Environmental Microbiology (2006) **8**(6), 1056–1063 doi:10.1111/j.1462-2920.2006.00997.x

Genetic variability in environmental isolates of *Legionella pneumophila* from Comunidad Valenciana (Spain)

Mireia Coscollá,[1] María José Gosalbes,[1] Vicente Catalán[2] and Fernando González-Candelas[1]*

[1]*Institut Cavanilles de Biodiversitat i Biologia Evolutiva and Departament de Genètica, Universitat de València, 46071 Valencia, Spain.*

[2]*LABAQUA, SA. Pol. Ind. Las Atalayas, C/Del Dracma, 16-18, 03114 Alicante, Spain.*

Summary

***Legionella pneumophila* is associated to recurrent outbreaks in several Comunidad Valenciana (Spain) localities, especially in Alcoi, where social and climatic conditions seem to provide an excellent environment for bacterial growth. We have analysed the nucleotide sequences of three loci from 25 environmental isolates from Alcoi and nearby locations sampled over 3 years. The analysis of these isolates has revealed a substantial level of genetic variation, with consistent patterns of variability across loci, and comparable to that found in a large, European-wide sampling of clinical isolates. Among the tree loci studied, *fliC* showed the highest level of nucleotide diversity. The analysis of isolates sampled in different years revealed a clear differentiation, with samples from 2001 being significantly distinct from those obtained in 2002 and 2003. Furthermore, although linkage disequilibrium measures indicate a clonal nature for population structure in this sample, the presence of some recombination events cannot be ruled out.**

Introduction

Legionella pneumophila is naturally found in fresh waters, where the bacteria parasitize within protozoa, and it also survives as free bacterial cells in water and biofilms. Human infection occurs mainly by inhalation of contaminated aerosols from air conditioning systems, cooling towers, natural hot spas and other water systems (Fields, 1996), and, occasionally, by aspiration of water containing the bacteria (Fields *et al*., 2002).

In order to fight and control emerging infectious diseases it is increasingly important to understand the population structure of bacterial pathogens (Feil, 2004). However, the population structures of microorganisms are complex and often controversial (Feil *et al*., 2001). The genetic structure of *L. pneumophila* has been described as clonal as many clones are apparently worldwide distributed (Selander *et al*., 1985). On the other hand, Ko and colleagues (2002) indicate that there is evidence of recombination among populations due to the incongruence observed in phylogenetic trees for different genes and the lack of correlation between serogroups and genotypes. These authors have also indicated that in spite of clonal proliferation there is evidence for recombination in pathogenicity islands of *L. pneumophila* (Ko *et al*., 2003). Consequently, the preservation of clonality in each subgroup and the congruence of subgrouping among isolates may reflect genetic barriers to gene flow between different subgroups within a population (Ko *et al*., 2003).

Differentiating accurately among strains of a species is essential in microevolutionary and epidemiological studies. The high levels and different sources of genetic diversity in bacterial species have resulted in a variety of typing methods to characterize and quantify it (Cooper and Feil, 2004). As nucleotide sequences provide direct genetic information, universal criteria for comparison and identify more variation than methods based on electrophoretic mobility, typing techniques based on nucleotide sequencing such as Multi-Locus Sequence Typing (MLST) provide results that are truly portable between laboratories electronically via internet and data from each species can be stored in an expanding global database (Maiden *et al*., 1998). In fact, a new MLST-based scheme has just been adopted by the European Working Group for *Legionella* Infections (EWGLI) (Gaia *et al*., 2003; 2005).

In the region around Alcoi (Alicante province, Comunidad Valenciana, Spain) an almost continuous outbreak of *Legionella* infections has occurred since 1999, with sporadic bouts affecting dozens of patients, and totalling over 300 affected people (Lopez *et al*., 2001; Fernandez *et al*., 2002; 2004). Despite intervention measures by public health authorities, recurrent bouts indicate either that control measures have failed or that risk installations are recolonized by bacteria from natural reservoirs that escape current control measures hence acting as sources

Received 19 August, 2005; accepted 19 December, 2005. *For correspondence. E-mail fernando.gonzalez@uv.es; Tel. (+34) 96 3543653; Fax (+34) 96 3543670.

for new infections. Consequently, in order to discriminate between these alternatives and to better implement adequate control measures that prevent and limit new *Legionella* infections, it is necessary to know the extent and nature of *Legionella* variation from both clinical and environmental sources.

The aim of this study is to explore the genetic variability of *L. pneumophila* in three loci, *fliC*, *proA* and *mompS*, previously characterized as the genetically most diverse genes (Gaia *et al*., 2003), in an array of samples from the reduced geographical area described above and to compare these results with those obtained for the same genetic markers from a large sampling area (Gaia *et al*., 2003). This analysis will allow us to understand the population genetic structure of *L. pneumophila* in the Alcoi area and to gain some insight on its temporal dynamics. This information will be useful in the prevention and control of the recurrent *Legionella* outbreaks produced in this region.

Results

Sequence typing profiles

Sequences of internal fragments of *fliC*, *proA* and *mompS* genes were obtained from 25 *L. pneumophila* environmental isolates from close geographical locations in Alicante province (Spain). Most alleles found in the loci were identical, in the common sequence fragments, to those previously described (Gaia *et al*., 2003). We also found five alleles in *fliC*, all identical in the common 182 nt segment to *flaA* alleles 1, 2, 5, 6 and 11 (Fig. 1). In locus *proA* we also found six alleles in the 25 samples, one of which differed from those previously reported (2003) in the common stretch of 405 bp (Fig. 1). Finally, locus *mompS* presented also six alleles, but three of these were different from the ones previously reported in the 352 bp fragment common to sequences from both studies. Next, allelic profiles were assigned for each isolate (Table 1) resulting in nine different profiles with frequencies ranging from 4% to 28%. Six of the nine profiles differed from those previously reported (Gaia *et al*., 2003) and resulted from the presence of new alleles in the newly determined sequences.

The most frequent allelic profile in the studied sample was (1,1,1), present in 7 of 25 isolates (Table 1). This was also the most common profile in different parts of Europe (Gaia *et al*., 2003), where it was found in 18 isolates from seven different countries. The second most common profile (5,10,6), present in five isolates, was also detected in several European samples (Gaia *et al*., 2003), thus pro-

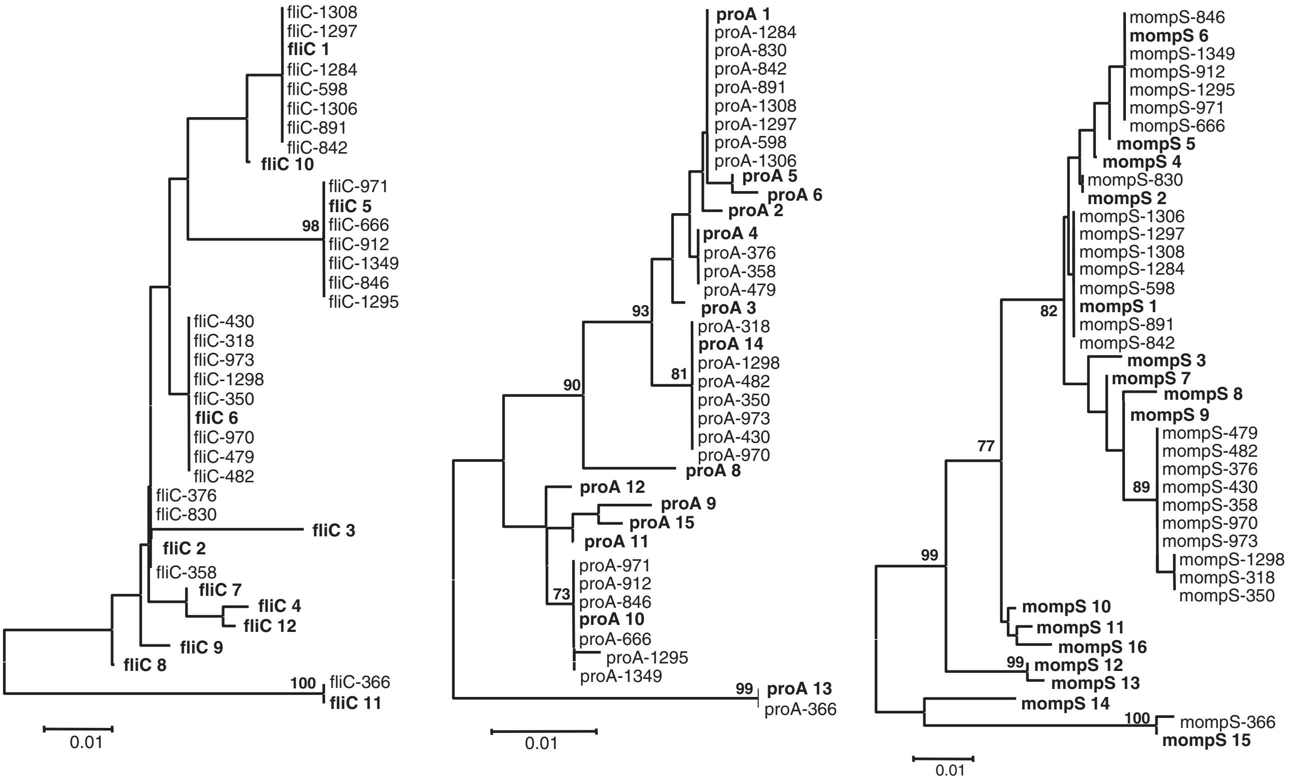

Fig. 1. Neighbour-joining trees obtained using sequences from environmental isolates from Alicante (Spain). Alleles found by Gaia and colleagues (2003) are shown in bold. Bootstrap support values larger than 70% are indicated next to the corresponding node.

Table 1. *Legionella pneumophila* environmental isolates from Alicante province (Spain) included in this study.

Sample	Geographical origin	Sampling date	Serogroup	Allelic profile[a]
L318	Cocentaina	6/3/01	6	6,14,A[b]
L350	Cocentaina	25/4/01	1	6,14,A
L358	Cocentaina	8/5/01	3	2,4,B
L366	Onil	17/5/01	2, 10, 13	11,13,C
L376	Cocentaina	22/5/01	3	2,4,B
L430	Muro de Alcoi	25/7/01	1	6,14,B
L479	Alcoi	10/10/01	2–14	6,4,B
L482	Alcoi	10/10/01	1	6,14,B
L598	Alcoi	7/2/02	1	1,1,1
L666	Cocentaina	14/5/02	2–14	5,10,6
L830	Cocentaina	6/8/02	1	2,1,2
L842	Cocentaina	26/8/02	1	1,1,1
L846	Cocentaina	26/8/02	1	5,10,6
L891	Alcoi	3/9/02	1	1,1,1
L912	Alcoi	10/10/02	1	5,10,6
L970	Cocentaina	26/11/02	1	6,14,B
L971	Alcoi	30/11/02	1	5,10,6
L973	Alcoi	5/12/02	1	6,14,B
L1284	Cocentaina	25/9/03	9	1,1,1
L1295	Ibi	1/10/03	9	5,A,6
L1297	Ibi	1/10/03	4, 5	1,1,1
L1298	Ibi	2/10/03	6	6,14,A
L1306	Alcoi	7/10/03	4, 5	1,1,1
L1308	Alcoi	7/10/03	1	1,1,1
L1349	Ibi	2/10/03	4, 5, 8	5,10,6

[a] Allelic profiles detected in loci *fliC*, *proA* and *mompS* are named as in Gaia and colleagues (2003).

[b] New alleles identified in *proA* and *mompS* are denoted by capital letters.

viding support for a global distribution of clonal complexes. The new profiles were present in frequencies from four to one single isolates.

In order to test for departures of random association of alleles from the three studied loci, a linkage disequilibrium measure was calculated for the 25 environmental isolates, resulting in an index of association (I_A) of 1.451 ± 0.14. This was significantly different from the expected value under equilibrium ($I_A = 0$).

Genetic variability of populations

Genetic variability was estimated in the three loci studied for the 25 environmental isolates from Alicante and 95 environmental and clinical isolates from 10 European countries (Gaia *et al.*, 2003) (Table 2). Although samples from Alicante showed fewer haplotypes than samples from Europe, in accordance with the smaller sample size in our study, other genetic diversity parameters that are not directly dependent on sample size, such as haplotype and nucleotide diversities, number of polymorphic nucleotide sites, theta (θ) per site from polymorphic sites (S) and average number of pairwise nucleotide differences (k), were similar in the two data sets for the three loci. Nevertheless, differences in nucleotide diversity (π) and

Table 2. Genetic variability in *fliC*, *proA* and *mompS* loci among 25 environmental isolates from Alicante of which 13 belonged to serogroup 1 (Sg.1), 95 serogroup 1 samples from Europe, some of which were sampled in Spain.

	fliC				*proA*				*mompS*			
	Alicante	Sg.1	Europe	Spain	Alicante	Sg.1	Europe	Spain	Alicante	Sg.1	Europe	Spain
Sequences, n	25	13	95	13	25	13	95	13	25	13	95	13
Sequence length, L	206	206	182	182	443	443	405	405	512	512	352	352
Haplotypes	5	4	12	8	6	3	15	5	6	5	16	8
Haplotype diversity, h	0.777	0.756	0.964	0.897	0.793	0.705	0.786	0.782	0.800	0.808	0.868	0.897
(standard deviation)	(0.039)	(0.070)	(0.017)	(0.067)	(0.063)	(0.064)	(0.030)	(0.079)	(0.038)	(0.066)	(0.018)	(0.067)
Nucleotide diversity, π	0.0215	0.0173	0.0235	0.0239	0.0158	0.0127	0.0181	0.0155	0.0158	0.0105	0.0291	0.0200
(standard deviation)	(0.0046)	(0.0021)	(0.0019)	(0.0037)	(0.0030)	(0.0022)	(0.0013)	(0.0021)	(0.0048)	(0.0010)	(0.0020)	(0.0047)
Polymorphic sites, S	20	8	23	14	29	13	32	17	45	12	42	26
θ (from S)	0.0253	0.0125	0.0246	0.0248	0.0173	0.0095	0.0154	0.0135	0.0234	0.0076	0.0233	0.0238
(standard deviation)	(0.0097)	(0.006)	(0.0078)	(0.0111)	(0.042)	(0.0043)	(0.0046)	(0.0059)	(0.0081)	(0.0035)	(0.0066)	(0.0100)
Pairwise differences, k	4.496	3.593	4.182	4.511	7.020	5.641	7.169	6.282	8.040	5.359	9.936	7.051
(standard deviation)	(2.292)	(1.947)	(2.097)	(2.303)	(3.412)	(2.894)	(3.391)	(3.188)	(3.864)	(2.392)	(4.585)	(3.541)
Silent mutations	20	8	21	11	28	13	30	16	32	9	29	17
Replacement mutations	0	0	3	3	1	0	2	1	14	3	8	8
dN/dS	0	0	0.091	0.0983	0.0039	0	0.021	0.0083	0.1365	0.065	0.168	0.14

polymorphic nucleotide sites (S) for *mompS* locus as well as in haplotype diversity (h) for *fliC* and *mompS* were found between the two studies.

Nucleotide substitutions accounting for differences among the different alleles were mostly synonymous for loci *proA* and *fliC*, in which we detected none and one replacement substitution respectively (Gaia *et al*., 2003). However, almost 30% (14 of 48) of all nucleotide changes detected in *mompS* were replacement substitutions. A similar proportion (8/37) of non-synonymous changes was detected by Gaia and colleagues (2003) in this locus. Furthermore, one sequence, L366, presented a codon insertion in this locus.

As the study by Gaia and colleagues (2003) included only sequences from serogroup 1 and our data set presented sequences from other serogroups, we also computed the same parameters for serogroup 1 sequences in our sample (Table 2). Similarly, and also for comparison, in Table 2 we provide estimates of the same parameters for the samples of Spanish origin in Gaia and colleagues (2003). Although occasionally the observed values in these subsets are slightly divergent from the corresponding whole sets, the general comments from the previous paragraphs still apply to the comparisons involving these subsets.

Next, we used the modified Nei–Gojobori method to compute the rates of synonymous and non-synonymous substitutions. Locus *mompS* showed a dN/dS ratio of 0.1365 for the 25 *L. pneumophila* sequences from Alicante, while the corresponding values for *fliC* and *proA* were 0 and 0.0039 respectively. The corresponding values for the 95 samples analysed by Gaia and colleagues (2003) were always larger, ranging from 0.021 in *proA* to 0.168 in *mompS*.

We combined the common alignment positions from the sequences obtained in this study and those previously reported (Gaia *et al*., 2003) to derive neighbour-joining trees for each locus. Trees were calculated using the Tamura–Nei nucleotide substitution model and are shown in Fig. 1. It is remarkable that most nodes have relatively low bootstrap support (BS < 70%), an indication of the close similarity at the nucleotide level among most alleles found in each locus. The comparison of the position of the sequences derived for each locus from the 25 environmental isolates in the corresponding phylogenetic trees revealed some minor incongruences, as sequences from several isolates grouped in different clusters for each gene. For instance, L376 grouped with L358 and L830 in *fliC*, with L479 and L358 in *proA*, and with L358, L430, L479, L482, L970 and L973 in *mompS*. One isolate, L366, very similar (identical in *fliC* and *proA*, with 1 nucleotide difference in *mompS* to alleles 11, 13 and 15 respectively) to EUL No. 40, 47 in Gaia and colleagues (2003), presents a relatively very large divergence from the remaining alleles in these three loci (Fig. 1).

Population genetic structure

To investigate the genetic structure of *L. pneumophila* populations, a hierarchical analysis of molecular variance (AMOVA) was performed for the 25 isolates from Alicante and the 95 isolates from Europe using the sequences of the three loci *fliC*, *proA* and *mompS* (Table 3). In all cases, the largest percentage of variation was found within populations, as this level accounted for from 77.01% (*mompS*) to 87.39% (*fliC*) of the total variation. The remaining variation was distributed in the 'among-groups' and 'among populations within groups' levels. The later was larger than the former for the three loci and the corresponding fixation indices (Fsc) were always significant at the $\alpha = 0.05$ level. The among-groups variation ranged from 3.85% (*proA*) to 10.57% (*mompS*), being marginally significant only for *fliC* and *mompS* (Table 3). These results indicate that there is

Table 3. Analyses of molecular variance (AMOVA) for loci *fliC*, *proA* and *mompS*.

Locus	Source of variation[a]	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation indices
fliC	Among groups	1	8.346	0.126 Va	5.55	$F_{ST} = 0.126^{**}$
	Among populations within groups	13	42.037	0.160 Vb	7.05	$F_{CT} = 0.056^{**}$
	Within populations	105	208.445	1.986 Vc	87.39	$F_{SC} = 0.075^{*}$
	Total	119	258.828	2.272		
proA	Among groups	1	14.194	0.147 Va	3.85	$F_{ST} = 0.201^{***}$
	Among populations within groups	13	102.726	0.621 Vb	16.22	$F_{CT} = 0.038^{ns}$
	Within populations	105	321.344	3.060 Vc	79.93	$F_{SC} = 0.169^{***}$
	Total	119	438.265	3.829		
mompS	Among groups	1	32.720	0.575 Va	10.57	$\mathrm{F}_{ST} = 0.230^{***}$
	Among populations within groups	13	122.894	0.675 Vb	12.42	$F_{CT} = 0.106^{*}$
	Within populations	105	439.837	4.189 Vc	77.01	$F_{SC} = 0.139^{***}$
	Total	119	595.452	5.343		

a. Two groups were considered for comparison, one comprising the 25 samples from five localities in Alicante province (Spain), the other including 95 samples from 10 European countries analysed in Gaia and colleagues (2003).
$^{ns}P > 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

less genetic differentiation between Alicante and Europe than among populations within these two regions and much fewer than within any of the populations in both regions. Similar values were obtained when only serogroup 1 sequences from our data set were used for comparison (data not shown).

We also investigated how genetic variation was distributed through time. For the 25 environmental samples, we analysed the levels of intra- and interannual genetic diversity for the years 2001, 2002 and 2003. As the sample taken in 2001 in Onil (L366) was markedly different from the rest for the three loci (Fig. 1) it was excluded from this analysis. Table 4 presents a summary of the net genetic differentiation among samples from each year for the three loci. Average numbers of nucleotide substitutions between year 2001 and years 2002 and 2003 were larger than the corresponding within-year values for *fliC* and *mompS*, but there were no significant differences between 2002 and 2003. This is an indication that samples taken in the last 2 years were quite homogeneous and different from those sampled in 2001. However, this was not the pattern for locus *proA*, for which nucleotide diversities within years were larger than those among them. For all three loci, nearest neighbour statistic was significant for comparisons involving year 2001 and 2002 or 2003 and not significant for the comparison between 2002 and 2003. This result further reinforces the previous observation of a significant differentiation between year 2001 and the rest.

Discussion

Recently, Gaia and colleagues (2003) introduced a nucleotide sequence-based method, sequence-based typing (SBT), for the analysis of *L. pneumophila* serogroup 1. After screening seven potentially useful loci they finally selected three for their high discriminating power. This article represents a landmark for the study of the extent and distribution of genetic variability in *L. pneumophila*, with important consequences for understanding its epidemiology and persistence in natural settings. Here we build upon this previous work and use their data for comparative purposes. We also introduce some analytical tools that exploit more extensively the information provided by nucleotide sequence data than in usual MLST or SBT analyses.

Our study of the levels and distribution of genetic variation in *L. pneumophila* samples from a reduced area in the Alicante province (Spain) has revealed substantial levels of heterogeneity both in space and time. The samples used in this study were obtained from five geographically close locations along three consecutive years (2001–2003). Although the sample size of this study is relatively reduced (25 isolates), the levels of genetic diversity detected are similar to those found in these same loci in a much larger sample (95 isolates from 10 European countries) (2003). Measures of genetic diversity that depend directly on sample size, such as the number of haplotypes and absolute numbers of polymorphic sites and substitutions, show more variation in the European isolates (2003), but those in which the number of nucleotides is used as weighting to obtain the final parameter, such as nucleotide diversity or the average number of pairwise differences, present similar values in the two data sets (Table 2). As Alicante samples include as many non-serogroup 1 isolates as those from this serogroup, it might be argued that this similarity results from comparing heterogeneous data sets. However, this is not the case for the following reasons. First, there is no genetic differentiation at the nucleotide level, at least in these loci, between isolates from serogroup 1 and those from other serogroups (Fig. 1), and second, genetic variability estimates are still similar when only the 13 serogroup 1 isolates from Alicante are compared with the 95 European isolates of this serogroup (Table 2).

We have analysed larger genome regions than Gaia and colleagues (2003), which were in all cases embedded in our sequenced genome fragments. Differences range

Table 4. Genetic differentiation between populations by sampling year.

Locus	Year (sample size)	2001	2002	2003
fliC	2001 ($n = 7$)	0.00229 (0.00082)	0.68627*	0.85714*
	2002 ($n = 10$)	0.00492 (0.00413)	0.01949 (0.0023)	0.43922^{ns}
	2003 ($n = 7$)	0.00832 (0.00543)	–0.00084 (0.00511)	0.01953 (0.00436)
proA	2001 ($n = 7$)	0.00519 (0.00109)	0.81176**	0.85714**
	2002 ($n = 10$)	0.0036 (0.00403)	0.01507 (0.00178)	0.44874^{ns}
	2003 ($n = 7$)	0.00237 (0.00444)	–0.0013 (0.00382)	0.01463 (0.00391)
mompS	2001 ($n = 7$)	0.00188 (0.00067)	0.80392**	0.85714**
	2002 ($n = 10$)	0.00768 (0.00347)	0.00978 (0.00154)	0.48403^{ns}
	2003 ($n = 7$)	0.00821 (0.00376)	–0.00051 (0.00257)	0.00926 (0.00275)

The main diagonal shows the within-year estimates (standard deviation) of the number of net nucleotide substitutions per site (Nei, 1987) for loci *fliC*, *proA* and *mompS*. The lower hemimatrix shows between-year estimates (standard deviation) of nucleotide diversity. The upper hemimatrix represents the nearest neighbour statistic (Hudson, 2000) estimates and the corresponding *P*-values obtained after 1000 resampling replicates. $^{ns}P > 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$.

from 27 (*fliC*) to 160 (*mompS*) nucleotides, and this larger region has provided three new alleles in locus *mompS* and one new allele for *proA*. For *mompS*, the larger size of the genome fragment used in this analysis provides additional polymorphic sites (25% more) with one-fourth the sample size used by Gaia and colleagues (2003). Nevertheless, this is not translated into larger estimates of genetic variation at the nucleotide level on a per site scale (Table 2). It is likely that additional efforts in the design of primers flanking these same target regions will provide additional polymorphic sites with the ensuing increase in discriminatory power with no or little extra experimental cost. This enterprise will benefit from the availability of three complete *L. pneumophila* genome sequences (Cazalet *et al*., 2004; Chien *et al*., 2004).

The largest proportion of the total genetic variability (about 80%) in the three *L. pneumophila* loci is attributable to intrapopulation differences. In consequence, two isolates from the same population are almost as likely to be as different as any two isolates from two different locations, regardless of their origin (country or population within country). According to the low portion of variation attributable to differences between European and Alicante samples, which varies from 4% to 10% in these three loci, we can conclude that the variability found in Alicante is almost as large as and representative of that found throughout Europe.

Most newly determined *fliC*, *mompS* and *proA* sequences are identical to previously reported variants and their combinations match some already described profiles (Gaia *et al*., 2003). Nevertheless, we have found four new alleles and six new profiles in these 25 samples. The combination of four new alleles into six new profiles and some incongruences among the phylogenetic trees corresponding to these three loci could be explained by the existence of intergenic recombination. However, we have detected departures from the expected equilibrium for the index of association. This result suggests lack of recombination in the *L. pneumophila* genome which would translate into a clonal structure at the population level, although the low sample size on which this result is based prevents us from drawing any strong conclusion. Clearly, more data are necessary to solve this and other similar issues.

As previous studies have already revealed (Ratcliff *et al*., 1998; Ko *et al*., 2002), there is no congruence between sequenced based groupings and serogroups. Identical sequences at the nucleotide level in these three loci correspond to different serogroups and samples from the same serogroup can be found in different clusters (Table 1, Fig. 1). Possibly the most remarkable example in this data set is provided by sample 366, assigned to serogroups 2, 3 or 10. This environmental sample, obtained in Onil in 2001, is almost identical to two epidemiologically related Italian isolates, one clinical and the other environmental, from serogroup 1. Despite this, these isolates are very different from sequences from the same serogroup and from any other *L. pneumophila* serogroup included in both studies.

Knowledge of the extent and nature of the genetic diversity of human pathogens in their natural habitats is essential not only for epidemiological studies or ascertainment of the sources of outbreaks but also for understanding the ecological and evolutionary forces that govern their distribution and population dynamics. These will impact profoundly on their interaction with the human hosts. The analysis of the spatial and temporal distribution of pathogens and their genetic variants allows monitoring of how pathogens respond to changing conditions imposed by our continuing fight against them. In consequence, we think that these analyses should be routinely included in surveillance schemes. The increasing availability of genome sequences for many pathogenic microorganisms will facilitate these tasks by providing new genetic markers that can be shared and compared by laboratories all over the world.

Experimental procedures

Isolates and DNA extraction

Twenty-five isolates *L. pneumophila* were obtained from cooling towers and irrigation hydrants (Table 1) from five localities (Fig. 2) of Alicante (Spain) from 2001 to 2003. The isolates were serogrouped by indirect immunofluorescence testing (Wilkinson *et al*., 1979). Bacterial colonies from pure cultures were resuspended in 200 μl of 20% Chelex 100 resin (Bio-Rad Laboratories, Richmond, CA). DNA was then extracted by three freeze–thaw cycles (–75°C for 10 min and 94°C for 10 min), and cellular debris was removed by pelleting at 10 000 *g* for 1 min. The quantity of genomic DNA was measured by spectrophotometry at 260 nm in triplicate, and DNA purity was checked using the A_{260}/A_{280} ratio. Purified DNA was stored at –20°C until used.

Polymerase chain reaction amplification and DNA sequencing

Polymerase chain reactions (PCRs) were performed to amplify internal fragments of three loci, *fliC* (positions 1479122–1478914 in the *L. pneumophila* Philadelphia genome sequence, Accession No. AE017354), which corresponds to *flaA* in Gaia and colleagues (2003), *proA* (510043–510483) and *mompS* (3351542–3351037). Primers used for amplification and sequencing were those reported in Gaia and colleagues (2005). Polymerase chain reactions were performed in a 50 μl volume containing 20 ng of genomic DNA, 1 U Taq DNA Polymerase (Promega), 200 μM of each dNTP, 10× Buffer Mg free, 2.5 mM $MgCl_2$, 0.2 μM of each pair of primers. Polymerase chain reaction products were purified using High Pure PCR Product Purification Kit (Roche Diag-

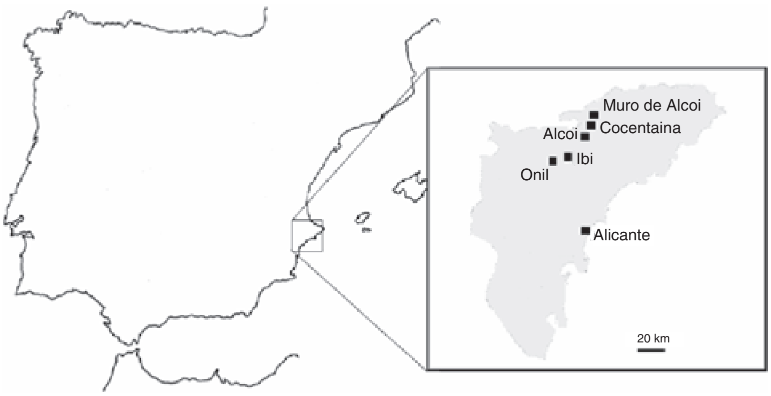

Fig. 2. Map showing the sampling locations of 25 *Legionella pneumophila* environmental isolates used in this study.

nostics GmbH, Mannheim, Germany). The purified DNA was directly sequenced by the dideoxy method using BigDye^{TM} Terminator v3.1 Sequencing Kit and analysed in an ABI PRISM 3700 sequencer (Applied Biosystems, Foster City, CA). Electropherograms were analysed using the Staden package (Staden *et al*., 2000). New sequences obtained in this project have been deposited in GenBank with Accession No.: AY941258–AY941262 (*fliC*), AY943939–AY943943, DQ026282 (*proA*) and AY943931–AY943935, AY934937 (*mompS*).

Sequence analysis

Sequence alignments were obtained using CLUSTALX (Thompson *et al*., 1997) and were further refined by visual inspection. Variability analyses were performed with DnaSP (Rozas *et al*., 2003). Estimates of the number of synonymous and non-synonymous substitutions among sequences (dN/dS) were calculated using the modified Nei–Gojobori method (Nei and Gojobori, 1986) implemented in MEGA 2.0 (Kumar *et al*., 2001). Phylogenetic trees were obtained by the neighbour-joining algorithm (Saitou and Nei, 1987) applied on Tamura–Nei pairwise nucleotide distances and support for the nodes was obtained by bootstrap resampling with 1000 replicates. The program MEGA (Kumar *et al*., 2001) was used for these analyses.

A linkage disequilibrium measure was calculated for the 25 environmental isolates. The index of association (I_A) (Maynard Smith *et al*., 1993) was obtained with the program START (Jolley *et al*., 2001) and calculated as $I_A = V_O/V_E - 1$, where V_O is the observed variance of K, the number of loci at which two individuals differ, and V_E is the expected variance of K. Clonal populations are identified by an I_A value significantly different from zero.

Population genetic structure

Hierarchical AMOVA was performed with program ARLEQUIN (Schneider *et al*., 2000) for the Alicante and European sequences of the three loci. This analysis provides estimates of variance components and F-statistics (Wright, 1931) analogues reflecting the correlation of haplotype diversity at different levels of the hierarchical subdivision. Unlike other approaches for partitioning genetic variation based on gene frequencies, AMOVA also takes into account the genetic relatedness between molecular haplotypes.

The hierarchical subdivision was made at three levels. At the upper level, the two groups considered were Alicante and Europe. Ten different countries in Europe and five different localities in Alicante were considered as populations within groups in the intermediate level. The third level corresponded to the different haplotypes found in each geographical population. AMOVA reports components of variance at the three levels under consideration (among groups, among populations within groups, and within populations) as well as F-statistics analogues. Under this scheme, F_{ST} is to be interpreted as the correlation of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from the whole species; F_{CT} as the correlation of random haplotypes within a group of populations, relative to that of random pairs of haplotypes drawn from the whole species; and F_{SC} as the correlation of the molecular diversity of random haplotypes within populations, relative to that of random pairs of haplotypes drawn from the group (Excoffier *et al*., 1992). The statistical significance of fixation indices was tested using a non-parametric permutation approach (Excoffier *et al*., 1992).

Finally, temporal structuring of genetic variation was tested by comparing genetic diversity between samples taken in three different years. The nearest neighbour statistic (S_{nn}) (Hudson, 2000) was used to test for genetic differentiation between pairs of samples. This statistic considers both the frequency and the nature of the haplotypes found in each sample. We used the implementation of this statistic in DnaSP (Rozas *et al*., 2003) with 1000 resampling replicates.

Acknowledgements

We are grateful to N.K. Fry for providing the nucleotide sequences reported in Gaia and colleagues (2003). This work was supported by Dirección General de Salud Pública, Conselleria de Sanidad, Generalitat Valenciana and project Grupos 03/204 from Agencia Valenciana de Ciencia y Tecnología-Programa AGGV. M.C. is recipient of graduate student fellowship from Generalitat Valenciana. Sequences were obtained at the Servicio de Secuenciación, SCSIE, Universitat de València.

References

Cazalet, C., Rusniok, C., Bruggemann, H., Zidane, N., Magnier, A., Ma, L., *et al.* (2004) Evidence in the

Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* **36:** 1165–1173.

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

Cooper, J.E., and Feil, E.J. (2004) Multilocus sequence typing – what is resolved? *Trends Microbiol* **12:** 373–377.

Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131:** 479–491.

Feil, E.J. (2004) Small change: keeping pace with microevolution. *Nat Rev Micro* **2:** 483–495.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.S., Day, N.P.J., Enright, M.C., *et al.* (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* **98:** 182–187.

Fernandez, J.A., Lopez, P., Orozco, D., and Merino, J. (2002) Clinical study of an outbreak of Legionnaire's disease in Alcoy, Southeastern Spain. *Eur J Clin Microbiol Infect Dis* **21:** 729–735.

Fernandez, J.A., Marco, T., Orozco, D., and Merino, J. (2004) El hospital ante un brote prolongado de legionelosis. *Gac Sanit* **18:** 335–337.

Fields, B.S. (1996) The molecular ecology of legionellae. *Trends Microbiol* **4:** 286–290.

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

Gaia, V., Fry, N.K., Harrison, T.G., and Peduzzi, R. (2003) Sequence-based typing of *Legionella pneumophila* serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. *J Clin Microbiol* **41:** 2932–2939.

Gaia, V., Fry, N.K., Afshar, B., Luck, P.C., Meugnier, H., Etienne, J., *et al.* (2005) Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. *J Clin Microbiol* **43:** 2047–2052.

Hudson, R.R. (2000) A new statistic for detecting genetic differentiation. *Genetics* **155:** 2011–2014.

Jolley, K.A., Feil, E.J., Chan, M.S., and Maiden, M.C.J. (2001) Sequence type analysis and recombinational tests (START). *Bioinformatics* **17:** 1230–1231.

Ko, K.S., Lee, H.K., Park, M.Y., Park, M.S., Lee, K.H., Woo, S.Y., *et al.* (2002) Population genetic structure of *Legionella pneumophila* inferred from RNA polymerase gene (*rpoB*) and DotA gene (*dotA*) sequences. *J Bacteriol* **184:** 2123–2130.

Ko, K.S., Hong, S.K., Lee, H.K., Park, M.Y., and Kook, Y.H. (2003) Molecular evolution of the *dotA* gene in *Legionella pneumophila*. *J Bacteriol* **185:** 6269–6277.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17:** 1244–1245.

Lopez, P., Chinchilla, A., Andreu, M., Pelaz, C., and Sastre, J. (2001) El laboratorio de microbiología clínica en el brote de *Legionella* spp. en la comarca de Alcoy: rentabilidad de las diferentes técnicas diagnósticas. *Enferm Infecc Microbiol Clin* **19:** 435–438.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* **95:** 3140–3145.

Maynard Smith, J., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993) How clonal are bacteria? *Proc Natl Acad Sci USA* **90:** 4384–4388.

Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, USA: Columbia University Press.

Nei, M., and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3:** 418–423.

Ratcliff, R.M., Lanser, J.A., Manning, P.A., and Heuzenroeder, M.W. (1998) Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. *J Clin Microbiol* **36:** 1560–1567.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19:** 2496–2497.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4:** 406–425.

Schneider, S., Roessli, D., and Excoffier, L. (2000) *Arlequin Ver. 2000: A Software for Population Genetics Data Analysis*. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Selander, R.K., McKinney, R.M., Whittam, T.S., Bibb, W.F., Brenner, D.J., Nolte, F.S., and Pattison, P.E. (1985) Genetic structure of populations of *Legionella pneumophila*. *J Bacteriol* **163:** 1021–1037.

Staden, R., Beal, K.F., and Bonfield, J.K. (2000) The Staden package, 1998. *Methods Mol Biol* **132:** 115–130.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876–4882.

Wilkinson, H.W., Fikes, B.J., and Cruce, D.D. (1979) Indirect immunofluorescence test for serodiagnosis of Legionnaires disease: evidence for serogroup diversity of Legionnaires disease bacterial antigens and for multiple specificity of human antibodies. *J Clin Microbiol* **9:** 397–383.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16:** 97–159.