Infection, Genetics and Evolution 43 (2016) 108–122

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Review

# Molecular epidemiology, phylogeny and evolution of *Legionella*

A. Khodr [a,b], E. Kay [c], L. Gomez-Valero [a,b], C. Ginevra [c,d], P. Doublet [c], C. Buchrieser [a,b], S. Jarraud [c,d]

[a] Institut Pasteur, Unité de Biologie des Bactéries Intracellulaires, France
[b] CNRS, UMR 3525, 28, Rue du Dr Roux, 75724 Paris, France
[c] CIRI, International Center for Infectiology Research, Inserm, U1111, CNRS, UMR 5308, Université Lyon 1, École Normale Supérieure de Lyon, Lyon F-69008, France
[d] French National Reference Center of Legionella, Institut des agents infectieux, Hospices Civils de Lyon, Lyon, France

ARTICLE INFO

*Article history:*
Received 14 February 2016
Received in revised form 29 April 2016
Accepted 30 April 2016
Available online 13 May 2016

*Keywords:*
*Legionella pneumophila*
Phylogeny
Epidemiology
Experimental evolution
ICEs
HGT
T4ASS

ABSTRACT

*Legionella* are opportunistic pathogens that develop in aquatic environments where they multiply in protozoa. When infected aerosols reach the human respiratory tract they may accidentally infect the alveolar macrophages leading to a severe pneumonia called Legionnaires' disease (LD). The ability of *Legionella* to survive within host-cells is strictly dependent on the Dot/Icm Type 4 Secretion System that translocates a large repertoire of effectors into the host cell cytosol. Although *Legionella* is a large genus comprising nearly 60 species that are worldwide distributed, only about half of them have been involved in LD cases. Strikingly, the species *Legionella pneumophila* alone is responsible for 90% of all LD cases. The present review summarizes the molecular approaches that are used for *L. pneumophila* genotyping with a major focus on the contribution of whole genome sequencing (WGS) to the investigation of local *L. pneumophila* outbreaks and global epidemiology studies. We report the newest knowledge regarding the phylogeny and the evolution of *Legionella* and then focus on virulence evolution of those *Legionella* species that are known to have the capacity to infect humans. Finally, we discuss the evolutionary forces and adaptation mechanisms acting on the Dot/Icm system itself as well as the role of mobile genetic elements (MGE) encoding T4ASSs and of gene duplications in the evolution of *Legionella* and its adaptation to different hosts and lifestyles.



**Contents**



http://dx.doi.org/10.1016/j.meegid.2016.04.033
1567-1348/

## 1. Introduction

*Legionella pneumophila* is a human pathogen that was recognized only 40 years ago after a large outbreak of pneumonia during a convention of the American Legion, thus named Legionnaires' disease (LD) (Fraser et al., 1977; McDade et al., 1977). Why was a bacterium that causes severe pneumonia only identified in 1977? Firstly, the development of man-made water systems producing aerosols (*e.g.* air conditioning systems, cooling towers, spas *etc.*) created conditions allowing the direct access of this opportunistic bacterium to human lungs. This fact is thought to have made LD an emerging disease since the 1970's. Secondly, the fastidious growth of *L. pneumophila* and the requirement of L-cysteine and iron for growth on axenic culture media would have made it difficult to isolate this organism earlier (Feeley et al., 1978). However, the magnitude of the first recognized outbreak of LD in Philadelphia in 1977, its mysterious character due to the long lasting search for the causative agent of this disease that was widely reported by the media, combined with the severity of the pneumonia and the group of patients that belonged mostly to the American Legion, made this outbreak an exceptional event. Thus, very significant resources were employed for the epidemiological investigation to finally identify the causative agent. Interestingly, only techniques routinely used in the field of *Rickettsia*, also an intracellular bacterium, such as inoculation of both guinea pigs and yolk sacs of egg embryos, allowed Joseph McDade and his collaborators to cultivate and identify *L. pneumophila* (McDade et al., 1977). Once the bacterium was identified and isolated, sero-epidemiological studies were done, which also allowed the recognition of earlier outbreaks of LD but also enabled description of Pontiac fever (Blaser, 1977; McDade et al., 1979). Pontiac fever is a milder form of legionellosis, classically described as an influenza-like illness without pneumonia.

Today, *L. pneumophila* has been identified as one of the three most common causes of severe community-acquired pneumonia (CAP): Legionnaires' disease accounts for 2%–8% of CAP cases (Bartlett, 2008, 2011; Roig and Rello, 2003). A review of 41 European studies of CAP identified *Legionella* as the causative agent in 1.9% of outpatients, 4.9% of hospitalised patients and 7.9% of ICU patients (Woodhead, 2002). The exact incidence of LD worldwide is unknown due to difference between countries in surveillance and reporting but also diagnosis of LD. The development of diagnostic tests for the easy detection of *L. pneumophila* serogroup 1-specific antigens in urine was a significant advance in the diagnosis of legionellosis. In 2014 the European prevalence was 13.5 cases per million inhabitants; most cases (78%) were confirmed by this urinary antigen test (European Centre for Disease Prevention and Control, 2016). The mortality rate remains high (8% to 30% in ICUs and for hospital-acquired LD) despite improved diagnostic and therapeutic management of patients. Known risk factors for LD include increasing age, male gender, smoking, chronic lung disease, diabetes and various conditions associated with immunodeficiency (Phin et al., 2014). LD occurs sporadically and in outbreaks. In Europe, most cases (approximately 80%) are community-acquired and sporadic (European Centre for Disease Prevention and Control, 2016). Most of the cases occur during summer and early autumn; the yearly incidence of LD seems to be associated with climate changes, such as increased precipitation (Cunha et al., 2015). The sources of infection of sporadic cases are more rarely investigated and identified. However several systems and matrices have been classified as confirmed sources of *Legionella* (van Heijnsbergen et al., 2015). Healthcare and travel-associated outbreaks are mainly related to contaminated water systems (showers and baths, respiratory therapy or air conditioning equipments, spa pools, and, less commonly, food display humidifiers). Community outbreaks are predominantly linked to contaminated aerosols from wet cooling systems. As mentioned above, *L. pneumophila* as several other species of *Legionella* can cause also Pontiac fever. However the causative bacteria have not been isolated yet from Pontiac fever patients. The diagnosis is performed by urinary antigen test or by using serology. While the disease progresses acutely and shows a high attack rate of approximately 95%, the mortality rate is zero (Fields et al., 2002).

Important knowledge about the epidemiology, the clinical presentations and the treatment of LD was readily gained and published soon after its recognition in 1976 (Fraser et al., 1977). However, advances in microbiology have now led to a better understanding of the ecological niches, the pathogenesis and the evolution of *L. pneumophila*. Here, we review the latest knowledge gained on the phylogeny and evolution of *L. pneumophila* through whole genome sequencing and different molecular approaches employed recently for its study.

## 2. Taxonomy and ecology of genus *Legionella*

Since Legionnaires' disease was recognized, isolation and identification of different strains and their characterization led to the establishment of the family *Legionellaceae* consisting of the single genus *Legionella* among the subdivision γ2-Proteobacteria (Brenner et al., 1979; Woese, 1987). On the basis of low DNA-DNA hybridization values between some *Legionella* species, a classification separating the family *Legionellaceae* in three genera - *Legionella*, *Fluoribacter* and *Tatlockia* was proposed (Garrity et al., 1980). However, additional studies showing that all legionellae studied have 16S ribosomal RNA sequences >95% identical did not support this division (Fry et al., 1991). Today, the genus *Legionella* comprises over 60 species (http://www.bacterio.net/legionella.html); all species have been isolated from environmental samples, and about half of the known species were also isolated at least once from patients and have thus been associated with infection. The type species is *Legionella pneumophila*, corresponding to the first bacterium of the genus described that is nowadays also the species responsible for nearly 95% of cases of LD diagnosed worldwide (Fraser et al., 1977; McDade et al., 1977). This species can be subdivided in 16 serogroups but the majority of culture-confirmed LD cases (84% worldwide, 80% in Europe) is caused by *L. pneumophila* serogroup 1 (Lp1) (Beaute et al., 2013; Fields et al., 2002; Yu et al., 2002). In 1988, Brenner et al. subdivided the species *Legionella pneumophila* in three subspecies (Brenner et al., 1988) - *L. pneumophila* subsp. *pneumophila*; *L. pneumophila* subsp. *fraseri*; *L. pneumophila* subsp. *pascullei*. Nevertheless, the tools to distinguish subspecies are not available in routine laboratories and thus these subdivisions are rarely reported. Other species of the genus *Legionella* also isolated from patients in 2014 in Europe are *Legionella longbeachae* (2%), *L. micdadei* (1%), *Legionella bozemanii* (<1%), *Legionella macaechernii* (<1%), *Legionella sainthelensi* (<1%), *L.* other *species* (<1%) and *L. species* not identified (1%) (European Centre for Disease Prevention and Control, 2016). The majority of confirmed infections involving non-*pneumophila Legionella* species occur in severely immune-compromised patients. Interestingly, Graham et al. reported a distinct epidemiological pattern of legionellosis in New Zealand and Australia, where *L. longbeachae* and *L. pneumophila* are similarly prevalent in LD (Graham et al., 2011). Infections with *L. longbeachae* are commonly associated with exposure to contaminated composts and potting soils, and have been increasingly reported in Europe in the past ten years (Amodeo et al., 2010; Currie and Beattie, 2015). The comparison of the *L. longbeachae* genome with that of Lp1 identified many species-specific differences that may account for the different environmental niches and disease epidemiology of these two species. Among these, *L. longbeachae* encodes several enzymes that might confer the ability to degrade plant material suggesting that *L. longbeachae* may also be able to interact with plants. Interestingly, *L. longbeachae* does not encode flagella, a major virulence factor for *L. pneumophila,* but is encapsulated (Cazalet et al., 2010). Regarding *L. pneumophila*, the high prevalence of Lp1 strains in human disease as mentioned above, does not seem to be due to its environmental distribution (Doleans et al., 2004). In order to explain this epidemiological predominance, comparative genome analyses of over 200 *Legionella* strains using macroarrays were performed. One major finding of this study was that the lipopolysaccharide (LPS) biosynthesis gene cluster of sg1 was the only common