Intracellular parasitism, the driving force of evolution of Legionella pneumophila and the genus Legionella

Laure Gomez-Valero, Carmen Buchrieser

Intracellular parasitism, the driving force of evolution of *Legionella pneumophila* and the genus *Legionella*

Laure Gomez-Valero and Carmen Buchrieser *

Institut Pasteur, Biologie des Bactéries Intracellulaires and CNRS UMR 3525, 75724, Paris, France

Running title: Co-evolution of *Legionella* and protozoa

*For correspondence:
Carmen Buchrieser
Institut Pasteur
Biologie des Bactéries Intracellulaires
28, rue du Dr. Roux, 75724 Paris Cedex 15, France
Tel: (33-1)-45-68-83-72
E-mail:***cbuch@pasteur.fr***

Abstract

Legionella pneumophila is an intracellular pathogen that causes a severe pneumonia called Legionnaires' disease that is often fatal when not promptly diagnosed and treated. However, *L. pneumophila* is mainly an environmental pathogen of protozoa. This bacterium parasitizes free-living amoeba and other aquatic protozoa with which it co-evolved over an evolutionary long time. Due to the close relationship between hosts and pathogens, their co-evolution leads to molecular interactions such as the exchange of genetic material through horizontal gene transfer (HGT). Those genes that confer an advantage to the bacteria were fixed in their genomes and help these pathogens to subvert host functions to their advantage. Genome sequencing of *L. pneumophila* and recently of the entire genus *Legionella* that comprises over 60 species revealed that Legionellae have co-opted genes and thus cellular functions from their eukaryotic hosts to a surprisingly high extent never observed before for an prokaryotic organism. Acquisition and loss of these eukaryotic-like genes and eukaryotic domains is an on-going process underlining the highly dynamic nature of the *Legionella* genomes. Although the large amount and diversity of HGT that occurred between *Legionella* and their protozoan hosts seems to be unique in the prokaryotic world the analyses of more and more genomes from environmental organisms and symbionts of amoeba revealed that such genetic exchanges occur among all amoeba associated bacteria and also among the different microorganisms that infect amoeba such as viruses. This dynamic reshuffling and gene-acquisition has led to the emergence of major human pathogens such as *Legionella* and may lead to the emergence of new human pathogens from the environment.

Introduction

Pathogenicity refers to the ability of an organism to cause disease and to harm the host. The more virulent a pathogen the higher the degree of host damage it can induce, but virulence evolves to the level that optimizes the pathogens reproduction and transmission rate. Pathogenicity and virulence developed through co-evolution of pathogens with its hosts, a major driver of evolution and biological innovation over millions of years. Host-pathogen co-evolution is very widespread across ecosystems, but perhaps the best studied is that occurring between plants, animals or humans and pathogenic parasites, fungi, viruses or bacteria. The result of this reciprocal selection lead to the evolution of sophisticated mechanisms to subvert host functions and shaped the immune defences in eukaryotic cells that should eliminate invading microorganisms.

Due to the close relationship between hosts and pathogens, their co-evolution leads to molecular interactions such as the exchange of genetic material through horizontal gene transfer (HGT). During evolution the sequences acquired can be adapted to the recipients species and thereby improve its fitness and affect the interaction between the pathogen and its host. Inter-bacterial HGT was first described in 1959 when the ability of *Shigella* to incorporate drug resistance genes from other *Shigella* strains and from *Escherichia coli* was discovered (1). Since then, it became clear that HGT is an important force driving the evolution of bacteria and archaea, as well as that of unicellular eukaryotes (2). It has now also been shown that prokaryotes cannot only exchange genetic material with other prokaryotes and viruses with viruses, but also between them and with eukaryotes. However, there are only few reports of eukaryote-to-prokaryote HGT.

One intriguing case where eukaryote-to-prokaryote HGT has been described is the co-evolution of *Legionella* with protozoa. *Legionella* are environmental bacteria belonging to the class of γ-proteobacteria. The genus contains over 60 species, among which *Legionella pneumophila* and *Legionella longbeachae* are major human pathogens that are known as the etiological agent of Legionnaires' disease, a severe pneumonia that is often fatal when not treated promptly (3, 4). *Legionella* are ubiquitous in fresh water reservoirs worldwide but certain species are also found in moist soil, where they parasitize within free-living protozoa (5). The finding that these bacteria replicate intracellularly in environmental protozoa such as *Acanthamoeba castellanii, Verbamoeba veriformis* or *Hartmanella veriformis* led to a new perception in microbiology: the ability of a bacterium to replicate within human monocytes and alveolar macrophages, may be derived from the conserved cell biology between amoeba, its natural host in aquatic environments, and human phagocytic cells (5-7).

Indeed, *L. pneumophila* encodes a type IV secretion system named Dot/Icm (8, 9) that is secreting proteins, which allow this bacterium to manipulate host functions in protozoan and in human cells. Furthermore, several of the traits that contribute to the fitness of *L. pneumophila* in the environment (protozoa) also facilitate its growth in alveolar macrophages (reviewed in (4, 10-12)). However, how the adaptation of *Legionella* to eukaryotic cells and the ability to replicate intracellularly may have evolved on the molecular level was not known.

The *L. pneumophila* genome sequence, a breakthrough in the understanding of *Legionella*-protozoa co-evolution

Legionella pneumophila was one of the human pathogens whose genome was completely sequenced only in 2004. The analysis of its genome sequence was key for a new understanding of the strategies employed by *Legionella* to subvert host functions as the genome sequence uncovered an intriguing feature of the *L. pneumophila* genome. It encodes an unmatched large number and diversity of bacterial proteins with eukaryotic-like properties (13). Among the about 3000 protein coding genes predicted in the genome, more than 150 proteins with high similarity to eukaryotic proteins or carrying eukaryotic motifs were predicted, representing about 5 % of its protein-coding capacity, a number that increased later when systematic searches were employed. Examples of protein domains that had been identified in the *L. pneumophila* genome are F-box and U-box domain proteins, SET-domains, Sel1-domains, STPK domains and Ankyrin domains (13). Examples for proteins homologue to eukaryotic proteins, which are proteins with ≥ 30% amino acid similarity over at least two thirds of the eukaryotic protein length are a eukaryotic glycoamylase, apyrases, or a sphingosine-1 phosphate lyase. This finding led to the suggestion that *L. pneumophila* secretes these proteins in the host cell to subvert eukaryotic signalling pathways by mimicking host cell functions (13). Indeed, also the first described Dot/Icm effector RalF that was identified before the genome was sequenced, encodes a eukaryotic Sec7 domain. Sec7 domains are components of Arf-specific guanine nucleotide exchange factors (GEFs). GEFs catalyse the nucleotide exchange of Arfs thereby converting them from an inactive state (GDP-bound) to the active one (GTP-bound). Like in a eukaryotic cell, following secretion into the host cell, RalF recruits Arf-1 and then functions like an Arf-1 specific GEF (14).

Based on the information gleaned from the genome sequence analyses many of the eukaryotic-like effectors were functionally analysed to learn whether they are bacterial weapons employed to subvert host functions as predicted. Indeed, each of the effectors analysed to date, encoded the predicted eukaryotic function and was shown to be part of a sophisticated effector network that evolved to manipulate the host cell. These effectors modulate a plethora of host cell processes including vesicular trafficking, apoptosis,

autophagy, protein synthesis, ubiquitination, epigenetic modifications, and induce many different post translational modifications (PTM) (15). They may induce the PTMs directly as do AnkB or LubX that contain an F-box or U-box motif, respectively and function as E-3 ligases that transfer ubiquitin moieties to host proteins (15-17) or they may recruit host enzymes such as the eukaryotic protein prenyl transferases to achieve membrane localization of the respective effector (18, 19).

More than 330 effectors secreted by the Dot/Icm type IV secretion system (T4SS) and over 25 proteins secreted by the type II secretion system (T2SS) have been described for *L. pneumpohila* (20-22). With a genome size of in average 3.2 Mb and 3100 protein coding genes this astonishing number of over 350 secreted proteins which represent over 10% of the *L. pneumophila* proteome is not matched by any other known bacterial pathogen. The closest comes *Coxiella burnetti,* which has a genome size of about 2Mb and about 2100 predicted protein coding genes (21, 22) and over 100 secreted effector proteins (23). Thus the question arises why does *Legionella* need that many effector proteins? This question becomes even more puzzling as many of the effectors studied to date do not show any or at least no strong intracellular growth defect when deleted nor does even the simultaneous deletion of over 60 effectors obtained through large chromosomal deletions that carry these genes (24). Based on the different data, it is thought that *L. pneumophila* encodes such a high number of secreted proteins, to fine-tune the host pathogen interactions to allow the replication in many different protozoan hosts. Thus the redundancy of effector functions observed in intracellular growth in human or mouse macrophages might be beneficial for *Legionella* when parasitizing protozoa in the environment as *L. pneumophila* may use different effector sets adapted to different protozoan species.

The genus *Legionella* co-opts eukaryotic functions to an unprecedented high number and diversity

Legionella pneumophila is part of a large genus of over 65 species of which most are harmless, environmental bacteria found in aquatic environments associated with amoeba. *Legionella longbeachae,* is the second species often found in human disease as it is a frequent cause of Legionnaires' disease in Australia, New Zealand and Southeast Asia but it emerges lately also in Europe and the United States (25). However most of the other *Legionella* species have been only rarely or never found in human disease and only little is known about them. Thus an exciting question to answer was, whether the presence of eukaryotic genes and eukaryotic domains is a general feature of the *Legionella* genomes. A first answer came from the analyses of the *L. longbeachae* genome, as indeed the effector repertoire seemed of similar size and a high number of eukaryotic domains and proteins had been identified (26). However the surprising finding was, that only about 34% of the

L. pneumophila effectors were conserved in the *L. longbeachae* genome, but 51 new, putative Dot/Icm substrates specific for *L. longbeachae* that encode eukaryotic-like domains were identified (26). Related to a different life style, *L longbeachae* is found in moist soil and potting soil, genes that might have been acquired from plants have been identified, such as proteins with pentatricopeptide repeat (PPR) domains, a family of proteins that is greatly expanded in plants.

Recently the nearly entire genus *Legionella* has been sequenced and analysed (27, 28). This disclosed a fascinating and unique feature of these bacteria. A highly dynamic and diverse effector repertoire of over 18 000 proteins that contain at least 137 different eukaryotic domains and over 200 different eukaryotic proteins was discovered (28). Comparative genome and evolutionary analyses brought evidence that *Legionella* species have acquired these eukaryotic-like proteins from all domains of life, plant, animal, fungal, and archaea (28). A particular exciting finding was the identification of 184 genes that are predicted to encode small GTPases, 71 of which are Rab GTPases. All have the best Blast hit with proteins form protozoan organisms such as *Entamoeba* or *Tetrahymena*. Furthermore phylogenetic analyses indicate that these proteins are indeed acquired from protozoan hosts (28). Thus RabGTPases are a unique feature of the genus *Legionella.*

Most interestingly, despite the enormous diversity of eukaryotic domains present in the *Legionella* effectors, it seems that certain signalling pathways are exploited by all species. Indeed, quasi all genomes contain U- and/or F-box proteins suggesting that the exploitation of ubiquitin signaling is of outmost importance to succeed replication inside eukaryotic host cells (28). Another example is the eukaryotic-type ecto-NTPDases (apyrases), which are conserved in all species analysed. It has been shown that this protein confers to *L. pneumophila* the ability to hydrolyse ATP, a function that seems necessary for optimal intracellular replication (29). Recently the structure of NTPDases from a legume plant revealed that these NTPDases could adopt two conformations depending on the molecule and co-factor bound in the active site (30). Interestingly this phenomenon had been previously described in *Rattus norvegicus*, *Toxoplasma gondii* NTPDaseIII and the *L. pneumophila* NTPDaseI suggesting a common catalytic mechanism across the domains of life. This structural similarity again supports the idea that *Legionella* have acquired these functions from eukaryotic organisms. Thaumatin domains that are considered a prototype for a pathogen-response protein domain in fungi, plants, and animals are also present in all *Legionella* genomes (28). Another interesting domain is the SET domain encoded by RomA of *L. pneumophila* where it has been shown to induce a unique host chromatin modification (31). This domain is present in nearly all *Legionella* species but *L. longbeachae* (28) suggesting that modification of histones is an important mechanism by which *Legionella* facilitate their intracellular survival. Thus although most surprisingly only a set of 8 conserved

core effectors was identified in the genus *Legionella* (27, 28) the identification of the presence of conserved domains suggests that one could perhaps define a core set of eukaryotic signalling pathways that intracellular bacteria need to modulate to replicate intracellularly.

Inter-domain horizontal gene transfer and the emergence of a human pathogen

The high number and wide variety of eukaryotic functions discovered in the *Legionella* genomes suggested that inter domain horizontal gene transfer may be the mechanism of acquisition and that these proteins and domains of eukaryotic origin witness the tight co-evolution between *Legionella* and its protozoan hosts (13, 26, 32-34). Many reviews on the functions of these different effectors of *L. pneumophila* and how they subvert host signalling pathways have been published in the last years (e.g. (15, 35-40)) thus we will here further detail only one "eukaryotic-like" effector protein, the *L. pneumophila* sphingosine-1 phosphate lyase named *Lp*Spl, as this protein is an excellent example of how these eukaryotic like proteins might have been acquired and evolved.

Sphingolipids are major components of all eukaryotic cellular membranes. They have important functions as signalling molecules in the eukaryotic cell by regulating processes such as the stress response, cell proliferation, apoptosis, angiogenesis, genetic diseases, and resistance to chemotherapy (41). Simplified, sphingomyelin, present in plasma membranes is hydrolysed by sphingomyelinase to ceramide that can also be *de novo* synthetized, which then is converted by ceramidase to sphingosine, which is phosphorylated by a sphingosine kinase to sphingosine-1 phosphate that can be converted by sphingosine-1 phosphate lyase (Spl) to hexadecanal + Ethanolamine-P (42). Interestingly, sphingolipid biosynthesis was shown to be conserved in *Acanthamoeba castellanii* and appears to be generally conserved among protozoa (43). Only very few bacteria such as *Sphingobacterium, Sphingomonas* and *Bacteroides* and *Bdellovibrio stolpii* are able to synthetize sphingolipids (42). Thus it was an intriguing finding that the *L. pneumophila* genome encodes several eukaryotic enzymes participating in the sphingolipid pathway, such as sphingosine kinase, sphingomyelinase and sphingosine-1 phosphate lyase (44).

The *L. pneumophila* sphingosine-1 phosphate lyase named *Lp*Spl, was further characterized. Its structural analyses showed that *Lp*Spl has a dimeric multidomain architecture that is very similar to the previously characterized SPL structures of the human (hSPL) and the yeast (Dpl1p) enzyme. Their comparison revealed that the active site of the enzyme was conserved among the *Lp*Spl and hSpl and activity analyses confirmed that the *L. pneumophila* Spl shows indeed sphingosine-1 phosphate lyase activity like its human counterpart (45). Furthermore metabolomics analyses of *L. pneumophila* infected human macrophages revealed that *L. pneumophila Lp*Spl targets the sphingolipid metabolism of the

host cell directly to modulate the levels of sphingosine and restrains autophagy (45). Thus, *Lp*Spl is an enzyme that modulates the host cell sphingolipid metabolism to the pathogens advantage.

The question arises "what is the origin of such an eukaryotic enzyme in an prokaryotic genome?". To answer this question we have undertaken phylogenetic analyses of this gene by recruiting homologous sequences from a database containing only completed genome sequences. Selected representatives of all eukaryotic groups and one representative of each bacterial species were included in the analyses. After Blastp only significant hits (e-value <10 × 10^{-4}) were retained, and only one hit for each species was included in the analysis. The resulting phylogenetic tree is shown in **Figure 1A**. Indeed, the *L. pneumophila Lp*Spl gene is embedded in the same clade as the eukaryotic sequences from *Entamoeba spp., Tetrahymena thermophila* and *Paramecium tetraurelia* Spl, which is suggesting that *Lp*Spl was acquired by horizontal gene transfer from a protist host as also suggested earlier (46, 47). The analyses of the distribution of the sphingosine-1 phosphate lyase in the genus *Legionella* reveals that this enzyme is present in 16 of the 58 *Legionella* species/subspecies analyzed **(Figure 1B**) suggesting that the remaining 42 species have evolved other ways to manipulate the host sphingolipid metabolism or employ different strategies to restrain autophagy. Indeed, even among different *L. pneumophila* species are differences in how they subvert the autophagy pathway. An example is RavZ, an effector of *L. pneumophila* strain Philadelphia that inhibits autophagosome maturation through irreversible ATG8 decongugation that is absent from strain Paris (48).

To better understand the evolutionary history of the sphingosine-1 phosphate lyase (*spl*) gene in the genus *Legionella*, we have analysed the phylogenetic relationship of the 16 *Legionella spl* genes. As shown in **Figure 2A,** the protein similarity ranges from 63-100% and five highly related groups can be distinguished. *L. pneumophila* subsp *pneumophila*, *L. pneumophila* subsp *pascuelleii, L. pneumophila* subsp *fraserii and L.* waltersii form one group where the Spl sequence shows 95-100% similarity to the *L. pneumophila Lp*Spl sequence. A second group with 70% sequence similarity is formed by *L. gresiliensis* and *L. busanensis,* a third group that shows 67-69% similarity to *Lp*Spl contains the species *L. hackeliae, L. jamestowniensis* and *L. brunensis* and finally the least conserved group contains five species that show 63-68% sequence similarity to *Lp*Spl (**Figure 2A**). Thus the phylogeny of the different Spl proteins in the genus *Legionella*, suggests either acquisition of an *spl* gene in a common ancestor and subsequent diversifying evolution and losses in many species or multiple acquisitions. To answer this question, we overlapped the distribution of the *spl* sequences on the phylogeny of the genus (**Figure 2B)** and carried out evolutionary analysis of presence/absence using GLOOME and stochastic mapping. These analyses showed that the *spl* gene has been acquired at least four times during the evolution of the

genus (green arrows) and has also been lost several times (red dots). Thus gene gain and loss seems to be an on-going process that shapes the *Legionella* genomes.

L. pneumophila was one of the first examples for evidence of eukaryote to prokaryote gene transfer. However, genome analyses from environmental bacteria including symbionts of amoeba showed that eukaryotic domains were also present in the amoeba symbiont *Amoebophilus asiaticus* (49). An analyses of 480 genomes of different prokaryotes revealed that eukaryotic domains are significantly enriched in the genomes of many amoeba-associated bacteria such as *Chlamydiae, Rickettsia bellii*, *Francisella tularensis,* or *Mycobacterium avium* (49). This indicates that phylogenetically and ecologically diverse bacteria, which thrive inside amoebae, exploit common mechanisms for interaction with their hosts and are all exchanges genetic material (49). Recently it was also proposed that amoeba-fungal interaction might select for traits that promote survival during animal infection and thereby contribute to virulence (50). Thus similar processes may contribute to the evolution of other amoeba-associated bacteria and fungi and may lead to the emergence of new human pathogens.

Horizontal gene transfer among amoeba associated bacteria or viruses within amoeba

The availability and comparison of genome sequences from organisms belonging to all domains of life and residing in different environmental niches brought evidence that HGT may occur between many organisms and not only between closely related species but even between different domains of life. In this context amoeba seem to be a privileged environment for DNA exchange. Indeed, *Legionella* seems to have exchanged genetic material also with viruses that infect amoeba, as it was reported that *L. pneumophila* encodes proteins homologous to proteins found in the mimivirus genome (51), a virus that infects *Acanthamoeba* (34, 52). Most interestingly, this situation seem to be reciprocal as intracellular bacteria appear to have transferred genes also to the mimiviral genome, some of which are involved in the parasitic adaptations of the mimivirus (52).

In addition to gene exchange between amoeba-associated bacteria such as *Legionella* with viruses, there is also evidence of gene exchange between different bacteria infecting amoeba. *Rickettsia* which also replicate in amoeba, contain genes encoding a putative conjugal DNA transfer system highly similar to that of *Protochlamydia amoebophila* UWE25, an obligate symbiont of amoebae and other genes highly similar to homologues in intracellular bacteria of amoebae (53). Indeed, one of the secreted effectors of *L. pnemophila*, RalF contains a eukaryotic Sec-7 domain (14). The analyses of the evolutionary history of this domain reveals that a similar domain is present in the *Rickettsia* genomes, and that both, *Rickettsia* and *Legionella* Sec-7 sequences are embedded within eukaryotic sequences suggesting that one of the bacteria acquired this domain from an

amoeba host and then the bacteria exchanged this domain among them (54). Another interesting report reveals that amoeba may also acquire genes from their bacterial parasites or symbionts. The anaerobic protist *Mastigamoeba balamuthi* encodes p-cresol- and indole-producing enzymes that most likely originated from phagocytized bacteria in the protist's anoxic habitat and allowed the eukaryotic recipient to produce the bacterial weapon *p*-cresol at bacteriostatic concentrations (55).

Thus gene exchange between many different organisms may take place and if the acquired DNA confers an advantage to the recipient, it will be fixed and will evolve further with the new genome. Thus evolutionary analyses might miss the real extent of these gene exchanges as there are likely genes *e.g.* in prokaryotes that originated from eukaryotic species but there are no identifiable eukaryotic homologs presumably due to substantial evolution of these proteins after their acquisition by the bacteria as suggested for the *Legionella* SH2 domain proteins (56). Another reason that makes it difficult to trace the evolutionary history of certain genes is due to the fact that we do not have enough sequencing data for environmental protozoa, fungi and bacteria. Once databases are enriched with such sequences we might see even more genetic exchange than thought.

Conclusion

Amoeba associated bacteria seem to strive in an environment that is prone to HGT. Gene exchange among amoeba associated bacteria such as *Legionellla, Chlamydia* or *Rickettsia* as well as between the amoebal host and the parasitizing bacteria or the viruses present in amoeba takes place (13, 26, 34, 47, 57-61). The investigation of the function of these horizontally acquired genes, suggests that they confer a selective advantage to the bacteria. Indeed, *Legionella* have transformed these proteins, using them as “tools of oppression” to hijack host cellular functions, in particular targeting signal transduction, protein turnover and chromatin modifying pathways. However, the finding that *Legionella* species have acquired eukaryotic-like proteins from all domains of life, plants, animals, fungi, and archaea, in an unprecedented high number and large diversity opens many new questions. One intriguing question is “what is the mechanism by which these transfers occur?” and “how is the foreign DNA integrated in the prokaryotic genomes?”. An interesting finding that may be related to the inter domain gene transfer is the identification of a gene predicted to encode a group II intron reverse transcriptase in the *L. pneumophila* genome. Thus a possibility is that *L. pneumophila* incorporates also RNA from its host, a fact that would explain why the eukaryotic genes in *Legionella* do not carry introns. The proof that RNA may be transferred horizontally would be the discovery of a new key mechanism for evolution and adaption of bacteria. Furthermore *Legionella* are able to develop competence for natural transformation (62), a major mechanism of HGT which may act in the intracellular environment of amoeba.

However, experimental proof is missing yet. Thus many exciting questions on the evolution of *Legionella* that may teach us also how new human pathogens may evolve from the environment remain to be answered. The knowledge on these evolutionary processes will be a precious help to avoid the emergence of new pathogens and gives an exciting outlook on future research.

Acknowledgements

Work in the CB laboratory is financed by the Institut Pasteur and has received funding from the French Government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant n°ANR-10-LABX-62-IBEID) and grant ANR 15 CE17 0014 03

Figure legends

Figure 1: The sphingosine -1 phosphate lyase gene was acquired by horizontal gene transfer from protozoa but is not conserved throughout the genus. A) Phylogeny of the sphingosine-1 phosphate lyse (Lpp2128) of *L. pneumophila* strain Paris and homologous sequences from eukaryotic and prokaryotic organisms. Homologous sequences were recruited from a database containing only completed genome sequences. Selected representatives of all eukaryotic groups and one representative of each bacterial species are included in the analyses. According to Blastp comparisons only significant hits were recruited (e-value $<10 \times 10^{-4}$), and only one hit for each species was retained. The alignment was performed with Muscle for Lpp2128, and followed by manual curation. The phylogeny was reconstructed using a distance method (NJ) with 1000 bootstrap replicates. The corresponding support values are shown in each node (values lower than 50 are not represented). Bars represent 20% and 10% of estimated phylogenetic divergence, respectively. B) Distribution of the orthologous genes of the sphingosine-1 phosphate lyse Lpp2128 in the genus *Legionella*. Orthology prediction has been done using the Pan Genome Ortholog Clustering Tool (PanOCT) with the following parameters: amino acid percentage identity cut-off 30%, BLAST e-value cutoff 10^{-5}, and minimum percentage match length of subject and query 65%

Figure 2: The sphingosine-1 phosphate lyase genes have been acquired and lost multiple times during the evolution of the genus *Legionella*. A) Phylogeny of the sphingosine-1 phosphate lyse (Lpp2128) of *L. pneumophila* strain Paris and the corresponding orthologous proteins in other *Legionella* species. The tree was constructed by likelihood using the software RAxML with 100 bootstrap replicates. The corresponding support values are shown in each node. B) Gain/loss prediction for the sphingosine-1 phosphate lyase in the different *Legionella* species. The prediction was done based on parsimony using the program GLOOME with double cost of gain versus loss. Green arrows represent acquisition events and red dots on the branches represent gain events

References

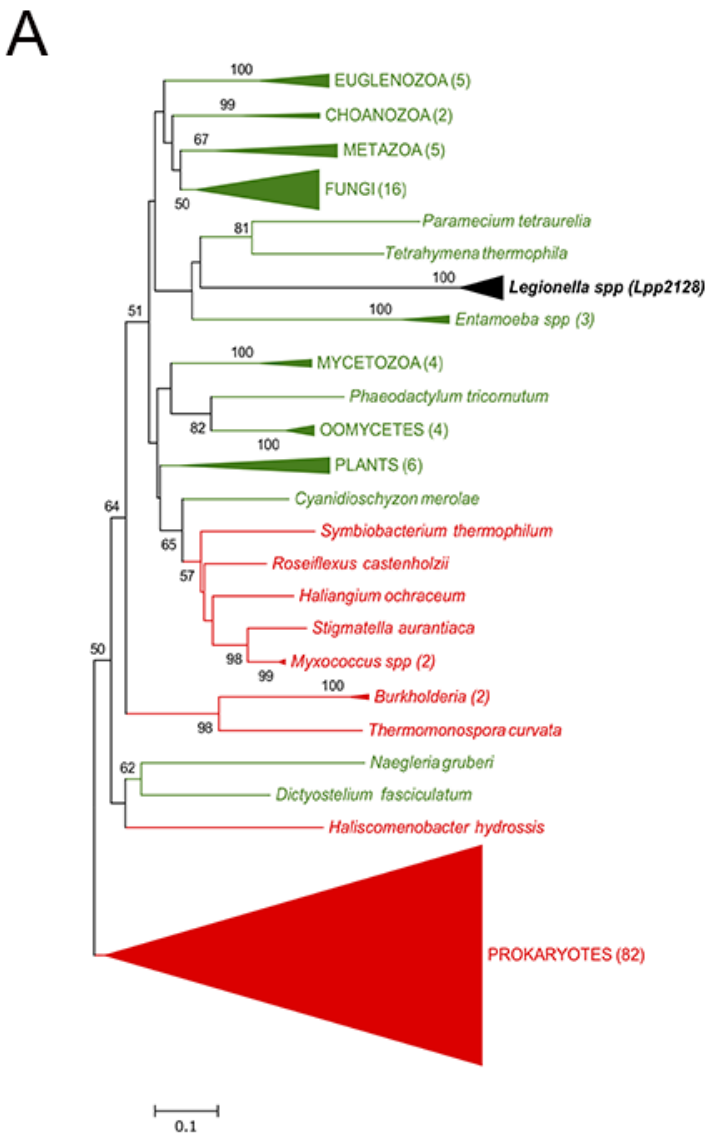
1. Ochiai K, Yamanaka T, Kimura K, Sawada O. Inheritance of drug resistance (and its transfer) between *Shigella* strains and between *Shigella* and *E. coli* strains". Hihon Iji Shimpor (in Japanese). 1959(1861):34.
2. Boto L. Horizontal gene transfer in evolution: facts and challenges. Proc Biol Sci. 2010;277(1683):819-27.
3. McDade JE, Shepard CC, Fraser DW, Tsai TR, Redus MA, Dowdle WR. Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. N Engl J Med. 1977;297(22):1197-203.
4. Newton HJ, Ang DK, van Driel IR, Hartland EL. Molecular pathogenesis of infections caused by *Legionella pneumophila*. Clin Microbiol Rev. 2010;23(2):274-98.
5. Rowbotham TJ. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. J Clin Pathol. 1980;33(12):1179-83.
6. Horwitz MA, Silverstein SC. Legionnaires' disease bacterium (*Legionella pneumophila*) multiples intracellularly in human monocytes. J Clin Invest. 1980;66(3):441-50.
7. Nash TW, Libby DM, Horwitz MA. Interaction between the legionnaires' disease bacterium (*Legionella pneumophila*) and human alveolar macrophages. Influence of antibody, lymphokines, and hydrocortisone. J Clin Invest. 1984;74(3):771-82.
8. Berger KH, Isberg RR. Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila.* Mol Microbiol. 1993;7(1):7-19.
9. Marra A, Blander SJ, Horwitz MA, Shuman HA. Identification of a *Legionella pneumophila* locus required for intracellular multiplication in human macrophages. Proc Natl Acad Sci U S A. 1992;89(20):9607-11.
10. Escoll P, Rolando M, Gomez-Valero L, Buchrieser C. From Amoeba to Macrophages: Exploring the Molecular Mechanisms of *Legionella pneumophila* Infection in Both Hosts. Curr Top Microbiol Immunol. 2013;376:1-34.
11. Hubber A, Roy CR. Modulation of host cell function by *Legionella pneumophila* type IV effectors. Annu Rev Cell Dev Biol. 2010;26:261-83.
12. Isberg RR, O'Connor TJ, Heidtman M. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. Nat Rev Microbiol. 2009;7(1):13-24.
13. Cazalet C, Rusniok C, Bruggemann H, Zidane N, Magnier A, Ma L, et al. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. Nat Genet. 2004;36(11):1165-73.
14. Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. Science. 2002;295(5555):679-82.
15. Rolando M, Buchrieser C. Post-translational modifications of host proteins by *Legionella pneumophila*: a sophisticated survival strategy. Future Microbiol. 2012;7(3):369-81.
16. Kubori T, Hyakutake A, Nagai H. *Legionella* translocates an E3 ubiquitin ligase that has multiple U-boxes with distinct functions. Mol Microbiol. 2008;67(6):1307-19.
17. Price CT, Al-Khodor S, Al-Quadan T, Santic M, Habyarimana F, Kalia A, et al. Molecular Mimicry by an F-Box Effector of *Legionella pneumophila* Hijacks a Conserved Polyubiquitination Machinery within Macrophages and Protozoa. PLoS Pathog. 2009;5(12):e1000704.
18. Ivanov SS, Charron G, Hang HC, Roy CR. Lipidation by the host prenyltransferase machinery facilitates membrane localization of *Legionella pneumophila* effector proteins. J Biol Chem. 2010;285(45):34686-98.
19. Price CT, Al-Quadan T, Santic M, Jones SC, Abu Kwaik Y. Exploitation of conserved eukaryotic host cell farnesylation machinery by an F-box effector of *Legionella pneumophila*. J Exp Med. 2010;207(8):1713-26.
20. DebRoy S, Dao J, Soderberg M, Rossier O, Cianciotto NP. *Legionella pneumophila* type II secretome reveals unique exoproteins and a chitinase that promotes bacterial persistence in the lung. Proc Natl Acad Sci U S A. 2006;103(50):19146-51.

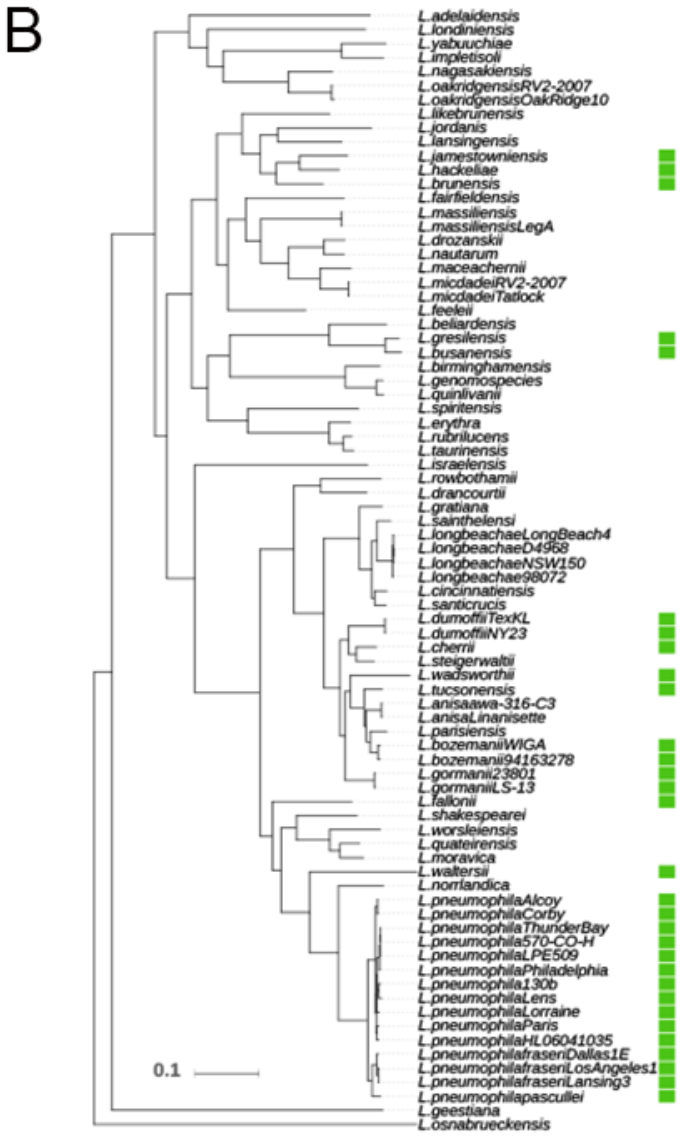
21. Lifshitz Z, Burstein D, Peeri M, Zusman T, Schwartz K, Shuman HA, et al. Computational modeling and experimental validation of the Legionella and Coxiella virulence-related type-IVB secretion signal. Proc Natl Acad Sci U S A. 2013;110(8):E707-15.
22. Zhu W, Banga S, Tan Y, Zheng C, Stephenson R, Gately J, et al. Comprehensive Identification of Protein Substrates of the Dot/Icm Type IV Transporter of *Legionella pneumophila*. PLoS One. 2011;6(3):e17638.
23. Crabill E, Schofield WB, Newton HJ, Goodman AL, Roy CR. Dot/Icm-Translocated Proteins Important for Biogenesis of the Coxiella burnetii-Containing Vacuole Identified by Screening of an Effector Mutant Sublibrary. Infect Immun. 2018;86(4).
24. O'Connor TJ, Adepoju Y, Boyd D, Isberg RR. Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. Proc Natl Acad Sci U S A. 2011;108(36):14733-40.
25. Bacigalupe R, Lindsay D, Edwards G, Fitzgerald JR. Population Genomics of *Legionella longbeachae* and Hidden Complexities of Infection Source Attribution. Emerg Infect Dis. 2017;23(5):750-7.
26. Cazalet C, Gomez-Valero L, Rusniok C, Lomma M, Dervins-Ravault D, Newton HJ, et al. Analysis of the *Legionella longbeachae* genome and transcriptome uncovers unique strategies to cause Legionnaires' disease. PLoS Genet. 2010;6(2):e1000851.
27. Burstein D, Amaro F, Zusman T, Lifshitz Z, Cohen O, Gilbert JA, et al. Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. Nat Genet. 2016;48(2):167-75.
28. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Perez-Cobas AE, Rolando M, et al. More than 18,000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. Proc Natl Acad Sci U S A. 2019:in press.
29. Riedmaier P, Sansom FM, Sofian T, Beddoe T, Schuelein R, Newton HJ, et al. Multiple ecto-nucleoside triphosphate diphosphohydrolases facilitate intracellular replication of *Legionella pneumophila*. Biochem J. 2014;462(2):279-89.
30. Summers EL, Cumming MH, Oulavallickal T, Roberts NJ, Arcus VL. Structures and kinetics for plant nucleoside triphosphate diphosphohydrolases support a domain motion catalytic mechanism. Protein Sci. 2017;26(8):1627-38.
31. Rolando M, Sanulli S, Rusniok C, Gomez-Valero L, Bertholet C, Sahr T, et al. *Legionella pneumophila* effector RomA uniquely modifies host chromatin to repress gene expression and promote intracellular bacterial replication. Cell Host Microbe. 2013;13(4):395-405.
32. de Felipe KS, Pampou S, Jovanovic OS, Pericone CD, Ye SF, Kalachikov S, et al. Evidence for acquisition of *Legionella* type IV secretion substrates via interdomain horizontal gene transfer. J Bacteriol. 2005;187(22):7716-26.
33. Gomez Valero L, Runsiok C, Cazalet C, C B. Comparative and functional genomics of *Legionella* identified eukaryotic like proteins as key players in host-pathogen interactions. Front Microbiol. 2011;2.
34. Lurie-Weinberger MN, Gomez-Valero L, Merault N, Glockner G, Buchrieser C, Gophna U. The origins of eukaryotic-like proteins in *Legionella pneumophila*. Int J Med Microbiol. 2010;300(7):470-81.
35. Ensminger AW. *Legionella pneumophila*, armed to the hilt: justifying the largest arsenal of effectors in the bacterial world. Curr Opin Microbiol. 2016;29:74-80.
36. Escoll P, Mondino S, Rolando M, Buchrieser C. Targeting of host organelles by pathogenic bacteria: a sophisticated subversion strategy. Nat Rev Microbiol. 2016;14(1):5-19.
37. Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C. Comparative and functional genomics of *Legionella* identified eukaryotic like proteins as key players in host-pathogen interactions. Front Microbiol. 2011;2:208.
38. Qiu J, Luo ZQ. *Legionella* and *Coxiella* effectors: strength in diversity and activity. Nat Rev Microbiol. 2017;15(10):591-605.

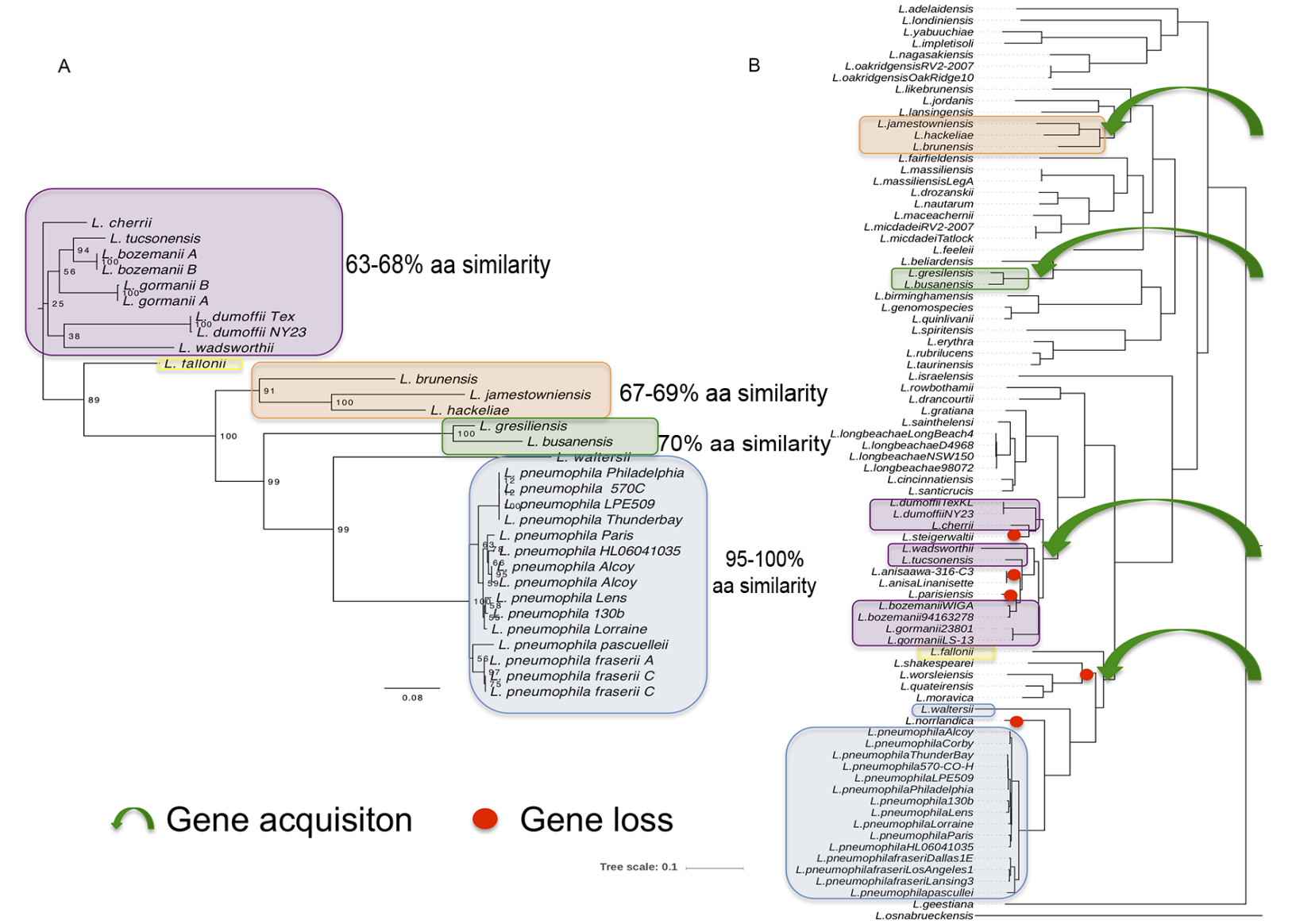
39. Rolando M, Buchrieser C. Legionella pneumophila type IV effectors hijack the transcription and translation machinery of the host cell. Trends Cell Biol. 2014;24(12):771-8.
40. Barlocher K, Welin A, Hilbi H. Formation of the *Legionella* Replicative Compartment at the Crossroads of Retrograde Trafficking. Frontiers in cellular and infection microbiology. 2017;7:482.
41. Merrill AH, Sandhoff K. Sphingolipids: metabolism and cell signalling In: Vance DE, Vance JE, editors. Biochemistry of Lipids, Lipoproteins and Membranes. 36. Elsevier Science, Amsterdam. ed2002. p. 373-407.
42. Heung JL, Luberto C, Del Poeta M. Role of Sphingolipids in Microbial Pathogenesis. Infect Immun. 2006;74(1):28-39.
43. Shabardina V, Kischka T, Kmita H, Suzuki Y, Makalowski W. Environmental adaptation of *Acanthamoeba castellanii* and *Entamoeba histolytica* at genome level as seen by comparative genomic analysis. Int J Biol Sci. 2018;14(3):306-20.
44. Rolando M, Escoll P, Buchrieser C. Legionella pneumophila restrains autophagy by modulating the host's sphingolipid metabolism. Autophagy. 2016;12(6):1053-4.
45. Rolando M, Escoll P, Nora T, Botti J, Boitez V, Bedia C, et al. Legionella pneumophila S1P-lyase targets host sphingolipid metabolism and restrains autophagy. Proc Natl Acad Sci U S A. 2016;113(7):1901-6.
46. Degtyar E, Zusman T, Ehrlich M, Segal G. A *Legionella* effector acquired from protozoa is involved in sphingolipids metabolism and is targeted to the host cell mitochondria. Cell Microbiol. 2009;11(8):1219-35.
47. Gomez-Valero L, Buchrieser C. Genome dynamics in *Legionella*: the basis of versatility and adaptation to intracellular replication. Cold Spring Harbor perspectives in medicine. 2013;3(6).
48. Choy A, Dancourt J, Mugo B, O'Connor TJ, Isberg RR, Melia TJ, et al. The *Legionella* effector RavZ inhibits host autophagy through irreversible Atg8 deconjugation. Science. 2012;338(6110):1072-6.
49. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, et al. The genome of the amoeba symbiont "Candidatus *Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. J Bacteriol. 2010;192(4):1045-57.
50. Casadevall A, Fu MS, Guimaraes AJ, Albuquerque P. The 'Amoeboid Predator-Fungal Animal Virulence' Hypothesis. J Fungi (Basel). 2019;5(1).
51. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306(5700):1344-50.
52. Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol Biol. 2008;8:12.
53. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, et al. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. PLoS Genet. 2006;2(5):e:76.
54. Gomez-Valero L, Neou Bonora M, Gribaldo S, Buchrieser C. Interdomain Horizontal Gene Transfer Shaped the Genomes of *Legionella pneumophila* and *Legionella longbeachae*. In: Gophna U, editor. Lateral Gene Transfer in Evolution. New York, NY: Springer; 2013. p. 199-219.
55. Nyvltova E, Sut'ak R, Zarsky V, Harant K, Hrdy I, Tachezy J. Lateral gene transfer of p-cresol- and indole-producing enzymes from environmental bacteria to Mastigamoeba balamuthi. Environ Microbiol. 2017;19(3):1091-102.
56. Kaneko T, Stogios PJ, Ruan X, Voss C, Evdokimova E, Skarina T, et al. Identification and characterization of a large family of superbinding bacterial SH2 domains. Nature communications. 2018;9(1):4549.
57. de Felipe KS, Glover RT, Charpentier X, Anderson OR, Reyes M, Pericone CD, et al. *Legionella* eukaryotic-like type IV substrates interfere with organelle trafficking. PLoS Pathog. 2008;4(8):e1000117.

557 58. Gimenez G, Bertelli C, Moliner C, Robert C, Raoult D, Fournier PE, et al. Insight into
558 cross-talk between intra-amoebal pathogens. BMC Genomics. 2011;12:542.
559 59. Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, et al. Extensive
560 recombination events and horizontal gene transfer shaped the *Legionella pneumophila*
561 genomes. BMC Genomics. 2011;12:536.
562 60. Wang Z, Wu M. Comparative Genomic Analysis of *Acanthamoeba* Endosymbionts
563 Highlights the Role of Amoebae as a "Melting Pot" Shaping the Rickettsiales Evolution.
564 Genome Biol Evol. 2017;9(11):3214-24.
565 61. Bertelli C, Greub G. Lateral gene exchanges shape the genomes of amoeba-resisting
566 microorganisms. Frontiers in cellular and infection microbiology. 2012;2:110.
567 62. Buchrieser C, Charpentier X. Induction of competence for natural transformation in
568 *Legionella pneumophila* and exploitation for mutant construction. Methods Mol Biol.
569 2013;954:183-95.

570


A
EUGLENOZOA (5)
CHOANOZOA (2)
METAZOA (5)
FUNGI (16)
Paramecium tetraurelia
Tetrahymena thermophila
Legionella spp (Lpp2128)
Entamoeba spp (3)
MYCETOZOA (4)
Phaeodactylum tricornutum
OOMYCETES (4)
PLANTS (6)
Cyanidioschyzon merolae
Symbiobacterium thermophilum
Roseiflexus castenholzii
Haliangium ochraceum
Stigmatella aurantiaca
Myxococcus spp (2)
Burkholderia (2)
Thermomonospora curvata
Naegleria gruberi
Dictyostelium fasciculatum
Haliscomenobacter hydrossis
PROKARYOTES (82)
0.1


B
L.adelaidensis
L.londiniensis
L.yabuuchiae
L.impletisoli
L.nagasakiensis
L.oakridgensisRV2-2007
L.oakridgensisOakRidge10
L.likebrunensis
L.jordanis
L.lansingensis
L.jamestowniensis
L.hackeliae
L.brunensis
L.fairfieldensis
L.massiliensis
L.massiliensisLegA
L.drozanskii
L.nautarum
L.maceachernii
L.micdadeiRV2-2007
L.micdadeiTatlock
L.feeleii
L.beliardensis
L.gresilensis
L.busanensis
L.birminghamensis
L.genomospecies
L.quinlivanii
L.spiritensis
L.erythra
L.rubrilucens
L.taurinensis
L.israelensis
L.rowbothamii
L.drancourtii
L.gratiana
L.sainthelensi
L.longbeachaeLongBeach4
L.longbeachaeD4968
L.longbeachaeNSW150
L.longbeachae98072
L.cincinnatiensis
L.santicrucis
L.dumoffiiTexKL
L.dumoffiiNY23
L.cherrii
L.steigerwaltii
L.wadsworthii
L.tucsonensis
L.anisaawa-316-C3
L.anisaLinanisette
L.parisiensis
L.bozemaniiWIGA
L.bozemanii94163278
L.gormanii23801
L.gormaniiLS-13
L.fallonii
L.shakespearei
L.worsleiensis
L.quateirensis
L.moravica
L.waltersii
L.norrlandica
L.pneumophilaAlcoy
L.pneumophilaCorby
L.pneumophilaThunderBay
L.pneumophila570-CO-H
L.pneumophilaLPE509
L.pneumophilaPhiladelphia
L.pneumophila130b
L.pneumophilaLens
L.pneumophilaLorraine
L.pneumophilaParis
L.pneumophilaHL06041035
L.pneumophilafraseriDallas1E
L.pneumophilafraseriLosAngeles1
L.pneumophilafraseriLansing3
L.pneumophilapascullei
L.geestiana
L.osnabrueckensis
0.1


A
L. cherrii
L. tucsonensis
L. bozemanii A
L. bozemanii B
L. gormanii B
L. gormanii A
L. dumoffii Tex
L. dumoffii NY23
L. wadsworthii
L. fallonii
L. brunensis
L. jamestowniensis
L. hackeliae
L. gresiliensis
L. busanensis
L. waltersii
L. pneumophila Philadelphia
L. pneumophila 570C
L. pneumophila LPE509
L. pneumophila Thunderbay
L. pneumophila Paris
L. pneumophila HL06041035
L. pneumophila Alcoy
L. pneumophila Alcoy
L. pneumophila Lens
L. pneumophila 130b
L. pneumophila Lorraine
L. pneumophila pascuelleii
L. pneumophila fraserii A
L. pneumophila fraserii C
L. pneumophila fraserii C
0.08
63-68% aa similarity
67-69% aa similarity
70% aa similarity
95-100%
aa similarity
B
L.adelaidensis
L.londiniensis
L.yabuuchiae
L.impletisoli
L.nagasakiensis
L.oakridgensisRV2-2007
L.oakridgensisOakRidge10
L.likebrunensis
L.jordanis
L.lansingensis
L.jamestowniensis
L.hackeliae
L.brunensis
L.fairfieldensis
L.massiliensis
L.massiliensisLegA
L.drozanskii
L.nautarum
L.maceachernii
L.micdadeiRV2-2007
L.micdadeiTatlock
L.feeleii
L.beliardensis
L.gresilensis
L.busanensis
L.birminghamensis
L.genomospecies
L.quinlivanii
L.spiritensis
L.erythra
L.rubrilucens
L.taurinensis
L.israelensis
L.rowbothamii
L.drancourtii
L.gratiana
L.sainthelensi
L.longbeachaeLongBeach4
L.longbeachaeD4968
L.longbeachaeNSW150
L.longbeachae98072
L.cincinnatiensis
L.santicrucis
L.dumoffiiTexKL
L.dumoffiiNY23
L.cherrii
L.steigerwaltii
L.wadsworthii
L.tucsonensis
L.anisaawa-316-C3
L.anisaLinanisette
L.parisiensis
L.bozemaniiWIGA
L.bozemanii94163278
L.gormanii23801
L.gormaniiLS-13
L.fallonii
L.shakespearei
L.worsleiensis
L.quateirensis
L.moravica
L.waltersii
L.norrlandica
L.pneumophilaAlcoy
L.pneumophilaCorby
L.pneumophilaThunderBay
L.pneumophila570-CO-H
L.pneumophilaLPE509
L.pneumophilaPhiladelphia
L.pneumophila130b
L.pneumophilaLens
L.pneumophilaLorraine
L.pneumophilaParis
L.pneumophilaHL06041035
L.pneumophilafraseriDallas1E
L.pneumophilafraseriLosAngeles1
L.pneumophilafraseriLansing3
L.pneumophilapascullei
L.geestiana
L.osnabrueckensis
Gene acquisiton
Gene loss
Tree scale: 0.1