.S., Wolfgang, M.C. & Lory, S. An adenylate cyclase-controlled signaling network regulates *Pseudomonas aeruginosa* virulence in a mouse model of acute pneumonia. *Infect. Immun.* **72**, 1677–1684 (2004).
34. Deng, W. *et al.* Comparative genomics of *Salmonella enterica* serovar typhi strain. *J. Bacteriol.* **185**, 2330–2337 (2003).
35. Rankin, S., Li, Z. & Isberg, R.R. Macrophage-induced genes of *Legionella pneumophila*: protection from reactive intermediates and solute imbalance during intracellular growth. *Infect. Immun.* **70**, 3637–3648 (2002).
36. Brassinga, A.K. *et al.* A 65-kilobase pathogenicity island is unique to Philadelphia-1 strains of *Legionella pneumophila*. *J. Bacteriol.* **185**, 4630–4637 (2003).
37. Segal, G., Russo, J.J. & Shuman, H.A. Relationships between a new type IV secretion system and the *icm/dot* virulence system of *Legionella pneumophila*. *Mol. Microbiol.* **34**, 799–809 (1999).
38. Luneberg, E. *et al.* Chromosomal insertion and excision of a 30 kb unstable genetic element is responsible for phase variation of lipopolysaccharide and other virulence determinants in *Legionella pneumophila*. *Mol. Microbiol.* **39**, 1259–1271 (2001).
39. Doyle, R.M. & Heuzenroeder, M.W. A mutation in an *ompR*-like gene on a *Legionella longbeachae* serogroup 1 plasmid attenuates virulence. *Int. J. Med. Microbiol.* **292**, 227–239 (2002).
40. Molofsky, A.B. & Swanson, M.S. *Legionella pneumophila* CsrA is a pivotal repressor of transmission traits and activator of replication. *Mol. Microbiol.* **50**, 445–461 (2003).
41. Bezanson, G., Fernandez, R., Haldane, D., Burbridge, S. & Marrie, T. Virulence of patient and water isolates of *Legionella pneumophila* in guinea pigs and mouse L929 cells varies with bacterial genotype. *Can. J. Microbiol.* **40**, 426–431 (1994).
42. Brown, A., Vickers, R.M., Elder, E.M., Lema, M. & Garrity, G.M. Plasmid and surface antigen markers of endemic and epidemic *Legionella pneumophila* strains. *J. Clin. Microbiol.* **16**, 230–235 (1982).
43. Cirillo, S.L., Bermudez, L.E., El-Etr, S.H., Duhamel, G.E. & Cirillo, J.D. *Legionella pneumophila* entry gene *rtxA* is involved in virulence. *Infect. Immun.* **69**, 508–517 (2001).
44. Cirillo, S.L., Yan, L., Littman, M., Samrakandi, M.M. & Cirillo, J.D. Role of the *Legionella pneumophila* *rtxA* gene in amoebae. *Microbiology* **148**, 1667–1677 (2002).
45. Stone, B.J. & Kwaik, Y.A. Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. *J. Bacteriol.* **181**, 1395–1402 (1999).
46. Vogel, J.P., Andrews, H.L., Wong, S.K. & Isberg, R.R. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* **279**, 873–876 (1998).
47. Miyamoto, H., Yoshida, S., Taniguchi, H. & Shuman, H.A. Virulence conversion of *Legionella pneumophila* by conjugal transfer of chromosomal DNA. *J. Bacteriol.* **185**, 6712–6718 (2003).
48. Glaser, P. *et al.* Comparative genomics of *Listeria* species. *Science* **294**, 849–852 (2001).
49. Buchrieser, C. *et al.* The 102-kilobase *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect. Immun.* **67**, 4851–4861 (1999).
50. Frangeul, L. *et al.* CAAT-box, contigs-assembly and annotation tool-box for genome sequencing projects. *Bioinformatics* **20**, 790–797 (2004).