# Mobile element integration reveals a chromosome dimer resolution system in Legionellales.

Beth Nicholson[a], Shayna R. Deecker[a] and Alexander W. Ensminger[a,b,*]

[a] Department of Biochemistry, University of Toronto, Toronto, Ontario, M5S 1A8 Canada.

[b] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, M5S 1A8 Canada.

Running title: Chromosome dimer resolution in Legionellales.

[*]Address correspondence to Alexander W. Ensminger, alex.ensminger@utoronto.ca

Classification: Biological Sciences; Microbiology

Keywords: *Legionella pneumophila*, *Coxiella burnetii*, chromosome dimer resolution, site-specific recombination, Xer, *dif,* IMEX

**ABSTRACT**

In bacteria, the mechanisms used to repair DNA lesions during genome replication include homologous recombination between sister chromosomes. This can lead to the formation of chromosome dimers if an odd number of crossover events occurs. The concatenated DNA must be resolved before cell separation to ensure genomic stability and cell viability. The broadly conserved *dif*/Xer system counteracts the formation of dimers by catalyzing one additional crossover event immediately prior to cell separation. While *dif*/Xer systems have been characterized or predicted in the vast majority of proteobacteria, no homologs to *dif* or *xer* have been identified in the order Legionellales. Here we report the discovery of a distinct single-recombinase *dif*/Xer system in the intracellular pathogen *Legionella pneumophila*. The *dif* site was uncovered by our analysis of *Legionella* mobile element-1 (LME-1), which harbors a *dif* site-mimic and integrates into the *L. pneumophila* genome via site-specific recombination. We demonstrate that *lpg1867* (herein named *xerL*) encodes a tyrosine recombinase that is necessary and sufficient for catalyzing recombination at the *dif* site, and that deletion of *dif* or *xerL* causes filamentation along with extracellular and intracellular growth defects. We show that the *dif*/XerL system is present throughout Legionellales and that *Coxiella burnetii* XerL and its cognate *dif* site can functionally substitute for the native system in *L. pneumophila*. Lastly, we describe an unexpected link between *C. burnetii dif*/Xer and the maintenance of its virulence plasmids.

**Significance Statement**

The maintenance of circular chromosomes depends on the ability to resolve aberrant chromosome dimers as they form. In most proteobacteria, broadly conserved Xer recombinases catalyze single crossovers at short, species-specific *dif* sites located near the replication terminus. Chromosomal

38 dimerization leads to the formation of two copies of *dif*, leading to rapid site-specific recombination
39 and elimination of the duplicated intervening sequence. The apparent absence of chromosome-dimer
40 resolution mechanisms in Legionellales has been a mystery to date. By studying a phage-like mobile
41 genetic element, LME-1, we have identified a previously unknown single-recombinase *dif*/Xer
42 system that is not only widespread across Legionellales but whose activity is linked to virulence in
43 two important human pathogens.
44

45 **INTRODUCTION**

46 The circular nature of bacterial chromosomes brings about a problem for the partitioning of
47 genetic information to daughter cells: the formation of chromosome dimers during replication.
48 Dimers are generated when homologous recombination-mediated repair of DNA lesions results in an
49 odd number of crossover events between sister chromosomes. The concatenated DNA must be
50 resolved into monomers for proper cell division to occur. The essential and broadly conserved
51 bacterial *dif*/Xer system overcomes this topological problem by catalyzing one additional crossover
52 event immediately prior to cell separation (1, 2). The *dif* (deletion-induced filamentation) site is a
53 conserved ~30-bp DNA element on the chromosome that contains binding sites for either two copies
54 of a single tyrosine recombinase or one copy of two different ones (3). In a highly regulated process,
55 a nucleoprotein complex containing the Xer tetramer and two aligned *dif* sites forms a recombination
56 synapse, and site-specific recombination between them deconcatenates the chromosome dimer into
57 monomers (4–6).

58 Dimer resolution via site-specific recombination by tyrosine recombinases was first
59 discovered in plasmids (7, 8) including ColE1, in which a DNA sequence named *cer* was found to be
60 essential for plasmid monomerization and stability (8). Recombination between *cer* sites was later
61 found to require catalysis by XerC (9) and XerD (10) encoded on the *E. coli* genome. These
62 findings, along with homology between the chromosomal *dif* site and *cer*, led to the discovery of the
63 *dif*/XerCD chromosome dimer resolution (CDR) pathway in *E. coli* (10–13) and later in several
64 other species of bacteria (14–19) and archaea (20, 21). Subsequent dissection of the *dif*/Xer
65 recombination machinery has been aided by another group of mobile genetic elements, named
66 integrative mobile elements exploiting Xer (IMEX) (22–27). IMEXs, which include lysogenic

67 phages, genomic islands, and plasmids, contain a *dif*-mimic sequence which allows them to integrate
68 into the chromosome at the *dif* site using the host's Xer recombination machinery (5, 28).
69 Homologs of *xerC* and *xerD* have been predicted to be present in the vast majority of
70 proteobacterial species (29). Exceptions to this include two groups of bacteria that utilize a single-
71 recombinase system. The *Streptococci*/*Lactococci* use a single recombinase XerS and an atypical *dif*
72 site (16) and a group of epsilon-proteobacteria, including *Helicobacter* and *Campylobacter* species,
73 uses a single-recombinase XerH system (14, 17, 29). While *dif*/Xer homologs have been detected in
74 almost 90% of proteobacteria, no *dif* or Xer homologs could be detected in the order Legionellales,
75 despite encoding RecA, RecBCD, and RecF, which are thought to be responsible for dimer
76 formation during replication (29). How then do the members of the order Legionellales overcome
77 the threat to genome stability that chromosome dimers impose? It is possible that they use an entirely
78 different method for CDR, or that its *dif*/Xer system is sufficiently divergent from others to avoid
79 detection by sequence-based homology searches.
80 Several strains of *Legionella pneumophila* harbor a phage-like integrative mobile genetic
81 element, named *Legionella* mobile element-1 (LME-1) (30–33). We previously found that
82 integration into the genome requires a 22-bp attachment site (*att*) in LME-1 that is identical to a
83 sequence on the *L. pneumophila* genome (33). Even though LME-1 is only integrated in a small
84 proportion of sequenced *L. pneumophila* isolates, the chromosomal *att* site is present in all *L.*
85 *pneumophila* isolates and *Legionella* species sequenced to date (Table S1), a level of conservation
86 that suggests an important function. We report here that the LME-1 *att* site on the chromosome
87 meets several criteria of the missing *dif* site in *Legionella*: size ~30 bp, low GC-content, some degree
88 of palindromicity, positioned close to the replication terminus, and in a non-coding region with

89 varying flanking genes. We identify a single recombinase that is both necessary and sufficient for
90 site-specific recombination at this site. Together, our results indicate that *L. pneumophila* uses a
91 distinct single-recombinase system for CDR and that LME-1 is an IMEX of that system. We also
92 report that *xerL* orthologs can be found across the Legionellales family. Lastly, we show that the
93 *Coxiellaceae dif* site and *xerL* ortholog can functionally substitute for the *Legionella dif*/XerL
94 system and propose that the migration of *C. burnetii* XerL off the chromosome has contributed to the
95 stability of virulence plasmids in this pathogen.
96

97 **RESULTS**

98 **The LME-1 attachment site is invariably close to the *Legionella* replication terminus.**

99 We previously determined that LME-1 integrates into the *L. pneumophila* genome via a 22-
100 bp *att* site (33). The corresponding *att* site on the chromosome is contained within a broader 29-bp
101 sequence that is conserved across all sequenced *L. pneumophila* strains and *Legionella* species (Fig.
102 1A, Table S1), suggesting that it performs an essential function. While the *att* site itself is conserved,
103 its genomic neighborhood varies (33) (Table S1), indicating that the intrinsic function of this
104 sequence may be unrelated to its flanking genes. In a search for attributes of the *att* site that may hint
105 at its function, we analyzed its chromosomal location in all *L. pneumophila* strains with available
106 genome sequences (Table S1). We found that despite the disparate gene neighborhoods, the location
107 of the *att* site was roughly opposite the origin of replication in all cases. In fact, GC-skew analysis of
108 each genome revealed that the *att* site was invariably close to the cumulative GC-skew maximum,
109 which occurs at the site of replication termination (34) (Fig. 1B, Table S1). The *att* site is present
110 within a small ~8-kb window between 35 kb and 43 kb from the GC-skew inflection point (Table

111 S1). This specificity of positioning extended to the *att* sites of other *Legionella* species, which are
112 located between 89 bp and 79 kb from the point of GC-skew inflection (Fig. 1B, Table S1).
113 The proximity of the LME-1 *att* site to the terminus region is reminiscent of bacterial *dif*
114 sites, so we next compared the *att* site to the *dif* sites of several bacteria with established *dif*/Xer
115 systems. In addition to its proximity to the terminus, the LME-1 *att* site shows several similarities to
116 known *dif* sites, including (i) intergenic location (Table S1), (ii) size of ~30 bp, (iii) low G+C
117 content, and (iv) a short central region flanked by left and right arms with some degree of dyad
118 symmetry (Fig. 1C). However, the LME-1 *att* site is distinguished from known *dif* sites by the
119 perfect dyad symmetry of its 10-bp arms and the short 4-bp central region separating them. The arms
120 do not show homology to any known Xer-binding motifs (Fig. 1C) (29).
121
122 **The *L. pneumophila dif* site supports highly efficient RecA-independent DNA recombination.**
123 The similarities between the chromosomal LME-1 *att* site (hereafter referred to as the
124 *Legionella dif* site) and established *dif* sites suggest that it likely contains two binding sites for one or
125 more tyrosine-type recombinases that catalyze site-specific recombination (3). To assess the ability
126 of the *L. pneumophila dif* site to undergo site-specific recombination, we used an intermolecular
127 recombination assay (Fig. 2A). In this assay, a non-replicative plasmid (pJB4648) containing a
128 gentamicin-resistance marker and a 22-bp portion of the *dif* site (corresponding to the LME-1 *att*
129 site) is transformed into *L. pneumophila* strain Lp02. If site-specific recombination occurs between
130 the *att* site on the plasmid and the chromosomal *dif* site, the entire plasmid will be integrated into the
131 genome and confer gentamicin resistance to the cell. We compared the number of transformants
132 resulting from the *att*-containing plasmid to one with a control 22-bp sequence derived from the *L.*
133 *pneumophila* genome near the *dif* site. To evaluate the efficiency of recombination supported by the

134 *dif* site, we also included a plasmid containing 2200 bp of unrelated intergenic sequence from *L.*
135 *pneumophila*, which can integrate into the genome via a single-crossover homologous recombination
136 event. Considering that chromosomal *dif* site-specific recombination is known to be RecA-
137 independent (13, 16, 35), we also assessed the integration of each of these plasmids in a *recA*
138 deletion strain (Fig. 2B). To control for any strain-to-strain differences in overall transformation
139 efficiency, we normalized the number of transformants for each plasmid to the number resulting
140 from transformation with a replicative plasmid (pJB1806_GmR). The *att* site-containing plasmid
141 integrated into the genome at a high efficiency that was 50-fold higher than that of the 2200-bp
142 control. In contrast, the 22-bp control sequence did not generate any gentamicin-resistant
143 recombinants, indicating that the number of integration events was lower than the detection limit of
144 1 CFU per ~$2 \times 10^9$ cells. In the Δ*recA* strain, no transformants were detected for the 2200-bp
145 control which relies on homologous recombination for integration, while the *att* site-containing
146 plasmid integrated at a level similar to that in the wildtype strain (Fig. 2B). In both the wildtype and
147 Δ*recA* strains, integration of the *att*-containing plasmid at the chromosomal *dif* site was confirmed
148 by PCR-amplification of the *dif* region and Sanger sequencing (data not shown). These data indicate
149 that, similar to other *dif* sites, site-specific recombination at the *L. pneumophila dif* site is efficient
150 and RecA-independent.
151
152 ***Legionella* Xer/*dif* is a single-recombinase system.**
153 We next aimed to identify the recombinase(s) that catalyzes recombination at the *Legionella*
154 *dif* site. No orthologs of XerC, XerD, XerS, or XerH have been identified in *Legionella* using
155 sequence homology (29). However, HHPred analysis (36) revealed that the proteins encoded by six
156 *L. pneumophila* genes present in strain Lp02 (*lpg0980*, *lpg0981*, *lpg1070*, *lpg1085*, *lpg1867*, and

157 *lpg2057*) show structural similarity to tyrosine-type recombinases (probability >99%). We therefore
158 repeated our intermolecular recombination assay in strains containing individual deletions of these
159 six genes, along with a strain lacking the *dif* site. As expected, deletion of the *dif* site resulted in no
160 detectable integration of the *att* site-containing plasmid. Deletion of *lpg1867* also specifically
161 abolished integration, while deletion of each of the other five recombinases had no significant effect
162 (Fig. 3A). We also performed rescue experiments using expression plasmids to confirm that the loss
163 of recombination at the *dif* site in the Δ*lpg1867* strain was due to loss of Lpg1867 protein function
164 and not any secondary mutations acquired during strain generation. Expression of Lpg1867 in the
165 Δ*lpg1867* strain resulted in near-wildtype levels of recombination (Fig. 3B). In contrast, no
166 recombination was detected with an empty vector control or when a point mutant of Lpg1867
167 (Y387F) was expressed. The Y387F mutant is predicted to be catalytically inactive based on the
168 equivalent catalytic tyrosine mutation in other tyrosine recombinases (10, 37, 38).
169 We next asked whether one of the other five putative tyrosine recombinases might be
170 involved along with Lpg1867 in recombination at the *dif* site, despite having no effect upon
171 individual deletion. To address this, we generated two multiple-deletion strains: one with five
172 putative recombinases deleted (all except *lpg1867*), and a pan-deletion strain with all six deleted. In
173 our intermolecular recombination assay, integration of the *att*-containing plasmid was similar to
174 wildtype levels in the 5-deletion strain and nonexistent in the pan-deletion strain (Fig. 4A).
175 Overexpression of Lpg1867, but not Lpg1867-Y387F, again rescued the recombination defect (Fig.
176 4B).
177 As chromosome dimer resolution involves recombination between two *dif* sites on the same
178 molecule, we also wanted to test whether Lpg1867 was sufficient to catalyze intramolecular
179 recombination between two *dif* sites. To this end, we developed an intramolecular recombination

180 assay by inserting a *dif* site-flanked kanamycin-resistance cassette at the normal *dif* locus of the
181 Lp02 pan-deletion strain, such that recombination between the two flanking *dif* sites would result in
182 loss of kanamycin resistance. To quantify excision of the reporter, we determined the proportion of
183 kanamycin-resistant cells in the population after transforming the cells with wildtype or mutant
184 Lpg1867. Overexpression of wildtype Lpg1867 in this strain resulted in a high level of cassette
185 excision, while overexpression of the catalytic mutant resulted in an excision level similar to
186 background level (Fig. 4C). Together, these data suggested that Lpg1867 does not require a partner
187 recombinase for recombination at the *dif* site. However, to rule out the possibility that another
188 unidentified recombinase or cofactor in *L. pneumophila* is also required, we performed a similar
189 intramolecular excision assay in yeast. In this case, a *dif* site-flanked KanMX cassette was inserted at
190 the HO locus of the *S. cerevisiae* BY4741 strain, and recombination between the two *dif* sites was
191 assessed by quantifying loss of geneticin resistance (Fig. 4C). Overexpression of Lpg1867, but not
192 Lpg1867-Y387F, again led to high levels of excision. Collectively, these data indicate that Lpg1867
193 is both necessary and sufficient to catalyze site-specific recombination at the *L. pneumophila dif* site.
194

195 **Deletion of *dif* or *lpg1867* induces filamentation and inhibits extracellular and intracellular**
196 **growth.**

197 Loss of a functional chromosome dimer resolution system causes defects in cell division,
198 which can result in a variety of phenotypes including slow growth and filamentation (the *dif* site was
199 named for its deletion-induced filamentation phenotype). Consistent with reported findings for *dif* or
200 *xer* deletions in other species (11, 13, 14, 18, 39, 40), we observed that loss of *dif* or Lpg1867 in *L.*
201 *pneumophila* led to filamentation in a subpopulation of cells (Fig. 5A). Notably, and consistent with

202 a reduction in chromosome dimer formation in the absence of RecA (2, 41), we observed a decrease
203 in filament formation upon deletion of *recA* in the Δ*lpg1867* background (Fig. 5A).
204 Deletion of *dif* or *lpg1867* also results in a visible reduction in colony growth on solid media
205 (Fig. 5B). Again, this defect is alleviated upon additional deletion of *recA* in the *Δlpg1867*
206 background (Fig. 5B), consistent with the growth inhibition being caused by an inability to resolve
207 chromosome dimers. The slow-growth phenotype is also present in the recombinase pan-deletion
208 strain, but not the 5-deletion strain, and the defect is rescued by expression of wildtype Lpg1867 but
209 not the catalytic mutant (Fig. 5C), mirroring the effects of these deletions on recombination (see Fig.
210 4A,B). Importantly, the *dif* and *lpg1867* deletions also result in diminished growth in human U937
211 macrophages (Fig. 5D), indicating that growth within host cells is also impacted by disruptions to
212 *dif*/Xer activity. Taken together, these data support a model in which a single site-specific
213 recombinase, Lpg1867 (which we designate here as XerL), catalyzes recombination between two *dif*
214 sites to resolve chromosome dimers in *Legionella*.
215
216 **LME-1 stability is modulated by the formation of a modified downstream *dif* site.**
217 We have previously shown that LME-1 integrates into the *L. pneumophila* chromosome at
218 what we now know to be the *dif* site (33). LME-1 is stably integrated into the genome of str. Murcia-
219 4983 (only excised in ~1% of the population), despite being flanked by two *att* sites (33). The
220 sequence of the LME-1 *att* site is identical to the *L. pneumophila dif* site, but only encompasses 22-
221 bp of it. Recombination between this 22-bp sequence and the chromosomal *dif* site results in *att*-site
222 duplication, which generates an intact *dif* site upstream and a modified *dif* site downstream of LME-
223 1 that contains three substitutions (Fig. 6A). We next asked whether these modifications to the
224 downstream *dif* site contribute to the stability of the integrated form of LME-1. We modified our

225 excision assay to use the kanamycin-resistance cassette as a proxy for LME-1, and determined the
226 effect of adding one or more of the LME-1 *att* substitutions to the downstream *dif* site (Fig. 6B). We
227 found that adding all three substitutions to the downstream *dif* site (to mimic integrated LME-1)
228 resulted in markedly reduced excision compared to the reporter flanked by two wildtype *dif* sites,
229 while one or two substitutions resulted in an intermediate level of excision (Fig. 6B). These data
230 suggest that the short 22-bp *att* site of LME-1 likely contributes to the stability of its integrated form
231 by preventing it from being flanked by two intact *dif* sites.
232
233 **Widespread distribution of *xerL* orthologs in *Legionella* and *Coxiella*.**
234 Our discovery of the *Legionella dif*/Xer pathway, which had been missed in homology-based
235 searches due to divergence from known CDR components, led us to next examine other bacteria in
236 which no *dif*/Xer system has been identified to date. One possibility is that several species have
237 XerL-like systems that were missed for the same reasons that they were missed in *Legionella*. To
238 explore this possibility, we performed BLASTp and tBLASTn analyses using the *L. pneumophila*
239 XerL protein as a query. These analyses identified several additional XerL orthologs across the order
240 Legionellales, which includes *Legionella* sp. as well as the select agent, *Coxiella burnetii* (Fig. 7).
241 Phylogenetic analysis of the XerL-like recombinases indicates widespread distribution across
242 Legionellales (Fig. S1). The amount of sequence divergence between XerL and the rest of the Xer
243 protein family likely explains how this machinery has been missed by prior analyses.
244
245 **The *Coxiella burnetii xerL* ortholog is encoded on a virulence plasmid.**
246 Despite the apparent vertical inheritance of *xerL* given its restriction within the order
247 Legionellales, the *Coxiella burnetii* orthologs of *xerL* (hereafter $xerL_{Cb}$) reside not on the circular

248 chromosome, but on a large ubiquitous plasmid that is critical for virulence (42–45). While Xer/*dif*
249 systems have established roles in resolving aberrant plasmid dimers, these are the only Xer
250 sequences within the *C. burnetii* genome. This raises the question as to whether this plasmid-based
251 XerL functions to resolve plasmid dimers, chromosome dimers, or both.

252 To search for putative *dif* sites within the *Coxiella* genome, we performed a sequence
253 similarity (BLASTn) search using the 29-nt *L. pneumophila dif* site. We identified a short sequence
254 of 30 nucleotides on the virulence plasmid immediately adjacent to $XerL_{Cb}$ (Fig. 8A). Using this
255 second sequence as a query, we identified an additional site on the chromosome (Fig. 8A). This
256 chromosomal sequence diverges from the plasmid site yet has several features consistent with a
257 functioning *dif* site. Like the *L. pneumophila dif* sequence, it is palindromic, highly conserved across
258 isolates, and has no apparent coding potential.

259 To further investigate the potential role of this chromosomal site as a functional *dif* site, we
260 examined its location relative to the point of inflection of GC skew (replication terminus). As we
261 observed for $dif_{Lp}$, the *Coxiella burnetii* chromosomal site is always located opposite the predicted
262 origin of replication and close to the cumulative GC-skew maximum (Fig. 8B, Table S2). Taken
263 together, this sequence not only shares homology to $dif_{Lp}$, but maintains several hallmarks of a *dif*
264 site involved in chromosome dimer resolution. Hereafter, we will refer to it as $dif_{Cb}$.

265

266 **The *Coxiella* XerL/*dif* system functionally complements a Δ*xerL*/*dif* mutant in *Legionella*.**

267 Having identified putative *dif* sites and *xerL* orthologs in *Coxiella burnetii*, we next wanted
268 to assay their potential for chromosome dimer resolution. Considering that colony size correlates
269 with *dif*/Xer function in *L. pneumophila* (Fig. 5), we generated a strain in which the endogenous *dif*
270 site and *xerL* gene were replaced by the *C. burnetii dif* site and *xerL*, and assessed growth on solid

271 media. If the *C. burnetii dif*/XerL were not able to resolve dimers in their surrogate chromosome,
272 then we would expect small colony sizes similar to the Δ*dif* or Δ*xerL* strains. However, the colony
273 size distribution of the strain containing $XerL_{Cb}$ and the *C. burnetii* chromosomal *dif* site ($dif_{Cb_chr}$)
274 was almost identical to that of wildtype *L. pneumophila* (Fig. 8C), while a strain encoding the
275 predicted catalytic mutant XerL-Y389F displayed the small colony sizes associated with
276 nonfunctional *dif*/Xer (Fig. 8C). This suggests that despite being encoded on the plasmid, $XerL_{Cb}$
277 may catalyze recombination at $dif_{Cb_chr}$ to resolve dimers of the *Coxiella* chromosome. Interestingly,
278 despite containing several substitutions (Fig. 8A) the *Coxiella* plasmid *dif* site ($dif_{Cb_pl}$), in concert
279 with $XerL_{Cb}$, is also able to complement the small colony phenotype (Fig. 8C). Similarly, in our
280 intramolecular recombination assay, the reporter cassette was excised when flanked by either
281 $dif_{Cb_chr}$ or $dif_{Cb_pl,}$ although excision for the cassette flanked by $dif_{Cb_pl}$ required overnight induction
282 of $XerL_{Cb}$ (Fig 8D).
283 Together, these findings suggest that $XerL_{Cb}$ may resolve both plasmid and chromosome
284 dimers in *Coxiella*, with the resolution of chromosome dimers uniquely dependent on plasmid-
285 encoded XerL. XerL is encoded by all four virulence plasmids described in *C. burnetii* and no
286 additional Xer orthologs can be found on the chromosome. This suggests that chromosome dimer
287 resolution and virulence plasmid maintenance are genetically linked in *C. burnetii*, with plasmid loss
288 leading to loss of the chromosome dimer resolution pathway. Consistent with this model, we note
289 that previously described instances of plasmidless strains – in which a subset of plasmid-like
290 sequences are instead integrated onto the chromosome (46–48) – contain *xerL* within the integrated
291 sequence. Our findings suggest that such integrants are likely selected for based on their ability to
292 escape the fitness costs that a plasmidless strain would otherwise incur. Despite the sequence
293 similarity between the *dif* sites located on the plasmid and chromosome, and their shared dependence

294 on XerL, the integration of plasmid-like sequences in the plasmidless strains is not at the
295 chromosomal *dif* site. In fact, the plasmid fragments are located ~1 Mb from the terminus region,
296 indicating that acquisition of the plasmid-like sequences did not originate with an IMEX-like
297 plasmid integration event. To see if this phenomenon would recapitulate in our intermolecular
298 recombination assay, we measured integration of a non-replicative plasmid containing $dif_{Cb_pl}$ into
299 the genome of a strain containing $dif_{Cb_chr}$ and found it to be almost 300-fold lower than when the
300 genome contained $dif_{Cb_pl}$ (Fig. 8E). Surprisingly, the efficiency of integration into the $dif_{Cb_chr}$-
301 containing genome was also very low when the non-replicative plasmid contained the identical
302 $dif_{Cb_chr}$ (Fig. 8E). One possible explanation is that highly efficient recombination between the two
303 $dif_{Cb_chr}$ sites leads to rapid re-excision after integration, resulting in no visible gentamicin-resistant
304 colonies. However, when we used a version of $dif_{Cb_chr}$ with a shortened region of dyad symmetry,
305 which is known to increase stability of integration for the LME-1 *dif* mimic (see Fig. 6),
306 transformation efficiency remained very low (Fig. 8E). These findings suggest that either shortened
307 dyad symmetry does not stabilize $dif_{Cb_chr}$ or the chromosomal *dif* site itself might be recalcitrant to
308 invasion. The latter notion is consistent with the fact that no sequenced *Coxiella burnetii* genomes
309 contain virulence plasmids as chromosomal integrants at the *dif* site.
310

311 **DISCUSSION**

312 The discovery of the *L. pneumophila* CDR system has solved two mysteries of the
313 pathogen's biology: the function of the conserved DNA sequence that is hijacked by LME-1, and
314 how *Legionella* species are able to resolve chromosome dimers despite having no homologs to
315 known *dif* sites and Xer recombinases. Similarly, no *dif*/Xer components had been identified
316 previously in members of the other Legionellales family, *Coxiellaceae* (29), which includes the

317 human pathogen *Coxiella burnetii*. It is rare that you uncover an overlooked, essential pathway in a
318 pathogen. We have identified one that is conserved across an entire order. Other extensively studied
319 dimensions of *Legionella* biology (e.g. translocated effectors) are not strong candidates for drug
320 targets due to their redundancy. In contrast, we show that disrupting one protein in this pathway
321 (XerL) is sufficient to dramatically restrict the growth of the pathogen.

322 With no apparent *dif* or Xer homologs in *Legionella*, the CDR system was instead unearthed
323 through our investigation of LME-1 and its attachment site. In turn, the discovery of the *Legionella*
324 *dif* site and XerL has allowed us to explore new LME-1 biology. Our findings are consistent with
325 LME-1 belonging to the IMEX class of integrative mobile elements, which exploit the host CDR
326 machinery for integration and in some cases excision (5, 28). The mechanistic details of LME-1
327 integration and excision, and how they relate to the strategies of other IMEXs, are interesting topics
328 of future study. One intriguing aspect of LME-1 is the stability of its integrated form. Our reporter
329 excision assay showed that the modified *dif* site and wildtype *dif* site that flank the integrated LME-1
330 recombine at a very low level, and that the stability of integration is influenced by the three
331 nucleotide substitutions present in the modified form. Another stably integrated IMEX, the
332 gonococcal genomic island (GGI) of *Neisseria gonorrhoeae*, generates a similarly modified *dif* site
333 ($dif_{GGI}$) at one end of its integrated form, while the wildtype *dif* ($dif_{Ng}$) remains intact at the other
334 (23). The *Neisseria* CDR system uses XerC and XerD homologs, and recombination takes place
335 when the FtsK DNA translocase pauses at the *dif*/Xer complex and activates XerD. Interestingly,
336 FtsK does not stop at the $dif_{GGI}$-XerC/D complex, but rather translocates through it, which likely
337 leads to disassembly of the complex and precludes recombination between the $dif_{NG}$ *and* $dif_{GGI}$ (23).
338 It is possible that this same FtsK-dependent process contributes to the stability of the integrated form

339 of LME-1, although we have yet to determine whether XerL requires FtsK for activation. Our future
340 investigative priorities include determining the involvement of FtsK in the *L. pneumophila* CDR
341 system and the stability of its associated IMEX, LME-1.

342 The *Legionella* and *Coxiella dif*/Xer systems are distinct in several ways from those
343 characterized previously in other bacterial species. The *dif* site arms consist of long inverted repeats
344 (10-bp and 12-bp, respectively) with no interruptions, whereas other *dif* sites show only partial dyad
345 symmetry between the two arms, even for single-recombinase systems. Additionally, in *L.*
346 *pneumophila*, *dif* and *xerL* are separated by >400 kb in the genome, whereas in the other bacterial
347 single-recombinase systems (*i.e.* $dif_{SL}$/XerS of *Streptococci* and *Lactococci* and $dif_{H}$/XerH of
348 *Campylobacter* and *Helicobacter*), the *dif* site is always near the *xer* gene, indicating they may have
349 been acquired as a single module (3, 16, 17, 29). The location of the *C. burnetii xerL* ortholog on its
350 virulence plasmid is also unusual. While many plasmids use Xer machinery for their own dimer
351 resolution, this is the first instance we are aware of where a plasmid-based Xer recombinase is
352 necessary for chromosome dimer resolution.

353 Given the general importance of chromosome dimer resolution to bacterial fitness and the
354 extent to which intracellular growth is compromised in *L. pneumophila* Δ*xerL* or *dif* mutants, we
355 anticipate that plasmid loss in *Coxiella* incurs a significant fitness cost through the loss of the
356 essential recombinase. Consistent with this model, isolates that have lost their plasmid appear to
357 maintain *dif*/XerL activity by including *xerL* within the fragments of plasmid sequence found to be
358 integrated on the chromosome. One interpretation of these results is that integrants that do not
359 maintain *xerL* chromosomally are rapidly selected against and not recovered. In light of our data, it
360 will be important to distinguish which phenotypes previously associated with directed plasmid-loss

361 (44) reflect virulence-specific defects (associated with loss of specific virulence factors, such as
362 effectors) and which phenotypes result from the loss of *dif*/XerL machinery.
363
364 **MATERIALS AND METHODS**
365 **Strains and plasmids**
366 The *L. pneumophila* Lp02 strain was used as the background for all *L. pneumophila* strains
367 generated in this study. The *L. pneumophila* Murcia-4983 strain, which harbors LME-1, is an
368 environmental isolate collected during the 2001 outbreak in Murcia, Spain (49). All *L. pneumophila*
369 strains were grown at 37°C in N-(2-acetamido)-2-aminoethanesulfonic acid (ACES) buffered yeast
370 extract supplemented with 100 ug/mL thymidine (AYET) and on charcoal AYET (CYET) agar
371 plates.
372 Deletion and mutant strains were made using a scar-free suicide-cassette method described
373 recently (50) with some minor modifications. Briefly, linear DNA containing a mazF-KanR cassette
374 flanked by ~1.5-kb homology arms was introduced into the parental strain via natural
375 transformation. Cells in which the mazF-KanR cassette had been integrated via homologous
376 recombination were selected by plating on media containing kanamycin. As mazF gene expression is
377 under an IPTG-inducible promoter, colonies can be screened on kanamycin- and IPTG-containing
378 plates to identify cells that are IPTG-sensitive and kanamycin-resistant. In a second natural
379 transformation, linear DNA containing only the homology arms with the deletion/mutation was
380 introduced, and cells in which the cassette was excised were selected on plates containing IPTG.
381 Colonies were then screened to identify those that were IPTG-resistant and kanamycin-sensitive, and
382 the deletion was verified by PCR screening and Sanger sequencing.

383 The reporter strains used for the intramolecular recombination (reporter excision) assays
384 were generated by introducing linear DNA containing the *dif*-KanR-*dif* cassette (flanked by ~1.5 kb
385 homology arms around the chromosomal *dif* site) into the Lp02 Δ*lpg1867* strain or pan-deletion
386 strain. Strains containing the cassette were selected on plates containing kanamycin, and the correct
387 integration of the cassettes was verified by Sanger sequencing. Similarly, for the yeast-based
388 excision assay, linear DNA containing the *dif*-KanMX-*dif* cassette (flanked by homology arms
389 around the HO locus) was transformed into *S. cerevisiae strain* BY4741 using the high efficiency
390 PEG/LiAc method (51). Integrants were selected on YPD agar plates containing geneticin and
391 verified by Sanger sequencing.
392 The pJB4648-based plasmids used for the intermolecular recombination assays were
393 constructed by cloning the 22-bp *att* site sequence, a control 22-bp sequence, or 2200 bp of
394 intergenic Lp02 sequence into pJB4648 digested with XhoI and ApaI. The gentamicin-resistant
395 plasmid pJB1806-GmR was generated by first amplifying the gentamicin resistance gene and
396 promoter from pJB4648 and cloning it in place of the chloramphenicol resistance gene in plasmid
397 pJB1806. Protein expression plasmids used for the recombination and colony size assays were made
398 by cloning the *lpg1867* gene (wildtype or modified to encode the Y387F mutation) into the multiple
399 cloning site of pNT562 (52). The tet-on-pJB1806-GmR-based expression plasmids used for the
400 reporter excision assays were generated by replacing the *lacIq* gene and tac promoter in pJB1806-
401 GmR with the *tetR* gene and tetracycline-responsive promoter from pLJR965 (53) and then cloning
402 in the wildtype or mutant Lpg1867 ORF immediately after the promoter, as described by Rock and
403 colleagues (53).
404
405 **GC-skew analysis**

406 The cumulative GC-skew maximum value for each *Legionella* genome (Table S1) or *C.*
407 *burnetii* genome (Table S2) was determined using the GenSkew java app
408 (https://genskew.csb.univie.ac.at) with default settings. The whole-genome GC-skew profiles and *dif*
409 site locations for seven *L. pneumophila* isolates, seven *Legionella* species, and seven *C. burnetii*
410 strains, selected to show a range of cGC-skew maxima locations relative to the origin, were plotted
411 using Circos (54). The GC-skew profiles of these genomes were normalized for genome length. The
412 selected *L. pneumophila* isolates and *Legionella* species are: *L. pneumophila*: Philadelphia_2
413 (GenBank accession number CP015929.1), Philadelphia_1 (AE017354.1), E7_O (CP015954.1),
414 Paris (CR628336.1), NY23 D-7705 (CP021261.1), NCTC11985 (LT906452.1), and subsp. fraseri
415 strain D-4058 (CP021277); *Legionella* species: *L. israelensis* L18-01051 (CP041668.1), *L. micdadei*
416 (LN614830.1), *L. spiritensis* NCTC11990 (LT906457.1), *L. sainthelensi* NCTC12450
417 (LR134178.1), *L. longbeachae* NSW150 (FN650140.1), *L. oakridgensis* NCTC11531
418 (LR134286.1), and *L. fallonii* LLAP-10 (LN614827.1). The selected *C. burnetii* isolates are: 3262
419 (CP013667.1), Heizburg (CP014561.1), Dugway FJ108-111 (CP000733.1), RSA 493 (AE016828.3),
420 CbuG_Q212 (CP001019.1), Schperling (CP014563.1), and CbuK_Q154 (CP001020.1).
421
422 **Recombination assays**
423 For the intermolecular recombination (plasmid integration) assays, each strain was
424 transformed with a pJB4648-based plasmid containing the indicated *att*, *dif*, or control sequences as
425 indicated. The strains were also transformed with the replicative plasmid pJB1806-GmR to control
426 for strain-to-strain differences in general transformation efficiency. Plasmid concentrations were
427 determined using a Quant-iT PicoGreen dsDNA kit (Invitrogen), and 200 ng of each plasmid was
428 electroporated into 2 ODU of *L. pneumophila* cells as described previously (55). After an 8-h

429 recovery in liquid AYET media shaking at 37°C, dilutions of each sample were plated onto CYET
430 plates supplemented with 15 µg/mL gentamicin and incubated at 37°C for 4 days. Colonies were
431 counted using GeneTools analysis software (Syngene), and the results were plotted using GraphPad
432 Prism version 9.2.0.

433 For the intramolecular recombination (excision) assay in *Legionella*, strains containing the
434 kanamycin-resistance cassette were transformed with tet-on-pJB1806-GmR containing wildtype
435 Lpg1867, Lpg1867-Y387F, $XerL_{Cb}$, $XerL_{Cb}$-Y389F, or empty vector by electroporation as described
436 above. Transformants were selected on CYET plates containing both gentamicin and kanamycin.
437 The transformant colonies were then collected in AYE and spread onto nonselective plates and
438 plates containing kanamycin to quantify the proportion of the population that had lost kanamycin
439 resistance. For the intramolecular recombination (excision) assay in yeast, *S. cerevisiae* strain
440 BY4741 containing the *dif*-KanMX-*dif* cassette was transformed with yeast vector pAG413GPD
441 (56) expressing wildtype or mutant Lpg1867, or empty vector. Percent excision was quantified by
442 determining the proportion of the transformants that had lost geneticin resistance by plating on
443 selective and nonselective plates.

444

445 **Fluorescence microscopy**

446 Overnight cultures of Lp02, Lp02*Δdif*, Lp02*Δlpg1867*, and Lp02*Δlpg1867ΔrecA* were grown
447 from a 2-day old patch in AYET medium. Bacteria were collected at post-exponential phase ($OD_{600}$
448 4.0 to 4.5). Approximately ~$1 \times 10^9$ cells (1 ODU) from each strain were washed once with 1 mL of
449 1× PBS before a final resuspension in 500 µL of 1× PBS. The bacteria were stained with 15 µg/mL
450 of DAPI (Roche) for 30 min at room temperature and imaged using a 63× oil immersion lens on a
451 Zeiss Axio Imager.M2 microscope.

452

453 **Colony size assay**

454 The distribution of colony size was determined by resuspending 2-day-old patches of the
455 indicated strains in AYET. The $OD_{600}$ of each cell suspension was determined, and the appropriate
456 dilutions made so that ~200-300 cells were added to each plate. Spread plates were incubated at
457 37°C for 4 days before imaging using a SynGene system with GeneSnap software (Syngene).
458 Colony diameter was determined using ColTapp automated image analysis software (57), and
459 plotted using GraphPad Prism version 9.2.0.

460

461 **Bacterial infection of U937 macrophages**

462 TPA-differentiated U937 cells were seeded in a 24-well plate at $4 \times 10^5$ cells per well in 500
463 µL of RPMI supplemented with glutamine and 10% heat-inactivated fetal bovine serum. Cells were
464 incubated overnight at 37°C with 5% $CO_2$. Overnight cultures of *Legionella* were grown to post-
465 exponential phase (OD 4.0–4.5 and motile) and used to inoculate the U937 cells at an MOI of 0.05.
466 After 2 h, cells were washed three times with media to remove extracellular bacteria before
467 incubating in fresh media at 37°C with 5% $CO_2$ for a further two days. At 2 h and 48 h post-
468 infection, cells were lysed with 0.02% saponin and then plated on CYET to determine the number of
469 bacterial colony-forming units.

470

471 **Bioinformatic analysis of XerL**

472 The amino acid sequence of XerL from *L. pneumophila* was used to uncover putative
473 orthologs of XerL. XerL was queried against the NCBI Legionellales database (taxid:118969) using
474 both BLASTp and tBLASTn with the default settings (58), in addition to a tBLASTn search of

475 additional *Legionella* species that were deposited in the NCBI Sequence Read Archive (59, 60). The
476 resulting putative XerL orthologs can be found in Table S3.
477 An unrooted phylogeny of Xer proteins was generated using the amino acid sequences from
478 representative XerL orthologs, in addition to reference Xer sequences (Table S3). The amino acid
479 sequences were aligned using MUSCLE (61) and the tree was generated with the FastTree (v 2.1.11)
480 plugin on Geneious Prime using the default settings (62). The resulting tree was visualized using the
481 Interactive Tree of Life (iTOL) server (v 6) (63).
482 A rooted phylogeny of all putative XerL orthologs was generated as described above, using
483 XerD from *Escherichia coli* as the root.
484
485 **Materials and Data Availability**
486 All strains and plasmids are available upon request. Sequences and locations of *dif* sites are
487 listed in figures and Tables S1 and S2. The accession numbers and amino acid sequences for Xer
488 proteins are listed in Table S3.
489
490 **ACKNOWLEDGEMENTS**
491 We thank members of the Ensminger laboratory for their suggestions and careful reading of the
492 manuscript, and Jordan Lin and John MacPherson for their help with the recombination assays and
493 HHPred analysis. We also thank Hayley Newton for the kind gift of genomic DNA from *C. burnetii*
494 strain nine mile phase II. SRD is supported by a fellowship from the Department of Biochemistry,
495 University of Toronto and an Ontario Graduate Scholarship. This work was supported by a Project
496 Grant from the Canadian Institutes of Health Research (PHT-148819).
497

498 **REFERENCES**


499 1. Barre FX, Søballe B, Michel B, Aroyo M, Robertson M, Sherratt D. 2001. Circles: the replication-
500 recombination-chromosome segregation connection. Proc Natl Acad Sci U S A 98:8189–8195.

501 2. Steiner WW, Kuempel PL. 1998. Sister Chromatid Exchange Frequencies in Escherichia coli
502 Analyzed by Recombination at the dif Resolvase Site. J BACTERIOL 180:7.

503 3. Castillo F, Benmohamed A, Szatmari G. 2017. Xer Site Specific Recombination: Double and
504 Single Recombinase Systems. Front Microbiol 8.

505 4. Diagne CT, Salhi M, Crozat E, Salomé L, Cornet F, Rousseau P, Tardin C. 2014. TPM analyses
506 reveal that FtsK contributes both to the assembly and the activation of the XerCD-dif
507 recombination synapse. Nucleic Acids Res 42:1721–1732.

508 5. Midonet C, Barre F-X. 2014. Xer Site-Specific Recombination: Promoting Vertical and
509 Horizontal Transmission of Genetic Information. Microbiol Spectr 2.

510 6. Thanbichler M. 2010. Synchronization of Chromosome Dynamics and Cell Division in Bacteria.
511 Cold Spring Harbor Perspectives in Biology 2:a000331–a000331.

512 7. Austin S, Ziese M, Sternberg N. 1981. A novel role for site-specific recombination in
513 maintenance of bacterial replicons. Cell 25:729–736.

514 8. Summers DK, Sherratt DJ. 1984. Multimerization of high copy number plasmids causes
515 instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability.
516 Cell 36:1097–1103.

517 9. Colloms SD, Sykora P, Szatmari G, Sherratt DJ. 1990. Recombination at ColE1 cer requires the
518 Escherichia coli xerC gene product, a member of the lambda integrase family of site-specific
519 recombinases. J Bacteriol 172:6973–6980.

520 10. Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, Sherratt DJ. 1993. Two
521 related recombinases are required for site-specific recombination at dif and cer in E. coli K12.
522 Cell 75:351–361.

523 11. Blakely G, Colloms S, May G, Burke M, Sherratt D. 1991. Escherichia coli XerC recombinase is
524 required for chromosomal segregation at cell division. New Biol 3:789–798.

525 12. Clerget M. 1991. Site-specific recombination promoted by a short DNA segment of plasmid R1
526 and by a homologous segment in the terminus region of the Escherichia coli chromosome.
527 New Biol 3:780–788.

528 13. Kuempel PL, Henson JM, Dircks L, Tecklenburg M, Lim DF. 1991. dif, a recA-independent
529 recombination site in the terminus region of the chromosome of Escherichia coli. New Biol
530 3:799–811.

531 14. Debowski AW, Carnoy C, Verbrugghe P, Nilsson H-O, Gauntlett JC, Fulurija A, Camilleri T, Berg
532 DE, Marshall BJ, Benghezal M. 2012. Xer recombinase and genome integrity in Helicobacter
533 pylori, a pathogen without topoisomerase IV. PLoS One 7:e33310.

534 15. Jensen RB. 2006. Analysis of the terminus region of the Caulobacter crescentus chromosome
535 and identification of the dif site. J Bacteriol 188:6016–6019.

536 16. Le Bourgeois P, Bugarel M, Campo N, Daveran-Mingot M-L, Labonté J, Lanfranchi D, Lautier T,
537 Pagès C, Ritzenthaler P. 2007. The Unconventional Xer Recombination Machinery of
538 Streptococci/Lactococci. PLoS Genet 3:e117.

539 17. Leroux M, Rezoug Z, Szatmari G. 2013. The Xer/dif site-specific recombination system of
540 Campylobacter jejuni. Mol Genet Genomics 288:495–502.

541 18. Sciochetti SA, Piggot PJ, Blakely GW. 2001. Identification and characterization of the dif Site
542 from Bacillus subtilis. J Bacteriol 183:1058–1068.

543 19. Val M-E, Kennedy SP, El Karoui M, Bonné L, Chevalier F, Barre F-X. 2008. FtsK-dependent
544 dimer resolution on multiple chromosomes in the pathogen Vibrio cholerae. PLoS Genet
545 4:e1000201.

546 20. Cortez D, Quevillon-Cheruel S, Gribaldo S, Desnoues N, Sezonov G, Forterre P, Serre M-C.
547 2010. Evidence for a Xer/dif System for Chromosome Resolution in Archaea. PLoS Genet
548 6:e1001166.

549 21. Duggin IG, Dubarry N, Bell SD. 2011. Replication termination and chromosome dimer
550 resolution in the archaeon Sulfolobus solfataricus. EMBO J 30:145–153.

551 22. Das B, Bischerour J, Barre F-X. 2011. VGJphi integration and excision mechanisms contribute
552 to the genetic diversity of Vibrio cholerae epidemic strains. Proc Natl Acad Sci U S A 108:2516–
553 2521.

554 23. Fournes F, Crozat E, Pages C, Tardin C, Salomé L, Cornet F, Rousseau P. 2016. FtsK
555 translocation permits discrimination between an endogenous and an imported Xer/ *dif*
556 recombination complex. Proc Natl Acad Sci USA 113:7882–7887.

557 24. Hassan F, Kamruzzaman M, Mekalanos JJ, Faruque SM. 2010. Satellite phage TLCϕ enables
558 toxigenic conversion by CTX phage through dif site alteration. Nature 467:982–985.

559 25. Huber KE, Waldor MK. 2002. Filamentous phage integration requires the host recombinases
560 XerC and XerD. Nature 417:656–659.

561 26. Midonet C, Das B, Paly E, Barre F-X. 2014. XerD-mediated FtsK-independent integration of
562 TLCφ into the Vibrio cholerae genome. Proc Natl Acad Sci U S A 111:16848–16853.

563 27. Val M-E, Bouvier M, Campos J, Sherratt D, Cornet F, Mazel D, Barre F-X. 2005. The single-
564 stranded genome of phage CTX is the form used for integration into the genome of Vibrio
565 cholerae. Mol Cell 19:559–566.

566 28. Das B, Martínez E, Midonet C, Barre F-X. 2013. Integrative mobile elements exploiting Xer
567 recombination. Trends in Microbiology 21:23–30.

568 29. Carnoy C, Roten C-A. 2009. The dif/Xer Recombination Systems in Proteobacteria. PLoS ONE
569 4:e6531.

570 30. Deecker SR, Urbanus ML, Nicholson B, Ensminger AW. 2021. Legionella pneumophila CRISPR-
571 Cas Suggests Recurrent Encounters with One or More Phages in the Family Microviridae. Appl
572 Environ Microbiol 87:e0046721.

573 31. Lüneberg E, Zetzmann N, Alber D, Knirel YA, Kooistra O, Zähringer U, Frosch M. 2000. Cloning
574 and functional characterization of a 30 kb gene locus required for lipopolysaccharide
575 biosynthesis in Legionella pneumophila. Int J Med Microbiol 290:37–49.

576 32. Lüneberg E, Mayer B, Daryab N, Kooistra O, Zähringer U, Rohde M, Swanson J, Frosch M. 2001.
577 Chromosomal insertion and excision of a 30 kb unstable genetic element is responsible for
578 phase variation of lipopolysaccharide and other virulence determinants in Legionella
579 pneumophila. Mol Microbiol 39:1259–1271.

580 33. Rao C, Guyard C, Pelaz C, Wasserscheid J, Bondy-Denomy J, Dewar K, Ensminger AW. 2016.
581 Active and adaptive *Legionella* CRISPR-Cas reveals a recurrent challenge to the pathogen.
582 Cellular Microbiology 18:1319–1338.

583 34. Grigoriev A. 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Research
584 26:2286–2290.

585 35. Domínguez NM, Hackett KT, Dillard JP. 2011. XerCD-Mediated Site-Specific Recombination
586 Leads to Loss of the 57-Kilobase Gonococcal Genetic Island. Journal of Bacteriology 193:377–
587 388.

588 36. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology
589 detection and structure prediction. Nucleic Acids Res 33:W244-248.

590 37. Gibb B, Gupta K, Ghosh K, Sharp R, Chen J, Van Duyne GD. 2010. Requirements for catalysis in
591 the Cre recombinase active site. Nucleic Acids Res 38:5817–5832.

592 38. Pan G, Luetke K, Sadowski PD. 1993. Mechanism of cleavage and ligation by FLP recombinase:
593 classification of mutations in FLP protein by in vitro complementation analysis. Mol Cell Biol
594 13:3167–3175.

595 39. Farrokhi A, Liu H, Szatmari G. 2019. Characterization of the Chromosome Dimer Resolution
596 Site in *Caulobacter crescentus*. J Bacteriol 201:e00391-19, /jb/201/24/JB.00391-19.atom.

597 40. Hendricks EC, Szerlong H, Hill T, Kuempel P. 2000. Cell division, guillotining of dimer
598 chromosomes and SOS induction in resolution mutants (dif, xerC and xerD) of Escherichia coli.
599 Molecular Microbiology 9.

600 41. Pérals K, Cornet F, Merlet Y, Delon I, Louarn JM. 2000. Functional polarization of the
601 Escherichia coli chromosome terminus: the dif site acts in chromosome dimer resolution only
602 when located between long stretches of opposite polarity. Mol Microbiol 36:33–43.

603 42. Beare PA, Samuel JE, Howe D, Virtaneva K, Porcella SF, Heinzen RA. 2006. Genetic diversity of
604 the Q fever agent, Coxiella burnetii, assessed by microarray-based whole-genome
605 comparisons. J Bacteriol 188:2309–2324.

606 43. Glazunova O, Roux V, Freylikman O, Sekeyova Z, Fournous G, Tyczka J, Tokarevich N, Kovacava
607 E, Marrie TJ, Raoult D. 2005. Coxiella burnetii genotyping. Emerg Infect Dis 11:1211–1217.

608 44. Luo S, Lu S, Fan H, Sun Z, Hu Y, Li R, An X, Uversky VN, Chen Z, Tong Y, Song L. 2021. The
609 Coxiella burnetii QpH1 plasmid is a virulence factor for colonizing bone marrow-derived
610 murine macrophages. J Bacteriol JB.00588-20.

611 45. Voth DE, Beare PA, Howe D, Sharma UM, Samoilis G, Cockrell DC, Omsland A, Heinzen RA.
612 2011. The Coxiella burnetii cryptic plasmid is enriched in genes encoding type IV secretion
613 system substrates. J Bacteriol 193:1493–1503.

614 46. Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, Williams KP, Sobral BW, Kupko
615 JJ, Porcella SF, Samuel JE, Heinzen RA. 2009. Comparative genomics reveal extensive
616 transposon-mediated genomic plasticity and diversity among potential effector proteins
617 within the genus Coxiella. Infect Immun 77:642–656.

618 47. Savinelli EA, Mallavia LP. 1990. Comparison of Coxiella burnetii Plasmids to Homologous
619 Chromosomal Sequences Present in a Plasmidless Endocarditis-Causing Isolatea. Annals of the
620 New York Academy of Sciences 590:523–533.

621 48. Willems H, Ritter M, Jäger C, Thiele D. 1997. Plasmid-homologous sequences in the
622 chromosome of plasmidless Coxiella burnetii Scurry Q217. J Bacteriol 179:3293–3297.

623 49. García-Fulgueiras A, Navarro C, Fenoll D, García J, González-Diego P, Jiménez-Buñuales T,
624 Rodriguez M, Lopez R, Pacheco F, Ruiz J, Segovia M, Balandrón B, Pelaz C. 2003. Legionnaires'
625 disease outbreak in Murcia, Spain. Emerg Infect Dis 9:915–921.

626 50. Bailo N, Kanaan H, Kay E, Charpentier X, Doublet P, Gilbert C. 2019. Scar-Free Genome Editing
627 in Legionella pneumophila. Methods Mol Biol 1921:93–105.

628 51. Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier
629 DNA/PEG method. Nat Protoc 2:31–34.

630 52. Nishida T, Watanabe K, Tachibana M, Shimizu T, Watarai M. 2017. Characterization of the
631 cryptic plasmid pOfk55 from Legionella pneumophila and construction of a pOfk55-derived
632 shuttle vector. Plasmid 90:30–37.

633 53. Rock JM, Hopkins FF, Chavez A, Diallo M, Chase MR, Gerrick ER, Pritchard JR, Church GM,
634 Rubin EJ, Sassetti CM, Schnappinger D, Fortune SM. 2017. Programmable transcriptional
635 repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol
636 2:16274.

637 54. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009.
638 Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645.

639 55. Rao C, Benhabib H, Ensminger AW. 2013. Phylogenetic reconstruction of the Legionella
640 pneumophila Philadelphia-1 laboratory strains through comparative genomics. PLoS One
641 8:e64129.

642 56. Alberti S, Gitler AD, Lindquist S. 2007. A suite of Gateway cloning vectors for high-throughput
643 genetic analysis in Saccharomyces cerevisiae. Yeast 24:913–919.

644 57. Bär J, Boumasmoud M, Kouyos RD, Zinkernagel AS, Vulin C. 2020. Efficient microbial colony
645 growth dynamics quantification with ColTapp, an automated image analysis application. Sci
646 Rep 10:16084.

647 58. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
648 BLAST+: architecture and applications. BMC Bioinformatics 10:421.

649 59. Burstein D, Amaro F, Zusman T, Lifshitz Z, Cohen O, Gilbert JA, Pupko T, Shuman HA, Segal G.
650 2016. Genomic analysis of 38 Legionella species identifies large and diverse effector
651 repertoires. Nat Genet 48:167–175.

652 60. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE, Rolando M, Pasricha S,
653 Reuter S, Demirtas J, Crumbach J, Descorps-Declere S, Hartland EL, Jarraud S, Dougan G,
654 Schroeder GN, Frankel G, Buchrieser C. 2019. More than 18,000 effectors in the Legionella
655 genus genome provide multiple, independent combinations for replication in human cells.
656 Proc Natl Acad Sci U S A 116:2265–2273.

657 61. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space
658 complexity. BMC Bioinformatics 5:113.

659 62. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for
660 large alignments. PLoS One 5:e9490.

661 63. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
662 display and annotation. Nucleic Acids Res 49:W293–W296.

663 64. Nolivos S, Pages C, Rousseau P, Le Bourgeois P, Cornet F. 2010. Are two better than one?
664 Analysis of an FtsK/Xer recombination system that uses a single recombinase. Nucleic Acids
665 Res 38:6477–6489.

666 65. Bebel A, Karaca E, Kumar B, Stark WM, Barabas O. 2016. Structural snapshots of Xer
667 recombination reveal activation by synaptic complex remodeling and DNA bending. Elife
668 5:e19706.

669 **FIGURE LEGENDS**

670 **Figure 1. The DNA sequence hijacked by LME-1 is invariably proximal to the site of GC-skew**
671 **inflection and resembles a *dif* site. A)** Sequence alignment of the LME-1 *att* site region in different
672 *L. pneumophila* strains (left panel) and other *Legionella* species (right panel). The dashed-line box
673 denotes the 22-nt sequence that corresponds to the *att* site present on LME-1. The gray arrows
674 indicate the inverted repeat portion of the sequence. **B)** Circos plots showing GC-skew across the
675 chromosome of several *L. pneumophila* strains (left panel) and other *Legionella* species (right
676 panel). Each ring corresponds to one genome and the black lines indicate the position of the *att* site.
677 *L. pneumophila* strains inside to outside: Philadelphia-2, Philadelphia-1, E7_O, Paris, NY23 (D-
678 7705), NCTC11985, Fraseri D-4058. *Legionella* species inside to outside: *L. israelensis*, *L.*
679 *micdadei, L. spiritensis*, *L. pneumophila*, *L. longbeachae*, *L. oakridgensis*, *L. fallonii*. **C)** Alignment
680 of the LME-1 *att* site with *dif* sites from bacterial species with established *dif*/Xer systems. The
681 binding sites for XerC and XerD homologs (named CodV and RipX, respectively in *B. subtilis*) (10,
682 11, 18), XerS (64), and XerH (17, 65) are underlined.

683

684 **Figure 2. Site-specific recombination at the *dif* site is highly efficient and does not require**
685 **RecA. A)** Schematic representation of the intermolecular recombination assay used to quantify *att*
686 site-specific recombination in *L. pneumophila*. A 5.5-kb non-replicating plasmid modified to include
687 the 22-bp LME-1 *att* site (or control sequences) is transformed into *L. pneumophila* strain Lp02. If
688 recombination between the plasmid *att* site and the chromosomal *dif* site takes place, the whole
689 plasmid is integrated into the genome and gentamicin resistance is conferred to the cell. Gentamicin-
690 resistant colonies are counted to quantify recombination. **B)** Transformation efficiency (relative to
691 that of a replicative plasmid control) of a non-replicative plasmid containing the 22-bp *att* site, a 22-

692 bp control sequence, or a 2200-bp intergenic region of the Lp02 genome, transformed into wildtype
693 or Δ*recA* Lp02 strains. Each dot shows the value obtained from a single experiment, and each
694 horizontal line represents the geometric mean of 3 independent experiments.
695
696 **Figure 3. The *L. pneumophila* Xer is encoded by *lpg1867*.** Dot plots showing levels of
697 intermolecular recombination between the chromosomal *dif* site and a plasmid containing the 22-bp
698 *att* site, as described for Figure 2. **A)** Intermolecular recombination in wildtype Lp02 and strains
699 containing deletions of the *dif* site or one of six tyrosine-type recombinases. Shown in pink are the
700 *dif*-deletion strain and *lpg1867*-deletion strain, which both displayed levels of recombination below
701 the limit of detection. **B)** Intermolecular recombination in the wildtype and Δ*lpg1867* strains after
702 transformation with plasmids expressing wildtype Lpg1867 or a predicted catalytic point mutant,
703 Y387F.
704
705 **Figure 4. The *Legionella* CDR system involves a single recombinase**
706 **A)** Intermolecular recombination in the wildtype Lp02 strain was compared to that in a 5-deletion
707 strain (*lpg0980*, *lpg0981*, *lpg1070*, *lpg1085*, and *lpg2057* deleted) and a pan-deletion strain, with all
708 six recombinases deleted (i.e, 5-deletion plus *lpg1867* deletion). **B)** Intermolecular recombination of
709 the wildtype strain and pan-deletion strain containing the indicated expression plasmids. **C)** Left
710 panel: Excision of a *dif* site-flanked kanamycin reporter in the Lp02 pan deletion strain to assess
711 intramolecular recombination between *dif* sites. The proportion of kanamycin-resistant cells in the
712 population after transformation with Lpg1867, the catalytic mutant Y387F, or empty vector control
713 was quantified by plating on selective and nonselective media. Right panel: Excision of a *dif* site-

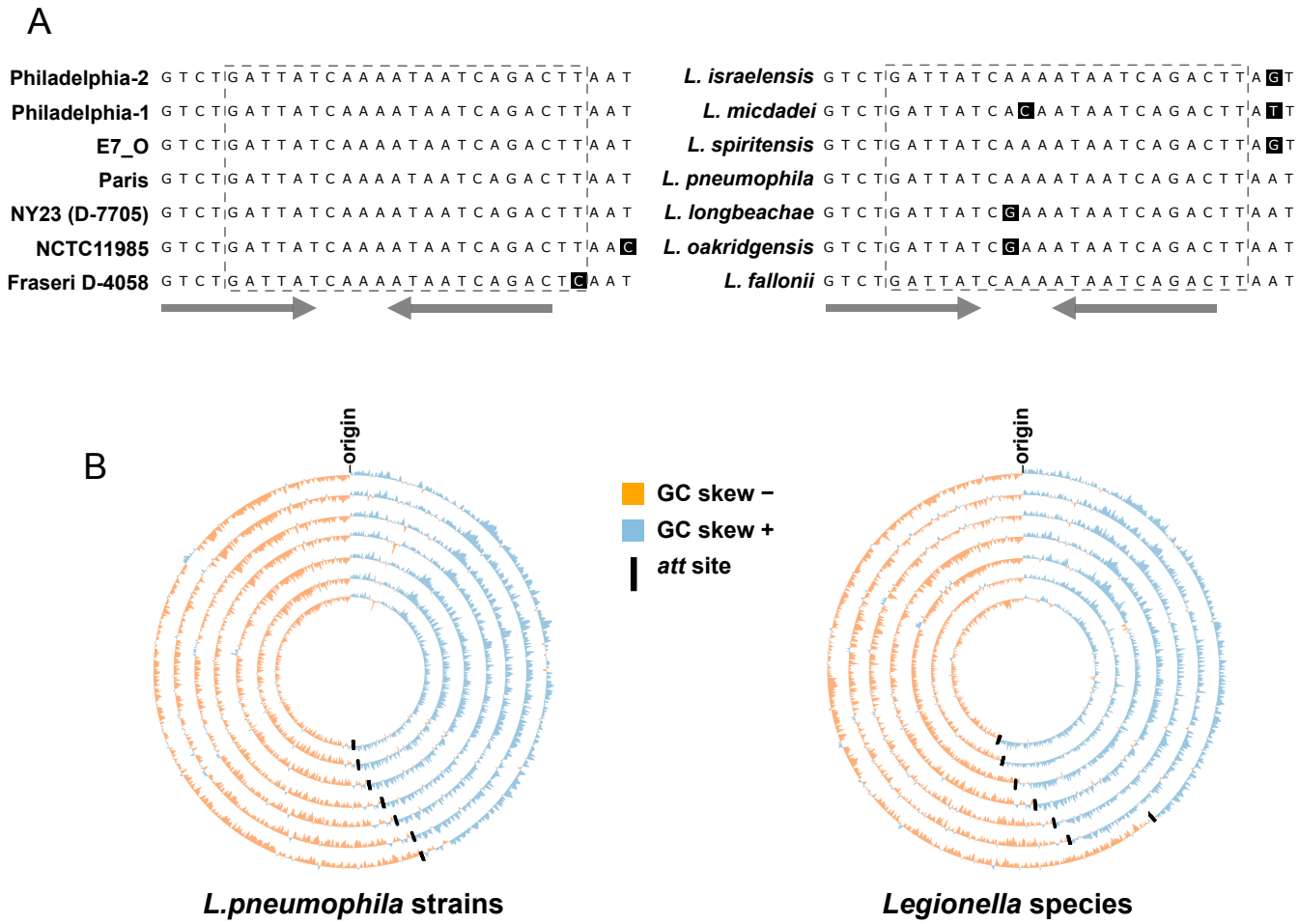
714 flanked KanMX reporter in *S. cerevisiae* BY4741. Bars indicate the mean of three independent
715 experiments, and error bars represent SEM.
716
717 **Figure 5. *dif- and lpg1867*-deletion strains exhibit filamentation and altered growth on solid**
718 **media and during infection of host cells. A)** DAPI-stained cells of the indicated *L. pneumophila*
719 strains, showing filamentation upon deletion of the *dif* site or *lpg1867* and rescue when *recA* is
720 deleted in the Δ*lpg1867* background. The scale bars represent 10 µm. **B)** Left panel: CYET spread
721 plates after 4 days of growth at 37°C with visible differences in colony size. Right panel: Each
722 colony radius is plotted as a separate data point, with the geometric mean shown by the purple line.
723 **C)** Distribution of colony sizes in the 5-deletion strain and pan-deletion strain (left panel) and in the
724 pan-deletion strain expressing wildtype or mutant Lpg1867 (right panel). Data are plotted as in B. **D)**
725 Intracellular growth of wildtype and deletion strains in U937 macrophages. The data are plotted as
726 the fold increase in colony-forming units recovered after lysis from host cells at 2 h and 48 h post
727 infection, relative to wildtype Lp02. Data show the average of three replicates for each strain from a
728 representative experiment. Error bars represent SEM.
729
730 **Figure 6. The sequence of the LME-1 *att* site contributes to its low level of excision, which**
731 **increases upon Lpg1867 (XerL) overexpression. A)** Comparison of the wildtype *dif* site in strain
732 Murcia-4983 and the sequences flanking LME-1 resulting from *att* site duplication upon integration
733 into the chromosome. The upstream copy of the *dif* site remains intact, but the downstream copy has
734 three nucleotide differences (pink letters). These differences are within the inverted repeat regions of
735 the *dif* site (denoted by the gray arrows), which presumably contain binding sites for Lpg1867
736 (XerL). **B)** A kanamycin-resistance reporter cassette flanked by a wildtype *dif* site upstream and the

737 indicated sequence downstream was introduced into the Lp02 Δ*lpg1867* strain. Excision of the
738 reporter (indicating recombination between the two flanking sites) was quantified by plating on
739 selective and nonselective media after transformation with empty vector or a plasmid expressing
740 Lpg1867.
741
742 **Figure 7. Phylogenetic analyses suggest that XerL forms a distinct clade from other Xer**
743 **recombinases.** XerL from *L. pneumophila* was used as a query to search for putative XerL orthologs
744 using BLASTp and tBLASTn. The amino acid sequences of recombinases from different Xer groups
745 were aligned with MUSCLE, with one representative per genera included for the XerL-like
746 recombinases. An unrooted phylogeny was generated from the alignment using FastTree. Xer group
747 designation is denoted with colored labels as indicated. The scale bar indicates the number of amino
748 acid substitutions per site.
749
750 **Figure 8. The *C. burnetii dif* site and XerL can functionally replace endogenous *dif*/Xer in *L.***
751 ***pneumophila*. A)** Sequence alignment of the *L. pneumophila dif* site with the plasmid and
752 chromosomal *dif* sites of *Coxiella burnetii*. The gray arrows indicate the inverted repeat portion of
753 the sequence. **B)** Circos plot showing GC-skew across the chromosome of several *C. burnetii* strains.
754 Each ring corresponds to one genome and the black lines indicate the position of the *dif* site. *C.*
755 *burnetii* strains, inside to outside: 3262, Heizburg, Dugway FJ108-111, RSA 493, CbuG_Q212,
756 Schperling, CbuK_Q154. **C)** Strains were generated in which the endogenous *L. pneumophila dif*
757 was replaced by the *C. burnetii* chromosomal *dif* site ($dif_{Cb_chr}$) or plasmid *dif* site ($dif_{Cb_pl}$), and the
758 endogenous XerL was replaced by *C. burnetii* XerL or a predicted catalytic point mutant (Y389F).
759 Colony sizes for each strain were quantified after 4 days of growth at 37°C on CYET spread plates.

760 Each colony radius is plotted as a separate data point, with the geometric mean shown by the purple
761 line. **D)** Intramolecular recombination assay performed as described for Figure 4C (uninduced) and
762 with an additional overnight incubation with 100 ng/mL anhydrotetracycline (induced). Percent
763 excision of a kanamycin reporter cassette flanked by the indicated *C. burnetii* chromosomal *dif* site
764 ($dif_{Cb_chr}$; left panel) or plasmid *dif* site ($dif_{Cb_pl}$; right panel). **E)** Intermolecular recombination assay
765 performed as described for Figure 2, showing the relative transformation efficiency of the indicated
766 non-replicative plasmids in strains containing the indicated *dif* site and *xerL*. Each dot shows the
767 value obtained from a single experiment, and each horizontal line represents the mean of 3
768 independent experiments. The Cb_chr* *dif* site is a truncated version of the chromosomal dif site,
769 with truncations made to mimic the shorter 22-bp LME-1 *att* site version of $dif_{Lp}$.
770
771 **Figure S1. Phylogenetic analysis of XerL-like recombinases indicates a widespread**
772 **distribution in Legionellales.** XerL from *L. pneumophila* was used as a query to search for putative
773 XerL orthologs using BLASTp and tBLASTn. The amino acid sequences of these putative orthologs
774 were aligned with MUSCLE. A phylogeny was generated from the alignment using FastTree, with
775 XerD from *E. coli* used as the root. Xer sequences that were also present in Figure 7 are denoted
776 with colored labels as indicated. The scale bar indicates the number of amino acid substitutions per
777 site.
778
779 **Table S1. The *att* site is close to the cumulative GC-skew maximum in all sequenced *L.***
780 ***pneumophila* strains and *Legionella* species.** The nucleotide positions of the *att* site and cumulative
781 GC-skew maximum (cGC skew max) are shown for each *L. pneumophila* strain and *Legionella*
782 species with a completed genome sequence available in Genbank. The cGC skew maximum

783 corresponds to the site of replication termination (34). The genes immediately upstream and
784 downstream of each *att* site are also indicated.
785
786 **Table S2. The putative *dif* site is close to the cumulative GC-skew maximum in all sequenced**
787 ***C. burnetii* strains.** The nucleotide positions of the *dif* site and cumulative GC-skew maximum are
788 shown for each *C. burnetii* strain with a completed genome sequence available in Genbank.
789
790 **Table S3. The accession numbers and amino acid sequences for the Xer proteins used in the**
791 **phylogenetic analyses.**
792

**FIGURE 1**




C

| | | | |
|---|---|---|---|
| | XerC/CodV | | XerD/RipX |
| *E. coli dif* | GGTGCGCATAA | TGTATA | TTATGTTAAAT |
| *N. gonorrheae dif* | AGTTCGCATTA | TGTATA | TTATGTTAAAT |
| *V. cholerae dif1* | AGTGCGCATTA | TGTATG | TTATGTTAAAT |
| *B. subtilis bsdif* | ACTTCCTAGAA | TATATA | TTATGTAAACT |
| | XerS | | XerS |
| *Streptococcus difSL* | ATTTTTCGCAA | AAACTATAAT | TTCTTGAAA |
| | XerH | | XerH |
| *Helicobacter difH* | AGTTATGAA | AACTGCACTT | TTCAAACT |
| | XerL? | | XerL? |
| *L. pneumophila* LME-1 *att* | GTCTGATTAT | CAAA | ATAATCAGACTT |

**FIGURE 2**

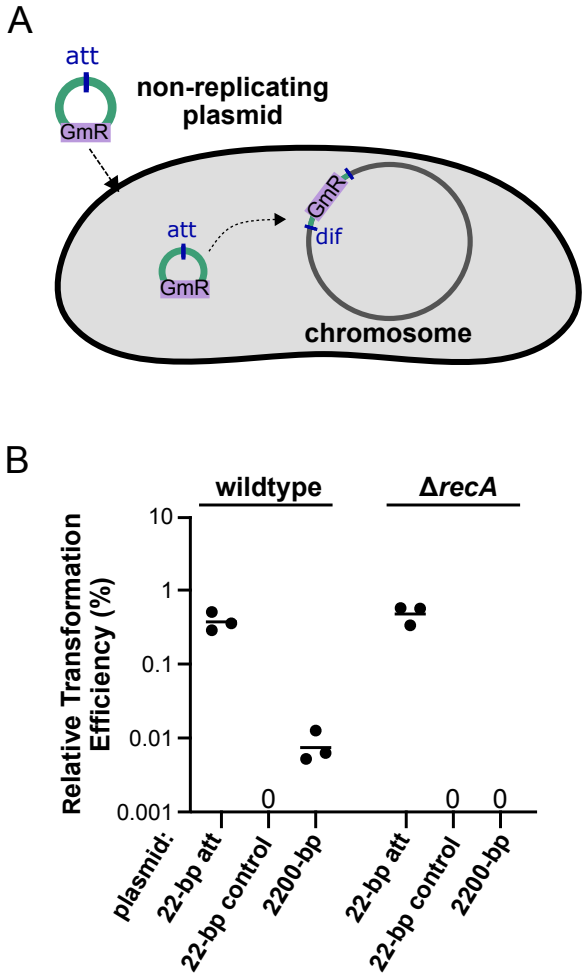

A
att
GmR
non-replicating
plasmid
att
GmR
GmR
dif
chromosome
B
wildtype
ΔrecA
Relative Transformation
Efficiency (%)
10
1
0.1
0.01
0.001
0
0
0
plasmid:
22-bp att
22-bp control
2200-bp
22-bp att
22-bp control
2200-bp

**FIGURE 3**

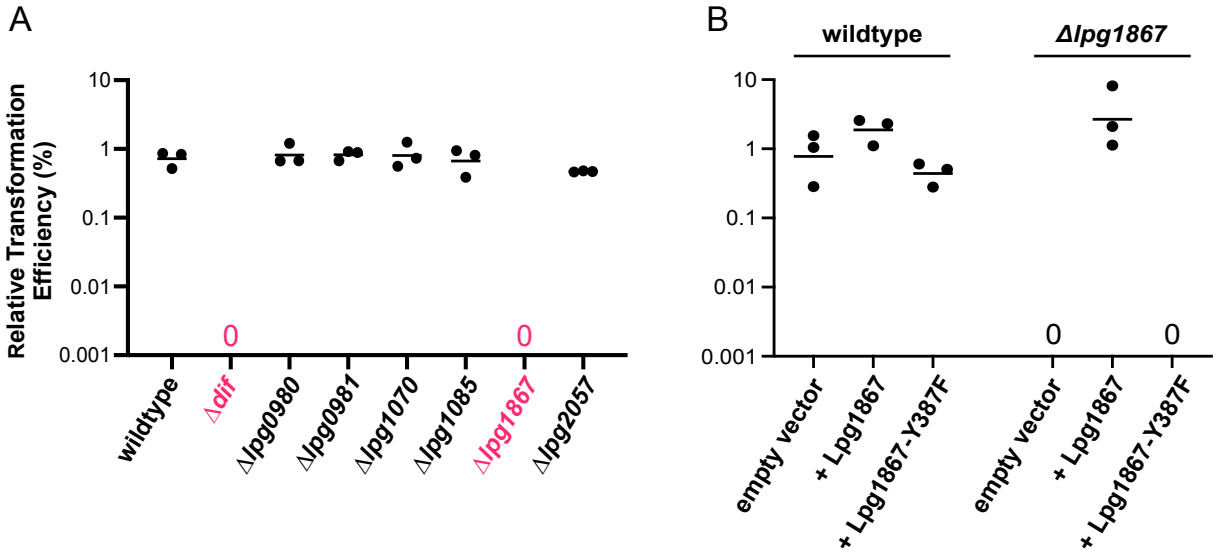

A
Relative Transformation Efficiency (%)
10
1
0.1
0.01
0.001
0
0
wildtype
Δdif
Δlpg0980
Δlpg0981
Δlpg1070
Δlpg1085
Δlpg1867
Δlpg2057
B
wildtype
Δlpg1867
10
1
0.1
0.01
0.001
0
0
empty vector
+ Lpg1867
+ Lpg1867-Y387F
empty vector
+ Lpg1867
+ Lpg1867-Y387F

**FIGURE 4**

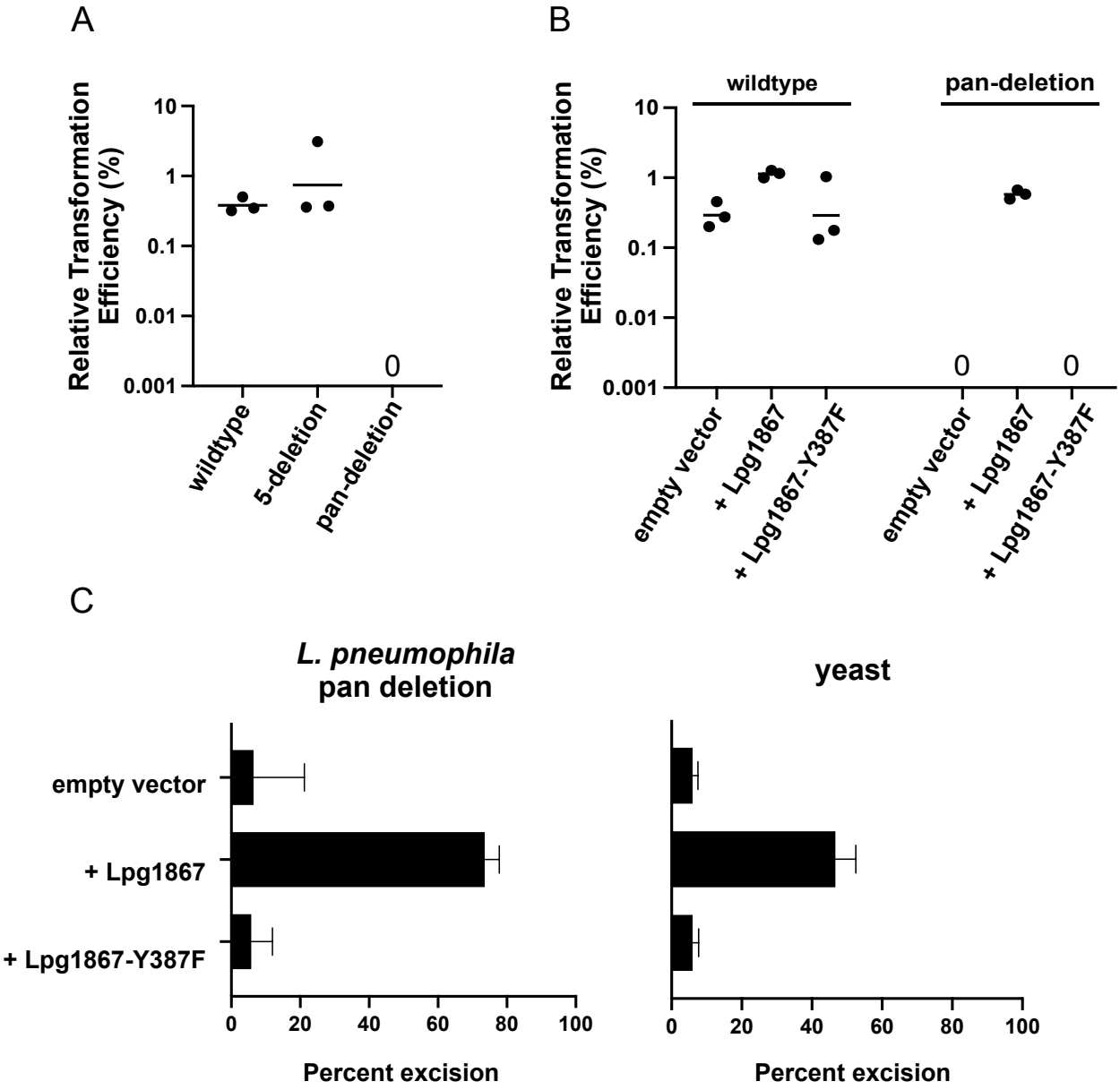
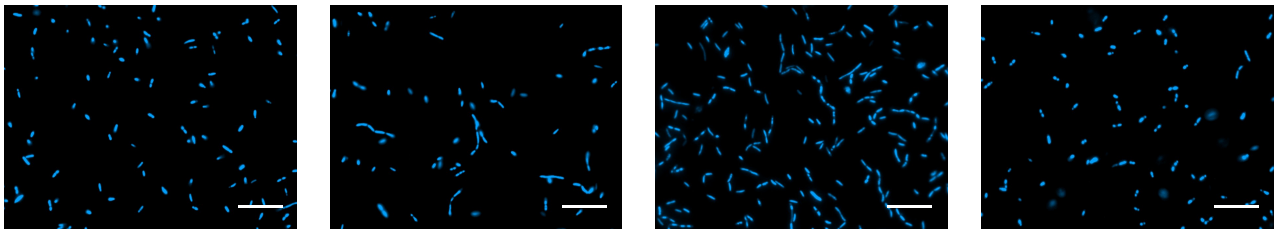

A
Relative Transformation Efficiency (%)
10
1
0.1
0.01
0.001
0
wildtype
5-deletion
pan-deletion
B
wildtype
pan-deletion
Relative Transformation Efficiency (%)
10
1
0.1
0.01
0.001
0
0
empty vector
+ Lpg1867
+ Lpg1867-Y387F
empty vector
+ Lpg1867
+ Lpg1867-Y387F
C
L. pneumophila pan deletion
yeast
empty vector
+ Lpg1867
+ Lpg1867-Y387F
0
20
40
60
80
100
Percent excision
0
20
40
60
80
100
Percent excision

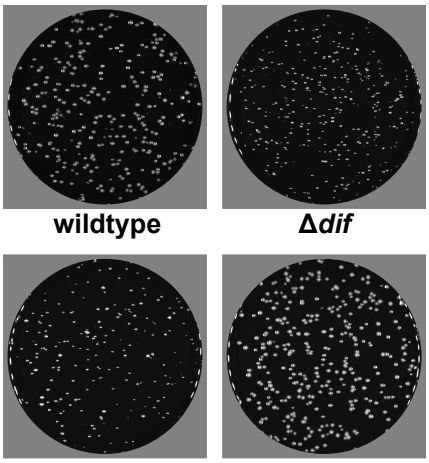
**FIGURE 5**

A

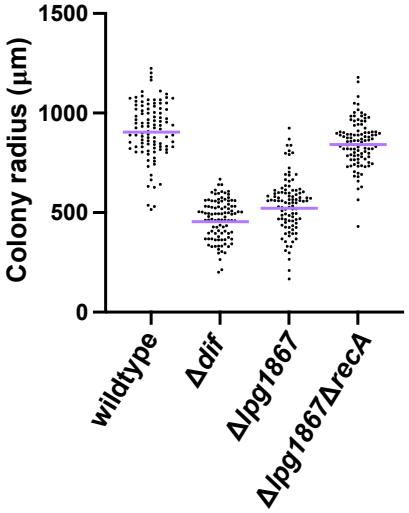


wildtype | Δ*dif* | Δ*lpg1867* | Δ*lpg1867*Δ*recA*

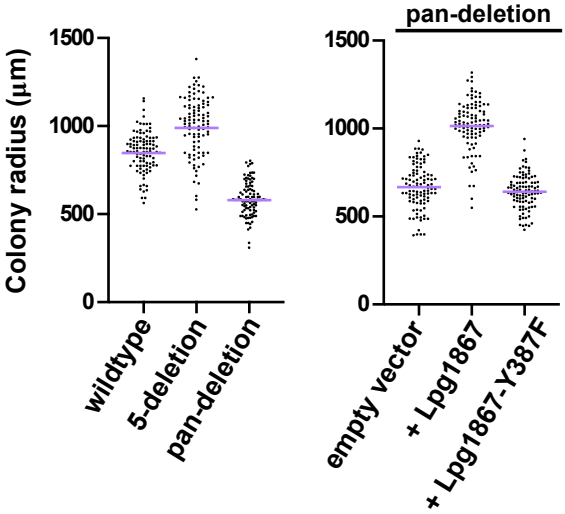
B



wildtype | Δ*dif* | Δ*lpg1867* | Δ*lpg1867*Δ*recA*




C




D

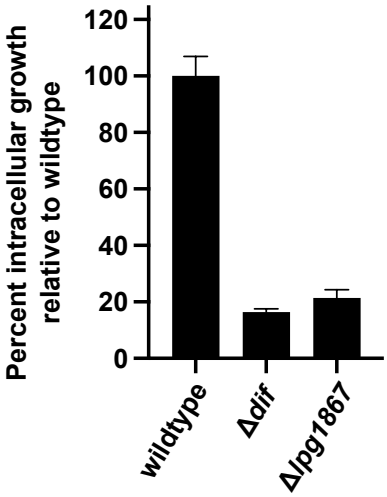


**FIGURE 6**

A

dif GTCTGATTATCAAAATAATCAGACTT
LME-1
upstream GTCTGATTATCAAAATAATCAGACTT
downstream TTACGATTATCAAAATAATCAGACTT

B

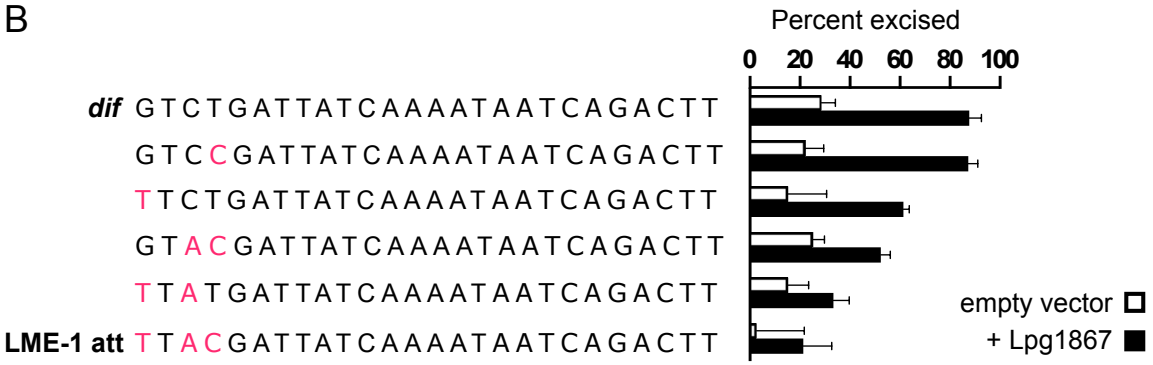
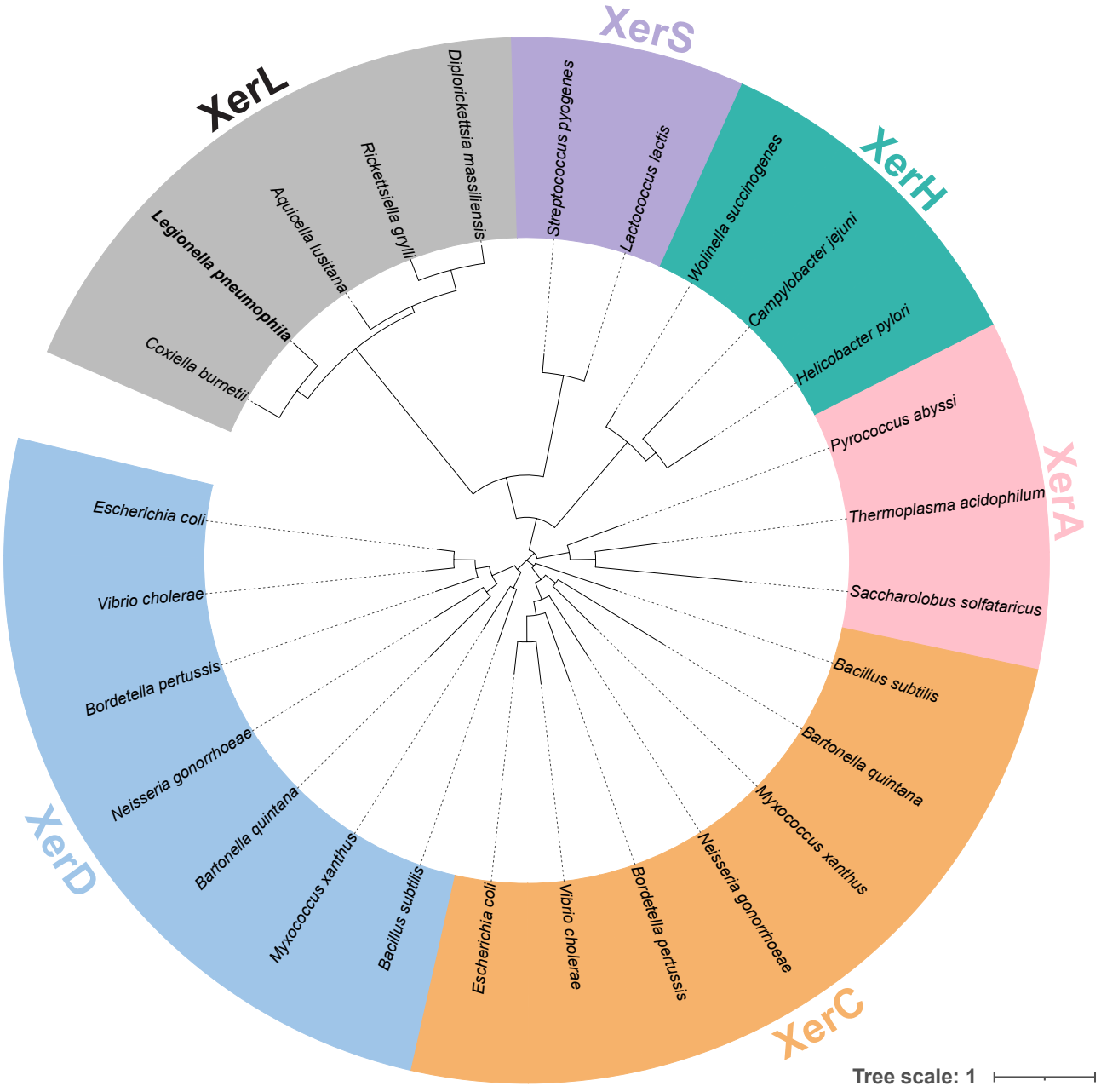
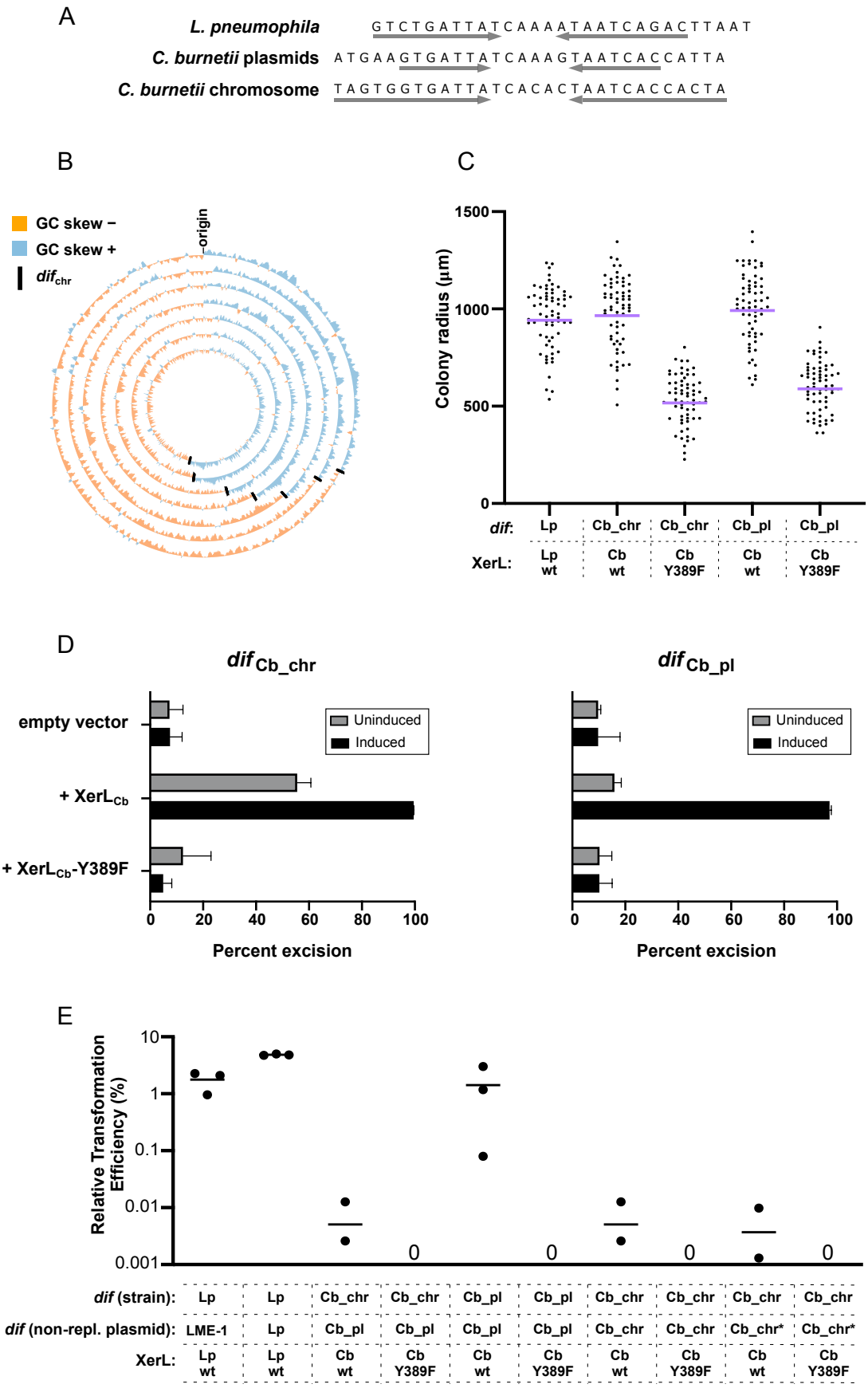

Percent excised
0 20 40 60 80 100
dif GTCTGATTATCAAAATAATCAGACTT
GTCCGATTATCAAAATAATCAGACTT
TTCTGATTATCAAAATAATCAGACTT
GTACGATTATCAAAATAATCAGACTT
TTATGATTATCAAAATAATCAGACTT
LME-1 att TTACGATTATCAAAATAATCAGACTT
empty vector
+ Lpg1867

**FIGURE 7**


XerL
XerS
XerH
XerA
XerD
XerC
Diplorickettsia massiliensis
Rickettsiella grylli
Aquicella lusitana
Legionella pneumophila
Coxiella burnetii
Streptococcus pyogenes
Lactococcus lactis
Wolinella succinogenes
Campylobacter jejuni
Helicobacter pylori
Pyrococcus abyssi
Thermoplasma acidophilum
Saccharolobus solfataricus
Bacillus subtilis
Bartonella quintana
Myxococcus xanthus
Neisseria gonorrhoeae
Bordetella pertussis
Vibrio cholerae
Escherichia coli
Escherichia coli
Vibrio cholerae
Bordetella pertussis
Neisseria gonorrhoeae
Bartonella quintana
Myxococcus xanthus
Bacillus subtilis
Tree scale: 1

# FIGURE 8


A
L. pneumophila
GTCTGATTATCAAAATAATCAGACTTAAT
C. burnetii plasmids
ATGAAGTGATTATCAAAGTAATCACCATTA
C. burnetii chromosome
TAGTGGTGATTATCACACTAATCACCACTA
B
GC skew −
GC skew +
dif chr
origin
C
Colony radius (µm)
1500
1000
500
0
dif:
Lp
Cb_chr
Cb_chr
Cb_pl
Cb_pl
XerL:
Lp wt
Cb wt
Cb Y389F
Cb wt
Cb Y389F
D
dif Cb_chr
dif Cb_pl
empty vector
+ XerL Cb
+ XerL Cb-Y389F
Uninduced
Induced
0
20
40
60
80
100
Percent excision
E
Relative Transformation Efficiency (%)
10
1
0.1
0.01
0.001
0
dif (strain):
Lp
Lp
Cb_chr
Cb_chr
Cb_pl
Cb_pl
Cb_chr
Cb_chr
Cb_chr
Cb_chr
dif (non-repl. plasmid):
LME-1
Lp
Cb_pl
Cb_pl
Cb_pl
Cb_pl
Cb_chr
Cb_chr
Cb_chr*
Cb_chr*
XerL:
Lp wt
Lp wt
Cb wt
Cb Y389F
Cb wt
Cb Y389F
Cb wt
Cb Y389F
Cb wt
Cb Y389F