



HAL
open science

Host-specific adaptation of *Legionella pneumophila* to single and multiple hosts

Anaísá Moreno, Kiran Paranjape, Martina Cederblom, Elisabeth Kay, Christian Dobre-Lereanu, Dan Andersson, Lionel Guy

► **To cite this version:**

Anaísá Moreno, Kiran Paranjape, Martina Cederblom, Elisabeth Kay, Christian Dobre-Lereanu, et al.. Host-specific adaptation of *Legionella pneumophila* to single and multiple hosts. 2024. hal-04792487

HAL Id: hal-04792487

<https://hal.science/hal-04792487v1>

Preprint submitted on 20 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Host-specific adaptation of *Legionella pneumophila* to
2 single and multiple hosts

3

4 Anaísa B. Moreno^{1,†}, Kiran Paranjape^{1,†}, Martina Cederblom¹, Elisabeth Kay²,
5 Christian Dobre-Lereanu¹, Dan I. Andersson¹, and Lionel Guy^{1,*}

6

7

8 ¹ Department of Medical Biochemistry and Microbiology, Science for Life
9 Laboratories, Uppsala University, Uppsala, Sweden

10 ² Centre International de Recherche en Infectiologie, Lyon, France

11

12 [†] These authors contributed equally to the study.

13 ^{*} Corresponding author (lionel.guy@imbim.uu.se)

14

15 ORCIDs. ABM: 0000-0002-3229-4029; KP: 0000-0002-3932-4083; EK: 0000-
16 0002-8763-7076; DIA: 0000-0001-6640-2174; LG: 0000-0001-8354-2398

17

18 Abstract

19 *Legionella pneumophila* is an endosymbiotic bacterial species able to infect and
20 reproduce in various protist and human hosts. Upon entry in human lungs, they may
21 infect lung macrophages, causing Legionnaires' disease (LD), an atypical
22 pneumonia, using similar mechanisms as in their protozoan hosts, despite the two
23 hosts being separated by a billion years of evolution. In this study, we used
24 experimental evolution to identify genes conferring host-specificity to *L.*
25 *pneumophila*. To this end, we passaged *L. pneumophila* in two different hosts -
26 *Acanthamoeba castellanii* and the human macrophage-like cells U937 - separately
27 and by switching between the hosts twice a week for a year. In total, we identified
28 1518 mutations present in at least 5% of the population at the time of sampling. Half
29 of these were localized in five groups of repeated sequences, likely to be
30 recombination hotspots. Forty-nine mutations were fixed in the 18 populations at
31 the end of the experiment, representing four different groups. The first two groups
32 involve adaptation to the specific selection conditions, including (i) two specific
33 mutations in the 30S ribosomal protein S12 (RpsL) that conferred resistance to
34 streptomycin, to which the bacteria were unexpectedly exposed to during serial
35 passage and (ii) two mutations, one in the 30S ribosomal protein S4 (RpsD), and one
36 in the chaperonin GroES, both of which are likely to be fitness-restoring
37 compensatory mutations to the original RpsL mutations. Two more interesting
38 groups of mutations included (iii) mutations in 4 different strain-specific genes
39 involved in LPS synthesis, found only in the lineages passaged with *A. castellanii* and
40 (iv) mutations in the gene coding for LerC, a key regulator of protein effector
41 expression, which was independently mutated in 6 lineages grown in presence of
42 the U937 cells. We propose that the mutations degrading the function of the
43 regulator LerC improve the fitness of *L. pneumophila* in human-derived cells, and
44 that modifications in the LPS are beneficial for growth in *A. castellanii*. This study is
45 a first step in further investigating determinants of host specificity in *L.*
46 *pneumophila*.

47

48 Introduction

49 To date, 65 species of *Legionella* have been discovered, half of which can cause
50 disease in humans. Globally, 85% of cases of legionellosis are caused by *Legionella*
51 *pneumophila* - with the exception of Australia and New Zealand where 30% of cases
52 are caused by *Legionella longbeachae* (Fields et al. 2002; Den Boer and Yzerman
53 2004). *Legionella* causes two forms of legionellosis in humans – Legionnaires’
54 disease (LD) and Pontiac fever. LD is presented as a severe pneumonia, with
55 mortality of up to 15%, whilst Pontiac fever induces a milder flu-like illness. The
56 worldwide occurrence of LD is difficult to calculate due to misdiagnosis and
57 underreporting, but it has been established as a major contributor to both
58 community- and hospital-acquired pneumonia (Falcó et al. 1991; Diederer 2008;
59 Lau and Ashbolt 2009).

60
61 *Legionella* has a long history of adaptation to eukaryotes. The order encompassing
62 *Legionella* and the related pathogen *Coxiella*, among others, is believed to have
63 started colonizing eukaryotes almost 2 billions years ago (Hugoson et al. 2022). The
64 ability of *Legionella* to infect humans has been associated with its long history of co-
65 evolution with protozoans. *Legionella* is able to control its host’s cellular processes,
66 such as vesicle trafficking. However, beyond the processes that are shared and
67 highly conserved among all eukaryotes, *Legionella* can also control others that are
68 absent in the protists it frequently encounters (Park et al. 2020). It is unlikely that
69 *Legionella* has evolved these abilities simply from accidental transmission to
70 humans and man-made aquatic environments in the past century, especially
71 because humans are presumably evolutionary dead-ends for *Legionella*, as it doesn’t
72 spread from patient to patient. The human host is not frequently met by *Legionella*,
73 and specific in-patient mutations occur, despite the short infection cycle (Leenheer
74 et al. 2023). This leads to a hypothesis where metazoan species are also natural hosts
75 of *Legionella*, and these have been instrumental in furthering the evolution of
76 *Legionella* to infect human macrophages (Best and Abu Kwaik 2019).

77
78 The life cycle of *Legionella* replication is similar in all eukaryotic cells, despite some
79 host-specific differences. It is characterized by two distinct phases, replicative and
80 transmissive (Oliva et al. 2018). Upon entry into the host cell, *Legionella* evades
81 phagosome-lysosome fusion by forming a *Legionella*-containing vacuole (LCV).
82 Shortly after the LCV is formed, it recruits host mitochondria and rough
83 endoplasmic reticulum to surround its surface. In the LCV, *Legionella* undergoes
84 exponential growth - the replicative stage. When the host’s nutrients are exhausted,
85 a new stage is triggered – the transmissive stage. There, *Legionella* is released from
86 the LCV to the host’s cytosol, before exiting the cell using several mechanisms. In
87 this transmissive state *Legionella* is able to infect alveolar macrophages and cause
88 disease in humans. In laboratory conditions, *Legionella* also displays these
89 replicative and transmissive characteristics, during the exponential and stationary
90 phases, respectively (Richards et al. 2013; Oliva et al. 2018).

91

92 The capacity of *Legionella* to evade lysosomes and modulate cellular processes in the
93 host is conferred by the Dot/Icm system (Isberg et al. 2009). The latter is classified
94 as a type IVB secretion system (T4BSS), and has two groups of loci - the *dot* (defect
95 in organelle trafficking) and the *icm* (intracellular multiplication). In *Legionella* this
96 secretion system injects over 300 effector proteins into the host cell. These effectors
97 play a role in entry of *Legionella* into the host cell, formation of LCV, and exit from
98 the host (Ensminger 2016). They also manipulate various cellular processes such as
99 vesicle trafficking, protein translation, and ubiquitination pathways. About 10% of
100 the coding capacity of *Legionella pneumophila* is used to express these effector
101 proteins (Richards et al. 2013).

102

103 A large number of effector proteins are functionally redundant. This has been
104 illustrated by experiments where deleting 31% of the effectors did not cause
105 significant impairment of intracellular replication of *Legionella* in mouse
106 macrophages (O'Connor et al. 2011). Many effector proteins contain eukaryotic
107 domains which are thought to facilitate modulation of host cells (Hubber and Roy
108 2010). These domains can combine to make new effectors, which could be one of the
109 reasons why *Legionella* can infect such a wide range of hosts (Gomez-Valero et al.
110 2019).

111

112 An important part of *Legionella*'s infection cycle is the coordination and regulation
113 of effector expression where four two-component systems (TCS) are central (Segal
114 2014). Together, they regulate the expression of over 100 genes (Feldheim et al.
115 2018) by sensing environmental stimuli and transducing the signal into the cell,
116 often via an autophosphorylation mechanism. By timing the expression of the right
117 effectors at the right time, they probably play an important role in host specificity.

118

119 The broad host-range of *Legionella* is thought to have arisen from its lifestyle of host
120 cycling, which selects against mutations that may decrease its fitness in the many
121 protozoans it naturally infects. This was tested in an experimental evolution set-up,
122 where *Legionella* was passaged in mouse macrophages for a year. The results showed
123 rapid parallel evolution of *Legionella*, and adaptive mutations that resulted in
124 improved replication in macrophages when compared to ancestral strains
125 (Ensminger et al. 2012). Another evolution experiment involving *Legionella* and
126 amoebae is the spontaneous endosymbiosis discovered by Jeon and Lorch (1967).
127 Following an unexpected infection of one of their cultures of *Amoeba proteus* by an
128 unknown bacterium (X-bacterium, now renamed '*Ca. Legionella jeonii*'), a few
129 amoebae survived, but kept the bacterium, which turned out to be essential for the
130 amoeba upon subsequent cultivation.

131

132 In this study, we aimed at exploring the selective pressures imposed on the multi-
133 host endosymbiont *L. pneumophila*. By using experimental evolution, we sought to

134 identify genetic determinants of host-specificity in this bacterium. We cycled *L.*
135 *pneumophila* in two different hosts separately and by repeated host switching. After
136 one year of biweekly passaging, we sequenced the evolved lineages at population
137 levels and tested the fitness effects of the most frequent genotypes.

138 Results

139 Distribution of mutations

140 Each of the six lineages for all three hosts were passaged in their hosts for up to one
141 year and sequenced at population level at two or three time points ([Supplementary](#)
142 [Table 1](#)). All lineages passaged with human macrophage-derived cells (U937 cells,
143 abbreviated as Mac) were sequenced at all three timepoints, while for the other two
144 conditions, passaged with *A. castellanii* (Acas) and switching between the two hosts
145 (alternation, Alt), only one population was sequenced at one of the two later time
146 points.

147
148 We identified in total 1518 mutations that exceeded a 5% frequency in any
149 population, after removing low-frequency mutations (<20%) in one sample (Alt_D,
150 t1) which was contaminated with another strain of *L. pneumophila*. The number of
151 mutations reaching 5% frequency ranged from 4 to 102 per population and time
152 point, with a median of 28.5 ([Supplementary Figure 1](#)). The number of mutations
153 appeared more variable and on average slightly higher at time point 1 than 3
154 ([Supplementary Figure 2A](#)). The fraction of non-fixed mutations is also higher at
155 time point 1 than 3 ([Supplementary Figure 2B](#)). Small indels accounted for only 4%
156 of the mutations (n = 62). As expected, among SNPs, transitions (987, 68%) were
157 more frequent than transversions (468, 32%). The number of AT → GC mutations
158 (N = 599, 50.4%) was very similar to the reverse (GC → AT; N = 589, 49.6%). The
159 numbers were also very similar when excluding repeated regions and recombination
160 hotspots prone to recombination (GC → AT: 270, 50.6%; AT → GC: 264, 49.4%). Very
161 few mutations (15) were found on the *L. pneumophila* Paris pLPP plasmid. All of them
162 appeared with frequencies < 15%. All further analysis focused on the chromosomal
163 mutations.

164

165 Mutational hotspots

166 Mutations were not uniformly distributed along the chromosome, and many
167 clustered around several hotspots ([Supplementary Figure 3](#)). In total, mutations in
168 these hotspots accounted for 751 or 49.4% of all mutations, but notably none of
169 them were fixed. Eleven of these hotspots corresponded to intergenic regions
170 upstream or downstream of GIY-YIG nuclease family protein, which is a homing
171 endonuclease/mobile element (Dunin-Horkawicz et al. 2006). In total, 97
172 mutations occurred in these 11 hotspots. A second hotspot for mutations was a gene
173 corresponding to locus tag lpp1100, coding for an ankyrin repeat-containing
174 protein, which displayed 142 mutations. The ankyrin repeats (19 repeats of 35
175 residues, [Supplementary Figure 4](#)) are presumably different enough so that the

176 variants that were called by BRESEQ represent true mutations and not mapping
177 errors. The third hotspot for mutations was one of the five tetratricopeptide repeat
178 (TPR; 19 mutations) proteins present in *L. pneumophila* str. Paris, lpp2912. The
179 fourth hotspot was a repeated protein containing a domain of unknown function
180 (DUF1566; 36 mutations) which is widely present in *Gammaproteobacteria*. Finally,
181 the fifth group of hotspots was constituted of three occurrences of a larger segment
182 containing a transposase and a type II-like restriction endonuclease (DUF559; 457
183 mutations).

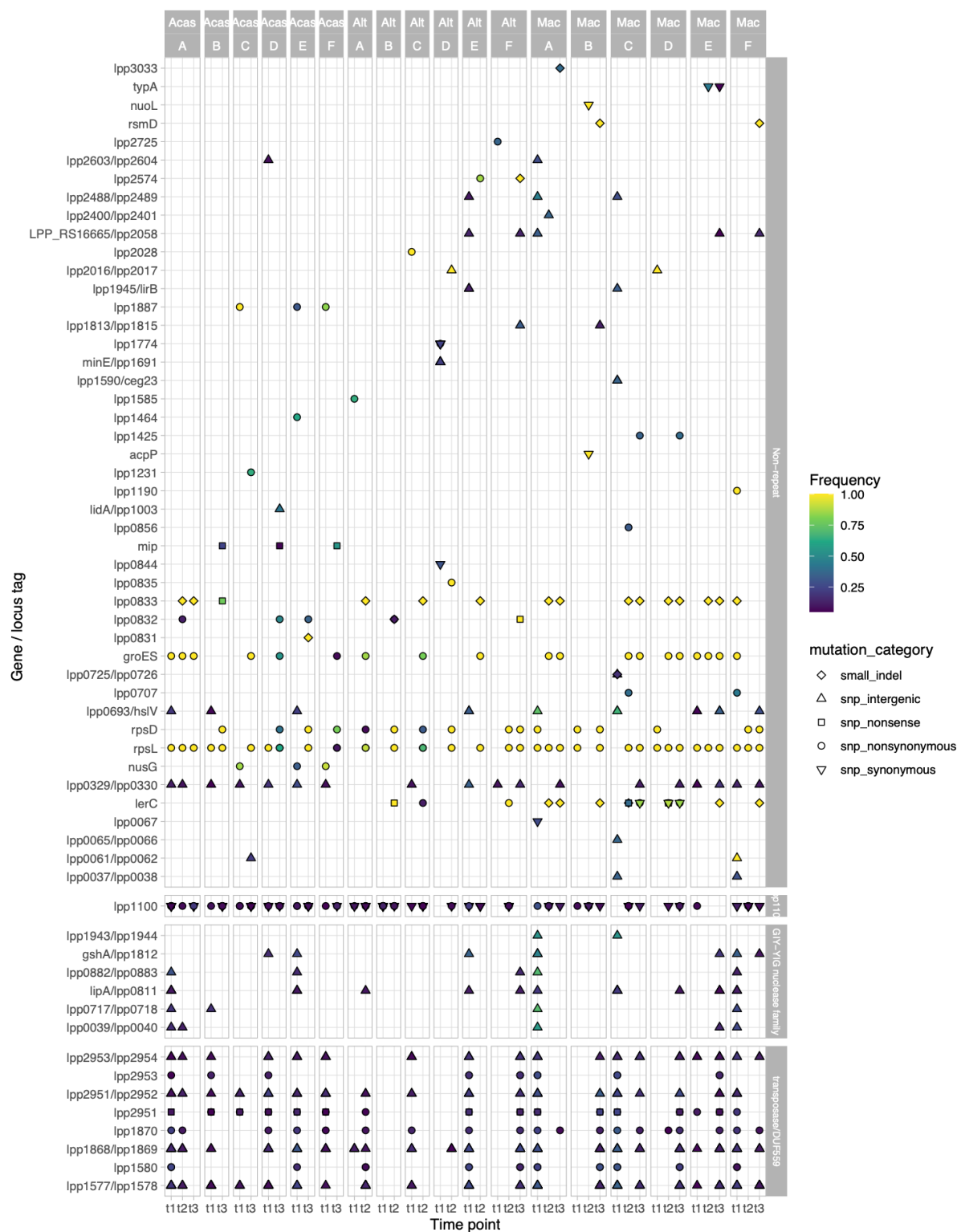
184

185 Frequent and fixed mutations

186 The frequency of variants varied greatly, but in total 20, 35 and 36 mutations
187 appeared fixed after t1, t2 and t3, respectively; this corresponded to 1.2, 2.7 and 2.8
188 mutations per population. After the last time point (t3 for Acas and Mac, and t2 for
189 most of the Alt populations), 10, 22, and 15 mutations were fixed in Acas, Mac, and
190 Alt conditions, respectively. In some populations (e.g. Mac_A and Alt_D at t1,
191 Acas_D and Mac_C at t3; [Supplementary Figure 2](#)), groups of variants were found
192 at a higher frequency but were not fixed, suggesting two sub-populations with
193 distinct mutations co-existing in the same culture.

194

195 In addition to the recombination hot-spots regions mentioned above, other genes
196 were also mutated several times with higher frequencies ([Figure 1](#)). Most fixed
197 mutations (92.3%) were small indels or non-synonymous mutations, while
198 intergenic or synonymous mutations were more common at lower frequencies
199 ([Table 1](#)), as expected if genetic drift is not prevalent. Remarkably, the number of
200 synonymous mutations dropped from 20.1% among all mutations to only 2.2% (n =
201 2) among fixed mutations. Many non-synonymous mutations occurred
202 independently in the same gene (e.g. *rpsL*, *lerC*, *lpp0832*) ([Figure 1](#)).



203
204
205
206
207
208
209
210

Figure 1: Frequency of mutations appearing at least once within a gene or an intergenic space, with a frequency in the population of at least 25%. Mutations occurring within a gene are labeled with the locus tag or the gene name (if available). Mutations occurring in an intergenic space are labeled with the two bordering gene names or locus tags. If a gene/intergenic space is mutated several times, with different types of mutations (e.g. *lerC*), the different symbols are stapled on top of each other. The shape of the points corresponds to the type of mutation. The color

211 of each point corresponds to the frequency of the mutation. Yellow dots are fixed
 212 mutations. Mutations occurring in non-repetitive regions are shown at the top, and
 213 repetitive regions at the bottom.

214

215 **Table 1: Number and frequency of mutations per category.** Mutations were counted
 216 for all three time points. The three pairs of columns represent the number and
 217 percentage of mutations with frequency over 5%, with frequency over 25%, and
 218 fixed in the population, respectively.

219

Mutation category	N (>5%)	%	N (> 25%)	%	N (fixed)	%
Small indel	62	4.1	29	3.7	23	25.3
SNP intergenic	611	40.4	404	52.1	3	3.3
SNP nonsense	42	2.8	21	2.7	2	2.2
SNP non-synonymous	492	32.6	240	30.9	61	67.0
SNP synonymous	304	20.1	82	10.6	2	2.2
<i>Total</i>	1511	100.0	776	100.0	91	100.0

220

221 In total, 48 mutations (representing 13 distinct mutations in 11 different genes) were
 222 fixed in the population after the last time point (**Table 2**). This corresponds to about
 223 a year of passaging in the three conditions: ~450 generations in U937 cells (Mac, t3),
 224 ~480 generations in the switching condition (Alt, t2), and ~800 generations in *A.*
 225 *castellanii* (Acas, t3) (**Supplementary Table 1**).

226

227 **Table 2: Fixed mutations.** Mutations shown are those fixed at the last time point for
 228 each lineage.

229

Gene	Product	Position ¹	Category ²	Pos ³	N fixed ⁴
lpp0833	N-acetylneuraminase synthase family protein	933 165	Small indel	182	8
<i>groES</i>	co-chaperone GroES	822 976	SNP nonsyn	45	7
<i>rpsL</i>	30S ribosomal protein S12	451 837	SNP nonsyn	43	7
		451 972	SNP nonsyn	88	7
<i>rpsD</i>	30S ribosomal protein S4	467 982	SNP nonsyn	205	7
<i>lerC</i>	response regulator LerC	119 826	Small indel	61	4
		119 903	SNP nonsense	46	1
<i>rsmD</i>	16S rRNA (guanine(966)-N(2))-methyltransferase RsmD	3 107 822	Small indel	42	2
lpp0831	hypothetical protein	930 446	Small indel	808	1
lpp0832	phosphotransferase	931 503	SNP nonsense	203	1

lpp0835	hypothetical protein	935 076	SNP nonsyn	228	1
lpp2016/ lpp2017	hypothetical protein/cation diffusion facilitator family transporter	2 283 953	SNP intergenic	162	1
lpp2574	HAMP domain-containing histidine kinase	2 928 346	Small indel	594	1

230

231 ¹ Position along the *L. pneumophila* Paris chromosome (NC_006368).

232 ² SNP: single-nucleotide polymorphism; (non-)syn: (non-)synonymous.

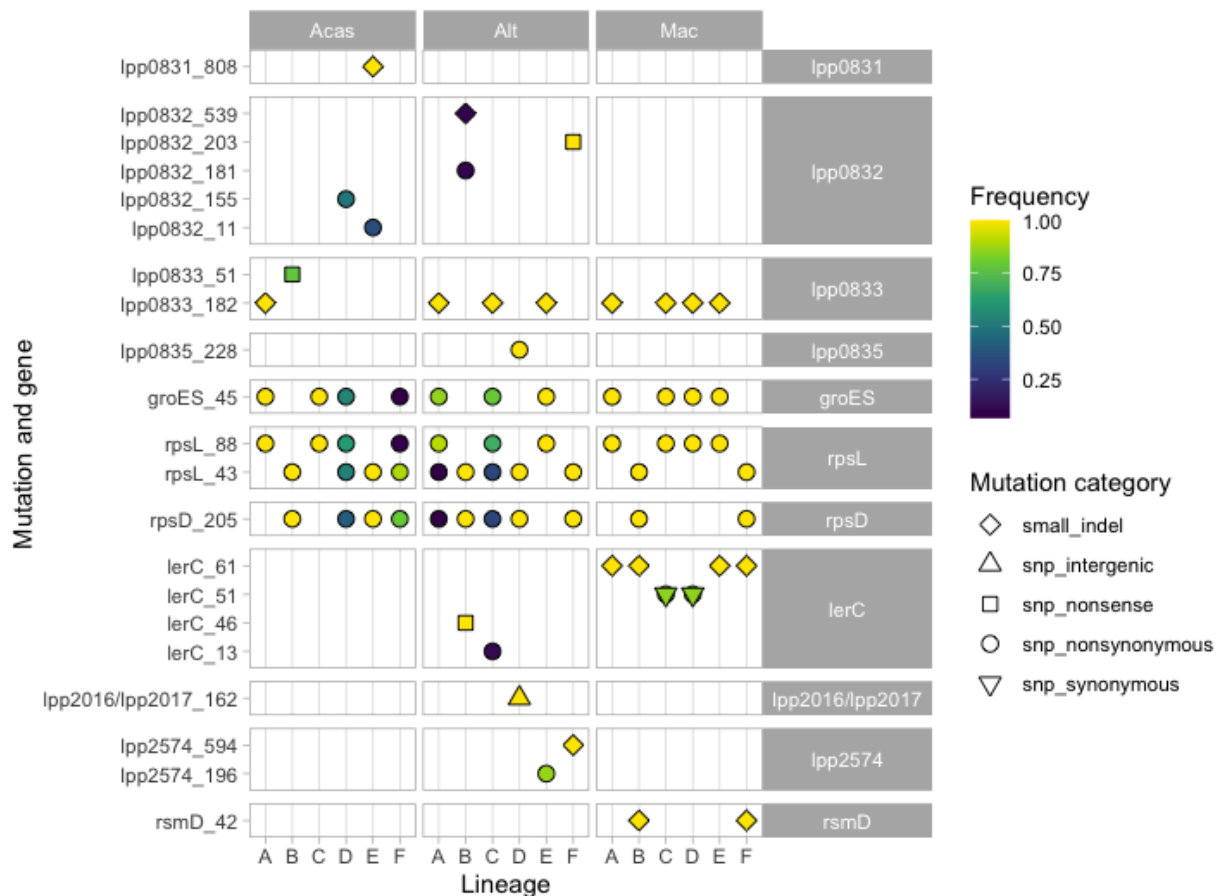
233 ³ Position along the protein sequence, or the nucleotide sequence (for indels and intergenic sequences).

234 ⁴ Number of fixed mutations across populations

235

236 As expected, most fixed mutations were non-synonymous or nonsense mutations,
 237 or small intragenic indels (Figure 2). Two synonymous mutations (in *acpP* and
 238 *nuoLL*, Figure 1) appeared to be fixed in the Mac_B population but were absent at the
 239 third time point. One intergenic mutation was found, 162 respectively 187
 240 nucleotides upstream of two genes, *lpp2016* and *lpp2017* (Figure 2). The latter gene
 241 is, according to in-silico predictions (Price et al. 2005), the first in a two-gene
 242 operon (*lpp2017* and *lpp2018*), both involved in ion transport. The former gene,
 243 *lpp2016*, has no described function, and appears to be strain-specific in *L.*
 244 *pneumophila*: it is present in subspecies *pascullei* and in strains Paris, Alcoy and
 245 Corby but absent in Philadelphia (Altschul et al. 1997).

246



247

248

249 **Figure 2: Genes with fixed mutations at the last sequenced time point under the**
250 **three conditions.** Mutations occurring within a gene are labeled with the locus tag
251 or the gene name (if available) on the right side of the figure; mutations occurring
252 in the same gene at different positions (*rpsL*, *lerC*) are shown on different lines, and
253 the position of the mutation is shown on the left side of the panel. For in-gene SNP
254 mutations, the position of the amino-acid whose codon has been mutated is
255 indicated. For the intergenic mutation, the position is the number of nucleotides
256 after the upstream gene; for the small indels, the position refers to the nucleotide
257 position within the gene for the start of the indel. Mutations occurring in an
258 intergenic space are labeled with the two bordering gene names or locus tags. The
259 shape of the points corresponds to the type of mutation. The *lerC*_51 mutation is a
260 double mutation in the same codon, at the second and third codon position,
261 resulting in a non-synonymous mutation. The color of each point corresponds to
262 the frequency of the mutation. Yellow points are fixed mutations.

263
264 Two pairs of mutations in ribosomal protein and chaperonin genes are mutually
265 exclusive

266 The distribution of mutations showed two distinct patterns of mutations, both of
267 which were present under all three conditions (**Figure 2**). In one (referred to as
268 RpsL43/RpsD), the 30S ribosomal protein S12 RpsL was mutated at position 43,
269 where the lysine is replaced by a threonine (K43T), while the 30S ribosomal protein
270 S4 RpsD is mutated at position 205, where a serine replaces a tyrosine (S205Y). In
271 the other pattern (RpsL88/GroES), RpsL has another mutation, where the lysine at
272 position 88 is replaced by an arginine (K88R), while the co-chaperone GroES has a
273 mutation at position 45 (A45T). Every population harbored one or the other pair of
274 mutations, or a mix of both, at the last time point. Strikingly, these two pairs of
275 mutations were never seen together on a single isolate. In the four cases where both
276 patterns were present (*Acas_D*, *Acas_F*, *Alt_A*, and *Alt_C*; **Figure 2**), the
277 frequencies of the two mutations within the pattern were very similar, and the sum
278 of the two patterns roughly equals one. An inspection of the reads spanning both
279 positions in the *rpsL* gene revealed that no single read displayed both mutations, but
280 that all reads carried one.

281
282 The two mutations in RpsL (K43T and K88R) have previously been shown to provide
283 resistance to streptomycin in *L. pneumophila* (Rao et al. 2013). We determined the
284 minimal inhibitory concentration to be over 1024 µg/ml in both cases. For both
285 mutations, a halo zone was observed on plates (**Supplementary Figure 5**) but the
286 halo was larger in the K88R-carrying mutant (down to ~96 µg/ml) than in the
287 K43T-carrying mutation (down to ~768 µg/ml).

288

289 Mutations in the LPS synthesis gene cluster

290 Four linked genes were harboring fixed mutations: lpp0831 (a small indel), lpp0832
291 (one fixed nonsense mutation and four non-fixed mutations: one small indel and
292 three non-synonymous mutations), lpp0833 (one fixed small indel in many
293 populations, as well as a non-fixed nonsense mutation in Acas_B), and lpp0835
294 (one non-synonymous mutation) (**Figure 2**). These four genes belong to the variable
295 segment of the lipopolysaccharide (LPS) synthesis gene cluster. The mutation in
296 lpp0833 was fixed in 8 out of 18 lineages, and only found in populations which
297 harbored the RpsL88/GroES genotype. In two of these lineages (Alt_A and Alt_D),
298 though, the lpp0833 mutation was found in 100% of the population whereas the two
299 genotypes RpsL88/GroES and RpsL43/RpsD coexisted. Thus, the lpp0833 was found
300 also in the RpsL43/RpsD background in these two lineages. The other mutations in
301 the LPS synthesis operon were found exclusively in the lineages passaged in Acas or
302 during switching conditions.

303

304 Mutations in LerC, a response regulator of the effector regulatory network

305 Four distinct mutations in LerC were found at the last time point: a A13D non-
306 synonymous change (lerC_13), a 13-nt insertion after nucleotide position 61 in the
307 gene (codon position 21; lerC_61), a C46* nonsense mutation (lerC_46), and a
308 double SNP, resulting in a non-synonymous V51G mutation (lerC_51). All Mac
309 lineages harbored one LerC mutation, and it was fixed in four of them; two of the Alt
310 lineages harbored one (B and C, fixed in B), but none was found in the Acas lineages
311 (**Figure 1, 2**). All these mutations were absent at the first time point.

312

313 The two non-synonymous lerC_13 and lerC_51 mutations seemed to have little
314 effect on the structure of the LerC protein (**Supplementary Figure 6A and D**). The
315 lerC_46 and lerC_61 mutations yielded a short peptide at the N-terminal (45 and
316 20 residues, respectively). However, longer proteins (88 and 122 residues,
317 respectively) can presumably be translated from alternative start codons, to the end
318 of the protein. The lerC_61 mutant protein lacks the first two β -strands and the first
319 α -helix (**Supplementary Figure 6B**), while the lerC_46 mutant protein lacks an
320 additional α -helix and the conserved aspartic acid residue at position 53
321 (**Supplementary Figure 6D**), which is essential for the phosphorylation and function
322 of LerC (Feldheim et al. 2018).

323

324 Other mutations

325 Three other regions harbored fixed mutations at the last time point (**Table 2; Figure**
326 **2**). The first one was an intergenic region, upstream of both lpp2016 and lpp2017, in
327 a location where the mutation could play a role in the regulation of either or both
328 genes. The second was the gene lpp2574, which encodes a HAMP domain-
329 containing histidine kinase. Two different mutations were found in this gene, a 1-nt
330 insertion in the middle of the gene (lpp2574_594), fixed in Alt_F, and a non-

331 synonymous R196S (lpp2574_196), present in 86% of the Alt_E population. The
332 third one was the gene coding for the rRNA methyltransferase RsmD ([Figure 3](#)),
333 which methylates the guanine 66 of the 16S ribosomal RNA during ribosome
334 assembly. The effect of this mutation (rsmD_42) is unclear but is likely to result in
335 a loss-of-function mutation, as the mutation consists in a 1-nt deletion early in the
336 gene.

337

338 Fitness effects

339 To estimate the fitness effects of the mutations occurring in the experiment, clones
340 were isolated from populations and whole-genome sequenced. Their growth was
341 compared to their ancestors' growth, in both host cell types (U937 cells or *A.*
342 *castellanii*). Since the two fluorescent markers (SYFP2 and dTomato) differ in many
343 aspects (intrinsic effect on organism fitness, brightness, folding time, half-life,
344 etc.), it is not possible to infer fitness differences between two clones carrying
345 different fluorescent markers directly from competition experiments. To correct for
346 this, a parallel experiment was run, competing the two ancestors, isogenic except
347 for the fluorescent marker. The fitness of the evolved clone, competed against the
348 ancestor with the other marker, can be compared to the fitness of the ancestor with
349 the same marker, competed against the ancestor with the other marker. Two
350 methods to estimate fitness were evaluated: the first one assessed growth rates by
351 fitting splines to the growth curves, the second one estimated total growth by
352 comparing the fluorescence in stationary phase to the fluorescence early in the
353 experiment. The first method proved difficult to reproduce consistently, and the
354 second method, which calculates a Competition Index (CI), similar to the method
355 used by Ensminger et al. (2012), was preferred.

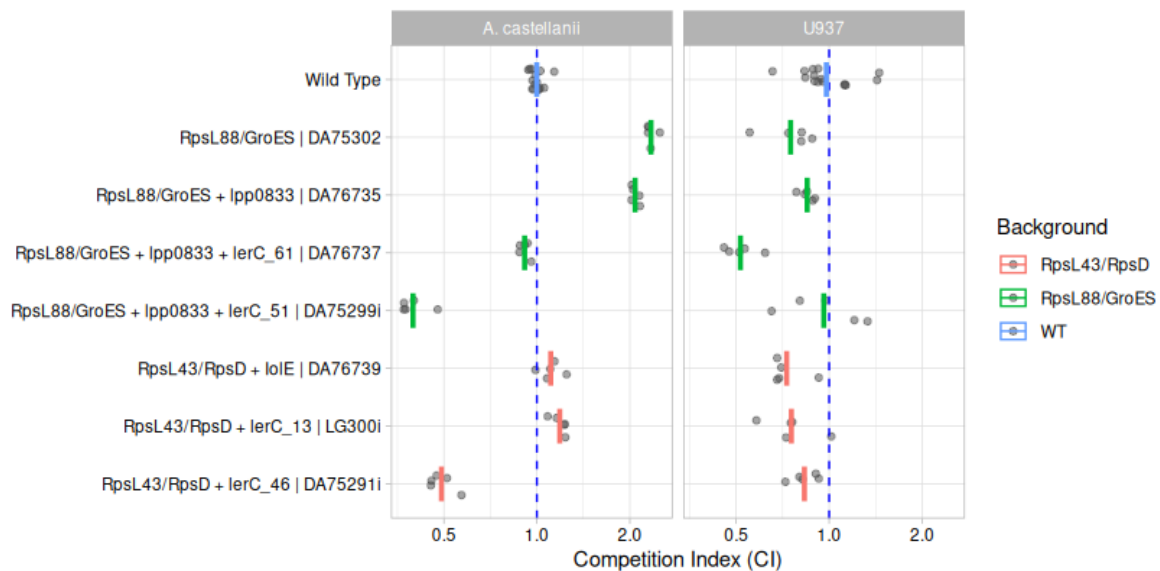
356

357 CI were calculated between the beginning of incubation (0h) and 64h of infection.
358 This timeframe represented the beginning and end of infections in our experimental
359 setup (see [Supplementary Figure 7](#)). Since a MOI of 10 was used, representing 10
360 bacterial CFUs for 1 host cell, it allowed for only one replication cycle. A CI of 1 is
361 equivalent to a neutral effect in terms of fitness between the evolved and the
362 ancestral strain. CI measures the total production of cells during one cycle of
363 replication of *Legionella* within its host. It captures the growth rate only indirectly,
364 and it does not capture differences in how bacteria exit the host cell. The results of
365 competition experiments are shown in [Figure 3](#).

366

367 The gene *rpsL*, encoding the 30S ribosomal protein S12, was mutated in all
368 populations, conferring streptomycin resistance to *L. pneumophila*. Two distinct
369 mutations (at position 43 and 88) were identified, and each was accompanied by
370 another mutation. The RpsL43/RpsD double mutation is found in 11 out of 18
371 lineages, and is fixed in 7 out of them. The other double mutation conferring
372 streptomycin resistance, RpsL88/GroES, is equally widespread, found in 11 out of 18
373 lineages, and fixed in 7. One isolate with the second genotype (RpsL88/GroES;

374 DA75302) showed a strong advantage in competitions in amoebae, but a slight
375 disadvantage in macrophages (CIs of 2.34 and 0.76, respectively) (**Figure 3A**). None
376 of our isolates had only the RpsL43/RpsD genotype, but one (DA76739) had these
377 two mutations and an extra G241V non-synonymous mutation in the IolE protein,
378 involved in myo-inositol metabolism. This isolate showed a similar trend, but not
379 as pronounced, as the RpsL88/GroES mutant: it grew slightly better than the wild
380 type in amoebae but worse in macrophages (CI values of 1.11 and 0.73, respectively)
381 (**Figure 3B**).
382



383
384 **Figure 3: Competition index (CI) of different isolated strains evolved during the**
385 **evolution experiment.** Evolved strains of *L. pneumophila* Paris were competed in the
386 U937-like macrophage cells and the amoeba *A. castellanii* against the ancestral
387 strain. Each replicate is represented by a grey dot and the average is shown as a red
388 bar. The blue dashed line is set at CI=1 (evolved clones grow as much as the
389 ancestors).
390

391 Another gene with several independent mutations was *lerC*, which encodes a single-
392 domain response regulator, involved in modulating the expression of the protein
393 effectors translocated into the host cytoplasm (Feldheim et al. 2018). Four isolates
394 contained mutations in their *lerC* gene: DA75291i with a nonsense mutation
395 (*lerC_46*, C46*); DA75299i with a non-synonymous double mutation (*lerC_51*,
396 V51G); LG300i with another non-synonymous mutation (*lerC_13*, A13D), and
397 DA76737 with a 13-bp deletion (*lerC_61*). The results of the competition with the
398 ancestors are shown in **Figure 3B**, while the predicted structure of the mutants is
399 shown in **Supplementary Figure 6**. All these isolates also contained streptomycin
400 resistance-conferring mutations: DA75291i (*lerC_46*, C46*) and LG300i (*lerC_13*,
401 A13D) also harbor the double mutation RpsL43/RpsD, while DA75299i (*lerC_51*,
402 V51G) and DA76737 (*lerC_61*, 13-bp deletion) harbor the other double mutation
403 RpsL88/GroES. The latter two isolates, DA75299i and DA76737, also harbor an extra

404 mutation in *lpp0833*, involved in LPS synthesis. The effects of the mutations are
405 thus compounded in the fitness experiments.

406
407 In *A. castellanii* cells, the RpsL88/GroES mutations provided the largest fitness gain
408 (DA75302; CI = 2.34). Adding a mutation in the *lpp0833* (involved in LPS synthesis),
409 the fitness was larger than for the wild type, but slightly lower than the latter mutant
410 (DA75735; CI = 2.08). The further addition of the 13-bp deletion in LerC (*lerC_61*)
411 results in a fitness similar to the wild type (DA75737; CI = 0.91) while the addition of
412 the non-synonymous *lerC_51* mutation resulted in a very large fitness drop
413 (DA75299i; CI = 0.40), suggesting that the net effect of both LerC mutations is a
414 large fitness loss. On the other hand, the RpsL43/RpsD (accompanied by the IolE
415 mutation) yielded a very slight fitness gain (DA76739; CI = 1.11). Removing the IolE
416 mutation but adding the non-synonymous *lerC_13* mutation resulted in a very
417 similar fitness (LG300i; CI = 1.19), while the addition of the nonsense *lerC_46*
418 mutation resulted in a large fitness drop (DA75291i; CI = 0.49) ([Figure 3](#)).
419 Unexpectedly, although mutations in the LPS synthesis cluster were widespread in
420 populations exposed to *A. castellanii*, the largest fitness increase seem to be provided
421 by the RpsL88/GroES mutations.

422
423 In macrophages, the mutations had in general a weaker effect, and most tested
424 mutations resulted in a slight fitness drop. The CI values were on average between
425 0.75 and 1 (with one exception), and showed a greater variance than in amoebae,
426 making comparisons less reliable. One isolate, DA76737, which has the four
427 mutations (in RpsL88/GroES, in *lpp0833*, and in *lerC_61*) that are fixed in two
428 lineages (Mac_A and Mac_E) had a markedly lower fitness compared to the
429 ancestral type, with an average CI of 0.52. This value can be compared to the CI of
430 DA76735, which has the same mutations but lacks the non-synonymous *lerC_61*
431 mutation (CI = 0.85). Despite the ubiquity of mutated LerC sequences in populations
432 exposed to macrophages, no LerC mutants seemed to show an increased fitness in
433 macrophages, at least in the experimental setting used here.

434
435 Mutation rates and diversity of streptomycin resistance-conferring mutations

436 To investigate the likelihood that the same mutations (in particular the ones
437 conferring resistance to streptomycin) occur many times in parallel, we measured
438 mutation rates to streptomycin using fluctuation assays for the three ancestral
439 strains (*L. pneumophila* Paris, and both SYFP2- and dTomato-marked descendants).
440 The results indicated that the mutation rates were low for all three strains (see
441 [Supplementary Table 2](#)), ranging from 5.5×10^{-10} to 7.6×10^{-10} . Considering an
442 average mutation rate of 6.5×10^{-10} and that the bacterial population size is 5×10^5
443 and 5×10^6 for the experiments with *A. castellanii* and U937 cells, respectively, there
444 is a 3.3×10^{-4} and 3.3×10^{-3} probability that a mutation conferring resistance to
445 streptomycin will appear in one cell at each generation. At t_1 , which corresponds to
446 128 and 113 generations ([Supplementary Table 1](#)), the probability that streptomycin

447 resistance would have occurred at least once in any given Acas and Mac population
448 reaches 4% and 30%, respectively. At t₃ (798 and 444 generations) the probabilities
449 were 23% and 76%, respectively. However, the streptomycin resistance was present
450 in half the Acas and 5 out of 6 Mac populations at t₁, and in all populations at t₃,
451 exceeding the calculated likelihoods, indicating that streptomycin resistance might
452 have introduced by contamination.

453
454 To investigate whether the only two RpsL mutations we obtained in our evolution
455 experiment (K43T and K88R) were representative of all streptomycin-resistance
456 conferring mutations, we sequenced the *rpsL* gene of 40 streptomycin-resistant
457 isolates obtained from the fluctuation assays. Although six other mutations were
458 identified, K43T (11 mutations) and K88R (21 mutations) accounted together for
459 80% of the mutations found in these isolates (n = 40) ([Supplementary Figure 8](#)).

460
461

462 Discussion

463 In this study, we experimentally evolved *Legionella pneumophila* in presence of one
464 of its natural hosts, the amoeba *Acanthamoeba castellanii*, and of the human
465 macrophage-derived cell line U937, to identify host-specific genetic adaptations.
466 We passaged *L. pneumophila* separately in each host and switching between them,
467 twice a week for a year. In the populations exposed to the amoebal host, we found
468 that the LPS-synthesis cluster was probably under selection whereas in the
469 populations exposed to the human macrophage-derived cell lines, LerC, a protein
470 playing a major role in regulating the temporal expression of protein effectors,
471 harbored potential adaptive mutations. We also found that most mutations occur in
472 mutational hotspots, presumably as a result of the high recombination rates in *L.*
473 *pneumophila*.

474

475 Host-specific adaptations in the LPS-synthesis cluster and in the LerC response
476 regulator

477 Most *L. pneumophila* populations exposed to *A. castellanii* displayed mutations in
478 genes of the LPS synthesis cluster. The *Legionella* LPS synthesis cluster contains ca.
479 30 genes (lpg0779-lpg0746 in the Philadelphia strain; lpp0812-lpp843 in Paris)
480 (Thürmer et al.). The genes that were mutated in our evolution experiment (lpp0831,
481 lpp0832, lpp0833 and lpp0835) correspond to ORF 11, 10, 9 and 7, respectively, and
482 belong to the serogroup 1 (SG1)-specific region of the LPS gene cluster (Mérault et
483 al. 2011; Petzold et al. 2013). Whereas lpp0831, lpp0832 and lpp0833 have homologs
484 in other serogroups (albeit with low similarity), lpp0835 is found only in SG1, and
485 has been proposed as a marker for that group (Mérault et al. 2011). The function of
486 the heavily mutated lpp0833 (lpg0768) is not fully understood, but it is homologous
487 to a sialic acid synthetase function (neuAc, neuB family), and is presumably involved
488 in late modifications of the O-antigen (Petzold et al. 2013). The 1-nt deletion in

489 lpp0833 (lpp0833_182) results in an out-of-frame mutation after 182 nucleotides
490 (of >1000), resulting in a very truncated protein. Although functional studies would
491 be required to firmly establish the effect of the truncation, it is likely that it results
492 in a non-functional protein.

493
494 The fact that 8 distinct mutations were found in the LPS operon, affecting only the
495 populations in contact with amoebae, strongly suggest that these are adaptive, and
496 beneficial for *Legionella* when infecting the *A. castellanii* strain used in this
497 experiment. One isolate with the 1-nt deletion (lpp0833_182) indeed showed a very
498 strong advantage in competitions inside *A. castellanii*. However, this isolate also
499 carries the RpsL88/GroES genotype, which itself provides an important advantage.
500 It is thus difficult to separate the effect of the two genotypes. Further research in the
501 function of these genes in different hosts might uncover the relationships between
502 LPS and host-specificity, but a reasonable hypothesis is that the LPS obtained in the
503 evolved *L. pneumophila* clones is different to the one in the wild-type. The modified
504 LPS thus potentially lowers the immune response of *A. castellanii*, providing an
505 advantage for this particular *Legionella* strain.

506
507 The response regulator LerC is another gene carrying multiple distinct mutations,
508 (namely four), two of which are fixed in 5 populations after the last time point. LerC
509 is involved in the network of regulation of effector proteins (Feldheim et al. 2018).
510 Primary regulation of the expression happens mainly through four distinct two-
511 component systems (TCS), which together regulate over 100 effectors (Segal 2014).
512 Among these, PmrAB regulates directly over 40 effectors (Al-Khodor et al. 2009),
513 while LetAS regulates another set of effectors, via the small RNAs RsmY and RsmZ
514 (Sahr et al. 2009). LerC was found to be a connector between these two systems: it is
515 activated by PmrAB, and inhibits LetAS (Feldheim et al. 2018). The repression of
516 LetAS seems to be growth-phase dependent, as PmrAB is activated under the
517 exponential growth phase, favoring cell divisions inside the *Legionella*-containing
518 vacuole (LCV). At some point, late in the exponential phase, the activation of LerC by
519 PmrAB decreases, which in turn relieves the repression on LetAS, which triggers the
520 cells to enter stationary phase (Feldheim et al. 2018).

521
522 In the study describing the function of LerC (Feldheim et al. 2018), the authors
523 noticed that knock-out mutants of *lerC* resulted in various colony types where one
524 of the types took over after a few passages. Sequencing the new dominant type
525 revealed, in addition to the *lerC* mutation, mutations in either *letA* or *letS* that
526 resulted in non-functional versions of the proteins. Although we found several non-
527 functional versions of LerC in our experiment, *letA* and *letS* were not mutated in any
528 of our strains. However, *lerC* mutants were grown on solid medium by Feldheim et
529 al., while our lineages were grown in presence of their hosts only. This indicates that
530 repression of LetAS might be important for growth on solid medium. It should also
531 be noted that Feldheim et al. used *L. pneumophila* JR32, derived from Philadelphia-

532 1, which are separated by ca. 70 000 SNPs from the Paris strain used here (David et
533 al. 2016; Wee et al. 2021): the regulation of effector secretion might be different
534 between the two strains, and could explain why we were able to obtain LerC mutants
535 without the compensating *letAS* mutations.

536
537 Mutations in LerC appeared to be generally deleterious in *A. castellanii*, as evidenced
538 by the absence of any LerC mutant in the populations exposed to this host only. This
539 is further supported by competition experiments: compared to the strain with the
540 same background except LerC, the LerC mutants had a lower CI. One strain with the
541 wild-type LerC (DA76735) had a CI of 2, while two other strains, harboring two
542 different LerC mutations (DA76737 with the *lerC*_61 deletion; DA75299i with the
543 non-synonymous V51G mutation) had much lower CIs, under 1 and under 0.4,
544 respectively. The comparisons for the other two mutants are more complicated to
545 interpret, as there were no isolates with exactly the same background, but the
546 mutant with the *lerC*_46 nonsense mutation (DA75291i) had a CI of 0.49, more than
547 half than the counterpart (DA76739) which had the wild-type LerC (CI = 1.11) but
548 also another mutation in *IolE*. The last *lerC* mutant (LG300i), carrying *lerC*_13, a
549 nonsynonymous mutation (A13D), had actually a slightly higher CI (1.19) than
550 DA76739 (1.11). In summary, both the distribution, diversity and competition results
551 show that mutations affecting LerC are deleterious for the survival and reproduction
552 of *L. pneumophila* in *A. castellanii*.

553
554 On the other hand, the mutations in LerC are likely to be adaptive in human
555 macrophage-derived cells: all six populations passaged in U937 cells only acquired
556 a mutation in the *lerC* gene, which got fixed in four of them. Two populations
557 passaged in alternation in U937 and *A. castellanii* cells showed mutations in LerC, of
558 which one got fixed. The expected advantage in macrophages was only partially
559 confirmed by our competition experiments. The comparison with the wild-type
560 strain is complicated by the fact that the LerC mutants in our collection of isolates
561 all have additional mutations. The most widespread mutation (*lerC*_61) had a lower
562 CI (0.52) than its counterpart without (0.82), which is quite unexpected. Conversely,
563 the isolate with the *lerC*_51 mutation (DA75299i), present in the majority of both
564 the Mac_C and Mac_D populations, had a higher CI than its counterpart without
565 (DA7535). The fitness for the other two LerC isolates (*lerC*_46, and *lerC*_13; also
566 harboring the RpsL43/RpsD mutations) is again difficult to evaluate, as there are no
567 isolates with only the RpsL43/RpsD mutations. However, compared to the
568 RpsL43/RpsD isolate that also has an *IolE* mutation, the *lerC*_46 and *lerC*_13
569 isolates have a slightly higher CI. In summary, the pervasive presence of LerC
570 mutants in the populations passaged in human macrophage-derived cells strongly
571 suggests that the wild-type LerC protein is deleterious to *L. pneumophila* Paris when
572 infecting these cells.

573

574 In summary, although the results of the competition experiments are somewhat
575 inconclusive, the pervasive presence of LerC mutations in populations exposed to
576 macrophages-like cells but not in populations exposed to *A. castellanii* only,
577 suggests that most LerC mutations are deleterious for growth in *A. castellanii*, while
578 they are likely advantageous for growth in human macrophage-like cells. The
579 variety of mutations found in the lineages exposed to macrophage-like cells,
580 including two that significantly shorten the protein, suggests that a loss or
581 reduction of the LerC function would be beneficial in the human host. The fact that
582 the LerC mutants are also found in lineages where *Legionella* was alternated between
583 the two hosts suggests that the positive selective value in macrophage-like cells is
584 large enough that it overcomes the negative selective values in *A. castellanii*. Thus,
585 LerC is a good candidate for a gene that confers a host-specific advantage, in this
586 case in human cells.

587

588 Adaptation to experimental growth conditions

589 In addition to the potentially host-specific adaptations, we identified groups of
590 mutations involved in adaptation to the specific growth conditions of our
591 experiment. Among them are two mutations in the ribosomal protein S12 RpsL
592 (K43T and K88R), known to provide resistance to streptomycin (Funatsu and
593 Wittmann 1972). Spontaneous resistance to streptomycin resistance in *Legionella* is
594 a common phenomenon, e.g. in *L. dumoffii* (Hubber et al. 2017) or in many laboratory
595 descendants of the *L. pneumophila* Philadelphia (Berger and Isberg 1993; Rao et al.
596 2013). Single-step resistance to streptomycin through mutation of *rpsL* has been
597 documented, among others, in *Escherichia coli* (Funatsu and Wittmann 1972) and in
598 *Mycobacterium tuberculosis* (Nair et al. 1993). In both organisms, mutation of codon
599 43 (42 in *E. coli*) provides resistance to streptomycin, while mutations in codon 88
600 (87) is found mostly in *E. coli* (Timms et al. 1992). Both mutations have been
601 identified in spontaneous streptomycin mutants in *L. pneumophila* (Rao et al. 2013).
602 Our *Legionella* lineages were never directly exposed to streptomycin, but early in the
603 experiment, both host cell lines were grown in the presence of penicillin-
604 streptomycin, to avoid contaminations. The host cells were thoroughly washed
605 before being infected with the *Legionella* lineages, precisely to prevent exposing
606 *Legionella* to the antibiotic. A possible explanation is that, despite washing,
607 streptomycin could accumulate in the host cytoplasm, where it gets in contact with
608 bacteria and increase the selective pressure for streptomycin resistance. Although it
609 is unlikely that the concentrations inside the washed host cells reached the
610 minimum inhibitory concentration (MIC), sub-MIC levels of antibiotic have been
611 shown to select for resistant mutants (Gullberg et al. 2011).

612

613 Each of the two RpsL mutations described above are almost in all cases accompanied
614 by another mutation. RpsL43 is paired with a mutation in the ribosomal protein S4,
615 RpsD, while RpsL88 is paired with a mutation in the co-chaperonin GroES. The
616 specific mutation found in RpsD is a known compensatory mutation, which, in *E.*

617 *coli*, restores an efficient translation to cells which also harbor the K43T mutation in
618 RpsL (Björkman et al. 1999). The specific role of the co-chaperonin GroES (also
619 referred to as HtpA in *Legionella*) in *Legionella* is not fully understood, but
620 chaperonins in *Legionella* have been shown to play very different roles in the
621 different phases of *Legionella* lifestyles. The chaperonin itself (GroEL/HtpB) appears
622 to be exposed on the surface of the bacterial cell during infection and can trigger
623 phagocytosis in non-phagocytosing cells, while it can redirect vesicular and
624 organelle trafficking in the host cytoplasm when inside cells (Garduno et al. 2011).
625 The function of the co-chaperonin GroES is less understood, but it is reasonable to
626 assume that it co-translocates with GroEL. That specific mutation might also simply
627 be compensatory, e.g. by helping to fold a potentially destabilized mutant RpsL
628 protein.

629
630 RpsL, besides its function as a ribosomal protein, might trigger macrophage death,
631 suggesting an adaptive role in host infection. Through pattern recognition receptors
632 (PRRs), mammalian cells recognize bacterial invaders, and trigger both cellular
633 apoptosis and inflammatory response (Akira et al. 2006). Many PRRs have known
634 ligands (e.g. flagellin, needle of the type III secretion system) but some are still
635 evasive. Experiments with mouse bone marrow-derived macrophages (BMDMs),
636 Zhu et al. (2015) suggests that RpsL is responsible for macrophage death caused by
637 *L. pneumophila*. Strikingly, in that study, only one of the two RpsL mutations found
638 in our experiment (K88R) did not trigger cell death, allowing *Legionella* to replicate.
639 Although we do not know how conserved this interaction is and whether human cells
640 or amoebae possess a similar mechanism, it is interesting to note that 4 out of 6 of
641 the Mac lineages display a fixed RpsL K88R mutation, while it is fixed in 2 Acas and
642 1 Alt lineages (other Acas lineages have two subpopulations with each one RpsL
643 mutation). This suggests that, even in human macrophages, the K88R RpsL mutant
644 might be less prone to trigger cell death. However, competition experiments do not
645 seem to corroborate this observation: the RspL88/GroES genotype provides an
646 important advantage in *A. castellanii* cells, but a slight disadvantage in U937 cells
647 (Figure 3).

648
649 The RpsL43/RpsD mutation occurred twice in the lineages exposed to macrophages.
650 We could not test the fitness of that genotype alone, since all the isolates we
651 collected had one or more extra mutations. An isolate with an extra IolE mutation
652 showed an increased fitness in *A. castellanii* cells but a decreased one in macrophages
653 (Figure 3), a pattern similar to the RpsL88/GroES genotype. It is counterintuitive
654 that a mutation with a fitness cost (in macrophages) would be fixed in all
655 populations. A potential explanation is that the *Legionella* populations might have
656 been exposed to streptomycin during the evolution experiment, setting a strong
657 evolutionary pressure, but streptomycin (and thus the pressure) was not present at
658 all during the competition experiment. These discordant results might be the result
659 of the RpsL mutations being always beneficial in presence of (even traces of)

660 streptomycin, and being beneficial in amoebae but detrimental in macrophages in
661 the absence of streptomycin.

662

663 An intergenic mutation in-between an ion-transport operon and a strain-specific
664 gene

665 One single intergenic mutation was found to be fixed, upstream of two genes,
666 lpp2016 and lpp2017 (**Figure 1**). The former gene, lpp2016, has no described
667 function, and appears to be strain-specific in *L. pneumophila*, while the latter is
668 involved in ion transport. From sequence analysis only it is not possible to
669 determine whether this mutation would yield a selective advantage. However, it is
670 possible that this mutation interferes with the regulation of one or both of the
671 downstream genes. It is also possible that this intergenic space, unusually large (ca.
672 350 nt) might encode a small, non-coding RNA (ncRNA), which might have
673 regulatory or even virulence functions. To the best of our knowledge, this specific
674 intergenic region has not been shown to contain a ncRNA (Sahr et al. 2012). Recently,
675 *L. pneumophila* has been shown to express and translocate sRNAs into its eukaryotic
676 host, where it regulates the host innate immune response (Sahr et al. 2022).

677

678 Recombination hotspots account for a large part of the mutations

679 *Legionella* genomes are known to experience extensive homologous recombination
680 (David et al. 2017). Early studies estimated that the recombination (r) and mutation
681 (m) rates should be similar ($r/m = 0.9$) (Coscollá and González-Candelas 2007; Vos
682 and Didelot 2008). However, later estimates based on larger datasets suggest a much
683 higher rate of recombination, with the vast majority of SNPs (96–99%, depending
684 on the groups) arising from recombination (David et al. 2017). In our study, we
685 identified five groups of homologous regions which harbor clusters of mutations.
686 These regions are likely to represent hotspots of intra-genomic recombinations:
687 eleven homing endonucleases, an ankyrin repeat-containing protein (lpp1100), one
688 tetratricopeptide repeat (TPR) protein, a repeated protein containing a domain of
689 unknown function (DUF1566), and three occurrences of a larger segment containing
690 a transposase and a restriction endonuclease type II-like (DUF559). It should be
691 noted that our experiment obviously excludes inter-genomic recombination, and
692 that intra-genomic recombination among identical chromosomes yield no
693 mutations. Only recombination between similar (but not identical) regions yields
694 mutations, leading to an underestimation of the recombination rates.

695

696 The first group of recombination hotspots consists of homing endonucleases -
697 found in all domains of life and in viruses - which are known to cause gene
698 conversions between mobile genetic elements (Dunin-Horkawicz et al. 2006). This
699 group of genes has been found to be enriched for in clinical sporadic strains (when
700 compared to environmental strains) in *L. pneumophila* sequence type (ST) 1,
701 suggesting that they contribute to the genetic diversity of ST1, the most prevalent

702 ST (Mercante et al. 2018). The second group is an ankyrin repeat-containing protein
703 (ANK). ANK motifs are typically eukaryotic motifs, and are found on many of the
704 *Legionella* effectors (Habyarimana et al. 2008). This specific protein is not
705 widespread in *Legionellales*, and is absent from most *L. pneumophila* strains, as well
706 as from *L. longbeachae* (Gomez-Valero et al. 2011). It has been shown to be highly
707 upregulated in transmission phase (Brüggemann et al. 2006; Jules and Buchrieser
708 2007). A distant homolog is however found in *L. drancourtii* LLAP12 ([Supplementary](#)
709 [Figure 9](#)). TPR motifs, the third group of recombination hotspots, are known to be
710 involved in protein-protein interactions in both eukaryotes and bacteria, and have
711 been shown to be involved in virulence in *L. pneumophila* (Bandyopadhyay et al.
712 2012). The fourth hotspot is a non-repeated protein containing a domain of
713 unknown function (DUF1566) which is widely present in Gammaproteobacteria.
714 Finally, the fifth group of hotspots is constituted of three occurrences of a larger
715 segment containing a transposase and a restriction endonuclease type II-like
716 (DUF559). The first and fifth groups were recognized as repeated regions (David et
717 al. 2017), and excluded from their recombination analysis. As for the other single-
718 protein hotspots, they were not deemed repeated but were not detected as
719 recombination hotspots.

720
721 About half (49.4%) of the mutations identified in our study occur in these five
722 groups of mutational hotspots, but none of them got fixed. The majority of them
723 occurred only at low frequency, and only mutations in three of the hotspot groups
724 (lpp1100, GIY-YIG homing endonucleases, transposase/DUF559) reached 25% in
725 frequency. However, the large number of intra-genomic recombination events that
726 can be detected probably provides variation in the genome architecture and
727 potentially contributes to the adaptability of *Legionella* genome.

728
729 The mutational patterns identified in this study follow the general principles of
730 molecular evolution, for example displaying about twice as many transitions than
731 transversions. Single-nucleotide polymorphisms are generally lowering the general
732 G+C content of bacterial genomes (e.g. Hershberg and Petrov 2010; Ely 2021), while
733 higher G+C content is generally believed to occur through GC-biased gene
734 conversion (gBGC) (Lassalle et al. 2015). However, in our study, the AT to GC
735 mutations unexpectedly accounted for almost half (49.4%) of the mutations
736 altering the GC content, even after removing mutations occurring in recombination
737 hotspots.

738
739 The pLPP plasmid found in *L. pneumophila* strain Paris (Cazalet et al. 2004) is
740 sparsely found in ST1-strains (Durieux et al. 2019). Strikingly, although the plasmid
741 is not essential, it has been retained in all of our lineages. This may be due to the
742 presence of a toxin/antitoxin (TA) system on the pLPP plasmid: at least one PHD-
743 RelE type TA system on the pLPP plasmid, according to the TADB (Shao et al. 2011)
744 and TASmania (Akarsu et al. 2019).

745

746 Fitness of evolved strains

747 A puzzling observation in the fitness measurements performed is the discordance
748 between our finding in the serial passage experiments of parallel and fixed
749 mutations, which indicates a strong selective advantage, versus the lack of an
750 advantage of the evolved clones (observed $CI < 1$) in competition experiments with
751 the ancestral strains. At present we do not understand the underlying reason(s) for
752 this discordance, but it should be noted that measuring fitness is a complex matter
753 and in an experiment like the one described here, selection is acting at different
754 levels. Compared to experiments in pure culture like the LTEE (Lenski 2017) where
755 the maximum growth rate in exponential phase is paramount, other aspects of the
756 life cycle of *Legionella* play a role. The time necessary to find a host, the ability to
757 trigger endocytosis, the ability to resist digestion and the efficiency to recruit
758 nutrients to the *Legionella*-containing vacuoles, the time required to exit lag phase
759 and enter exponential phase, the efficiency to exit the host cells, are important in
760 how many of each genotype make it to the next passage. Thus, the CI method used
761 here is probably too crude to capture all these aspects, which might explain the
762 sometimes-discordant results presented above. It is also possible that the fitness
763 changes are non-transitive, and that competition experiments of the evolved
764 strains versus the ancestral strain are not properly capturing such dynamics.
765 Additional measures of fitness might help better understand the phenotypic effects
766 of the mutations.

767

768 Comparisons with other studies

769 Other studies have identified mutations specifically occurring within-host. In the
770 first and - to our knowledge - only published evolution experiment with *L.*
771 *pneumophila*, Ensminger *et al.* (2012) passage *L. pneumophila* in mouse macrophages
772 for a year. This study identified, among others, mutations in the lysine synthesis
773 pathway and in the flagellar regulation gene *fleN*. Surprisingly, there was no overlap
774 between the mutations found in this study and ours. This might however also reflect
775 the differences between the hosts and *Legionella* strains used in the two studies:
776 while we used human cell lines and *A. castellanii* as hosts and *L. pneumophila* Paris,
777 the other study cycled *L. pneumophila* LP01, derived from the Philadelphia strain,
778 cycled in mouse-derived macrophages. A comparison with another study
779 identifying mutation occurring in-patient (Leenheer *et al.* 2023) yielded one single
780 match: in this study, the gene *lidB* (*lpp2223/lpg2269*) was found to be mutated twice
781 independently in different outbreaks, while this gene is mutated in 13% of the
782 population *Mac_A* at t2. The protein is an ATP-dependent
783 helicase/deoxyribonuclease, possibly involved in replication. However, the evidence
784 in both studies is too sheer to conclude a possible selective advantage of a mutated
785 version of this protein without further functional studies.

786

787 Early fixed mutations

788 An unexpected finding was the presence of multiple, apparently fixed mutations at
789 the first time point (i.e. after 10–15 passages), some of which subsequently
790 disappeared (**Figure 1**). We first discuss the probabilities of the same mutations
791 appearing independently in various lineages, and then the fact that some of them
792 disappeared, despite being apparently fixed.

793
794 The analysis of mutation rates and population sizes shows that parallel evolution in
795 our setup is not unlikely. The mutation rates to streptomycin (mutations per cell per
796 generation) measured by fluctuation assays on our ancestral strains (*L. pneumophila*
797 Paris and the two fluorescently labeled strains) were in the range of 6.5×10^{-10} . In
798 comparison, mutation rates for bacteria range from 6×10^{-11} to 3×10^{-8} (Lynch et al.
799 2023), putting our isolates in the middle of the range. We calculated the probability
800 of the appearance of a mutation providing resistance to streptomycin, given the
801 number of generations and the mutation rates above, assuming no selection.
802 Considering that streptomycin resistance is provided in most cases by two single
803 RpsL mutations (K88R and K43T), the probability of gaining any given mutation is
804 close to the probability of gaining a streptomycin resistance mutation. For an
805 amoeba-exposed or a human cell-exposed population, the probability was 4% and
806 39% at the first time point (t1), while it reached 23% and 76% at the last time point
807 (t3), respectively. This shows that it is not unlikely that single, identical mutations
808 would appear parallelly in different populations. Given a strong enough selection,
809 these mutations could have been quickly fixed in the population.

810
811 However, although other recurring mutations (e.g. in LerC or in lpp0833) are very
812 likely to have occurred independently in different lines, the two paired mutations
813 (RpsL43/RpsD and RpsL88/GroES), which always occurred together and in the same
814 genetic background (SYFP2 and dTomato, respectively), have a higher likelihood to
815 be the result of a spread of these two genotypes very early in the experiment. We had
816 established very rigorous procedures to minimize the risks of cross contaminations:
817 bacteria were grown in separate flasks and manipulated individually, always in a
818 laminar flow cabinet, using filter tips when pipetting, and applying general sterility
819 precautions when manipulating the strains. These two genotypes have rapidly taken
820 over most of the populations, emphasizing their probable high selective value.
821 Although a possible early contamination across lineages might have caused the
822 spread of the RpsL mutants, we believe that the latter would be the only ones to be
823 spread that way, and thus that the conclusions of the paper would not be affected.

824
825 To explain the fact that a few apparently fixed mutations disappeared at a later time
826 point, we consider three possible explanations: contaminations across lineages,
827 reversions, or not deep-enough sequencing. As mentioned above, we have
828 established very rigorous procedures to minimize the risks of cross contaminations
829 and thus favor the latter two hypotheses, either that the mutations reverted or that

830 the sequencing missed the presence of the original allele. The apparently high
831 selective value of some of the later mutations might explain these reversions or re-
832 emergence of low-frequency wild-type alleles. An analysis of the reads mapping to
833 the genes encoding the fluorescent markers (SYFP2 and dTomato) shows that in the
834 populations harboring mutations considered fixed, and which therefore should
835 contain only the reads mapping to only one of the fluorescent genes, a few reads of
836 the other gene can be found ([Supplementary Figure 10](#), see e.g. Acas_D or Alt_E).
837 This suggests that the population still contains at least low frequencies of both
838 populations, which might explain how apparently fixed mutations might disappear,
839 should a more favorable mutation appear on the genome of the low-frequency
840 genotype. It should be noted, however, that this later hypothesis is not particularly
841 probable, as these favorable mutations are more likely to occur on the high-
842 frequency genomes.

843

844 Conclusion

845 In summary, the exact role, fate and histories of the mutations that occurred in the
846 18 lineages of *L. pneumophila* investigated here remains to be established. However,
847 for both hosts, we identified potential adaptive, host-specific mutations. Lineages
848 exposed to the amoebal host *A. castellanii* harbored mutations in the host-specific
849 part of the LPS synthesis cluster. Lineages exposed to the human macrophage-
850 derived cells displayed mutations in the LerC protein, a protein regulating the
851 expression of protein effectors, potentially highlighting the different temporal
852 regulation of protein effector secretion in the human host compared to *L.*
853 *pneumophila*'s natural hosts. These results also show that experimental evolution
854 can reveal bacterial genes specifically involved in the adaptation to certain hosts.

855

856 Material and Methods

857 Bacterial strains, host cell culture and media

858 *Legionella pneumophila* Paris was cultured in charcoal yeast extract (CYE) (1% ACES,
859 1% yeast extract, 0.2% charcoal, 1.5% agar, 0.025% Iron (III) pyrophosphate, 0.04%
860 L-cysteine, pH 6.9) plates or ACES yeast extract (AYE) (1% ACES, 1% yeast extract,
861 0.025% Iron (III) pyrophosphate, 0.04% L-cysteine, pH 6.9) broth at 37°C, unless
862 otherwise stated. When necessary, *Legionella* GVPC selective supplement (Oxoid) or
863 0.5 mM IPTG were also added to the media. The *L. pneumophila* strains used were
864 tagged with a SYFP2 or dTomato fluorescence gene, under an IPTG-inducible
865 promoter. A *L. pneumophila* Paris mutant harboring a deletion in the *dotA* gene was
866 used as a negative control in infection experiments, as this strain is not able to
867 replicate in host cells. The $\Delta dotA$ mutant was kindly gifted to us by Elizabeth Kay.

868 *Acanthamoeba castellanii* strain Neff (ATCC 30010) was cultured in Peptone Yeast
869 Glucose (PYG) medium (ATCC 712 medium) in culture flasks at 30°C. For infections,

870 *A. castellanii* was harvested, pelleted, and resuspended in LoFlo medium
871 (Formedium, Norfolk, UK), which does not support *Legionella* growth.

872 Human monocyte-like U937 cells (ATCC-CRL-3253) were maintained in
873 RPMI1640+GlutaMAX™ (Gibco) supplemented with 10% heat-inactivated fetal
874 bovine serum (FBS) (Gibco) and 1% Penicillin-Streptomycin (Pen-Strep; Gibco), in
875 a 37°C incubator with 5% CO₂. To induce differentiation into macrophage-like cells,
876 the U937 culture was centrifuged at 200 × *g* for 5 min, resuspended in growth
877 medium containing 50 ng ml⁻¹ of phorbol 12-myristate 13-acetate (PMA) and
878 incubated for 48h. The medium was replaced and cells were incubated for a further
879 48h. Before infection, cells were resuspended in RPMI 1640 without phenol red
880 (Gibco) supplemented with 10% heat inactivated FBS and 1% GlutaMAX™ (Gibco).
881 This medium also does not support *Legionella* growth.

882

883 Passaging

884 *L. pneumophila* SYFP2 and dTomato patched on CYE for 48h, were resuspended in
885 dH₂O, the OD₆₀₀ was measured and the cultures were diluted, before being mixed in
886 equal proportion. Host cells were prepared as above, 2 × 10⁶ cells of *A. castellanii* or
887 U937 were seeded in six T25 flasks, and challenged at an MOI of 0.25 (*A. castellanii*)
888 and 2.5 (U937 macrophages). After inoculation, flasks were incubated at 30°C, or
889 37°C for 3 days. At the end of the infection period, flasks were vortexed, and an
890 aliquot of the culture was collected. These aliquots were centrifuged for 5 min at 200
891 × *g* to pellet the host cells. The endpoint concentration of *L. pneumophila* for each
892 infection was estimated by measuring the OD₆₀₀, and calculating the CFU based on a
893 standard curve (OD₆₀₀ vs CFU/ml) made for this purpose. The volume of bacteria
894 needed to infect the hosts at the same MOIs was used to inoculate fresh host
895 cultures. Thus, each new passage of *A. castellanii* was inoculated with 5 × 10⁵ CFUs,
896 whereas the macrophages lineages were inoculated with 5 × 10⁶ CFUs. To confirm
897 there was no contamination of the cultures, 10 µL of the cultures were plated on LB
898 agar and CYE plates. *L. pneumophila* was passaged in this manner every 3-4 days,
899 and at every 5th passage glycerol stocks were made by collecting all the infection
900 supernatant (after pelleting the hosts) and centrifuging at 7000 × *g* for 15 minutes
901 to pellet the bacteria. The resulting pellet was resuspended in AYE with 50%
902 glycerol, and kept at -80°C.

903

904 Measure of growth rates

905 To isolate individual clones, frozen stocks of the selected populations were streaked
906 on CYE+GVPC plates and incubated for 72h. The bacterial lawn was scraped,
907 resuspended in ultrapure water, and serial dilutions were made and plated on
908 CYE+IPTG plates to obtain single colonies. Two colonies of each population were re-
909 streaked, and glycerol stocks were made.

910 Fitness assays of the ancestor and evolved strains were performed both in AYE broth
911 and in hosts. For growth in broth, a 24h patch of each colony was resuspended in

912 AYE, the density (OD₆₀₀) was measured and adjusted to 0.1 (ca. 2×10^8 CFU/ml) with
913 AYE+IPTG. Each culture was aliquoted in triplicate into a black 96 well plate, the
914 growth rate was tracked by measuring the OD₆₀₀, and fluorescence of SYFP2
915 (excitation: 508 nm, emission: 555 nm) or dTomato (excitation: 554 nm, emission:
916 635 nm) every 30 minutes for 24h hours using a Tecan Spark™ equipped with a
917 monochromator. For in-host growth assays, bacteria and hosts were prepared as
918 before, but here the 1×10^5 host cells were seeded per well in a black 96-well plate
919 and challenged at an MOI of 50. The infection media was supplemented with 0.5 mM
920 IPTG. Intracellular replication was tracked by measuring the fluorescence every 30
921 mins, for 72h, as described above.

922

923 Competition assays

924 Competition assays were performed in human monocyte-like cells (U937) and *A.*
925 *castellanii* amoebae using a similar protocol to the one used by Ensminger et al.
926 (2012). For in-amoeba competitions, evolved strains to be assayed were streaked
927 onto CYE plates with IPTG (1 mM) and incubated at 37°C for three days. Biomass was
928 taken from the lawn created by each strain and resuspended into either LoFlo
929 medium (Formedium, Norfolk, UK) with 1mM IPTG (for infections in amoebae) or
930 in RPMI without phenol red and with 1mM of IPTG (for infections in U937 cells).
931 These solutions were diluted to an OD₆₀₀ of 0.50 (around 5×10^8 CFU/ml) with a Tecan
932 Spark™ 10M plate reader. Each tested strain was mixed in equal parts with the
933 ancestral strain (*L. pneumophila* Paris) with the other fluorescent tag: dTomato-
934 tagged ancestral strain with SYFP2-tagged evolved strains, and SYFP2-tagged
935 ancestral strains with dTomato-tagged evolved strain. The solutions were serially
936 diluted down to 1/10th and a volume, depending on host species, was transferred to
937 corresponding wells of 24-well plates containing either *A. castellanii* or U937 cells.

938

939 For competitions in amoebae, 200 µl of the mix of bacteria were incubated with $5 \times$
940 10^5 *A. castellanii* cells in 2 ml of LoFlo medium with 1 mM of IPTG, to reach a
941 multiplicity of infection (MOI) of 10. This MOI was chosen as it exceeds the minimal
942 level of fluorescence detectable by the plate reader at the beginning of the assay. The
943 plates were incubated at 30°C.

944

945 For competitions in macrophages, the U937 differentiated macrophages-like cells
946 were seeded in the wells at 2.5×10^5 cells per well, adding 100 µl of mixed
947 evolved/ancestral bacterial strain solution to reach a MOI of 10. RPMI without
948 phenol red but with 1mM of IPTG was used as medium for the infection of the
949 bacterial inoculum solutions. Finally, the 24-well plates were incubated at 37°C with
950 5% CO₂.

951

952 For both competitions, the ancestral dTomato- and SYFP2-tagged *L. pneumophila*
953 were also competed on each plate, as a reference. All competitions were done in 5
954 replicates, for each pair of strains tested. The following controls were also

955 performed on each plate: growing amoebae or U937 cells only (blank control),
956 ancestral SYFP2-tagged strain and amoeba (SYFP2 positive control), ancestral
957 dTomato-tagged strain and amoeba (dTomato positive control), and $\Delta dotA$ mutant
958 (infection negative control, infection-defective). The $\Delta dotA$ was used as infection-
959 negative control, CFU/ml counts were measured at the beginning and at the end of
960 the competition experiment.

961
962 The fluorescence was measured with the Tecan 10M spark spectrofluorometer at an
963 excitation/emission wavelength of 508/555 nm and 554/635 nm to measure SYFP2
964 and dTomato fluorescence, respectively. Fluorescence was measured at 0 hours, 40
965 hours, and 64 hours post infection. These time points correspond to the beginning
966 of the infection, to the beginning of the exponential phase, and the beginning of the
967 post-exponential phase of growth, respectively, in our experimental setup. The time
968 points were determined by measuring growth curves and optimizing the conditions
969 (see [Supplementary Figure 7](#)). Fluorescence data was used to calculate a
970 competition index (Ensminger et al. 2012). The fluorescence measurements were
971 blanked with the fluorescence values of the amoeba-LoFlo or U937-RPMI control
972 wells. Competition index (CI) is calculated as the ratio of the fluorescence of the
973 mutant (evolved) strain to the fluorescence of the wild type (ancestor) strain at t_1 ,
974 divided by the same ratio at t_0 :

$$975 \text{ CI} = [F(\text{mut})_{t_1}/F(\text{WT})_{t_1}]/[F(\text{mut})_{t_0}/F(\text{WT})_{t_0}]$$

977
978 Where:

979 $F(\text{mut})_{t_1}$ = Fluorescence of mutant at end time point (t_1)

980 $F(\text{WT})_{t_1}$ = Fluorescence of wildtype at end time point (t_1)

981 $F(\text{mut})_{t_0}$ = Fluorescence of mutant at beginning time point (t_0)

982 $F(\text{WT})_{t_0}$ = Fluorescence of wildtype at beginning time point (t_0)

983
984 We normalized the CI data of each strain tested by dividing their CIs by the average
985 CI of the SYFP2-tagged ancestor against the dTomato-tagged ancestor. This
986 allowed us to compare CIs across plates. The competition index reflects the ratio of
987 the number of mutant cells to the number of ancestral cells, produced between t_0
988 and t_1 .

989
990 Fluctuation assays to determine mutation rates

991 Fluctuation assays were carried out according to the recommendations of Lang
992 (2018) and Rosche and Foster (2000). Briefly, the tested strains (*L. pneumophila*
993 Paris WT and the two SYFP2 and dTomato-tagged mutants) were inoculated in AYE
994 broth and diluted to reach an OD_{600} of 0.50, corresponding to ca. 5×10^8 CFU/ml. The
995 solutions were diluted 1/1000, by adding 5 μ l of this solution to 5 ml of fresh AYE,
996 and vortexed. From these, 100 μ l were used to inoculate 2 ml of AYE into 50-ml
997 Falcon tubes. This was repeated 72 times for each strain tested. The tubes were

998 incubated at 37°C for three days. After growth, 150 µl of the solution was used to
999 count CFU/ml numbers using the serial-dilution spot plate method (Wang et al.
1000 2017). The rest of the solution was centrifuged at 6000 × *g* and most of the
1001 supernatant removed. The bacterial pellet was resuspended in the remaining
1002 supernatant and plated on CYE containing 6 µg/ml streptomycin, which is around
1003 8x the minimum inhibitory concentration (MIC). The plates were incubated for 3-4
1004 days and total and resistant colonies were counted. Mutation rates were calculated
1005 using the web-based Fluctuation AnaLysis CalculatOR (FALCOR) (Hall et al. 2009).
1006 The MSS Maximum Likelihood Method was used to calculate mutation rates using
1007 the “Group all data” option.
1008

1009 Identification of *rpsL* mutations in streptomycin-resistant strains

1010 To survey the various mutations contributing to streptomycin resistance, we picked
1011 a total of 40 streptomycin-resistant colonies from the fluctuation assay and
1012 sequenced their *rpsL* gene, known to confer resistance mutations to streptomycin
1013 (Zhu et al. 2015). This was done by amplifying a 300-bp segment of the *rpsL* gene
1014 using the following primers: forward 5’-AAGAAAGCCTCGTGTGGACG-3 and reverse
1015 5’-TCGGTCGTTCACTCCTGAAG-3’. The DreamTAq Green PCR Master mix
1016 (ThermoFisher Scientific) was used to amplify the gene and the manufacturer’s
1017 instructions were used for PCR reactions with an annealing temperature of 59.6°C
1018 and 25 cycles. PCR reactions were checked on 1% agarose gel, and then purified using
1019 the GeneJET DNA Cleanup Micro kit (ThermoScientific™). The purified fragments
1020 were sent for Sanger sequencing to identify mutations.
1021

1022 Streptomycin susceptibility testing

1023 The level of streptomycin resistance of the ancestral strains and evolved isolates was
1024 evaluated using an E-test on CYE+IPTG agar plates. The minimum inhibitory
1025 concentration (MIC) was recorded as the lowest antibiotic concentration at which
1026 the zone of inhibition intersected the E-test strip; for many isolates there was a zone
1027 with haze of growth around the strip, here the MIC was read the same way, but the
1028 MIC at the start at the hazy zone was also recorded ([Supplementary Figure 5](#)).
1029

1030 The selected *Legionella* strains were grown on CYE+GVPC plates for 48-72h; a loop
1031 of bacteria was suspended in dH₂O, and the density was adjusted to OD₆₀₀ = 0.05 (ca.
1032 1 × 10⁸ CFU). Then 1 ml of the *Legionella* suspension was added to 5 ml of 0.5% (soft)
1033 agar solution, the mixture was gently vortexed, to avoid the formation of bubbles,
1034 and poured onto a CYE+IPTG agar plate. The plates were left to solidify for 10-15 min
1035 before applying the streptomycin E-test strip (AB-BIODISK, bioMérieux). The
1036 plates were then incubated at 37°C, and the results were taken at 48h.
1037

1038 Population and clone sequencing

1039 In order to identify the mutations that arose from the evolution experiment in the
1040 different lineages, we sequenced populations at passages 10 and 65 for the *A.*
1041 *castellanii* lineage, passages 10 and 49 for the alternating lineage, and passages 15
1042 and 38 for the macrophage (U937) lineage ([Supplementary Table 1](#)). All replicates
1043 from each lineage were sequenced. All clones involved in competition experiments
1044 were also sequenced, to confirm that other mutations did not occur ([Supplementary](#)
1045 [Table 3](#)).

1046
1047 Briefly, the frozen stocks of the different lineages were plated on CYE agar plates
1048 and incubated at 37°C for 3 days. After incubation, the bacterial growth was scraped
1049 off from the plate and suspended in 5 ml of distilled water. The cultures were
1050 centrifuged at 6000 × *g* for 5 minutes and the supernatant was removed.
1051 Centrifugation and supernatant removal were repeated. The DNA was extracted
1052 using the DNeasy blood and tissue kit (Qiagen), using the pretreatment for Gram-
1053 negative bacteria, following the manufacturer's instructions, with the following
1054 exceptions: the lysis incubation period was increased to three hours, 4 µl of 10
1055 mg/ml of RNase A were used, and the DNA was eluted in water. Next, the DNA
1056 samples were prepared for barcoding and sequencing using the Nextera XT kit from
1057 Illumina. The manufacturer's instructions were followed to prepare the libraries.

1058

1059 Variant calling

1060 Reads were processed for quality using fastp version 0.23.2 (Chen et al. 2018).
1061 Adapters were trimmed from the reads using the `--detect_adapter_for_pe` option
1062 and the `--overrepresentation_analysis`, `--correction`, and `--cut_right` options
1063 were used for quality trimming. Trimmed reads were deposited at ENA under study
1064 accession number PRJEB82630. Variant calling using the clean and processed reads
1065 was done using breseq version 0.36.1 (Deatherage and Barrick 2014). In brief, breseq
1066 calculates a Bayesian score that all disagreements with the consensus at each
1067 position derive from sequencing errors only. It also filters out potential mutations
1068 where there is a strand bias, a quality bias, and positions in homopolymeric
1069 stretches. Finally, it only calls variants with at least 5% frequency in the population.
1070 For population sequencing, the `--polymorphisms-prediction` option was used for
1071 variant calling.

1072

1073 One population where the hosts were alternated (Alt_D) displayed a large excess of
1074 mutations at passage 10 (t1): while the median number of mutations is 28.5 per
1075 population, Alt_D displayed 989 mutations. A transient contamination from
1076 another *L. pneumophila* strain could be identified in this population. The distribution
1077 of the mutation frequency showed a clear peak around 6% ([Supplementary Figure](#)
1078 [11](#)). In this population only, all mutations with a frequency lower than 20% were
1079 filtered out, thus retaining only 15 mutations, of which none were fixed.

1080

1081 Mutations occurring in regions containing many tandem repeats were filtered out.

1082

1083 For coverage analysis of the reads mapping to the genes encoding fluorescent genes,
1084 the trimmed reads were mapped to the sequence of either gene using bowtie2 v2.4.4
1085 (Langmead and Salzberg 2012), removing bad reads with --qc-filter.

1086

1087 Statistical analysis and visualization

1088 All statistics and most figures in this contribution were performed in R (R
1089 Development Core Team 2024), using the tidyverse (Wickham et al. 2019) and, in
1090 particular, the ggplot2 package (Wickham 2009: 2).

1091

1092 Structure prediction of the LerC mutants

1093 Alternative gene prediction of the *lerC* mutants (*lerC_46* and *lerC_61*) was
1094 performed by inputting the mutant sequence of the gene in prodigal 2.6.3 (Hyatt et
1095 al. 2010), with default options. Structures were predicted using AlphaFold3
1096 (Abramson et al. 2024) then superimposed using PyMol v2.5.4 (Schrödinger, LLC
1097 2015).

1098

1099 Competing interest

1100 The authors declare no competing interest.

1101 Data availability

1102 Population sequencing data is available at ENA under Study accession number
1103 PRJEB82630.

1104 Acknowledgments

1105 The authors would like to thank Chayan Kumar Saha, for helping with the LerC
1106 mutant structures.

1107 This study was supported by grants from the Swedish Research Council [2017-
1108 03709 to L.G.], the Carl Trygger Foundation [CTS 15:184, CTS 17:178 to L.G.], the
1109 Fonds de recherche du Québec [FRQNT B3X Postdoctoral Scholarship to K.P.] and the
1110 Natural Sciences and Engineering Research Council of Canada (NSERC)
1111 [Postdoctoral Scholarship to K.P]. The funders had no role in the design of the study.

1112

1113 References

1114 Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore
1115 L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of
1116 biomolecular interactions with AlphaFold 3. *Nature* 630:493–500.

1117 Akarsu H, Bordes P, Mansour M, Bigot D-J, Genevoux P, Falquet L. 2019. TASmania:
1118 A bacterial Toxin-Antitoxin Systems database. *PLOS Comput. Biol.*
1119 15:e1006946.

1120 Akira S, Uematsu S, Takeuchi O. 2006. Pathogen Recognition and Innate Immunity.
1121 *Cell* 124:783–801.

- 1122 Al-Khodor S, Kalachikov S, Morozova I, Price CT, Abu Kwaik Y. 2009. The
1123 PmrA/PmrB Two-Component System of *Legionella pneumophila* Is a Global
1124 Regulator Required for Intracellular Replication within Macrophages and
1125 Protozoa. *Infect. Immun.* 77:374–386.
- 1126 Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997.
1127 Gapped BLAST and PSI-BLAST: a new generation of protein database search
1128 programs. *Nucleic Acids Res.* 25:3389–3402.
- 1129 Bandyopadhyay P, Sumer EU, Jayakumar D, Liu S, Xiao H, Steinman HM. 2012.
1130 Implication of Proteins Containing Tetratricopeptide Repeats in Conditional
1131 Virulence Phenotypes of *Legionella pneumophila*. *J. Bacteriol.* 194:3579–3588.
- 1132 Berger KH, Isberg RR. 1993. Two distinct defects in intracellular growth
1133 complemented by a single genetic locus in *Legionella pneumophila*. *Mol.*
1134 *Microbiol.* 7:7–19.
- 1135 Best AM, Abu Kwaik Y. 2019. Evasion of phagotrophic predation by protist hosts and
1136 innate immunity of metazoan hosts by *Legionella pneumophila*. *Cell.*
1137 *Microbiol.* 21:e12971.
- 1138 Björkman J, Samuelsson P, Andersson DI, Hughes D. 1999. Novel ribosomal
1139 mutations affecting translational accuracy, antibiotic resistance and
1140 virulence of *Salmonella typhimurium*. *Mol. Microbiol.* 31:53–58.
- 1141 Brüggemann H, Hagman A, Jules M, Sismeiro O, Dillies M-A, Gouyette C, Kunst F,
1142 Steinert M, Heuner K, Coppée J-Y, et al. 2006. Virulence strategies for
1143 infecting phagocytes deduced from the in vivo transcriptional program of
1144 *Legionella pneumophila*. *Cell. Microbiol.* 8:1228–1240.
- 1145 Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L, Tichit M, Jarraud
1146 S, Bouchier C, Vandenesch F, et al. 2004. Evidence in the *Legionella*
1147 *pneumophila* genome for exploitation of host cell functions and high genome
1148 plasticity. *Nat. Genet.* 36:1165–1173.
- 1149 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ
1150 preprocessor. *Bioinformatics* 34:i884–i890.
- 1151 Coscollá M, González-Candelas F. 2007. Population structure and recombination in
1152 environmental isolates of *Legionella pneumophila*. *Environ. Microbiol.* 9:643–
1153 656.
- 1154 David S, Rusniok C, Mentasti M, Gomez-Valero L, Harris SR, Lechat P, Lees J,
1155 Ginevra C, Glaser P, Ma L, et al. 2016. Multiple major disease-associated
1156 clones of *Legionella pneumophila* have emerged recently and independently.
1157 *Genome Res.* 26:1555–1564.
- 1158 David S, Sanchez-Buso L, Harris SR, Marttinen P, Rusniok C, Buchrieser C, Harrison
1159 TG, Parkhill J. 2017. Dynamics and impact of homologous recombination on
1160 the evolution of *Legionella pneumophila*. *PLoS Genet* 13:e1006855.
- 1161 Deatherage DE, Barrick JE. 2014. Identification of Mutations in Laboratory-Evolved
1162 Microbes from Next-Generation Sequencing Data Using breseq. In: Sun L,
1163 Shou W, editors. Engineering and Analyzing Multicellular Systems: Methods
1164 and Protocols. Methods in Molecular Biology. New York, NY: Springer. p. 165–
1165 188. Available from: https://doi.org/10.1007/978-1-4939-0554-6_12
- 1166 Den Boer JW, Yzerman EPF. 2004. Diagnosis of *Legionella* infection in Legionnaires'
1167 disease. *Eur. J. Clin. Microbiol. Infect. Dis.* 23:871–878.
- 1168 Diederer BM. 2008. *Legionella* spp. and Legionnaires' disease. *J. Infect.* 56:1–12.
- 1169 Dunin-Horkawicz S, Feder M, Bujnicki JM. 2006. Phylogenomic analysis of the GIY-
1170 YIG nuclease superfamily. *BMC Genomics* 7:98.

- 1171 Durieux I, Ginevra C, Attaiech L, Picq K, Juan P-A, Jarraud S, Charpentier X. 2019.
1172 Diverse conjugative elements silence natural transformation in *Legionella*
1173 species. *Proc. Natl. Acad. Sci.* 116:18613–18618.
- 1174 Ely B. 2021. Genomic GC content drifts downward in most bacterial genomes. *PLOS*
1175 *ONE* 16:e0244163.
- 1176 Ensminger AW. 2016. *Legionella pneumophila*, armed to the hilt: justifying the
1177 largest arsenal of effectors in the bacterial world. *Curr. Opin. Microbiol.* 29:74–
1178 80.
- 1179 Ensminger AW, Yassin Y, Miron A, Isberg RR. 2012. Experimental evolution of
1180 *Legionella pneumophila* in mouse macrophages leads to strains with altered
1181 determinants of environmental survival. *PLoS Pathog.* 8:e1002731.
- 1182 Falcó V, de Sevilla TF, Alegre J, Ferrer A, Vázquez JMM. 1991. *Legionella*
1183 *pneumophila*: A Cause of Severe Community-acquired Pneumonia. *Chest*
1184 100:1007–1011.
- 1185 Feldheim YS, Zusman T, Kapach A, Segal G. 2018. The single-domain response
1186 regulator LerC functions as a connector protein in the *Legionella*
1187 *pneumophila* effectors regulatory network. *Mol. Microbiol.* 110:741–760.
- 1188 Fields BS, Benson RF, Besser RE. 2002. *Legionella* and Legionnaires' disease: 25
1189 years of investigation. *Clin. Microbiol. Rev.* 15:506–526.
- 1190 Funatsu G, Wittmann HG. 1972. Ribosomal proteins: XXXIII. Location of amino-acid
1191 replacements in protein S12 isolated from *Escherichia coli* mutants resistant
1192 to streptomycin. *J. Mol. Biol.* 68:547–550.
- 1193 Garduno RA, Chong A, Nasrallah GK, Allan DS. 2011. The *Legionella pneumophila*
1194 Chaperonin – An Unusual Multifunctional Protein in Unusual Locations.
1195 *Front. Microbiol.* 2:122.
- 1196 Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, Medigue C,
1197 Etienne J, Buchrieser C. 2011. Extensive recombination events and horizontal
1198 gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics*
1199 12:536.
- 1200 Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, Andersson DI. 2011.
1201 Selection of Resistant Bacteria at Very Low Antibiotic Concentrations. *PLOS*
1202 *Pathog.* 7:e1002158.
- 1203 Habyarimana F, Al-khodori S, Kalia A, Graham JE, Price CT, Garcia MT, Kwaik YA.
1204 2008. Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila*
1205 in parasitism of protozoan hosts and human macrophages. *Environ. Microbiol.*
1206 10:1460–1474.
- 1207 Hall BM, Ma C-X, Liang P, Singh KK. 2009. Fluctuation AnaLysis CalculatOR: a web
1208 tool for the determination of mutation rate using Luria–Delbrück fluctuation
1209 analysis. *Bioinformatics* 25:1564–1565.
- 1210 Heger A, Holm L. 2000. Rapid automatic detection and alignment of repeats in
1211 protein sequences. *Proteins Struct. Funct. Bioinforma.* 41:224–237.
- 1212 Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards
1213 AT in bacteria. *PLoS Genet* 6:e1001115.
- 1214 Hubber A, Kubori T, Coban C, Matsuzawa T, Ogawa M, Kawabata T, Yoshimori T,
1215 Nagai H. 2017. Bacterial secretion system skews the fate of *Legionella*-
1216 containing vacuoles towards LC3-associated phagocytosis. *Sci. Rep.* 7:44795.
- 1217 Hubber A, Roy CR. 2010. Modulation of Host Cell Function by *Legionella*
1218 *pneumophila* Type IV Effectors. *Annu. Rev. Cell Dev. Biol.* 26:261–283.
- 1219 Hugoson E, Guliaev A, Ammunét T, Guy L. 2022. Host Adaptation in Legionellales Is
1220 1.9 Ga, Coincident with Eukaryogenesis. *Mol. Biol. Evol.* 39:msac037.

- 1221 Hyatt D, Chen G-L, LoCascio P, Land M, Larimer F, Hauser L. 2010. Prodigal:
1222 prokaryotic gene recognition and translation initiation site identification.
1223 *BMC Bioinformatics* 11:119.
- 1224 Isberg RR, O'Connor TJ, Heidtman M. 2009. The Legionella pneumophila replication
1225 vacuole: making a cosy niche inside host cells. *Nat. Rev. Microbiol.* 7:12–24.
- 1226 Jeon KW, Lorch IJ. 1967. Unusual intra-cellular bacterial infection in large, free-
1227 living amoebae. *Exp. Cell Res.* 48:236–240.
- 1228 Jules M, Buchrieser C. 2007. Legionella pneumophila adaptation to intracellular life
1229 and the host response: Clues from genomics and transcriptomics. *FEBS Lett.*
1230 581:2829–2838.
- 1231 Lang GI. 2018. Measuring Mutation Rates Using the Luria-Delbrück Fluctuation
1232 Assay. In: Muzi-Falconi M, Brown GW, editors. *Genome Instability: Methods*
1233 *and Protocols*. New York, NY: Springer. p. 21–31. Available from:
1234 https://doi.org/10.1007/978-1-4939-7306-4_3
- 1235 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat.*
1236 *Methods* 9:357–359.
- 1237 Lassalle F, Perian S, Bataillon T, Nesme X, Duret L, Daubin V. 2015. GC-Content
1238 evolution in bacterial genomes: the biased gene conversion hypothesis
1239 expands. *PLoS Genet* 11:e1004941.
- 1240 Lau HY, Ashbolt NJ. 2009. The role of biofilms and protozoa in Legionella
1241 pathogenesis: implications for drinking water. *J. Appl. Microbiol.* 107:368–
1242 378.
- 1243 Leenheer D, Moreno AB, Paranjape K, Murray S, Jarraud S, Ginevra C, Guy L. 2023.
1244 Rapid adaptations of Legionella pneumophila to the human host. *Microb.*
1245 *Genomics* 9:000958.
- 1246 Lenski RE. 2017. Experimental evolution and the dynamics of adaptation and
1247 genome evolution in microbial populations. *ISME J.* 11:2181–2194.
- 1248 Lynch M, Ali F, Lin T, Wang Y, Ni J, Long H. 2023. The divergence of mutation rates
1249 and spectra across the Tree of Life. *EMBO Rep.* 24:e57561.
- 1250 Mérault N, Rusniok C, Jarraud S, Gomez-Valero L, Cazalet C, Marin M, Brachet E,
1251 Aegerter P, Gaillard JL, Etienne J, et al. 2011. Specific Real-Time PCR for
1252 Simultaneous Detection and Identification of Legionella pneumophila
1253 Serogroup 1 in Water and Clinical Samples. *Appl. Environ. Microbiol.* 77:1708–
1254 1717.
- 1255 Mercante JW, Caravas JA, Ishaq MK, Kozak-Muiznieks NA, Raphael BH, Winchell JM.
1256 2018. Genomic heterogeneity differentiates clinical and environmental
1257 subgroups of Legionella pneumophila sequence type 1. *PLOS ONE* 13:e0206110.
- 1258 Nair J, Rouse DA, Bai G-H, Morris SL. 1993. The rpsL gene and streptomycin
1259 resistance in single and multiple drug-resistant strains of Mycobacterium
1260 tuberculosis. *Mol. Microbiol.* 10:521–527.
- 1261 O'Connor TJ, Adepoju Y, Boyd D, Isberg RR. 2011. Minimization of the Legionella
1262 pneumophila genome reveals chromosomal regions involved in host range
1263 expansion. *Proc. Natl. Acad. Sci.* 108:14733–14740.
- 1264 Oliva G, Sahr T, Buchrieser C. 2018. The Life Cycle of L. pneumophila: Cellular
1265 Differentiation Is Linked to Virulence and Metabolism. *Front. Cell. Infect.*
1266 *Microbiol.* 8:3.
- 1267 Park JM, Ghosh S, O'Connor TJ. 2020. Combinatorial selection in amoebal hosts
1268 drives the evolution of the human pathogen Legionella pneumophila. *Nat.*
1269 *Microbiol.*:1–11.

- 1270 Petzold M, Thürmer A, Menzel S, Mouton JW, Heuner K, Lück C. 2013. A structural
1271 comparison of lipopolysaccharide biosynthesis loci of *Legionella*
1272 *pneumophila* serogroup 1 strains. *BMC Microbiol.* 13:198.
- 1273 Price MN, Huang KH, Alm EJ, Arkin AP. 2005. A novel method for accurate operon
1274 predictions in all sequenced prokaryotes. *Nucleic Acids Res.* 33:880–892.
- 1275 R Development Core Team. 2024. R: A language and environment for statistical
1276 computing. Vienna, Austria: R Foundation for Statistical Computing
- 1277 Rao C, Benhabib H, Ensminger AW. 2013. Phylogenetic Reconstruction of the
1278 *Legionella pneumophila* Philadelphia-1 Laboratory Strains through
1279 Comparative Genomics. *PLOS ONE* 8:e64129.
- 1280 Richards AM, Von Dwingelo JE, Price CT, Abu Kwaik Y. 2013. Cellular microbiology
1281 and molecular ecology of *Legionella*-amoeba interaction. *Virulence* 4:307–
1282 314.
- 1283 Rosche WA, Foster PL. 2000. Determining Mutation Rates in Bacterial Populations.
1284 *Methods* 20:4–17.
- 1285 Sahr T, Brüggemann H, Jules M, Lomma M, Albert-Weissenberger C, Cazalet C,
1286 Buchrieser C. 2009. Two small ncRNAs jointly govern virulence and
1287 transmission in *Legionella pneumophila*. *Mol. Microbiol.* 72:741–762.
- 1288 Sahr T, Escoll P, Rusniok C, Bui S, Pehau-Arnaudet G, Lavieu G, Buchrieser C. 2022.
1289 Translocated *Legionella pneumophila* small RNAs mimic eukaryotic
1290 microRNAs targeting the host immune response. *Nat. Commun.* 13:762.
- 1291 Sahr T, Rusniok C, Dervins-Ravault D, Sismeiro O, Coppee J-Y, Buchrieser C. 2012.
1292 Deep sequencing defines the transcriptional map of *L. pneumophila* and
1293 identifies growth phase-dependent regulated ncRNAs implicated in
1294 virulence. *RNA Biol.* 9:503–519.
- 1295 Schrödinger, LLC. 2015. The PyMOL Molecular Graphics System, Version 1.8.
- 1296 Segal G. 2014. The *Legionella pneumophila* Two-Component Regulatory Systems
1297 that Participate in the Regulation of Icm/Dot Effectors. In: Hilbi H, editor.
1298 Molecular Mechanisms in *Legionella* Pathogenesis. Current Topics in
1299 Microbiology and Immunology. Berlin, Heidelberg: Springer. p. 35–52.
1300 Available from: https://doi.org/10.1007/82_2013_346
- 1301 Shao Y, Harrison EM, Bi D, Tai C, He X, Ou H-Y, Rajakumar K, Deng Z. 2011. TADB: a
1302 web-based resource for Type 2 toxin–antitoxin loci in bacteria and archaea.
1303 *Nucleic Acids Res.* 39:D606–D611.
- 1304 Thürmer A, Helbig JH, Jacobs E, Lück PC 2009. PCR-based ‘serotyping’ of *Legionella*
1305 *pneumophila*. *J. Med. Microbiol.* 58:588–595.
- 1306 Timms AR, Steingrimsdottir H, Lehmann AR, Bridges BA. 1992. Mutant sequences
1307 in the *rpsL* gene of *Escherichia coli* B/r: Mechanistic implications for
1308 spontaneous and ultraviolet light mutagenesis. *Mol. Gen. Genet. MGG* 232:89–
1309 96.
- 1310 Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G,
1311 Scarpelli C, Medigue C. 2006. MaGe: a microbial genome annotation system
1312 supported by synteny results. *Nucleic Acids Res.* 34:53–65.
- 1313 Vos M, Didelot X. 2008. A comparison of homologous recombination rates in
1314 bacteria and archaea. *ISME J.* 3:199–208.
- 1315 Wang J, Woo M, Yan C. 2017. Spot Plating Assay for the Determination of Survival
1316 and Plating Efficiency of *Escherichia coli* in sub-MIC Levels of Antibiotics |
1317 jemi.microbiology.ubc.ca. *Undergrad. J. Exp. Microbiol. Immunol.* 1:26–29.
- 1318 Wee BA, Alves J, Lindsay DSJ, Klatt A-B, Sargison FA, Cameron RL, Pickering A,
1319 Gorzynski J, Corander J, Marttinen P, et al. 2021. Population analysis of

1320 Legionella pneumophila reveals a basis for resistance to complement-
 1321 mediated killing. *Nat. Commun.* 12:7165.
 1322 Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-
 1323 Verlag Available from: <https://www.springer.com/gp/book/9780387981413>
 1324 Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G,
 1325 Hayes A, Henry L, Hester J, et al. 2019. Welcome to the Tidyverse. *J. Open*
 1326 *Source Softw.* 4:1686.
 1327 Zhu W, Tao L, Quick ML, Joyce JA, Qu J-M, Luo Z-Q. 2015. Sensing Cytosolic RpsL by
 1328 Macrophages Induces Lysosomal Cell Death and Termination of Bacterial
 1329 Infection. *PLOS Pathog.* 11:e1004704.

1330
 1331

1332 Supplementary Material

1333 Supplementary Tables

1334

1335 [Supplementary Table 1: Population sequencing and number of generations in the](#) 1336 [different hosts.](#)

1337

Host	Abbrev.	t1 Passages	N generations	t2 Passages	N generations	t3 Passages	N generations
<i>A. castellanii</i>	Acas	p10	128 +/- 7	p40	517 +/- 9	p65	798 +/- 9
Alternation	Alt	p10	124 +/- 5	p49	481 +/- 13	p85	799 +/- 13
U937 cells	Mac	p15	113 +/- 5	p38	270 +/- 10	p65	444 +/- 40

1338

1339 [Supplementary Table 2: Mutation rates \(mutations per cell per cell division or](#) 1340 [generation\) measured for *L. pneumophila* Paris, dTomato-, and SYFP2-labeled](#) 1341 [strains. Mutation rates were measured using streptomycin as the selection](#) 1342 [antibiotic and then calculated using MSS Maximum Likelihood method. Mutations](#) 1343 [rates are expressed per cell per generation per billion. CI, Confidence Interval.](#)

1344

Strain	Mutation rate $\times 10^{-10}$	95% CI range $\times 10^{-10}$	
		Upper Bound	Lower Bound
<i>L. pneumophila</i> Paris	6.5	8.5	4.7
dTomato	7.6	9.9	5.5
SYFP	5.5	7.3	3.8

1345

1346

1347 [Supplementary Table 3: Clones and mutations involved in competition experiments.](#) 1348 [Genes and mutations are color-coded.](#)

1349

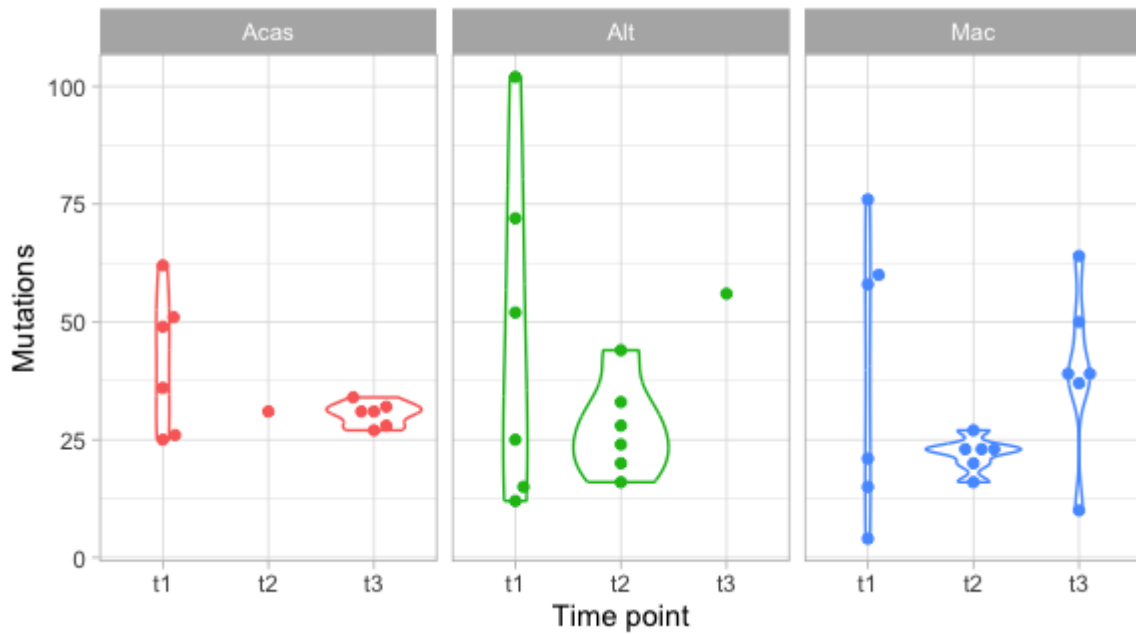
Strain	Genotype
DA75302	RpsL(K88R)/GroES(A45T)
DA76735	RpsL(K88R)/GroES(A45T) + lpp0833(Δ 1bp_1092nt)
DA76737	RpsL(K88R)/GroES(A45T) + lpp0833(Δ 1bp_1092nt) + LerC_61(Δ 13bp_61-73/453nt)
DA75299i	RpsL(K88R)/GroES(A45T) + lpp0833(Δ 1bp_1092nt) + LerC_51(V51G)
DA76739	RpsL(K43T)/RpsD(S205Y) + IolE(G241V)
DA75291i	RpsL(K43T)/RpsD(S205Y) + LerC_46(C46*)

LG300i *RpsL(K43T)/RpsD(S205Y) + LerC_13(A13D)*

1350

1351 Supplementary Figures

1352 Supplementary Figure 1



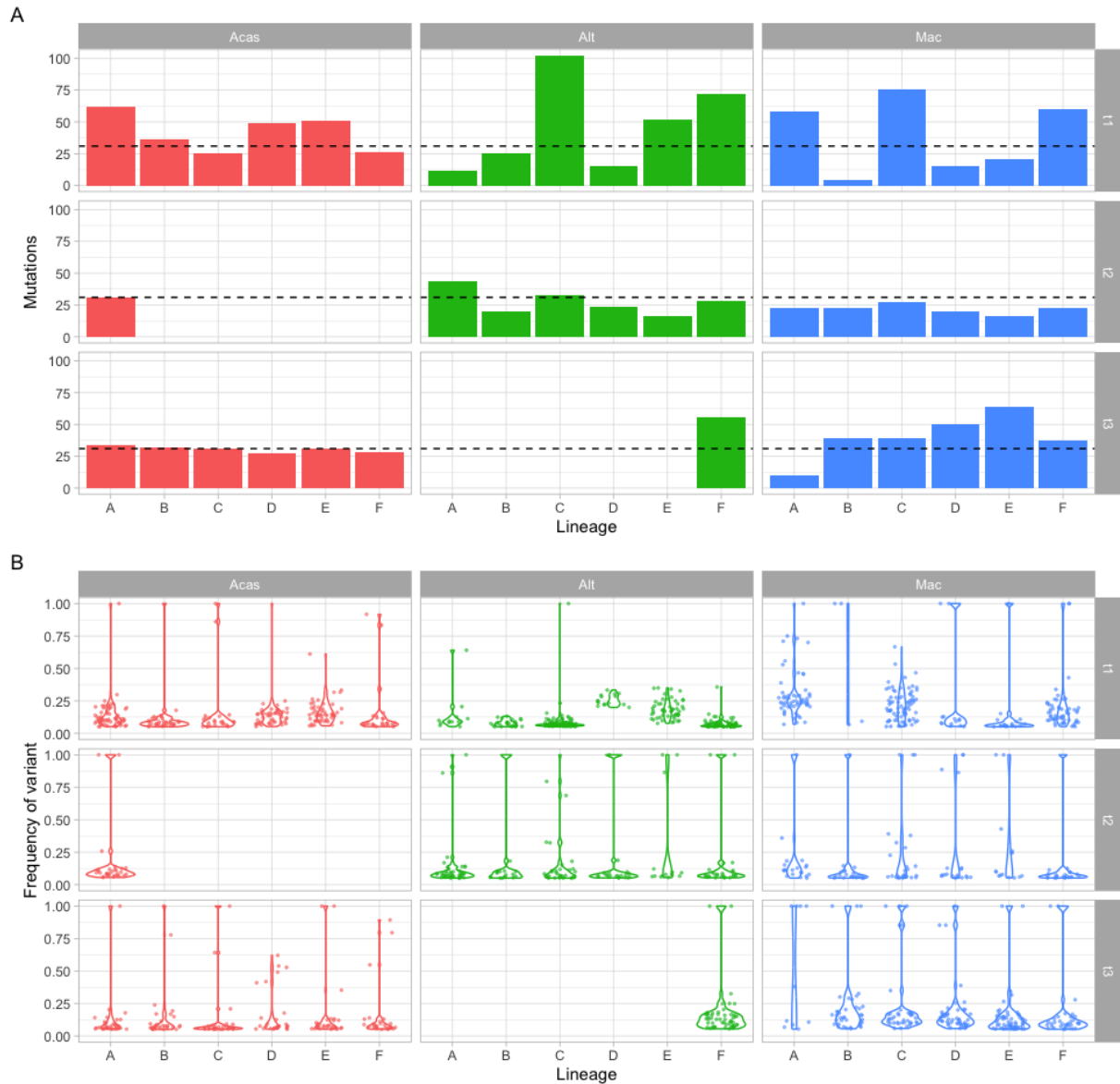
1353

1354 **Supplementary Figure 1:** Distribution of mutations per host and time point. Each
1355 point is a lineage. Only one Acas population and was sequenced at t2, and only one
1356 Alt population was sequenced at t3.

1357

1358 Supplementary Figure 2

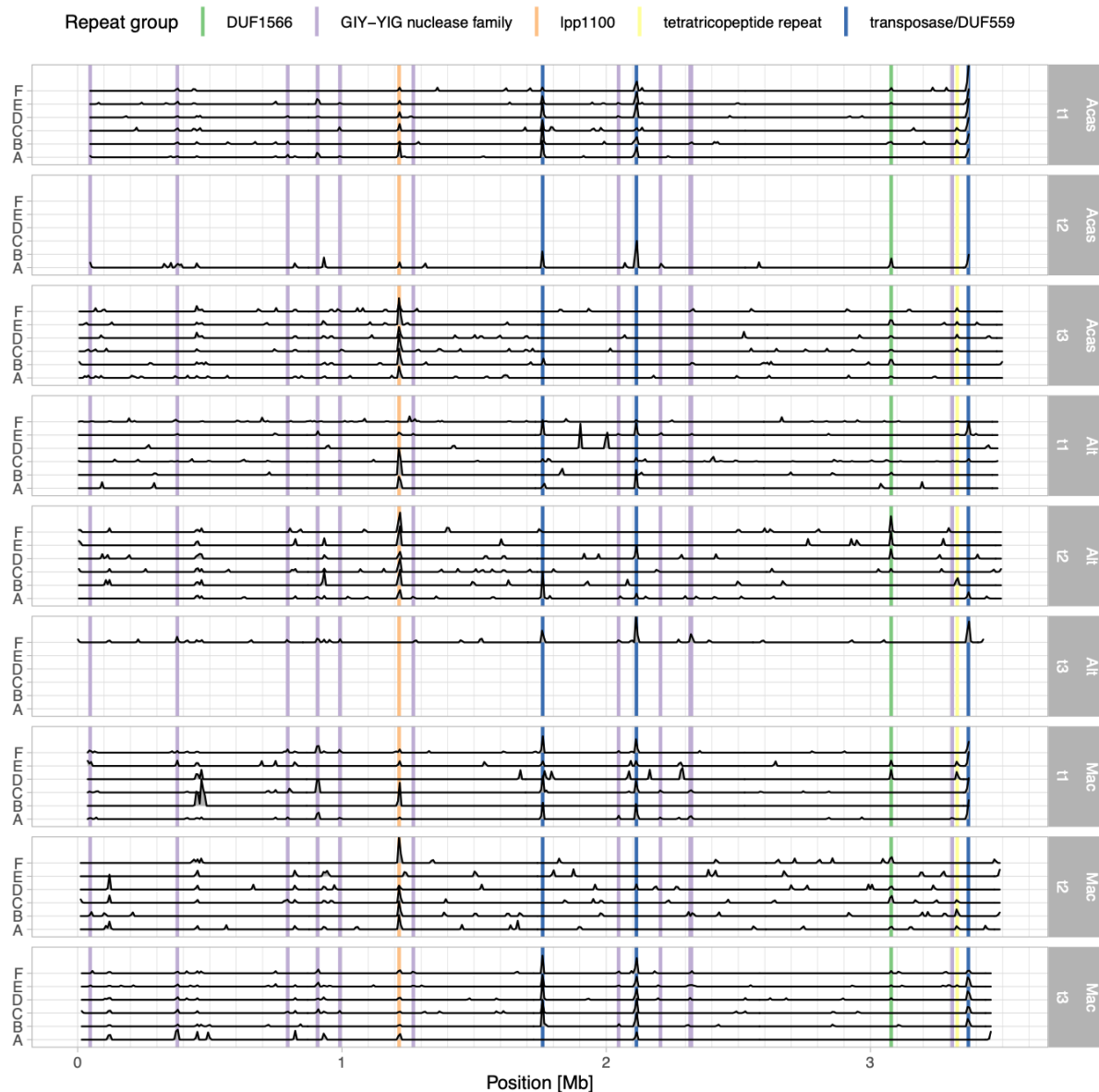
1359



1360
1361
1362
1363
1364
1365
1366
1367
1368

Supplementary Figure 2: Distribution of mutation numbers and frequencies. (A): Number of mutations with frequency > 5% per host (Acas, red; Alt, green; Mac, blue), lineage (A-F, x-axis) and time point (t1-t3). The median number of mutations per population (32.5) is shown with a dashed black line. **(B):** Frequency (y-axis) of mutations per host, lineage and time point. Each dot represents a mutation, and its position on the y-axis represents its frequency. Violin plots are superimposed on the dots to provide a representation of the distribution.

1369 Supplementary Figure 3



1370
1371
1372
1373
1374
1375

Supplementary Figure 3: Distribution of mutations along the *L. pneumophila* genome. Each line corresponds to a separate population. The host and timepoint is indicated on the right, the lineage on the left. The position of genes containing repeats or being present in different parts of the genome are indicated with lines.

1376 Supplementary Figure 4

No. of Repeats	Total Score	Length	Diagonal	BW-From	BW-To	Level
19	1211.97	35	35	1075	1109	1
630- 664	(54.25/30.86)		LGALSVEQRVELL.VVQDQHRNTVLHKTLS.RPELLK			
667- 701	(60.83/35.53)		LEGLSLEQRVELL.KVKGSDGNIVLHTVAS.HPKPLK			
704- 738	(66.27/39.39)		LEGLSVEQRVELL.KVKGSDGNTLLHFAAS.NPKSLK			
741- 776	(51.78/29.11)		LEGLPEESRVELLtTLRRKDGASVLDFAAT.HPKSLK			
779- 813	(62.56/36.76)		WESLSEEKRIEVL.EMNDEHGNTVLHWAVS.RPKSLE			
816- 850	(59.21/34.38)		LERLSDEQRFKLL.WMKVQFGNTVLHLAAN.SPETLK			
853- 887	(59.88/34.86)		LEGLPEEKRIEVL.TILNQYGDTVIHRAAH.FPASLK			
890- 924	(67.66/40.37)		WEGLEPEEKRIELL.RMKGNYNNTLLHNAAS.HPESLK			
927- 961	(58.64/33.97)		LEGLSEQQRAELL.RINGSIGNTVLHCALG.YLESK			
964- 998	(72.67/43.93)		LEGLSEESRFQLL.GMKNQDDNTVLYLAAS.RPESLK			
1001- 1035	(70.30/42.25)		LEDLSVENRVQLL.GMKNQYGNNTVLHEALS.QPKSLK			
1038- 1072	(73.90/44.81)		LEGLSEESRFQLL.RMKNQYGDTVLHEALS.QPKSLK			
1075- 1109	(75.50/45.94)		LEGLSEESRVQLL.GMKNQDGYTVLHKALS.LPESLK			
1112- 1146	(75.50/45.94)		LEGLSEESRVQLL.GMKNQDGYTVLHKALS.LPESLK			
1149- 1183	(73.09/44.23)		LEGLSEESRVQLL.GMKDYHGNTLLHEALS.RPKSLK			
1186- 1220	(75.46/45.91)		LEGLSEESRFQLL.GMKNQDGNNTVLYLAAS.RPESLK			
1223- 1257	(66.54/39.58)		LEGLSEESRVQLL.GMKDYGNNTALRCVAVP.HPASLK			
1260- 1295	(60.00/34.94)		LEGLSEEKRIELL.SAKNQHGDSVLRALScYPESLG			
1298- 1329	(27.93/12.17)		LEMLCETTQRQVLeDLGCLRGNVFLGK....PEILE			

1377

1378

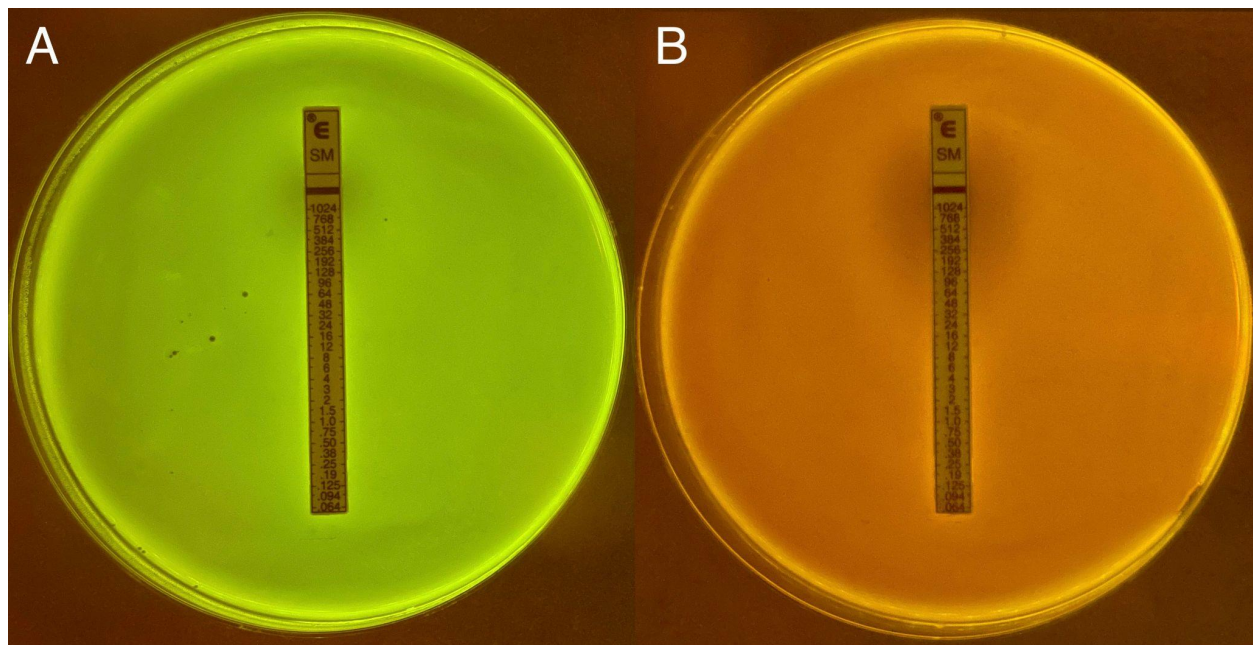
1379

1380

1381

Supplementary Figure 4: Repeats in ankyrin-repeat containing protein lpp1100. The analysis was performed using RADAR (Heger and Holm 2000).

Supplementary Figure 5



1382

1383

1384

1385

1386

1387

1388

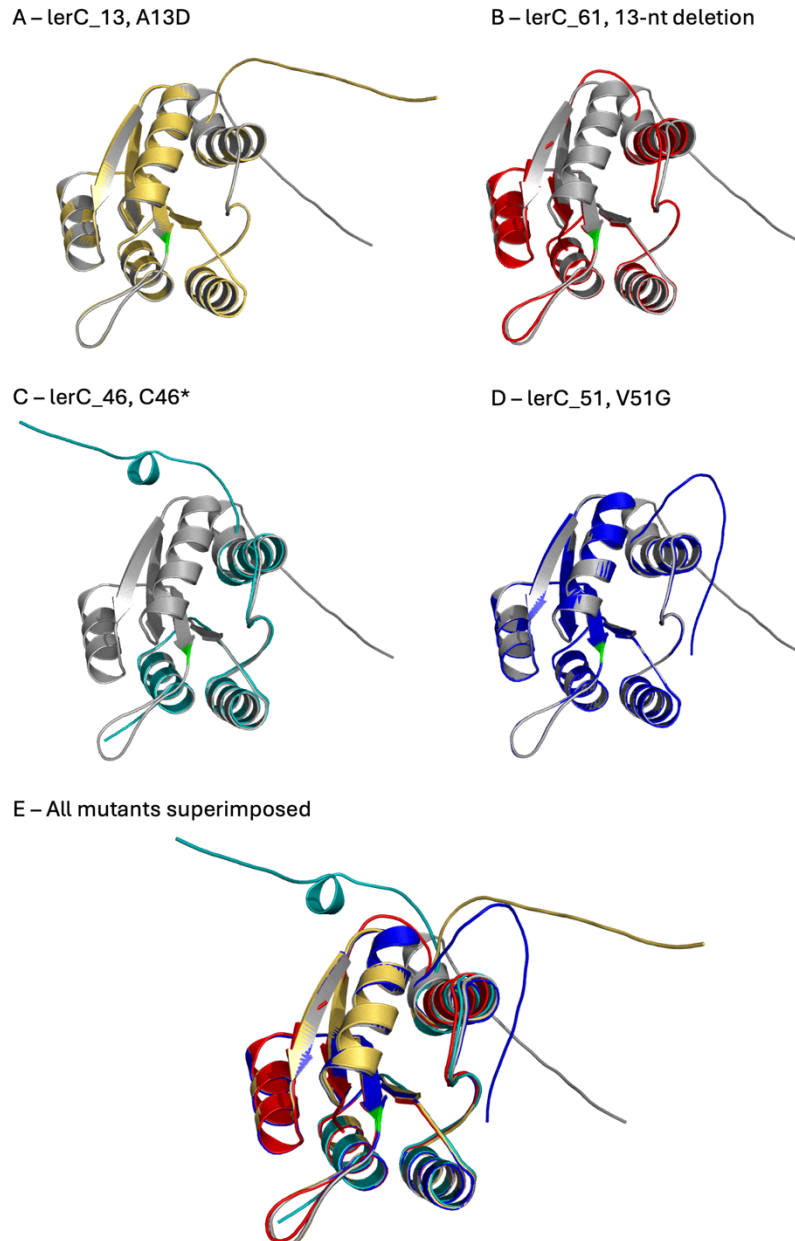
1389

1390

1391

Supplementary Figure 5: Resistance to streptomycin of two *Legionella* clones bearing different mutations in the *rpsL* gene. (A) *L. pneumophila* with the Rpsl43/RpsD genotype, in a SYFP2 background. (B) *L. pneumophila* with the Rpsl88/GroES genotype, in a dTomato background. Although the bacteria grow on the whole plate, including at the maximum streptomycin concentration (1024 $\mu\text{g/ml}$), a larger halo zone can be observed in B (to $\sim 96 \mu\text{g/ml}$) than in A (to $\sim 768 \mu\text{g/ml}$).

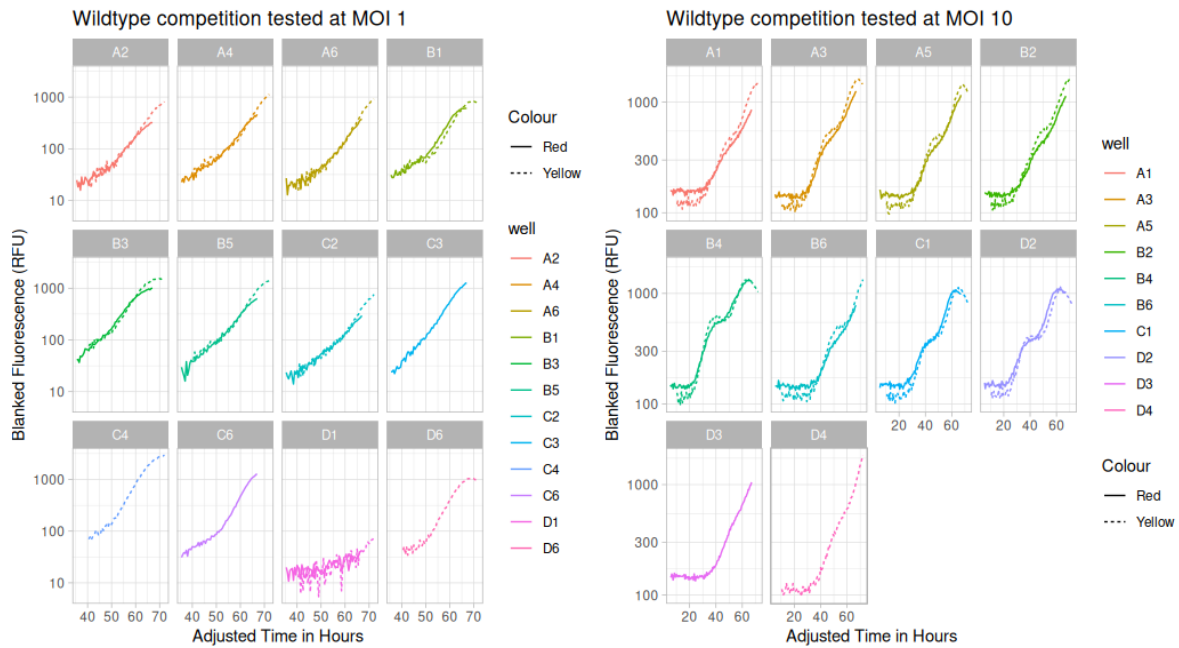
Supplementary Figure 6



1392
1393
1394
1395
1396
1397

Supplementary Figure 6: Structure of the *lerC* mutants, as predicted by AlphaFold3. The predicted wild-type LerC protein is shown in grey on all panels, *lerC*_13 in yellow (A, E), *lerC*_61 in red (B, E), *lerC*_46 in teal (C, E), and *lerC*_51 in blue (D, E). The location of the conserved aspartic acid (D53) is shown in green.

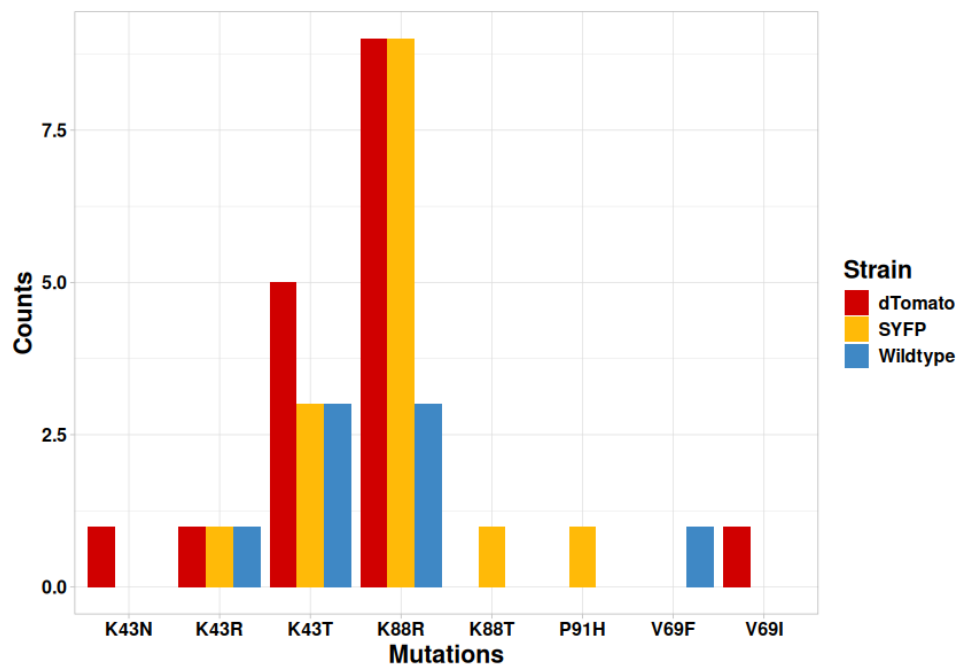
1398 Supplementary Figure 7



1399
1400
1401
1402
1403
1404
1405

Supplementary Figure 7: Competition experiments between selected isolates. Competitions of *L. pneumophila* strain Paris marked with dTomato (Red) and SYFP2 (Yellow) in an amoeba infection model at MOI 1 and MOI 10.

Supplementary Figure 8



1406
1407
1408
1409
1410

Supplementary Figure 8: Distribution and diversity of non-synonymous mutations in the *rpsL* gene for the three ancestral strains (*L. pneumophila* Paris, SYFP, and dTomato). The x-axis represents the amino acid residues switched through the mutation and their position in the RpsL sequence.

1411

1412 **Supplementary Figure 9**

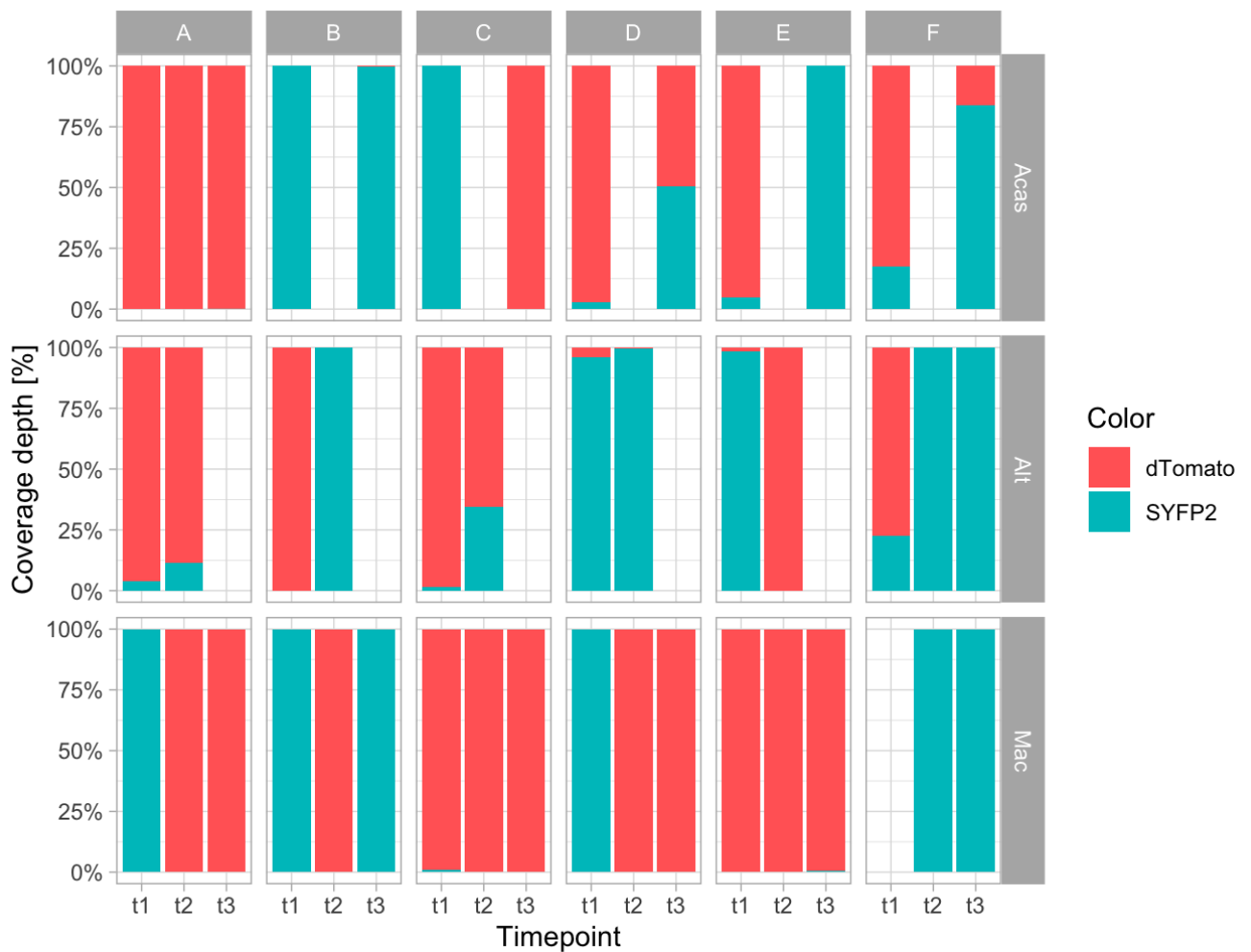


1413

1414 **Supplementary Figure 9: Synteny plot of *Legionella* and *Coxiella* species, centered**
1415 **around *lpp1100* (LPP_RS0510) in *L. pneumophila* Paris (upper row). The other rows**
1416 **Colors indicate groups of genes colocalized in a specific genome. Figure obtained**
1417 **with MAGE (Vallenet et al. 2006).**

1418

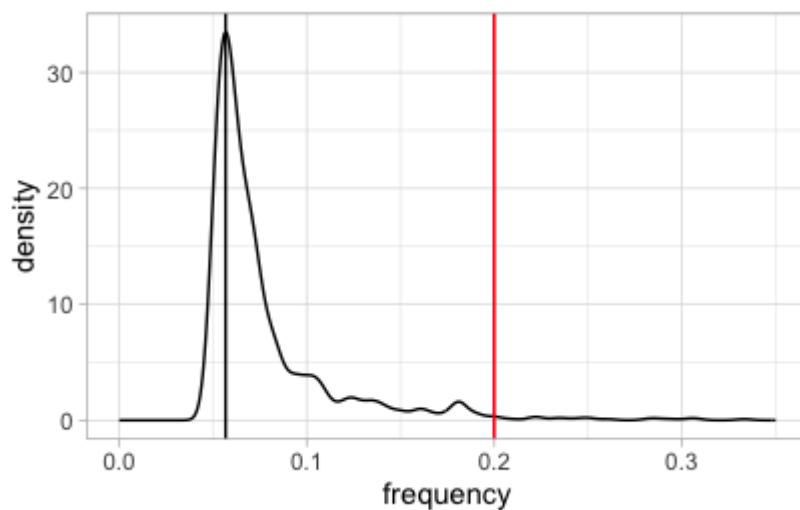
1419 Supplementary Figure 10



1420
1421
1422
1423
1424
1425
1426

Supplementary Figure 10: Proportion of SYFP2- and dTomato-marked *L. pneumophila* in each lineage at the 3 time points. At t0, each population consisted of an equal mix of each ancestor. The proportion at the other time points was estimated by mapping the reads obtained from population sequencing to the genes encoding the two fluorescent proteins (dTomato, red; SYFP2, green).

1427 Supplementary Figure 11



1428
1429 **Supplementary Figure 11:** Distribution of mutation frequencies in the Alt D
1430 population, which counted 991 mutations. The black vertical line marks the
1431 maximum of the density curve, at 5.7%, while the red line marks the chosen 20%
1432 cut-off, under which mutations were discarded.
1433
1434