Type I-F CRISPR-Cas distribution and array dynamics in *Legionella pneumophila.*

Shayna R. Deecker[1] and Alexander W. Ensminger[1,2]

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

[2] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

Running title: *L. pneumophila* type I-F CRISPR-Cas

Abstract

In bacteria and archaea, several distinct types of CRISPR-Cas systems provide adaptive immunity through broadly similar mechanisms: short nucleic acid sequences derived from foreign DNA, known as spacers, engage in complementary base pairing with invasive genetic elements setting the stage for nucleases to degrade the target DNA. A hallmark of type I CRISPR-Cas systems is their ability to acquire spacers in response to both new and previously encountered invaders (naïve and primed acquisition, respectively). Our phylogenetic analyses of 47 *L. pneumophila* type I-F CRISPR-Cas systems and their resident genomes suggest that many of these systems have been horizontally acquired. These systems are frequently encoded on plasmids and can co-occur with nearly identical chromosomal loci. We show that two such co-occurring systems are highly protective and undergo efficient primed acquisition in the lab. Furthermore, we observe that targeting by one system's array can prime spacer acquisition in the other. Lastly, we provide experimental and genomic evidence for a model in which primed acquisition can efficiently replenish a depleted type I CRISPR array following a mass spacer deletion event.

24

25 **Introduction**

26 Microorganisms have evolved over millions of years to survive in harsh environments,
27 backed in part by immune strategies that protect against foreign genetic elements, such as viral
28 phages and foreign DNA elements (van Houte *et al.* 2016). Clustered regularly interspaced short
29 palindromic repeats (CRISPR) coupled with associated *cas* genes form a potent adaptive immune
30 response in numerous prokaryotic species (Barrangou *et al.* 2007; Brouns *et al.* 2008; Marraffini
31 and Sontheimer 2008). These systems have been classified into six major types, which are
32 further divided into various sub-types, based on their mechanism of action and Cas protein
33 content (Makarova *et al.* 2011; 2015; Koonin *et al.* 2017).

34 A CRISPR-Cas response to invading DNA occurs in three distinct phases: adaptation,
35 expression and interference (Barrangou *et al.* 2007; Brouns *et al.* 2008; Marraffini and
36 Sontheimer 2008). In the adaptation phase, the CRISPR-Cas system acquires a DNA sequence
37 (spacer) from the invader and integrates it into an array of spacers interspersed with repetitive
38 sequences (Ishino *et al.* 1987; Mojica *et al.* 2000; Jansen, van Embden, Gaastra, and Schouls
39 2002b; Barrangou *et al.* 2007; Jackson *et al.* 2017). The spacers are generally derived from
40 foreign elements whose infection was unsuccessful, such as defunct phage (Hynes *et al.* 2014),
41 and form the basis of immunological memory for the bacterium. During the expression phase,
42 the array is transcribed and processed to form CRISPR RNA (crRNA) molecules which form a
43 surveillance complex with Cas proteins (Brouns *et al.* 2008; Haurwitz *et al.* 2010). Infection by
44 an invading genetic element initiates the interference step, wherein the surveillance complex
45 recognizes and binds the foreign DNA via base-pairing with the complementary crRNA and

46 cleaves it, effectively neutralizing the threat to the host (Brouns *et al.* 2008; Jore *et al.* 2011;
47 Wiedenheft *et al.* 2011; Westra *et al.* 2012).

48 Despite the sophistication of CRISPR-Cas systems, phages and other foreign DNA
49 elements can still escape CRISPR-Cas targeting (Samson *et al.* 2013; Stanley and Maxwell
50 2018). A common mechanism of escape is the accumulation of random mutations within the
51 foreign element which prevent complementary base pairing with crRNAs during interference
52 (Deveau *et al.* 2008; Semenova *et al.* 2011; Datsenko *et al.* 2012). CRISPR-Cas systems can
53 overcome this escape by acquiring new spacers; in fact, imperfect CRISPR targeting often leads
54 to a highly efficient "primed" acquisition response, providing an intrinsic mechanism to protect
55 against mutational escape (Swarts *et al.* 2012; Datsenko *et al.* 2012; Savitskaya *et al.* 2013;
56 Fineran *et al.* 2014; Richter *et al.* 2014). Primed acquisition has been studied in type I-B (Li,
57 Wang, Zhao, *et al.* 2014; Li, Wang, and Xiang 2014; Li *et al.* 2017), I-C (Rao *et al.* 2017), I-E
58 (Swarts *et al.* 2012; Datsenko *et al.* 2012; Savitskaya *et al.* 2013; Fineran *et al.* 2014; Xue *et al.*
59 2015; Jackson *et al.* 2019) and I-F (Richter *et al.* 2014; Vorontsova *et al.* 2015; Staals *et al.*
60 2016; Heussler *et al.* 2016; Jackson *et al.* 2019) CRISPR-Cas systems. Interference-driven
61 acquisition, or targeted acquisition, has also been observed in type I-C (Rao *et al.* 2017) and I-F
62 systems (Staals *et al.* 2016), wherein a primed acquisition response occurs against a target with a
63 perfect match to a spacer already within the array.

64 *Legionella pneumophila* is a gram-negative bacterium and the causative agent of
65 Legionnaires' disease (Brenner *et al.* 1979). Most isolates possess any of three different
66 CRISPR-Cas systems: type I-C, I-F and/or II-B (D'Auria *et al.* 2010; Rao *et al.* 2016). Our lab
67 has recently shown that all three types of CRISPR-Cas systems found in *L. pneumophila* isolates
68 are active (Rao *et al.* 2016) and we have characterized the targeted acquisition response for the

69 type I-C system (Rao *et al.* 2017). While much of our work to date has focused on the I-C
70 systems of *L. pneumophila*, the type I-F systems of this pathogen are highly protective,
71 remarkably diverse with respect to spacer content, and are frequently found on plasmids –
72 suggesting that they may be circulated via horizontal gene transfer (Rao *et al.* 2016). In this
73 study, we perform the first comprehensive phylogenetic analysis of the *L. pneumophila* type I-F
74 systems and test a model by which horizontal acquisition of a mobile type I-F CRISPR-Cas
75 system could replenish a collapsed chromosomal array.
76
77 **Results**
78 **Phylogenetic analyses suggest widespread horizontal exchange of *L. pneumophila* type I-F**
79 **systems.**
80 In order to explore the hypothesis that plasmid-based type I-F CRISPR-Cas systems in *L.*
81 *pneumophila* could be circulated via horizontal gene transfer, we bioinformatically examined the
82 diversity of type I-F CRISPR-Cas systems within this species. Leveraging both
83 CRISPRCasFinder (Couvin *et al.* 2018) and CRISPRDetect (Biswas *et al.* 2016), we surveyed
84 525 draft and 6 completed *L. pneumophila* genomes. In total, we identified 47 *L. pneumophila*
85 isolates that possessed type I-F systems (Table S1), including 5 that we had described previously
86 (Rao *et al.* 2016). We next performed two types of phylogenetic analysis: *cas1* phylogeny (Fig
87 1A), which placed each CRISPR-Cas system into one of six phylotypes; and core-genome
88 phylogeny (Fig 1B), which reflects the overall relatedness between each of the 47 isolates. A
89 comparison of the two trees indicates a clear phylogenetic incongruence suggesting that
90 horizontal acquisition has impacted the distribution of type I-F CRISPR-Cas systems within the
91 species.

92 Given the results of these phylogenetic analyses, we next examined the spacer
93 distribution across each of the arrays to determine the level of array diversification within each
94 of the six *L. pneumophila* type I-F *cas1* phylotypes. We aligned, clustered, and visualized each
95 of the 26 distinct *L. pneumophila* type I-F CRISPR arrays using CRISPRStudio (Dion *et al.*
96 2018) (Fig. 2). These analyses revealed patterns consistent with both spacer acquisition and
97 spacer loss, suggesting that both processes contribute to *L. pneumophila* type I-F CRISPR array
98 diversity.
99
100 **The plasmid and chromosomal Lens CRISPR-Cas systems are active and adaptive**
101 Given the evidence for spacer gain and loss that we observe in *L. pneumophila* type I-F
102 arrays, (Fig. 2), we next decided to examine array dynamics experimentally. We focused on *L.*
103 *pneumophila* str. Lens, which possesses two type I-F CRISPR-Cas systems: one on its
104 chromosome and one on an endogenous 60 Kb plasmid, pLPL (D'Auria *et al.* 2010; Rao *et al.*
105 2016). The two systems have a 97.6% Cas protein identity and the repeat units between the
106 spacers in the CRISPR array differ by only a single nucleotide (Rao *et al.* 2016) (Fig. 3). The
107 CRISPR arrays themselves are of different lengths (64 spacers for the chromosomal system and
108 53 for that contained on the pLPL plasmid). Despite a high degree of identity between the Cas
109 proteins, each array contains a completely unique set of spacers (D'Auria *et al.* 2010; Rao *et al.*
110 2016). The presence of two remarkably similar I-F systems in *L. pneumophila* str. Lens provided
111 us with an opportunity to examine targeted spacer acquisition in both of these largely
112 uncharacterized CRISPR-Cas systems and the interplay between them.
113 To assess CRISPR-Cas activity in both Lens type I-F systems, we performed an
114 established transformation efficiency assay (Marraffini and Sontheimer 2008) using two different

115 targeted protospacer sequences: one matching the most recently acquired spacer and one
116 matching a spacer from the middle of the array (chromosomal spacer 23 and pLPL spacer 50).
117 Consistent with active CRISPR-Cas protection, each of the protospacer-containing plasmids
118 exhibited significant reductions in transformation efficiencies relative to a scrambled protospacer
119 control (Fig. S1). These relative transformation efficiencies ranged from $1x10^{-2}$ to $1x10^{-4}$, with
120 the most recently acquired spacers providing ~100-fold greater protection than spacers located in
121 the middle of each array.

122 To determine whether spacer acquisition occurs within the context of a perfectly matched
123 protospacer target, we pooled the transformed populations, passaged them on an automated
124 liquid handler for 20 generations without selection, extracted their genomic DNA, and screened
125 the leader end of the CRISPR array by PCR and agarose gel electrophoresis. While the
126 populations transformed with plasmids encoding either protospacer 23 (chromosome) or
127 protospacer 50 (pLPL plasmid) exhibited spacer acquisition in both Lens systems (Fig. 4), the
128 populations transformed with protospacer 1 plasmids exhibited spacer loss, with spacer
129 acquisition undetectable on a gel (Fig. 4). These data support our bioinformatic analyses, which
130 indicated that both spacer acquisition and spacer loss contribute to type I-F CRISPR array
131 diversity in *L. pneumophila* isolates (Fig. 2).

132 Our observation that the mid-array spacers of each type I-F system provided relatively
133 modest protection led us to ask whether these spacers could nevertheless drive primed
134 acquisition of new, more protective spacer sequences. To characterize the patterns of targeted
135 spacer acquisition in the chromosomal and pLPL CRISPR-Cas systems, we amplified the leader-
136 proximal region of each CRISPR array from wild-type populations that had been transformed
137 with the plasmids targeted by their relatively permissive mid-array spacers (chromosomal: spacer

138 23; pLPL: spacer 50). We Illumina sequenced these PCR products and used an established
139 bioinformatics pipeline (Rao *et al.* 2017) to identify newly acquired spacer sequences within
140 each read (Table 1). We then mapped the target of each new spacer to the priming plasmid
141 (Krzywinski *et al.* 2009) (Fig. 5).

142 Both the Lens type I-F CRISPR-Cas systems exhibited a biased distribution of acquired
143 spacers (Fig. 5, S2) consistent with what has been seen previously for the type I-F systems of
144 *Pectobacterium astrosepticum* (Richter *et al.* 2014; Staals *et al.* 2016) and *Pseudomonas*
145 *aeruginosa* (Vorontsova *et al.* 2015; Heussler *et al.* 2016). The majority of the newly targeted
146 protospacers clustered around the priming sequence on the targeted plasmid, with the non-
147 primed strand of DNA (the plus (+) strand) containing ~75% of these protospacers (Fig. 5).
148 Consistent with observations in other type I-F systems (Richter *et al.* 2014; Staals *et al.* 2016;
149 Heussler *et al.* 2016), switching the target sequence to the opposite strand led to an acquisition
150 pattern that mirrored the original distribution observed when the (-) strand contained the targeted
151 protospacer (Fig. S3). The spacer length distribution and PAM usage were also consistent with
152 previous observations of other type I-F systems (Mojica *et al.* 2009; Cady *et al.* 2012; Richter *et*
153 *al.* 2014; Vorontsova *et al.* 2015; Staals *et al.* 2016) (Fig. S3B, C, S4, S5). Taken together, these
154 data suggest that spacer acquisition is qualitatively similar between the chromosomal and pLPL
155 CRISPR-Cas systems.
156

157 **Permissive targeting by one system can lead to primed acquisition in the other system**

158 Since the chromosomal and pLPL CRISPR-Cas systems function in a very similar
159 manner during targeted acquisition and share a high degree of homology within their Cas genes
160 and repeat sequences, we speculated that priming of one system might lead to spacer acquisition

161 in the other. Specifically, we tested whether introducing a protospacer-containing plasmid
162 targeted by one CRISPR-Cas system would initiate a primed acquisition response in the second
163 system. Indeed, this led to efficient spacer acquisition on the second array (Fig. 6), with patterns
164 largely indistinguishable from what we previously observed on the cognate array (Fig. 5).
165

166 **Primed repopulation of collapsed arrays**

167 Based on our observations suggesting widespread horizontal inheritance of *L. pneumophila* type
168 I-F systems (Fig. 1), the diversity of spacer sequences (Fig. 2), and the ability of closely related
169 systems to prime each other (Fig. 6), we next asked whether coincident CRISPR-Cas might
170 provide a mechanism for replenishing collapsed chromosomal arrays. Spacer loss is one of
171 several outcomes when the targeting of a particular sequence becomes detrimental to bacterial
172 survival. In the lab, this occurs when we artificially "force" the coexistence of an efficiently
173 targeted plasmid and an active CRISPR-Cas system through selection (Fig. 4). Similar events are
174 also likely to occur randomly or when CRISPR-Cas systems acquire self-targeting spacers at a
175 low, but detectable rate (Fig. 7) (Yosef *et al.* 2012; Datsenko *et al.* 2012; Savitskaya *et al.* 2013;
176 Vorontsova *et al.* 2015; Staals *et al.* 2016; Rao *et al.* 2017). In the most extreme instances, an
177 entire array of spacers could be lost through recombination between the first and last repeats.
178 Normally, such a loss could only be reversed by the relatively inefficient mechanism of naïve
179 spacer acquisition (Yosef *et al.* 2012; Datsenko *et al.* 2012; Savitskaya *et al.* 2013). However, if
180 such a collapsed array could be restored through primed acquisition driven by a coincident
181 system, strains with multiple arrays would be inherently more resistant to the catastrophic loss of
182 CRISPR-Cas protection. Such a model makes a number of predictions (Fig. 7), which we sought
183 to bioinformatically and experimentally test.

184 First, we used allelic replacement to generate an *L. pneumophila* Lens strain in which the
185 entire chromosomal array was replaced by a single copy of its last repeat, mimicking what would
186 occur after complete spacer loss. Next, we transformed two independently derived array deletion
187 strains with the pLPL type I-F priming plasmid as above. Using PCR and Illumina sequencing,
188 we observed robust spacer acquisition in the formerly depleted chromosomal CRISPR array (Fig
189 8, Table S4), indicating that primed acquisition can replenish the completely collapsed array of a
190 coincidental type I-F system. As expected, the consensus repeat sequence of this replenished
191 array adopted the same alternate sequence as the last repeat of the array.
192 We next examined the CRISPR (repeat) sequences in each of the isolates used in our
193 earlier bioinformatic analyses for similar evidence of complete collapse followed by
194 replenishment. In many of the arrays, there is a consensus repeat that is found throughout the
195 majority of the array, with the last repeat in the array carrying a mutation (Table 2). This has
196 been observed before in other systems (Jansen, van Embden, Gaastra, and Schouls 2002a; 2002b;
197 Horvath *et al.* 2008; Lopez-Sanchez *et al.* 2012). However, for a number of the *L. pneumophila*
198 isolates, including Alcoy, JFIM01, LBAN01, and LBAV01, this is not the case. In these isolates,
199 their consensus repeat (found throughout the CRISPR array) is identical to their last repeat.
200 Intriguingly, the sequence that remains is identical to the last repeat found in other type I-F
201 isolates – what one would predict if they were the product of complete array collapse followed
202 by subsequent replenishment (Fig. 8).
203
204 **Discussion**
205 Horizontal gene transfer is a driving force in shaping bacterial biology and pathogenicity
206 (Ochman *et al.* 2000; Juhas 2013; Hall *et al.* 2017). With respect to *L. pneumophila*, comparative

genomic studies have highlighted the importance of horizontal gene transfer, mobile genetic elements, and homologous recombination in shaping the bacterium's evolutionary trajectory (Cazalet *et al.* 2004; 2008; D'Auria *et al.* 2010; Gomez-Valero *et al.* 2011; Sánchez-Busó *et al.* 2014; Gomez-Valero *et al.* 2014; McAdam *et al.* 2014; Burstein *et al.* 2016; David *et al.* 2017). This aspect of *L. pneumophila* biology extends to the presence and maintenance of plasmid-based CRISPR-Cas systems (D'Auria *et al.* 2010; Rao *et al.* 2016). Our bioinformatic analyses suggest that CRISPR-Cas type I-F systems are horizontally distributed in this species (Fig. 1). These CRISPR arrays have also undergone extensive spacer acquisition and some spacer loss (Fig. 2). Horizontal acquisition of type I-F systems is well-established in other species. For instance, in *Vibrio* species, 97% of identified type I-F systems are encoded on mobile genetic elements (McDonald *et al.* 2019). Like *L. pneumophila*, it's hypothesized these systems have been acquired by their host through horizontal gene transfer (McDonald *et al.* 2019). Our data, however, suggest that such transfer is not merely a mechanism by which isolates acquire CRISPR-Cas protection, but also a potential mechanism to maintain existing defensive capabilities through the unique spacer dynamics provided by two inter-priming arrays. These inter-priming arrays can be part of two different systems, as demonstrated by our data, but could also occur in a system where two (or more) different arrays share a set of *cas* genes (Swarts *et al.* 2012; Datsenko *et al.* 2012; Staals *et al.* 2013; 2014; Majumdar *et al.* 2015; Elmore *et al.* 2015; Silas *et al.* 2017). For example, in *E. coli* str. K12, it has been shown that two different CRISPR arrays can be populated by the same set of *cas* genes (Swarts *et al.* 2012; Datsenko *et al.* 2012).

We propose that when a bacterium acquires a second, closely related CRISPR-Cas system, it gains a mechanism by which depleted CRISPR arrays can be repopulated (Fig. 7). We have modeled such an event in the lab, made predictions about what the signatures of such

230 events would be, and provide genomic data to suggest it may have occurred on several occasions
231 within our collection of sequenced isolates. One obvious line of future investigation would be to
232 observe whether such patterns are present in other species with mobilized type I CRISPR-Cas
233 and perhaps absent in instances where CRISPR-Cas is acquired primarily through vertical
234 inheritance.

235 Lastly, while we think of spacer loss as a predominantly negative event (loss of
236 protection), it is likely to play a more nuanced role in the maintenance of CRISPR-Cas activity.
237 Clearly, array length is the product of a dynamic process whose impact on adaptation,
238 expression, and interference remains largely unexplored. Many of the type I-F systems in *L.*
239 *pneumophila* have different array lengths, ranging from 8 spacers to 129 spacers, with an average
240 length of 61 spacers (Table 2). Toms and Barrangou have performed a global analysis of class I
241 CRISPR arrays and found that the average array length for type I-F systems was 33 spacers, with
242 statistically significant differences between the array lengths of different type I subtypes (Toms
243 and Barrangou 2017). Accordingly, if spacer acquisition is a driving force in array divergence, it
244 is likely coupled to spacer loss. Close examination of the mechanisms driving spacer loss in
245 these systems - and their subsequent impact on CRISPR-Cas functionality - will be crucial to
246 further testing the model of array diversification in *L. pneumophila*.

247

248 **Methods and Materials**

249 **Bioinformatic analyses**

250 Bioinformatic analyses of the Illumina sequence data were performed as described
251 previously (Rao *et al.* 2017). Briefly, the raw paired-end reads were merged using FLASH
252 (Magoc and Salzberg 2011), and any unpaired reads were subsequently quality trimmed using

253 Trimmomatic (Bolger *et al.* 2014). These processed reads were then combined and analyzed
254 using a Perl script (available upon request) that annotated existing spacers (S), newly acquired
255 spacers (X), repetitive sequences (R) and the downstream sequence (D). The newly acquired
256 spacers were aligned to the priming plasmid, the *L. pneumophila* str. Lens chromosome or the *L.*
257 *pneumophila* str. Lens plasmid using BLASTN (Altschul *et al.* 1990). The results from the
258 BLASTN alignment for the priming plasmid were then processed to obtain coverage per
259 nucleotide, and plotted on the reference sequence using Circos (Krzywinski *et al.* 2009).
260 For the bioinformatic analyses of the *L. pneumophila* type I-F system diversity, *L.*
261 *pneumophila* draft genomes and completed genomes were downloaded from the European
262 Nucleotide Archive and NCBI respectively. Type I-F CRISPR-Cas systems were identified using
263 CRISPRCasFinder (Couvin *et al.* 2018) and CRISPRDetect (Biswas *et al.* 2016). All genomes
264 with type I-F systems present were annotated using Prokka (Seemann 2014). For the core
265 genome phylogeny, pan-genome analysis was performed with Roary(Page *et al.* 2015). Both the
266 *cas1* and core genome phylogenies were created with RAxML (Stamatakis 2014) and the
267 RAxML trees were condensed with MEGA7 (Kumar *et al.* 2016). CRISPR array alignments,
268 clustering and visualization were performed with CRISPRStudio (Dion *et al.* 2018).
269

270 **Bacterial strains, plasmids and oligos used**

271 The bacterial strains and plasmids used in this study are listed in supplementary table 2,
272 and the oligos used in this study are listed in supplementary table 3.
273 The priming plasmids were created by annealing oligos (see Table S3) to create the
274 protospacer insert with the canonical GG PAM (Mojica *et al.* 2009; Cady *et al.* 2012; Richter *et*
275 *al.* 2014; Vorontsova *et al.* 2015; Staals *et al.* 2016) and subsequently ligating the insert into an

276 ApaI/PstI-digested pMMB207 vector (Solomon *et al.* 2000). The scrambled control plasmid was
277 created in the same manner, except it contained a 32-nt scrambled sequence in place of a targeted
278 protospacer sequence.

279 Lens (chromsome) array deletion mutants were generated through allelic replacement.
280 Briefly, 1 kb of DNA upstream and 2 kb downstream of the CRISPR array were amplified by
281 PCR and stitched together to create an insert where the entire array, save for the last repeat, was
282 deleted. The insert was ligated into a pJB4648 plasmid. Overnight cultures of *L. pneumophila* str.
283 Lens were grown in ACES-buffered yeast extract (AYE) medium to an OD_{600} of ~4.0 using two-
284 day patches that were grown on charcoal-buffered ACES yeast extract (CYE) plates. Pellets from
285 4.0 ODU of culture underwent three washing steps: twice with 1 mL of ice-cold ultrapure water
286 and once with 1 mL of ice-cold 10% glycerol. The pellet was then re-suspended in 200 uL of ice-
287 cold 10% glycerol and 400 ng of plasmid was added to the sample. The solution was transferred
288 to an ice-cold electroporation cuvette with a 2 mm gap and electroporated with the following
289 settings: 2500 kV, 600 Ω and 25 mF. After electroporation, 800 uL of AYE medium was added
290 to each sample and the samples recovered for 3 hours at 37°C at 600 RPM in a shaking
291 incubator. The samples were plated on CYE plates supplemented with 15 μg mL^{-1} of gentamycin
292 and incubated at 37°C for 3 days. Surviving colonies were patched onto CYE + gentamycin
293 plates and grown at 37°C for 2 days. Patches were subsequently struck onto CYE plates
294 supplemented with sucrose and incubated at 37°C for 3 days. Surviving colonies were patched
295 onto CYE + sucrose plates, grown at 37°C for 2 days and screened by PCR to confirm the
296 deletion. Two independent clones were Illumina sequenced and used for subsequent
297 replenishment assays.
298

299 **Transformation efficiency assay and population pool generation**

300 The transformation efficiency assay was performed as we have described previously (Rao
301 *et al.* 2016) with some modifications. Briefly, *L. pneumophila* str. Lens was electroporated as
302 described above. The samples were plated in a dilution series on CYE plates supplemented with
303 5 μg mL^{-1} of chloramphenicol and incubated at 37°C for 3 days. The relative transformation
304 efficiency for each targeted plasmid was calculated as a percentage of the transformation
305 efficiency obtained from the scrambled control plasmid. Three biological replicates were
306 performed for each transformation efficiency assay.

307 Population pools for spacer acquisition experiments were generated by mixing together ≥
308 50 colonies per population from a newly transformed wild-type strain on CYE plates
309 supplemented with 5 μg mL^{-1} of chloramphenicol using AYE medium supplemented with 5 μg
310 mL^{-1} of chloramphenicol. Population pools were made in triplicate for each transformed plasmid.
311

312 **Serial passaging on an automated liquid handler**

313 The serial passaging of transformed *L. pneumophila* str. Lens populations was performed
314 as described previously (Rao *et al.* 2016). Briefly, overnight cultures of the population pools in
315 AYE medium supplemented with 5 μg mL^{-1} of chloramphenicol for plasmid maintenance were
316 grown to an OD_{600} of ~2.0. The culture was then back diluted to an OD_{600} of ~0.0625 and grown
317 in a flat-bottom 48-well plate (Greiner) in a shaking incubator at 37°C. A Freedom Evo 100
318 liquid handler (Tecan) connected to an Infinite M200 Pro plate reader (Tecan) measured the
319 optical density of the plate every 20 minutes, until an OD_{600} of ~2.0 was reached. The cultures
320 were then automatically back diluted to an OD_{600} of ~0.0625 in the adjacent well to continue
321 growth, and the remaining culture was transferred to a 48-well plate that was kept at 4°C. In this

322 manner, each saved culture represented ~5 generations of growth. The passaging was done
323 without selection in AYE medium to allow for plasmid loss during passaging.
324
325 **Genomic DNA extraction, PCR and agarose gel screening**
326 Genomic DNA was extracted from the passaged cultures and the parental chromosome
327 Lens array deletion strains using a Machery-Nagel Nucleospin Tissue kit according to the
328 manufacturer's protocol. The extracted samples from passaged cultures were used as a template
329 in a 30-cycle PCR reaction with Econotaq Polymerase (Lucigen) to amplify the leader end of the
330 CRISPR array using primers listed in Table S3. The PCR products were then separated on a 3%
331 agarose gel to determine if spacer acquisition (or spacer loss) had occurred based on the presence
332 of an upper (or lower) band relative to the control sample.
333
334 **Nextera library prep and Illumina sequencing**
335 The extracted genomic DNA from passaged cultures was prepared for leader-end array
336 sequencing by performing a 20-cycle PCR using Kapa HiFi Polymerase (Kapa Biosystems) and
337 the primers listed in Table S3. The PCR products were purified using a Machery-Nagel
338 Nucleospin Gel and PCR Clean-up kit as per the manufacturer's instructions and normalized to 1
339 ng using the Invitrogen Quant-iT PicoGreen dsDNA assay. The DNA was then tagmented using
340 a Nextera XT tagmentation kit as per the manufacturer's instructions. The tagmented products
341 were sequenced with a paired-end (2 x 150 bp) sequencing run on an Illumina NextSeq platform
342 at the Centre for the Analysis of Genome Evolution and Function (CAGEF) at the University of
343 Toronto.

344 The genomic DNA from the parental chromosome Lens array deletion strains was
345 normalized to 1 ng using PicoGreen. The DNA was then tagmented using a Nextera XT
346 tagmentation kit as per the manufacturer's instructions. The tagmented products were sequenced
347 with paired-end (2 x 150 bp) sequencing in-house on an Illumina MiniSeq platform.
348

349 **Data Accessibility**

350 The raw Illumina reads have been deposited into the NCBI sequence read archive under
351 the BioProject PRJNA433194.
352

353 **Acknowledgements**

354 The authors thank Griffin Deecker (a volunteer high school student) for his assistance in
355 bioinformatically examining the diversity of I-F repeat sequences in *L. pneumophila*, Chitong
356 Rao for his contributions to experimental design and Kamran Rizzolo for discussions regarding
357 phylogenetic analyses. We also thank the Center for the Analysis of Genome Evolution and
358 Function (CAGEF) at the University of Toronto for performing Illumina sequencing. We thank
359 members of the Ensminger laboratory for their suggestions and careful reading of the
360 manuscript, in particular Beth Nicholson and Malene Urbanus. SRD is supported by a fellowship
361 from the Department of Biochemistry, University of Toronto. This work was supported by a
362 Project Grant from the Canadian Institutes of Health Research (PHT-148819), the Connaught
363 Fund (NR-2015-16), and an infrastructure grant from the Canada Foundation for Innovation and
364 the Ontario Research Fund (30364) to AWE.
365

366 **Figures**

367 **Figure 1 | Phylogenetic analysis of type I-F CRISPR-Cas system diversity in *L.***
368 ***pneumophila* reveals a horizontal distribution across isolates.** *L. pneumophila* draft and
369 completed genomes were analyzed using CRISPRDetect and CRISPRCasFinder to identify type
370 I-F systems present within the genomes. Isolates that possessed type I-F systems were subjected
371 to phylogenetic analyses of their *cas1* gene and core genome. The core genome alignment for the
372 examined isolates was determined by Roary. Isolate names are colour-coded based on the *cas1*
373 gene phylogeny to allow for comparison between the analyses. The isolates used in this analysis
374 and their accession numbers can be found in supplementary table 1. **A)** The *cas1* gene phylogeny
375 for *L. pneumophila* isolates with type I-F systems reveals six different cas1 groups. **B)** The core
376 genome phylogeny of the examined isolates is not congruent with the *cas1* phylogeny,
377 suggesting that the type I-F system was horizontally distributed rather than vertically inherited.
378 Note that Lens possesses two type I-F CRISPR-Cas systems, one on a plasmid (group Θ) and
379 one on its chromosome (group Φ). Also note that isolates with 100% nucleotide identity in their
380 core genome, as well as a shared *cas* gene group and redundant CRISPR array, have been
381 collapsed to one representative in the phylogeny. Isolates with unique CRISPR arrays are listed.
382 Collapsed isolates include:
383 +: FJEK01, FJEL01, FJEM01, FJEN01, FJEO01, FJEP01, FJEQ01, FJER01, FJES01, FJET01,
384 FJEU01, FJEV01, FJEX01, FJEY01, FJEZ01, FJFA01
385 ++: FJBI01, FJBU01
386 +++: LAXR01
387 ++++: FJEI01
388 +++++: FFI105, FFI337

389 **Figure 2 | CRISPR array analysis suggests that spacer acquisition and spacer loss has**
390 **contributed to array diversification in *L. pneumophila* type I-F systems.** *L. pneumophila*
391 isolates were subjected to a CRISPRStudio analysis to look at the spacer composition of their
392 CRISPR arrays. Grey boxes denote unique spacers, coloured boxes denote shared spacers and a
393 dashed line denotes spacer loss. The isolate colour coding scheme is based on the *cas1* grouping
394 from figure 1. Denoted isolates within the same *cas* gene group have 100% nucleotide identity in
395 their core genomes. Note that *cas1* group β appears to have spacer rearrangements present,
396 resulting in an imperfect alignment between the arrays.
397
398 **Figure 3 | A comparison of the Lens chromosome and pLPL CRISPR-Cas systems. A)** The
399 overall pairwise amino acid identity across the Cas proteins of the two systems is approximately
400 97%, with individual pairwise Cas protein identities ranging from 96% to 99%. **B)** A
401 CRISPRStudio alignment of the arrays for two systems shows completely unique spacer content
402 and a differing number of spacers. **C)** Analysis of the repeat sequences shows that the Lens
403 pLPL and chromosome systems have one SNP between their consensus repeats, in addition to
404 possessing a mutated last repeat in their arrays. Mutations are denoted in red.
405
406 **Figure 4 | Spacer acquisition and spacer loss in both Lens CRISPR-Cas systems.** Spacer
407 acquisition and loss were analyzed using a PCR-based screen where the leader-end of the
408 CRISPR array for both the control samples and the transformed samples was amplified with
409 system-specific primers to differentiate between the chromosomal Lens and the plasmid Lens
410 arrays and visualized on an agarose gel. Products from the transformed samples were compared

411 to the control, which contained untransformed genomic DNA. Bands representing spacer
412 acquisition and loss are indicated.
413
414 **Figure 5 | Characterization of self-primed spacer acquisition in the two Lens CRISPR-Cas**
415 **systems.** Bacterial transformants with targeted plasmids were passaged for 20 generations
416 without antibiotic selection to enrich for spacer acquisition; the leader end of the CRISPR array
417 was amplified and the amplicons were Illumina sequenced. Newly targeted protospacers were
418 obtained from the raw reads using an in-house bioinformatics pipeline and visualized with
419 Circos. The arrows in the simplified array schematic show the primer location for PCR
420 amplification prior to sequencing, while the star denotes the priming spacer. All data are the
421 average of three biological replicates. **A)** The distribution of newly targeted protospacers mapped
422 to the priming plasmid on the Circos plot reveals a strand bias in self-primed spacer acquisition
423 within the Lens (pLPL) CRISPR-Cas system. The height of the bars indicates the number of
424 spacers mapped to the position on the plasmid, up to 5% of total acquired spacers. **B)** The
425 distribution of newly targeted protospacers mapped to the priming plasmid shows a similar
426 pattern of self-primed spacer acquisition within the Lens (chromosome) CRISPR-Cas system to
427 that of its pLPL CRISPR-Cas counterpart. Labelling as in **(A)**.
428
429 **Figure 6 | Characterization of cross-primed spacer acquisition in the two Lens CRISPR-**
430 **Cas systems.** The experimental set-up is the same as described in figure 5. All data are the
431 average of three biological replicates. The distribution of newly targeted protospacers mapped to
432 the priming plasmid reveals cross-priming between the Lens (chromosome) and Lens (pLPL)

433 CRISPR-Cas systems. **(A)** shows Lens (chromosome) primed, Lens (pLPL) array examined
434 while **(B)** shows Lens (pLPL) primed, Lens (chromosome) array examined.
435
436 **Figure 7 | Cross-priming between two similar CRISPR-Cas systems can repopulate a**
437 **collapsed CRISPR array.** A bacterium with a CRISPR-Cas system could undergo a mass
438 spacer deletion event through homologous recombination between the first (black) and last
439 (white) repeat sequences. The remaining locus carries a single repeat (white). While such
440 "catastrophic collapse" events are likely to occur randomly at a certain rate, a driver of such a
441 collapse could be the acquisition of a self-targeting spacer (yellow), selecting for spacer loss.
442 Horizontal acquisition of a second CRISPR-Cas array (e.g. on a plasmid) is a first step towards
443 replenishing the primary array. If cross-priming can occur between this secondary array and the
444 collapsed array, the original CRISPR array is replenished, but bears an observable molecular scar
445 - conversion of all the repeats to the sequence of the last repeat (white).
446
447 **Figure 8 | An experimentally depleted chromosomal CRISPR array can be replenished**
448 **through the activity of a plasmid-based array. A)** Replenishment of a depleted array in the
449 lab. Allelic replacement was used to remove the entire array from the chromosomal Lens I-F
450 system, leaving behind a single, last repeat sequence (see materials and methods). This strain
451 was then transformed with a plasmid targeted by the pLPL (Lens plasmid-based) I-F array
452 previously shown to drive primed acquisition (pLPL protospacer 50; see Figure 4). Spacer
453 acquisition by the empty chromosomal array was analyzed using a PCR based screen where the
454 leader-end of the CRISPR array was amplified and visualized on an agarose gel. Products from
455 the transformed samples (samples 2 and 4) were compared to untransformed controls (samples 1

456 and 3). Samples 1 and 2 are from depleted array clone #1 and samples 3 and 4 are from depleted
457 array clone #2. **B)** Repeat signatures of depletion/replenishment. The repeat structure of the
458 experimentally replenished Lens CRISPR arrays resembles that of *L. pneumophila* str. Alcoy,
459 suggesting a similar array depletion/replenishment event may have occurred within the Alcoy
460 lineage. The frequency of acquiring one new spacer versus two new spacers following
461 replenishment in the Lens array depletion isolates was determined by Illumina sequencing.

462 **References**

463 Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic Local
464 Alignment Search Tool. Journal of Molecular Biology 215: 403–410.

465 Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval *et al.*, 2007 CRISPR provides
466 acquired resistance against viruses in prokaryotes. Science 315: 1709–1712.

467 Biswas, A., R. H. J. Staals, S. E. Morales, P. C. Fineran, and C. M. Brown, 2016 CRISPRDetect:
468 A flexible algorithm to define CRISPR arrays. BMC Genomics 1–14.

469 Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina
470 sequence data. Bioinformatics 30: 2114–2120.

471 Brenner, D. J., A. G. Steigerwalt, and J. E. McDade, 1979 Classification of the Legionnaires'
472 Disease Bacterium: Legionella pneumophila, genus novum, species nova, of the Family
473 Legionellaceae, familia nova. Annals of Internal Medicine 90: 656–658.

474 Brouns, S. J. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis *et al.*, 2008 Small
475 CRISPR RNAs guide antiviral defense in prokaryotes. Science 321: 960–964.

476 Burstein, D., F. Amaro, T. Zusman, Z. Lifshitz, O. Cohen *et al.*, 2016 Genomic analysis of 38
477 Legionella species identifies large and diverse effector repertoires. Nature Publishing Group
478 48: 167–175.

479 Cady, K. C., J. Bondy-Denomy, G. E. Heussler, and A. R. Davidson, 2012 The CRISPR/Cas
480 adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally
481 occurring and engineered phages. Journal of Bacteriology 194: 5728–5738.

482 Cazalet, C., S. Jarraud, Y. Ghavi-Helm, F. Kunst, P. Glaser *et al.*, 2008 Multigenome analysis
483 identifies a worldwide distributed epidemic Legionella pneumophila clone that emerged
484 within a highly diverse species. Genome Research 18: 431–441.

485 Cazalet, C., C. Rusniok, H. Brüggemann, N. Zidane, A. Magnier *et al.*, 2004 Evidence in the
486 Legionella pneumophila genome for exploitation of host cell functions and high genome
487 plasticity. Nat Genet 36: 1165–1173.

488 Couvin, D., A. Bernheim, C. Toffano-Nioche, M. Touchon, J. Michalik *et al.*, 2018
489 CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced
490 performance and integrates search for Cas proteins. Nucleic Acids Research 99: 7536–6.

491 D'Auria, G., N. Jimenez-Hernandez, F. Peris-Bondia, A. Moya, and A. Latorre, 2010 Legionella
492 pneumophila pangenome reveals strain-specific virulence factors. BMC Genomics 11: 181.

493 Datsenko, K. A., K. Pougach, A. Tikhonov, B. L. Wanner, K. V. Severinov *et al.*, 2012
494 Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity
495 system. Nature Communications 3: 945–7.

496 David, S., L. Sánchez-Busó, S. R. Harris, P. Marttinen, C. Rusniok *et al.*, 2017 Dynamics and
497 impact of homologous recombination on the evolution of Legionella pneumophila (X.
498 Didelot, Ed.). PLoS Genet 13: e1006855–21.

499 Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux *et al.*, 2008 Phage response to
500 CRISPR-encoded resistance in *Streptococcus thermophilus*. Journal of Bacteriology 190:
501 1390–1400.

502 Dion, M., S. Labrie, S. Shah, and S. Moineau, 2018 CRISPRStudio: A User-Friendly Software
503 for Rapid CRISPR Array Visualization. Viruses 10: 602–11.

504 Elmore, J., T. Deighan, J. Westpheling, R. M. Terns, and M. P. Terns, 2015 DNA targeting by
505 the type I-G and type I-A CRISPR–Cas systems of Pyrococcus furiosus. Nucleic Acids
506 Research gkv1140–11.

507 Fineran, P. C., M. J. H. Gerritzen, M. Suarez-Diez, T. Kunne, J. Boekhorst *et al.*, 2014
508 Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci USA
509 111: E1629–E1638.

510 Gomez-Valero, L., C. Rusniok, S. Jarraud, B. Vacherie, Z. Rouy *et al.*, 2011 Extensive
511 recombination events and horizontal gene transfer shaped the Legionella pneumophila
512 genomes. BMC Genomics 12: 536.

513 Gomez-Valero, L., C. Rusniok, M. Rolando, M. Neou, D. Dervins-Ravault *et al.*, 2014
514 Comparative analyses of Legionella species identifies genetic features of strains causing
515 Legionnaires disease. Genome Biology 1–21.

516 Hall, J. P. J., M. A. Brockhurst, and E. Harrison, 2017 Sampling the mobile gene pool:
517 innovation via horizontal gene transfer in bacteria. Phil. Trans. R. Soc. B 372: 20160424–10.

518 Haurwitz, R. E., M. Jinek, B. Wiedenheft, K. Zhou, and J. A. Doudna, 2010 Sequence- and
519 structure-specific RNA processing by a CRISPR endonuclease. Science 329: 1355–1358.

520 Heussler, G. E., J. L. Miller, C. E. Price, A. J. Collins, and G. A. O'Toole, 2016 Requirements
521 for Pseudomonas aeruginosa Type I-F CRISPR-Cas Adaptation Determined Using a Biofilm
522 Enrichment Assay (T. J. Silhavy, Ed.). Journal of Bacteriology 198: 3080–3090.

523 Horvath, P., D. A. Romero, A. C. Coute-Monvoisin, M. Richards, H. Deveau *et al.*, 2008
524 Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. Journal of
525 Bacteriology 190: 1401–1412.

526 Hynes, A. P., M. Villion, and S. Moineau, 2014 Adaptation in bacterial CRISPR-Cas immunity
527 can be driven by defective phages. Nature Communications 5: 1–6.

528 Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata, 1987 Nucleotide sequence
529 of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*,
530 and identification of the gene product. Journal of Bacteriology 169: 5429–5433.

531 Jackson, S. A., N. Birkholz, L. M. Malone, and P. C. Fineran, 2019 Imprecise Spacer
532 Acquisition Generates CRISPR-Cas Immune Diversity through Primed Adaptation. Cell
533 Host and Microbe 1–16.

534 Jackson, S. A., R. E. McKenzie, R. D. Fagerlund, S. N. Kieper, P. C. Fineran *et al.*, 2017
535 CRISPR-Cas: Adapting to change. Science 356: eaal5056–11.

536 Jansen, R., J. D. van Embden, W. Gaastra, and L. M. Schouls, 2002a Identification of a novel
537 family of sequence repeats among prokaryotes. OMICS 6: 23–33.

538 Jansen, R., J. D. van Embden, W. Gaastra, and L. M. Schouls, 2002b Identification of genes that
539 are associated with DNA repeats in prokaryotes. Molecular Microbiology 43: 1565–1575.

540 Jore, M. M., M. Lundgren, E. van Duijn, J. B. Bultema, E. R. Westra *et al.*, 2011 Structural basis
541 for CRISPR RNA-guided DNA recognition by Cascade. Nature Publishing Group 18: 529–
542 536.

543 Juhas, M., 2013 Horizontal gene transfer in human pathogens. Critical Reviews in Microbiology
544 41: 101–108.

545 Koonin, E. V., K. S. Makarova, and F. Zhang, 2017 Diversity, classification and evolution of
546 CRISPR-Cas systems. Current Opinion in Microbiology 37: 67–78.

547 Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: an information
548 aesthetic for comparative genomics. Genome Research 19: 1639–1645.

549 Kumar, S., G. Stecher, and K. Tamura, 2016 MEGA7: Molecular Evolutionary Genetics
550 Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33: 1870–1874.

551 Li, M., L. Gong, D. Zhao, J. Zhou, and H. Xiang, 2017 The spacer size of I-B CRISPR is
552 modulated by the terminal sequence of the protospacer. Nucleic Acids Research 45: 4642–
553 4654.

554 Li, M., R. Wang, and H. Xiang, 2014 *Haloarcula hispanica* CRISPR authenticates PAM of a
555 target sequence to prime discriminative adaptation. Nucleic Acids Research 42: 7226–7235.

556 Li, M., R. Wang, D. Zhao, and H. Xiang, 2014 Adaptation of the *Haloarcula hispanica* CRISPR-
557 Cas system to a purified virus strictly requires a priming process. Nucleic Acids Research
558 42: 2483–2492.

559 Lopez-Sanchez, M.-J., E. Sauvage, V. Da Cunha, D. Clermont, E. Ratsima Hariniaina *et al.*,
560 2012 The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the
561 diversity of its mobilome. Molecular Microbiology 85: 1057–1071.

562 Magoc, T., and S. L. Salzberg, 2011 FLASH: fast length adjustment of short reads to improve
563 genome assemblies. Bioinformatics 27: 2957–2963.

564 Majumdar, S., P. Zhao, N. T. Pfister, M. Compton, S. Olson *et al.*, 2015 Three CRISPR-Cas
565 immune effector complexes coexist in Pyrococcus furiosus. RNA 21: 1147–1158.

566 Makarova, K. S., D. H. Haft, R. Barrangou, S. J. J. Brouns, E. Charpentier *et al.*, 2011 Evolution
567 and classification of the CRISPR–Cas systems. Nature Reviews Microbiology 9: 467–477.

568 Makarova, K. S., Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah *et al.*, 2015 An updated
569 evolutionary classification of CRISPR–Cas systems. Nature Reviews Microbiology 13: 722–
570 736.

571 Marraffini, L. A., and E. J. Sontheimer, 2008 CRISPR interference limits horizontal gene
572 transfer in Staphylococci by targeting DNA. Science 322: 1843–1845.

573 McAdam, P. R., C. W. Vander Broek, D. S. Lindsay, M. J. Ward, M. F. Hanson *et al.*, 2014
574 Gene flow in environmental Legionella pneumophila leads to genetic and pathogenic
575 heterogeneity within a Legionnaires disease outbreak. Genome Biology 15: 1–10.

576 McDonald, N. D., A. Regmi, D. P. Morreale, J. D. Borowski, and E. F. Boyd, 2019 CRISPR-Cas
577 systems are present predominantly on mobile genetic elements in Vibrio species. 1–23.

578 Mojica, F. J. M., C. Díez-Villaseñor, J. García-Martínez, and C. Almendros, 2009 Short motif
579 sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology
580 155: 733–740.

581 Mojica, F. J. M., C. Díez-Villaseñor, E. Soria, and G. Juez, 2000 Biological significance of a
582 family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria.
583 Molecular Microbiology 36: 244–246.

584 Ochman, H., J. G. Lawrence, and E. A. Groisman, 2000 Lateral gene transfer and the nature of
585 bacterial innovation. Nature 405: 299–304.

586 Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter *et al.*, 2015 Roary: rapid large-
587 scale prokaryote pan genome analysis. Bioinformatics 31: 3691–3693.

588 Rao, C., D. Chin, and A. W. Ensminger, 2017 Priming in a permissive type I-C CRISPR–Cas
589 system reveals distinct dynamics of spacer acquisition and loss. RNA 23: 1525–1538.

590 Rao, C., C. Guyard, C. Pelaz, J. Wasserscheid, J. Bondy-Denomy *et al.*, 2016 Active and
591 adaptive *Legionella* CRISPR-Cas reveals a recurrent challenge to the pathogen. Cellular
592 Microbiology 18: 1319–1338.

593 Richter, C., R. L. Dy, R. E. McKenzie, B. N. J. Watson, C. Taylor *et al.*, 2014 Priming in the
594 Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-
595 directionally from the primed protospacer. Nucleic Acids Research 42: 8516–8526.

596 Samson, J. E., A. H. Magadan, M. Sabri, and S. Moineau, 2013 Revenge of the phages: defeating
597 bacterial defences. Nature Publishing Group 11: 675–687.

598 Savitskaya, E. E., E. Semenova, V. Dedkov, A. Metlitskaya, and K. V. Severinov, 2013 High-
599 throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. RNA Biology 10:
600 716–725.

601 Sánchez-Busó, L., I. Comas, G. Jorques, and F. González-Candelas, 2014 Recombination drives
602 genome evolution in outbreak-related Legionella pneumophila isolates. Nat Genet 46: 1205–
603 1211.

604 Seemann, T., 2014 Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068–2069.

605 Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra *et al.*, 2011 Interference
606 by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a
607 seed sequence. Proc Natl Acad Sci USA 108: 10098–10103.

608 Silas, S., P. Lucas-Elio, S. A. Jackson, A. A.-C. Elife, 2017, 2017 Type III CRISPR-Cas systems
609 can provide redundancy to counteract viral escape from type I systems. eLife e27601.

610 Solomon, J. M., A. Rupper, J. A. Cardelli, and R. R. Isberg, 2000 Intracellular growth of
611 *Legionella pneumophila* in *Dictyostelium discoideum*, a system for genetic analysis of host-
612 pathogen interactions. Infect. Immun. 68: 2939–2947.

613 Staals, R. H. J., Y. Agari, S. Maki-Yonekura, Y. Zhu, D. W. Taylor *et al.*, 2013 Structure and
614 Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus.
615 Molecular Cell 52: 135–145.

616 Staals, R. H. J., S. A. Jackson, A. Biswas, S. J. J. Brouns, C. M. Brown *et al.*, 2016 Interference-
617 driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-
618 Cas system. Nature Communications 7: 1–13.

619 Staals, R. H. J., Y. Zhu, D. W. Taylor, J. E. Kornfeld, K. Sharma *et al.*, 2014 RNA Targeting by
620 the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. Molecular Cell 56:
621 518–530.

622 Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of
623 large phylogenies. Bioinformatics 30: 1312–1313.

624 Stanley, S. Y., and K. L. Maxwell, 2018 Phage-Encoded Anti-CRISPR Defenses. Annu. Rev.
625 Genet. 52: 445–464.

626 Swarts, D. C., C. Mosterd, M. W. J. van Passel, and S. J. J. Brouns, 2012 CRISPR interference
627 directs strand specific spacer acquisition (I. Mokrousov, Ed.). PLoS ONE 7: e35888–7.

628 Toms, A., and R. Barrangou, 2017 On the global CRISPR array behavior in class I systems.
629 Biology Direct 12: 1–12.

630 van Houte, S., A. Buckling, and E. R. Westra, 2016 Evolutionary ecology of prokaryotic immune
631 mechanisms. Microbiol. Mol. Biol. Rev. 80: 745–763.

632 Vorontsova, D., K. A. Datsenko, S. Medvedeva, J. Bondy-Denomy, E. E. Savitskaya *et al.*, 2015
633 Foreign DNA acquisition by the I-F CRISPR–Cas system requires all components of the
634 interference machinery. Nucleic Acids Research 43: 10848–10860.

635 Westra, E. R., P. B. G. van Erp, T. Künne, S. P. Wong, R. H. J. Staals *et al.*, 2012 CRISPR
636 immunity relies on the consecutive binding and degradation of negatively supercoiled
637 invader DNA by Cascade and Cas3. Molecular Cell 46: 595–605.

638 Wiedenheft, B., E. van Duijn, J. B. Bultema, S. P. Waghmare, K. Zhou *et al.*, 2011 RNA-guided
639 complex from a bacterial immune system enhances target recognition through seed sequence
640 interactions. Proc Natl Acad Sci USA 108: 10092–10097.

641 Xue, C., A. S. Seetharam, O. Musharova, K. V. Severinov, S. J. J. Brouns *et al.*, 2015 CRISPR
642 interference and priming varies with individual spacer sequences. Nucleic Acids Research
643 43: 10831–10847.

644 Yosef, I., M. G. Goren, and U. Qimron, 2012 Proteins and DNA elements essential for the
645 CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Research 40: 5569–5576.

TABLE 1

The number of acquired spacers for the Lens (pLPL) and Lens (chromosome) CRISPR-Cas systems and their respective targets during self-priming.

System	Number of acquired spacers[a]	% map to priming plasmid[a]	% map to Lens (chromosome)[a]	% map to Lens (pLPL)[a]	% unknown[a]
Lens (pLPL)	78,753	98.38	0.05	0.01	1.57
Lens (chromsosome)	86,789	98.66	0.01	0.00	1.33

[a] Data are the average of 3 biological replicates.

TABLE 2

The repeat sequences of *L. pneumophila* type I-F CRISPR-Cas systems.[a]

Strain	# of spacers in the array	Consensus Repeat (5'-3')[b]	Mutated last repeat (5'-3')[b]
Lens (chromosome)	64	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**
FFI104	55	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**
Alcoy	56	GTTCACTGCCGCACAGG CAGCTTAGAA**G**	--
Lens (pLPL)	53	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
Missisauga-2006 (plasmid)	74	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
C8_S (plasmid)	24	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
Calgary2017	8	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
FJAB01	59	GTTCACTGCCGCACAGG CAGCTTAGAAA	--[c]
FJAD01	48	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
FJAJ01	55	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
FJAM01	72	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
FJBW01	69	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
FJDB01	107	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**
FJDP01	55	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
FJEH01	91	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**
FJEJ01	56	GTTCACTGCCGCACAGG CAGCTTAGAAA	--[c]
FJMD01	47	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
JFID01	92	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCATAGG CAG**TGGGTTG**
JFIM01	39	GTTCACTGCCGCACAGG CAGCTTAGAA**G**	--
LAYB01	94	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**

LBAK01	53	GTTCACTGCCG**T**ACAGG CAGCTTAGAAA	GTTCACTGCCG**T**ACAGG CAGCTTAGAA**G**
LBAN01	54	GTTCACTGCCGCACAGG CAGCTTAGAA**G**	--
LBAV01	34	GTTCACTGCCGCACAGG CAGCTTAGAA**G**	--
UGOL01	129	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**
UGOM01	53	GTTCACTGCCGCACAGG CAGCTTAGAAA	GTTCACTGCCGCACAGG CAGCTTAGAA**G**

[a] The isolates listed in this table reflect those present in the phylogeny from figure 1B

[b] Mutations in the last repeat relative to the consensus repeat, and in the two primary consensus repeats relative to each other, are denoted in red

[c] While these strains do not possess a mutated last repeat, their 3rd last repeat is mutated relative to the consensus repeat

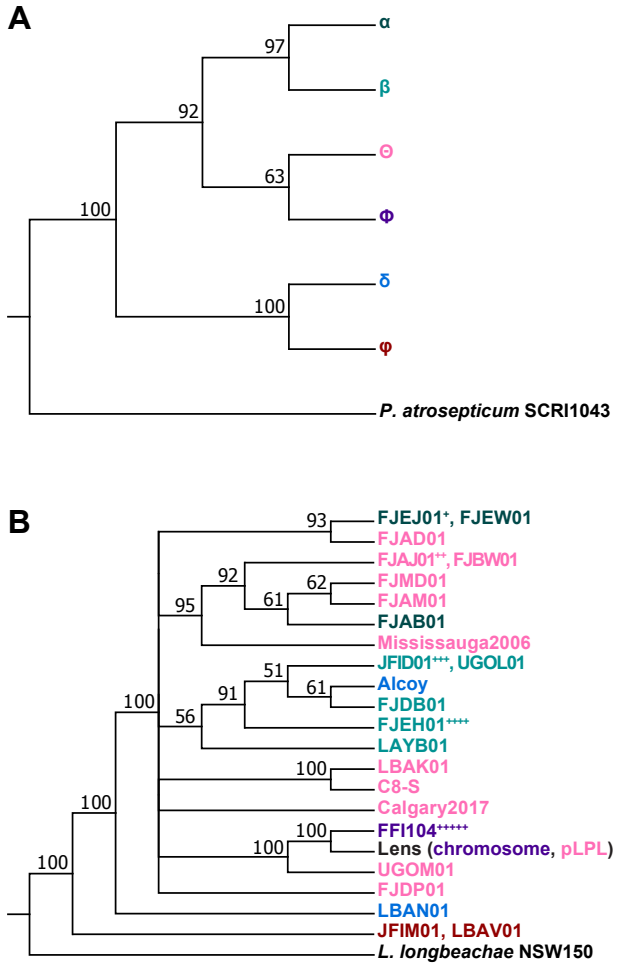

Figure 1 | Phylogenetic analysis of type I-F CRISPR-Cas system diversity in *L. pneumophila* reveals a horizontal distribution across isolates. *L. pneumophila* draft and completed genomes were analyzed using CRISPRDetect and CRISPRCasFinder to identify type I-F systems present within the genomes. Isolates that possessed type I-F systems were subjected to phylogenetic analyses of their *cas1* gene and core genome. The core genome alignment for the examined isolates was determined by Roary. Isolate names are colour-coded based on the *cas1* gene phylogeny to allow for comparison between the analyses. The isolates used in this analysis and their accession numbers can be found in supplementary table 1. **A)** The *cas1* gene phylogeny for *L. pneumophila* isolates with type I-F systems reveals six different *cas1* groups. **B)** The core genome phylogeny of the examined isolates is not congruent with the *cas1* phylogeny, suggesting that the type I-F system was horizontally distributed rather than vertically inherited. Note that Lens possesses two type I-F CRISPR-Cas systems, one on a plasmid (group Θ) and one on its chromosome (group Φ). Also note that isolates with 100% nucleotide identity in their core genome, as well as a shared *cas* gene group and redundant CRISPR array, have been collapsed to one representative in the phylogeny. Isolates with unique CRISPR arrays are listed. Collapsed isolates include:
+: FJEK01, FJEL01, FJEM01, FJEN01, FJEO01, FJEP01, FJEQ01, FJER01, FJES01, FJET01, FJEU01, FJEV01, FJEX01, FJEY01, FJEZ01, FJFA01
++: FJBI01, FJBU01
+++: LAXR01
++++: FJEI01
+++++: FFI105, FFI337

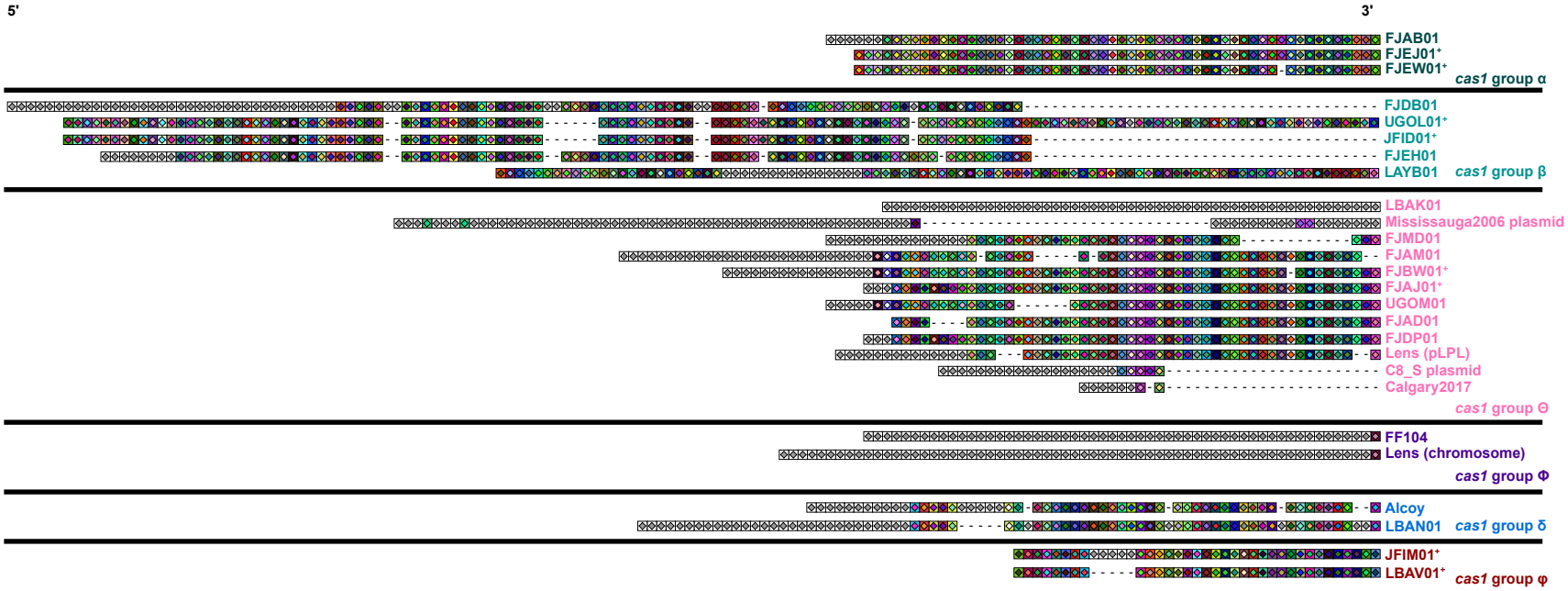

Figure 2 | CRISPR array analysis suggests that spacer acquisition and spacer loss has contributed to array diversification in *L. pneumophila* type I-F systems. *L. pneumophila* isolates were subjected to a CRISPRStudio analysis to look at the spacer composition of their CRISPR arrays. Grey boxes denote unique spacers, coloured boxes denote shared spacers and a dashed line denotes spacer loss. The isolate colour coding scheme is based on the *cas1* grouping from figure 1. Denoted isolates within the same *cas* gene group have 100% nucleotide identity in their core genomes. Note that *cas1* group β appears to have spacer rearrangements present, resulting in an imperfect alignment between the arrays.

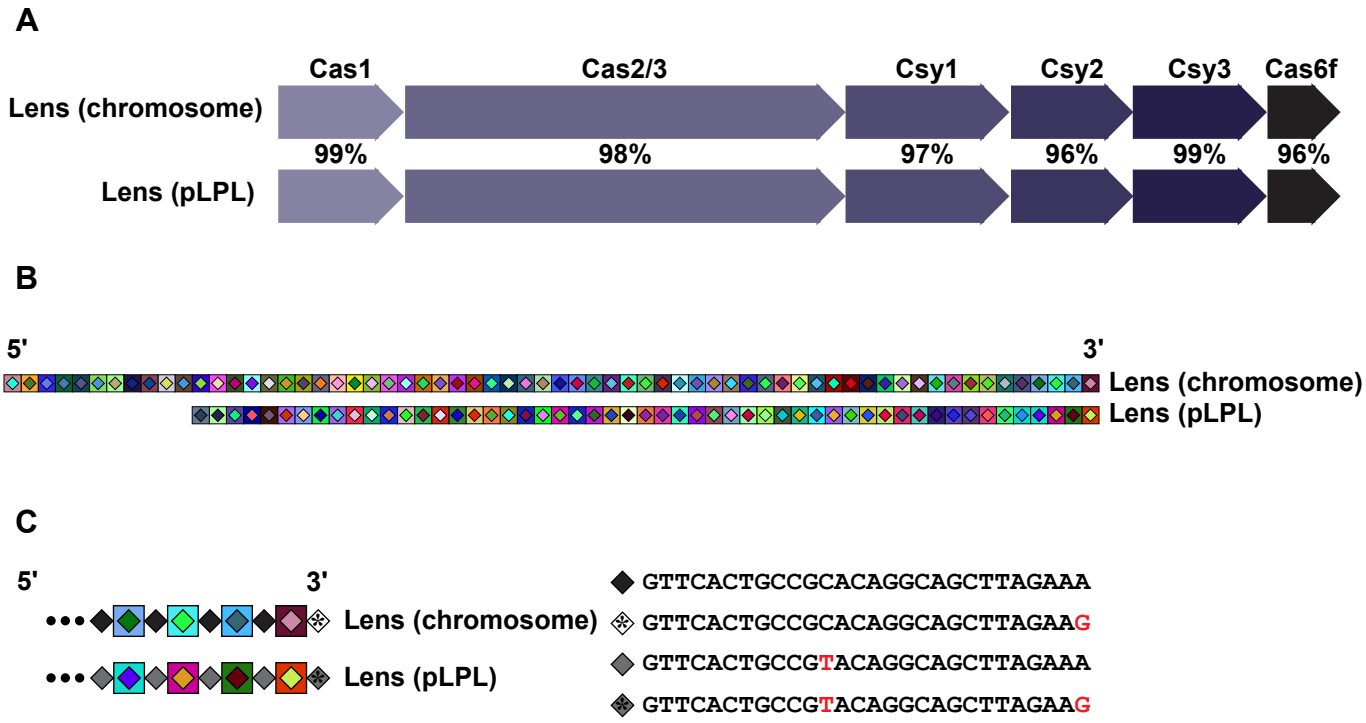

Figure 3 | A comparison of the Lens chromosome and pLPL CRISPR-Cas systems. A) The overall pairwise amino acid identity across the Cas proteins of the two systems is approximately 97%, with individual pairwise Cas protein identities ranging from 96% to 99%. **B)** A CRISPRStudio alignment of the arrays for two systems shows completely unique spacer content and a differing number of spacers. **C)** Analysis of the repeat sequences shows that the Lens pLPL and chromosome systems have one SNP between their consensus repeats, in addition to possessing a mutated last repeat in their arrays. Mutations are denoted in red.

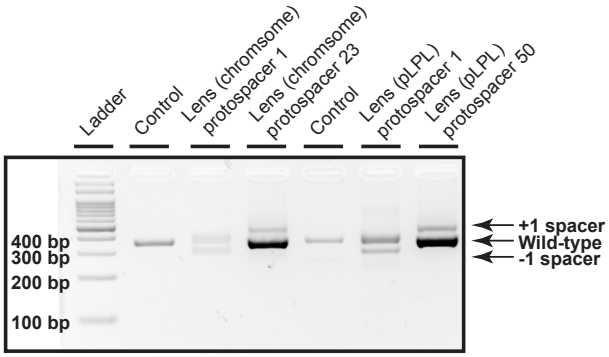

Figure 4 | Spacer acquisition and spacer loss in both Lens CRISPR-Cas systems. Spacer acquisition and loss were analyzed using a PCR-based screen where the leader-end of the CRISPR array for both the control samples and the transformed samples was amplified with system-specific primers to differentiate between the chromosomal Lens and the plasmid Lens arrays and visualized on an agarose gel. Products from the transformed samples were compared to the control, which contained untransformed genomic DNA. Bands representing spacer acquisition and loss are indicated.

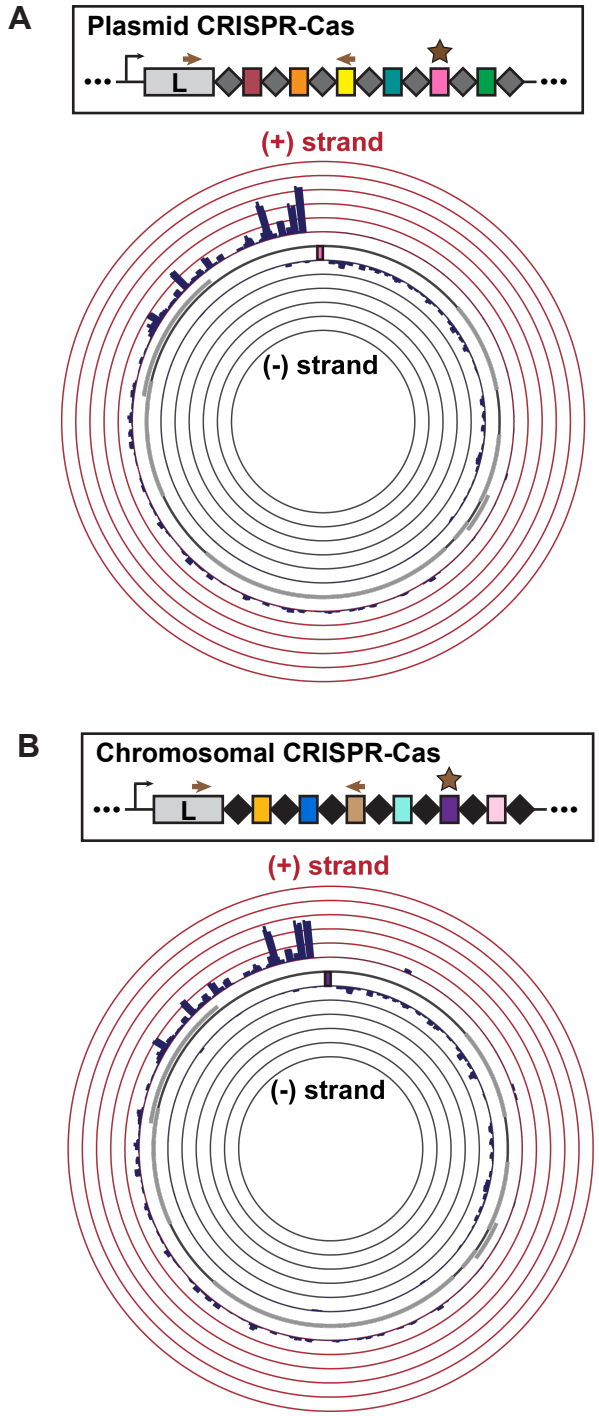

Figure 5 | Characterization of self-primed spacer acquisition in the two Lens CRISPR-Cas systems. Bacterial transformants with targeted plasmids were passaged for 20 generations without antibiotic selection to enrich for spacer acquisition; the leader end of the CRISPR array was amplified and the amplicons were Illumina sequenced. Newly targeted protospacers were obtained from the raw reads using an in-house bioinformatics pipeline and visualized with Circos. The arrows in the simplified array schematic show the primer location for PCR amplification prior to sequencing, while the star denotes the priming spacer. All data are the average of three biological replicates. **A)** The distribution of newly targeted protospacers mapped to the priming plasmid on the Circos plot reveals a strand bias in self-primed spacer acquisition within the Lens (pLPL) CRISPR-Cas system. The height of the bars indicates the number of spacers mapped to the position on the plasmid, up to 5% of total acquired spacers. **B)** The distribution of newly targeted protospacers mapped to the priming plasmid shows a similar pattern of self-primed spacer acquisition within the Lens (chromosome) CRISPR-Cas system to that of its pLPL CRISPR-Cas counterpart. Labelling as in **(A)**.

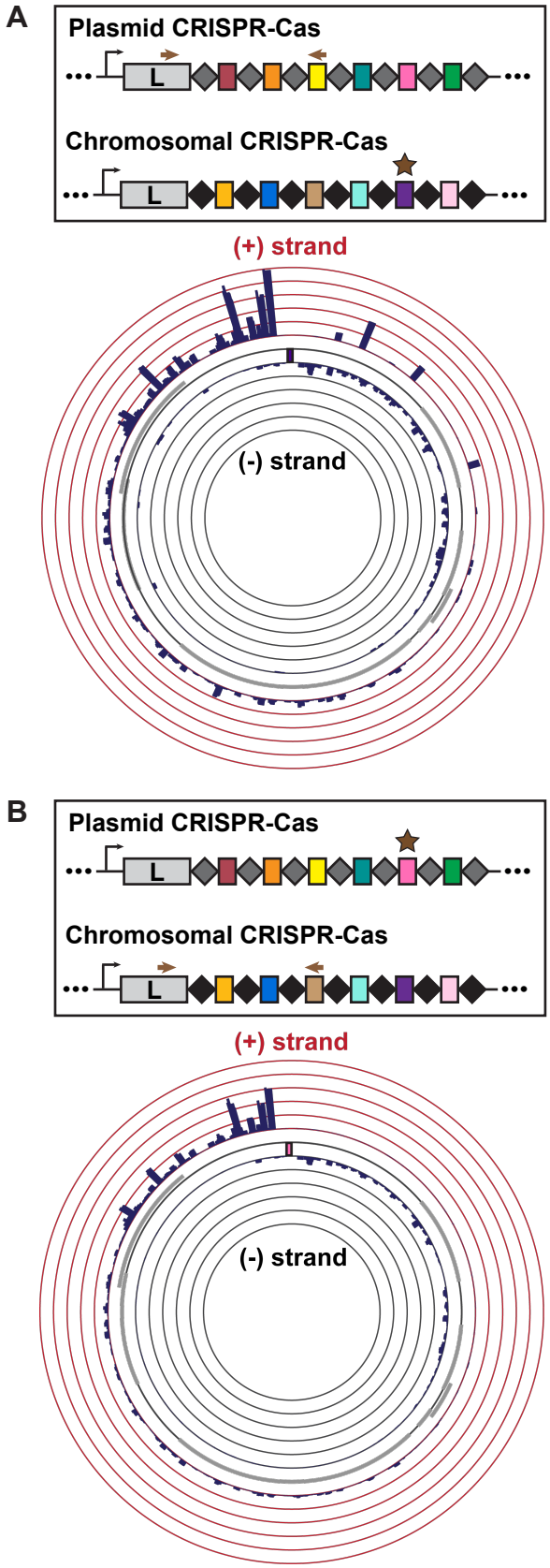

Figure 6 | Characterization of cross-primed spacer acquisition in the two Lens CRISPR-Cas systems. The experimental set-up is the same as described in figure 5. All data are the average of three biological replicates. The distribution of newly targeted protospacers mapped to the priming plasmid reveals cross-priming between the Lens (chromosome) and Lens (pLPL) CRISPR-Cas systems. **(A)** shows Lens (chromosome) primed, Lens (pLPL) array examined while **(B)** shows Lens (pLPL) primed, Lens (chromosome) array examined.

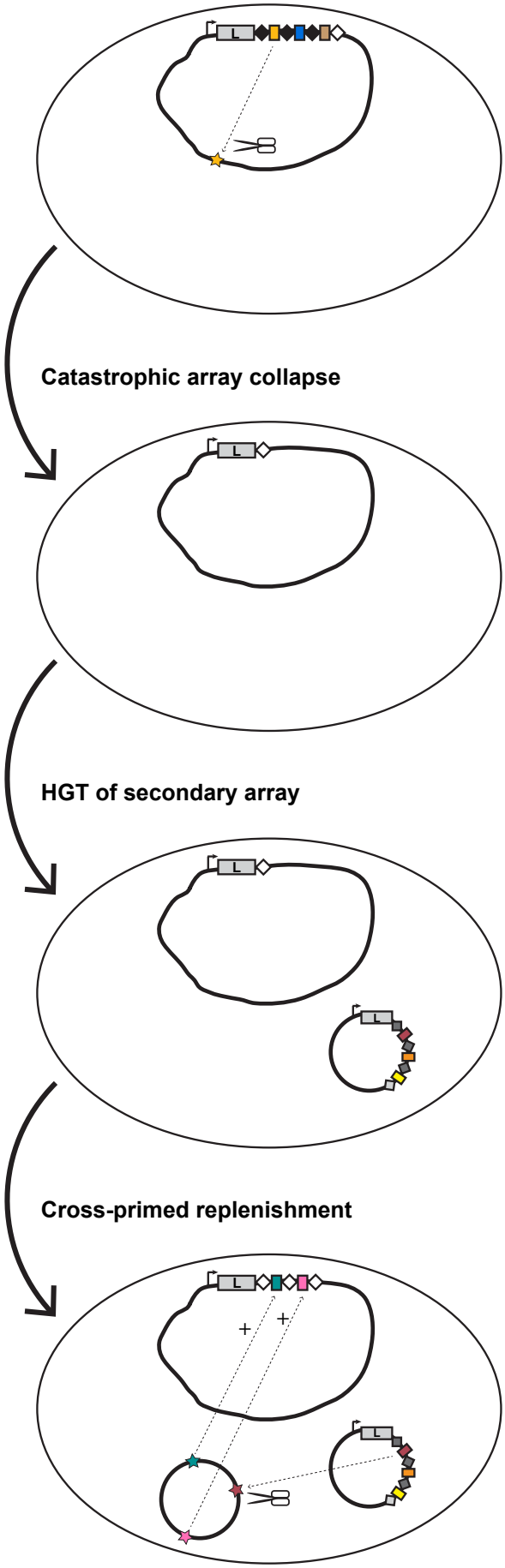

Figure 7 | Cross-priming between two similar CRISPR-Cas systems can repopulate a collapsed CRISPR array. A bacterium with a CRISPR-Cas system could undergo a mass spacer deletion event through homologous recombination between the first (black) and last (white) repeat sequences. The remaining locus carries a single repeat (white). While such "catastrophic collapse" events are likely to occur randomly at a certain rate, a driver of such a collapse could be the acquisition of a self-targeting spacer (yellow), selecting for spacer loss. Horizontal acquisition of a second CRISPR-Cas array (e.g. on a plasmid) is a first step towards replenishing the primary array. If cross-priming can occur between this secondary array and the collapsed array, the original CRISPR array is replenished, but bears an observable molecular scar - conversion of all the repeats to the sequence of the last repeat (white).

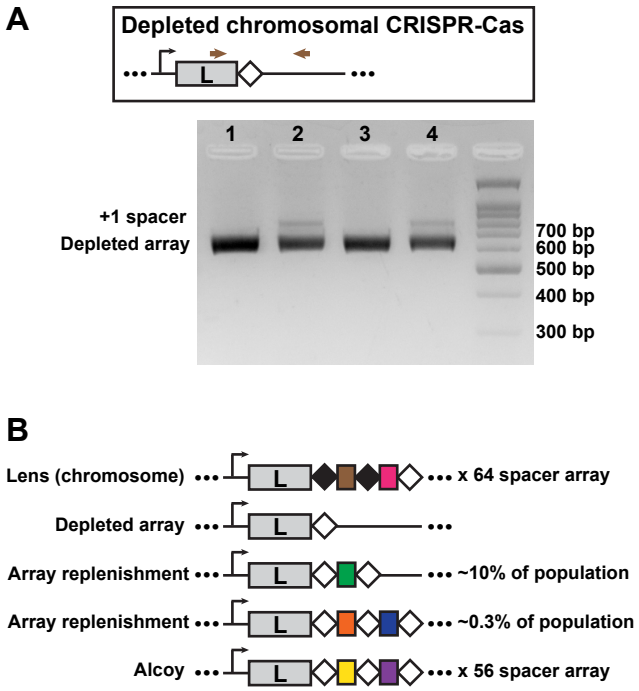

Figure 8 | An experimentally depleted chromosomal CRISPR array can be replenished through the activity of a plasmid-based array. A) Replenishment of a depleted array in the lab. Allelic replacement was used to remove the entire array from the chromosomal Lens I-F system, leaving behind a single, last repeat sequence (see materials and methods). This strain was then transformed with a plasmid targeted by the pLPL (Lens plasmid-based) I-F array previously shown to drive primed acquisition (pLPL protospacer 50; see Figure 4). Spacer acquisition by the empty chromosomal array was analyzed using a PCR based screen where the leader-end of the CRISPR array was amplified and visualized on an agarose gel. Products from the transformed samples (samples 2 and 4) were compared to untransformed controls (samples 1 and 3). Samples 1 and 2 are from depleted array clone #1 and samples 3 and 4 are from depleted array clone #2. **B)** Repeat signatures of depletion/replenishment. The repeat structure of the experimentally replenished Lens CRISPR arrays resembles that of *L. pneumophila* str. Alcoy, suggesting a similar array depletion/replenishment event may have occurred within the Alcoy lineage. The frequency of acquiring one new spacer versus two new spacers following replenishment in the Lens array depletion isolates was determined by Illumina sequencing.