The use of 16S ribosomal RNA analyses to investigate the phylogeny of the family *Legionellaceae*

Norman K. Fry, Simon Warwick, Nicholas A. Saunders and T. Martin Embley

Molecular Biology Unit, DMRQC, 61 Colindale Avenue, London NW9 5HT, UK
Microbial Technology Research Unit, Polytechnic of East London, Romford Road, London E15 4LZ, UK

(Received 10 October 1990; revised 7 January 1991; accepted 16 January 1991)

The 16S ribosomal RNA sequences of *Legionella pneumophila*, *L. erythra*, *L. hackeliae*, *L. spiritensis*, *L. longbeachae*, *L. bozemanii* (*Fluoribacter bozemanii*) and *L. micdadei* (*Tatlockia micdadei*) were determined using reverse transcriptase. The sequences were compared with published sequences for Gram-negative bacteria and phylogenetic trees were constructed. The data confirm previous work which showed that the family *Legionellaceae* forms a monophyletic subgroup within the gamma subdivision of the Proteobacteria. The data show that all of the legionellae studied are highly related (>95%) on the basis of 16S rRNA sequences and do not support the division of the family *Legionellaceae* into three genera.

**Introduction**

The family *Legionellaceae* was initially proposed for a single genus *Legionella* and species *L. pneumophila*, the causal agent of Legionnaires' disease (Brenner et al., 1979). Subsequently large numbers of new species of *Legionella* have been isolated from clinical and environmental samples (Brenner, 1986; Brenner et al., 1985, 1988). All of these isolates share a common set of phenotypic characteristics that include a non-fermentative metabolism, the requirement for L-cysteine and iron salts for growth, ubiquinones with 10 or more isooprene units, and predominantly branched-chain cellular fatty acids (Karr et al., 1982; Brenner et al., 1984; Brenner, 1986). The identification of legionellae at the species level is more difficult as there are relatively few suitable phenotypic tests (Brenner, 1986; Brenner et al., 1984; Fox et al., 1984, 1990; Vesey et al., 1988). New species are identified by DNA pairing measurements, and this method is the only means by which some species may be distinguished (Brenner, 1986; Brenner et al., 1988). However, DNA pairing values between species within the genus *Legionella* may vary between 0% and 67% under optimum reassocation conditions (Garrity et al., 1980; Brown et al., 1981; Brenner, 1986). The very low DNA hybridization values that are obtained between some species of *Legionella* have been interpreted differently. One group (Garrity et al., 1980; Brown et al., 1981) has advocated that the family *Legionellaceae* be subdivided into three genera. The genus *Fluoribacter* was proposed for strains of *Legionella bozemanii* which exhibit blue–white autofluorescence and show less than 14% DNA relatedness to *L. pneumophila* (Garrity et al., 1980; Brown et al., 1981). A further genus, *Tatlockia*, was proposed for *L. micdadei* (Hébert et al., 1980b), which shows less than 5% DNA relatedness to *L. pneumophila* (Garrity et al., 1980; Brown et al., 1981). In contrast to these proposals, others have argued that all legionellae are sufficiently similar and unusual in phenotype to be classified in a single genus (Brenner et al., 1980, 1984; Hébert et al., 1980a, b; Morris et al., 1980; Brenner, 1986). At present there is no clear agreement as to which position is correct and the same strains appear in the Approved Lists of Bacterial Names under two different names.

The sequence of the genes that code for ribosomal RNA (rRNA) are among the most highly conserved yet identified (Brenner, 1986; Woese, 1987). Consequently they can be used to determine taxonomic relationships between species which show little DNA interrelatedness (Brenner, 1986). The accuracy of phylogenetic inferences from rRNA sequences depends upon the number of bases compared, and to be fully effective at least 1000 bases should be considered for each organism (Murray et al., 1990). Ludwig & Stackebrandt (1983) investigated the phylogeny of *Legionella pneumophila*, *Legionella bozemanii* (*F. bozemanii*) and *Legionella micdadei* (*T. micdadei*) using oligonucleotide cataloguing of 16S rRNA, which recovers about 40% of the 16S rRNA sequence. The results of that study showed that all three
organisms were related and well separated from other Gram-negative bacteria but did not resolve whether there should be one or three genera. Fox & Brown (1989) used reverse transcriptase to sequence short stretches (350–450 bases) of 16S rRNA from three strains of *L. pneumophila* and three of *L. micdadei* (*T. micdadei*). However, they were unable to draw any conclusions about the taxonomic structure of the family *Legionellaceae*. In the current investigation we have attempted to determine the detailed relationships of the family *Legionellaceae*. About the taxonomic structure of the family *Legionellaceae* including *L. micdadei*, *L. erythra*, *L. hackeliae*, *L. spiritensis*, *L. longbeachae*, *L. bozemanii* (*F. bozemanae*) and *L. micdadei* (*T. micdadei*)

**Methods**

**Bacterial strains and cultivation.** The test strains (Table 1) were inoculated onto buffered charcoal yeast-extract agar (BCYE) and incubated for 48 h at 37 °C. The cells were removed from the surface of the plates using sterile saline (0.85% w/v), pelleted by centrifugation at 10000 g, and stored at −70 °C.

**Isolation of RNA.** The method used was modified from that described by Saunders & Grant (1984). Approximately 2 g wet weight of cells was resuspended in 5 ml of a buffer containing 25% (w/v) sucrose, 50 mM-Tris/HCl (pH 8.0) and 3-3 mg lysozyme ml⁻¹, and incubated for 30 min at 37 °C. The cells were recovered by centrifugation (14000 g), resuspended in 20 ml 10 mM-Tris/HCl (pH 7.5) buffer containing 1% (w/v) SDS, 5 mM-EDTA and 200 μg proteinase K ml⁻¹ (Sigma), and incubated for 1 h at 45 °C. The lysate was extracted twice with equilibrated phenol (Maniatis et al., 1982) and then twice with chloroform/isoamyl alcohol (24:1, v/v). The aqueous phases were pooled and adjusted to 0.3 M with respect to sodium acetate (pH 6.0). Total nucleic acids were precipitated by incubation with 2 vols ethanol at −20 °C. The nucleic acid pellet was resuspended in 20 ml sodium acetate (pH 6.0) and the DNA was sheared by passing the suspension repeatedly through a sterile 23 gauge needle until the viscosity decreased significantly. The suspension was then centrifuged again and the dried pellet resuspended in 0.5 ml TE buffer (10 mM-Tris/HCl, pH 8.0, 1 mM-EDTA) containing 6 mM-MgCl₂, RNAase-free DNAase (37.5 units, Pharmacia LKB) was then added and the mixture incubated for 30 min at 20 °C. The RNA was precipitated with 2 vols ethanol at −20 °C, collected by centrifugation (10000 g) and the dried pellet was resuspended in 50 μl sterile distilled water to which 1.5 ml 3 M-sodium acetate (pH 6.0) was added. After incubation for 2 h at 4 °C the RNA was collected by centrifugation (10000 g) and the pellet washed with 0.1 M-sodium acetate (pH 6.0)/70% (v/v) ethanol prior to redissolving in 100 μl sterile distilled water. The integrity of the rRNA was determined by electrophoresis in an agarose (1%, w/v) mini-gel prepared using TBE buffer (0.1 M-Tris, 83 mM-boric acid, 1 mM-EDTA, pH 8.6) and containing 5 μg ethidium bromide ml⁻¹. The concentration of the RNA solution was adjusted to 3 μg μl⁻¹ and stored at −70 °C.

**RNA sequencing.** Eight primers were routinely used to generate sequence data. The sequences and provenance of primers have been given previously (Embley et al., 1988). The method used for sequencing 16S rRNA was that devised by Lane et al. (1985, 1988) with slight modifications as follows. The reverse transcription reaction mixture contained 7 units of AMV reverse transcriptase (Stratagene) and 10 μCi (370 kBq) deoxyadenosine 5'-[β-32P]thiotriphosphate (Du Pont; specific activity 1300 Ci mmol⁻¹, 48 TBq mmol⁻¹). The nucleotide mixes were prepared using individual ultrapure nucleotide solutions (Pharmacia LKB). Stock mixes were determined empirically to give about 300–350 bases of readable sequence and comprised: A mix, dATP 20 μM, dCTP dGTP dTTP each at 1.7 mM, ddATP 3.2 μM, C mix, dATP 8.5 μM, dCTP dGTP and dTTP each 0.4 mM, ddCTP 2.5 μM; G mix, dATP 8 μM, dCTP and dTTP each 1.6 mM, dGTP 0.4 mM, ddGTP 5.8 μM; T mix, dATP 9 μM, dCTP and dGTP each 1.8 mM, dTTP 0.3 mM, ddTTP 6.7 μM. The chase mix contained each dNTP at 0.5 mM. The products of the sequencing reactions were separated on 55 cm wedge-shaped (0.2–0.6 mm) 6% (w/v) polyacrylamide denaturing (7 M-urea) gels using an LKB Macrophor system

**Table 1. Legionella strains included in this study**

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain designation</th>
<th>Source</th>
<th>GenBank no. of rRNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. pneumophila</em></td>
<td>Philadelphia 1</td>
<td>NCTC 11192*</td>
<td>M36023</td>
</tr>
<tr>
<td>subsp. pneumophila</td>
<td>Knoxiville 1</td>
<td>NCTC 11286</td>
<td>M36024</td>
</tr>
<tr>
<td><em>L. pneumophila</em></td>
<td>Los Angeles 1</td>
<td>NCTC 11233*</td>
<td>M36025</td>
</tr>
<tr>
<td>subsp. fraseri</td>
<td>Dallas 1E</td>
<td>NCTC 11405</td>
<td>M36026</td>
</tr>
<tr>
<td><em>L. erythra</em></td>
<td>SE 32A C8</td>
<td>NCTC 11977*</td>
<td>M36027</td>
</tr>
<tr>
<td><em>L. hackeliae</em></td>
<td>Lansing 2</td>
<td>NCTC 11979*</td>
<td>M36028</td>
</tr>
<tr>
<td><em>L. longbeachae</em></td>
<td>Long Beach 4</td>
<td>NCTC 11477*</td>
<td>M36029</td>
</tr>
<tr>
<td><em>L. spiritensis</em></td>
<td>Mt St Helens 9</td>
<td>NCTC 11990*</td>
<td>M36030</td>
</tr>
<tr>
<td><em>L. bozemanii</em></td>
<td>WIGA</td>
<td>NCTC 11368*</td>
<td>M36031</td>
</tr>
<tr>
<td>(<em>Fluoribacter bozemaniae</em>)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>L. micdadei</em></td>
<td>TATLOCK</td>
<td>NCTC 11371*</td>
<td>M36032</td>
</tr>
<tr>
<td>(<em>Tatlockia micdadei</em>)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Denotes type strain.
operated at 50 W per gel and at 60 °C. Prior to loading the samples, and after 10 min pre-electrophoresis, sodium acetate (3 M in TBE buffer) was added to the lower chamber to give a final concentration of 0.33 M (Brown & Lund, 1988). After a further 10 min pre-electrophoresis the samples were loaded and allowed to run until the bromophenol blue tracking dye reached the bottom of the gel (about 4-5 h). Urea was removed by washing (twice, 10 min) in a solution comprising 10% (v/v) methanol, 10% (v/v) acetic acid and 2% (w/v) glycerol, followed by a rinse in distilled water. The gel was dried overnight at 80 °C in a standard oven. Autoradiography was for 24-48 hours using Fuji RX film.

Data analysis. Sequences were aligned and similarity values calculated using the ALIGN program (Wilbur & Lipman, 1983) from the DNASTAR suite of programs. The Legionella strains and 13 sequences of Gram-negative bacteria representing the major subgroups of the Proteobacteria were included in these calculations (Dams et al., 1988; Rossau et al., 1988; Stackebrandt et al., 1988; Dewhirst et al., 1989; Dorsch et al., 1989; Weisburg et al., 1989; L. M. Manchester, S. Warwick, W. D. Grant & T. M. Embley, unpublished data). Positions where a nucleotide could not be determined unambiguously (designated by an IUB uncertainty code) were not included in the alignment. Regions of extreme variation were removed to eliminate noise due to multiple mutations at a single position, and to ensure that only homologous positions were compared. Seven long stretches of sequence (about 1350 bases) were compared (Escherichia coli positions; Brosius et al., 1978): 2-68, 99-190, 220-840, 846-998, 1004-1023, 1037-1132, 1143-1444. The homology values between the sequences were converted to evolutionary distance or $K_{nu}$ values (Horii, 1975; Jukes & Cantor, 1969). These values were used to produce phylogenetic trees using the algorithm of Fitch & Margoliash (1967) contained in the PHYLIP 3.22 program (J. Felsenstein, University of Washington). The relationships between the Legionellaceae, Coxiella burnetii, E. coli and Wolbachia persica (as outgroups) were also analysed using the program DNAROOT from the PHYLIP 3.22 program. DNAROOT computes a consensus phylogenetic tree using parsimony and the bootstrap method of sampling sequence data.

Results and Discussion

The sequencing strategy used in the current investigation generated between 1446 and 1474 bases (about 95%) of 16S rRNA sequence for the ten strains investigated. The primary sequence data have been deposited in GenBank and assigned the accession numbers shown in Table 1. Fig. 1 shows the matrix of homology values and corresponding $K_{nu}$ values obtained from a comparison of approximately 1350 bases of sequence from the legionellae and representatives of the major groupings within the Proteobacteria (Stackebrandt et al., 1988). Fig. 2 is a phylogenetic tree showing the position of the family Legionellaceae. The family forms a subline within the gamma subdivision of the Proteobacteria. This confirms the results of earlier phylogenetic studies which showed the legionellae to be distant relatives of this group (Ludwig & Stackebrandt, 1983; Weisburg et al., 1989). The relationship is supported by the presence of the signature oligonucleotide CUACUYYG, which is present at position 510 (E. coli numbering) in all members of the gamma subgroup (Stackebrandt et al., 1988). The two representatives of the Rickettsiaceae, Coxiella burnetii and Wolbachia persica, showed the closest relationship to the Legionellaceae. The common evolutionary origin of L. pneumophila and C. burnetii, the causal agent of Q fever, was previously established by Weisburg et al. (1989). In this respect it is interesting that some of the bacteria classified as legionellae were originally described as rickettsia-like organisms because of their biological characteristics (Tatlock, 1944; Bozeman et al., 1968).

Fig. 3 shows the internal relationships within the family Legionellaceae based on an analysis of 1350 bases. The tree shown is rooted using Wolbachia persica as an outgroup. The consensus tree obtained using the bootstrap parsimony method had exactly the same topology as this tree. The strains of Legionella show at least 95% similarity over the 1350 bases of rRNA sequence compared. These values are higher than the similarities (about 90-6%) based on approximately 400 bases of sequence which were detected between L. micdadei (T. micdadei) and L. pneumophila by Fox & Brown (1989). The discrepancies are due to the fact that the short stretches of sequence which they compared contained a high proportion of the highly variable regions of RNA sequence. In contrast, the data presented here are based upon an analysis of approximately 90% of the molecule. The representatives of Legionella, including L. bozemanii (F. bozemanae) and L. micdadei (T. micdadei), form a cluster which is of a similar depth to the branch between E. coli and Hafnia alvei (Fig. 2). It is, however, much shallower than published values between species of genera such as Mycobacterium (Smida et al., 1988), Pseudomonas (Johnson & Palleroni, 1989), Saccharopolyspora (Bowen et al., 1989) and Streptomyces (Witt et al., 1989). Fig. 3 shows that L. bozemanii (F. bozemanae) is closely related (97.7%) to L. longbeachae and forms a subbranch of the line leading to the subspecies of L. pneumophila. This relationship was recovered irrespective of which reference organism, i.e. C. burnetii, E. coli or W. persica, was used as an outgroup. It was also recovered in the bootstrap parsimony analysis (not shown). The classification of L. bozemanii (F. bozemanae) has been the subject of considerable debate (Garrity et al., 1980; Ludwig & Stackebrandt, 1983; Brenner, 1986; Brenner et al., 1988). The name L. bozemanii was effectively published for a Legionella-like organism isolated from diseased lung tissue and which exhibited blue-white autofluorescence under long-wave UV light (Brenner et al., 1980). This strain showed less than 13% DNA relatedness to L. pneumophila (Brenner et al., 1980; Garrity et al., 1980). Subsequently, the new genus Fluoribacter was validly proposed for the same strain (Garrity et al., 1980). Both names currently appear in the Approved Lists of Bacterial Names and it has been left to
Fig. 1. Homology values (lower left) and $K_{mec}$ values (upper right) for 1350 bases of 16S rRNA from representative Proteobacteria and members of the family *Legionellaceae*. 
Fig. 2. Unrooted phylogenetic tree showing the position of the family Legionellaceae within the confines of the gamma subdivision of the Proteobacteria. Genera are abbreviated as follows: A., Agrobacterium; B., Brucella; Cx., Coxella; D., Desulfovibrio; E., Escherichia; H., Hafnia; M., Myxococcus; N., Neisseria; P., Proteus; Ps., Pseudomonas; R., Ruminobacter; Ro., Rochalimaea; Wolb., Wolbachia; Woli., Wolinella.

Fig. 3. Phylogenetic tree showing the relationships within the family Legionellaceae. The tree is rooted by reference to Wolbachia persica.

individuals to decide which name they wish to use. A critical argument in the proposal for Fluoribacter bozemaniae (Garrity et al., 1980) was a reference to published DNA data for relationships between genera in the family Enterobacteriaceae (Brenner, 1973). This suggested that members of the same genus should show at least 25% DNA relatedness under optimum conditions. However, a number of workers have argued that the genetic relationships between members of the Enterobacteriaceae are uniquely high among eubacterial genera, and are thus an inappropriate model for defining relationships in other taxa (Ludwig & Stackebrandt, 1983; Brenner, 1986; Brenner et al., 1988). Indeed there are many examples of good phenotypic species which
show low levels of DNA relatedness in genera such as Streptomyces (Mordarski et al., 1986) and Staphylococcus (Freney et al., 1988). In a further publication, Garrity and coworkers (Brown et al., 1981) transferred two additional species of autofluorescing legionellae to the genus Fluoribacter, as F. dumoffii and F. gormanii. With the inclusion of these strains the level of DNA relatedness within the genus Fluoribacter dropped to as little as 4% between some strains (Brown et al., 1981). In contrast, DNA relatedness between some fluorobacters and the non-autofluorescing L. longbeachae and L. wadsworthii became as high as 17% and 30%, respectively (Brenner et al., 1985; Brenner, 1986). The results of DNA pairing studies therefore fail to separate Fluoribacter from non-autofluorescing strains of Legionella (Brenner et al., 1985). In our investigation, L. bozemani (F. bozemanae) clearly falls within the phylogenetic radiation defined by non-autofluorescent species of Legionella.

The genus Tatlockia was proposed (Brown et al., 1981) for a strain previously classified as Legionella micdadei (Hébert et al., 1980b). The main rationale behind this proposal was the inability to detect any DNA homology between this strain and L. pneumophila (Brown et al., 1981). Subsequent work has shown that L. micdadei (T. micdadi) shows moderate levels of relatedness to L. maceachernii (23%) and L. spiritensis (16%) (Brenner et al., 1985). In the current investigation L. micdadei (T. micdadi) showed 95–96% similarity in rRNA sequence (over 1350 bases) to strains of L. pneumophila, L. spiritensis, L. erythra, L. hackeliae and L. bozemani (F. bozemanae) (Fig. 1). In addition, L. micdadei was found to lie within the confines of the radiation defined by Legionella in all of the phylogenetic trees constructed from our data.

The strains of L. pneumophila were recovered as two closely related (>99% similarity) clusters which represented the two subspecies pneumophila and fraseri. The two strains of L. pneumophila subsp. pneumophila were indistinguishable for the entire 1470 bases of 16S rRNA sequence considered. The two strains of L. pneumophila subsp. fraseri differed at only two positions over 1470 bases. The two subspecies could be differentiated by seven base changes and a single deletion. All of the L. pneumophila isolates had a characteristic oligonucleotide UUAUAUGAAAUAA at position 837 (E. coli numbering), which served to differentiate them from the other legionellae. L. erythra, L. spiritensis and L. hackeliae defined the periphery of the family Legionellaceae in all of the phylogenetic trees. However, the juxtaposition of the three species showed slight variation depending upon the choice of outgroup. Very little can therefore be concluded concerning their specific relationships.

It has recently been suggested that the integrated use of phylogenetic and phenotypic characters is necessary for the delineation of bacterial taxa at all levels (Murray et al., 1990). L. bozemanii (F. bozemanae) and L. micdadei (T. micdadi) can be distinguished from other species of Legionella using detailed analysis of carbohydrate composition (Fox et al., 1984, 1990), and quantitative fatty acid and ubiquinone composition (Karr et al., 1982; Lambert & Moss, 1989). However, the spectrum of variation of these features is of the same magnitude as found between other species of Legionella and is therefore insufficient to justify separate genera for these strains (Brenner et al., 1984; Lambert & Moss, 1989; Fox et al., 1990). The major evidence for the separation of the Legionellaceae into different genera stems from the interpretation of the low DNA pairing values which are obtained between some species (Garrity et al., 1980; Brown et al., 1981; Brenner et al., 1985). At the moment these cannot be consistently interpreted to recognize phenotypically coherent groups. In contrast to the low DNA pairing values published for legionellae, all of the species examined in the present study were found to be highly related at the rRNA sequence level. Furthermore, L. bozemanii (F. bozemanae) and L. micdadei (T. micdadi) clearly fall within the radiation defined by Legionella. It is therefore phylogenetically invalid to place these species in separate genera. In the absence of compelling phenotypic differences (Brenner, 1986; Murray et al., 1990) we conclude that all of the species examined in the current investigation should be regarded as members of the genus Legionella (Brenner et al., 1980; Hébert et al., 1980b; Morris et al., 1980).

The authors would like to thank Dr T. Harrison for help and advice on the choice of strains. This work was supported in part by a research grant from the PCFC Selective Research Initiative to T. M. E. and S. W.

References


