# Working together: cross-priming in two *Legionella pneumophila* type I-F CRISPR-Cas systems.

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: Cross-priming in *L. pneumophila* 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 against 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). In this work, we leverage the power of *Legionella pneumophila*, a genetically tractable, gram-negative bacterium and the causative agent of Legionnaires disease, to examine CRISPR array dynamics and the interplay between two extremely similar type I-F systems present in a single isolate. Using an established transformation efficiency assay, we show that the type I-F system in *L. pneumophila* is a highly protective system, with prominent spacer loss occurring in some transformed populations for both plasmid and chromosomal systems. Turning to next-generation sequencing, we demonstrate that, during a primed acquisition response, both systems acquire spacers in a strand-biased and

directional manner, consistent with the patterns observed for previously studied type I-F systems in other bacterial species. We also show that the two systems can undergo cross-priming, whereby a target for one system can stimulate a primed acquisition response in the second. Finally, we combine these experimental data with bioinformatic analyses to propose a model in which cross-priming may replenish a depleted CRISPR array following a mass spacer deletion event.

**IMPORTANCE:** *Legionella pneumophila* is an aquatic bacterium that causes Legionnaires' disease, an often-fatal pneumonia. Many *L. pneumophila* strains possess one or more bacterial immune systems (CRISPR-Cas) that protect them from potentially harmful genetic elements. The genetic tractability of *L. pneumophila*, together with the diversity of CRISPR-Cas systems found within the species, make these bacteria attractive model systems within which to study bacterial defenses. In particular, key strengths are the ability to compare the functionality of different systems in otherwise identical genetic backgrounds and the cross-talk between multiple systems present within a single isolate. In this work, we characterized two nearly identical systems in a single *L. pneumophila* isolate and propose a model whereby cross-talk may restore functionality to otherwise defenseless systems.

## Introduction

Microorganisms have evolved over millions of years to survive in harsh environments, and their prosperity can be attributed in part to immune strategies that protect against antagonistic genetic elements, such as viral phages and foreign DNA elements (1). Clustered regularly interspaced short palindromic repeats (CRISPR) when coupled with associated *cas*

47 genes form a potent adaptive immune response in numerous prokaryotic species (2-4). These
48 systems have been classified into six major types, which are further divided into various sub-
49 types, based on their mechanism of action and Cas protein content (5-7).

50 A CRISPR response to invading DNA occurs in three distinct phases: adaptation,
51 expression and interference (2-4). In the adaptation phase, the CRISPR-Cas system acquires a
52 DNA sequence (spacer) from the invader and integrates it into an array of spacers interspersed
53 with repetitive sequences (2, 8-11). The spacers are generally derived from foreign elements
54 whose infection was unsuccessful, such as defunct phage (12), and form the basis of
55 immunological memory for the bacterium. During the expression phase, the array is transcribed
56 and processed to form CRISPR RNA (crRNA) molecules that recruit Cas proteins to form a
57 surveillance complex (3, 13). Infection by a previously encountered invader initiates the
58 interference step, wherein the surveillance complex recognizes and binds the foreign DNA via
59 base-pairing with the complementary crRNA, and cleaves it using a double stranded break,
60 effectively neutralizing the threat to the host (3, 14-16).

61 Although there are many differences between CRISPR-Cas systems, Cas1 and Cas2 are
62 present in all known systems (5-7), and are the only Cas proteins necessary for adaptation in
63 *Escherichia coli* type I-E systems (17, 18). When an invading element has not been previously
64 encountered by the bacterium, "naïve" acquisition occurs (17, 19), in which Cas1 and Cas2 form
65 a complex (20-22) that binds a dsDNA "pre-spacer" substrate (23), which is processed and
66 integrated into the CRISPR array on the leader-proximal end (23, 24).

67 Despite the sophistication of CRISPR-Cas systems, phages and foreign DNA elements
68 can still escape CRISPR-Cas targeting. A common mechanism of escape is the accumulation of
69 random mutations, which can prevent complementary base pairing with crRNAs during

70 interference (18, 25, 26). Although effective, the CRISPR-Cas system can overcome this
71 challenge by simply acquiring a new spacer; in fact, imperfect CRISPR targeting often leads to a
72 highly efficient "primed" acquisition response, providing an intrinsic mechanism to protect
73 against mutational escape (18, 27-30). Primed acquisition has been studied in type I-B (31-33), I-
74 C (34), I-E (18, 27-29, 35) and I-F (30, 36, 37) CRISPR-Cas systems, and a model has been
75 proposed in which the interference complex is recruited to the targeted sequence and
76 subsequently "slides" away from the site in a 3'- 5' direction (30, 31, 37). When it recognizes an
77 appropriate protospacer adjacent motif (PAM) sequence, the complex recruits Cas1 and Cas2 to
78 extract the spacer and integrate it into the array (30, 31, 37). Interference-driven acquisition, or
79 targeted acquisition, has also been observed in type I-C (34) and I-F systems (37), wherein a
80 primed acquisition response occurs against a target with a perfect match to a spacer already
81 within the array.

82 Most isolates of *Legionella pneumophila*, a genetically tractable gram-negative bacterium
83 and the causative agent of Legionnaires' disease, possess any of three different CRISPR-Cas
84 systems: types I-C, I-F and/or II-B (38, 39). We recently showed that the type I-C system
85 actively acquire spacers to protect against invasion (39) and characterized its targeted acquisition
86 response (34). One strength of *L. pneumophila* as a model is the frequent presence of multiple
87 CRISPR-Cas loci in one isolate, allowing for the study of interplay between different systems.
88 For instance, in *L. pneumophila* str. Lens, two type I-F CRISPR-Cas systems are present: one on
89 its chromosome and one on an endogenous 60 Kb plasmid (38, 39). The two systems have a
90 97.6% Cas protein identity and the repeat units between the spacers in the CRISPR array differ
91 by only a single nucleotide (39). The CRISPR arrays themselves are of different lengths (64
92 spacers for chromosomal Lens and 53 spacers for plasmid Lens) and each array contains a set of

non-overlapping, unique spacer sequences (38, 39). The presence of two remarkably similar I-F systems in *L. pneumophila* str. Lens provided us with an opportunity to examine targeted spacer acquisition in both of these largely uncharacterized CRISPR-Cas systems and the interplay between them.

## Results

### The two type I-F CRISPR-Cas systems in *L. pneumophila* str. Lens can undergo targeted spacer acquisition and spacer loss

In previous studies, we established that *L. pneumophila* type I-C CRISPR-Cas systems are active (39), and that it is a relatively permissive system that allows for targeted spacer acquisition when challenged with the most recently acquired spacer in the CRISPR array (34). To similarly lay the groundwork for type I-F study in *L. pneumophila*, we sought to determine the appropriateness of perfectly matched protospacer containing plasmids for driving spacer acquisition in these systems. As a first step, we performed an established transformation efficiency assay (4) to assess CRISPR-Cas activity in both Lens systems using two different targeted protospacer sequences: one matching the most recently acquired spacer and one matching a spacer from the middle of the array. (Unless otherwise stated, all targeted protospacer sequences used to investigate spacer acquisition were located on the DNA minus (-) strand.) When normalized to a scrambled plasmid control transformation, the protospacer matching the most recently acquired spacer (spacer 1) exhibited a ~100-fold reduction in transformation efficiency compared to the protospacer matching a mid-array spacer (chromosomal spacer 23 and plasmid spacer 50) (Fig. 1).

115 To determine whether spacer acquisition occurs within the context of a perfectly matched
116 protospacer target, we pooled the transformed populations, passaged them on an automated
117 liquid handler for 20 generations without selection, extracted their genomic DNA and screened
118 the leader end of the CRISPR array by PCR and agarose gel electrophoresis. Notably, while the
119 populations transformed with plasmids encoding either protospacer 23 (chromosome) or
120 protospacer 50 (plasmid) exhibited spacer acquisition in both Lens systems (Fig. 2), the
121 populations transformed with protospacer 1 plasmids exhibited spacer loss, with spacer
122 acquisition undetectable on a gel. While spacer loss has been noted previously in the literature
123 (34, 40-44), its prominence in our populations stand in stark contrast to our observations on the
124 *L. pneumophila* type I-C system, which is relatively permissive and highly adaptive - even in the
125 context of a perfectly matched protospacer (34). Given this observation, we proceeded to use the
126 mid-array targeted protospacer sequences for the remainder of our experiments on *L.*
127 *pneumophila* type I-F adaptation.
128
129 **Targeted spacer acquisition in the plasmid Lens CRISPR-Cas system**
130 To characterize the patterns of targeted spacer acquisition in the plasmid Lens CRISPR-
131 Cas system, we amplified the leader-proximal region of the plasmid Lens CRISPR array from
132 the populations transformed with the protospacer 50 plasmid. We Illumina sequenced these PCR
133 products and used an established bioinformatics pipeline (34) to identify newly acquired spacer
134 sequences within each read (Table 1). We mapped the protospacer locations on the priming
135 plasmid for the newly acquired spacers and visualized these patterns with Circos (45) using an
136 average of three replicates (Fig. 3A), although the individual distributions for all three replicates
137 were consistent (Fig. S1). Similar to the patterns of primed and targeted spacer acquisition

138 observed in the *Pectobacterium atrosepticum* type I-F CRISPR-Cas system (30, 37), the plasmid
139 Lens CRISPR-Cas system exhibited a biased distribution of acquired spacers. The majority of
140 the acquired protospacers clustered around the priming sequence on the plasmid (Fig. 3A).
141 Furthermore, the non-primed strand of DNA, in this case the plus (+) strand, contained ~3/4 of
142 the newly targeted protospacers. A similar distribution skew was observed moving in the 3' and
143 5' directions from the priming protospacer, as the 3' direction contained ~2/3 of the new
144 protospacers, consistent with the aforementioned sliding model (30, 31, 37). One prediction of
145 the sliding model is that swapping the strand on which the protospacer resides should result in a
146 "mirror-reflection" pattern of acquisition (30, 37). To test this prediction, we repeated the above
147 experiment with a protospacer 50 plasmid that targeted the (+) strand instead of the (-) strand. As
148 expected, we observed the distribution of new protospacers mirrored the distribution observed
149 when the (-) strand contained the targeted protospacer (Fig. 3B)

150 We next sought to determine the length distribution of the acquired spacers and the PAM
151 sequences associated with the new protospacers. When the (-) strand contained the targeted
152 protospacer, the predominant length for the acquired spacers was 32 nt (~95%), which is the only
153 spacer length found in the wild-type plasmid Lens CRISPR array (Fig. 3C). The most prevalent
154 PAM for the new protospacers was the canonical GG PAM found in type I-F systems (30, 36,
155 37, 46, 47), which accounted for ~95% of new protospacer PAMs (Fig. 3D and 3E). In the
156 mirrored (+) strand targeted samples, the spacer length and PAM distributions are comparable
157 with those of the (-) strand targeted samples (Fig. S2). Taken together, these data suggest
158 distribution bias of new protospacers is influenced by the strand containing the targeted
159 protospacer, while the spacer length and PAM distributions are not in the plasmid Lens CRISPR-

160 Cas system, the, consistent with the results reported by Staals and colleagues for targeted
161 acquisition in a *P. atrosepticum* type I-F CRISPR-Cas system (37).
162
163 **Targeted spacer acquisition in the chromosomal Lens CRISPR-Cas system**
164 After surveying the plasmid Lens CRISPR-Cas system for targeted spacer acquisition, we
165 turned our attention to exploring this phenomenon in the chromosomal Lens CRISPR-Cas
166 system. We amplified the leader-proximal region of chromosomal Lens CRISPR array from the
167 populations transformed with the protospacer 23 plasmid, and subsequently analyzed targeted
168 spacer acquisition as described for the plasmid Lens system.
169 Unsurprisingly, given how similar the chromosomal Lens and plasmid Lens systems are
170 on a Cas protein sequence level, the distribution of new protospacers for the chromosomal Lens
171 system resembled that of the plasmid Lens system (Table 1, Fig. 4A). The predominant spacer
172 length was 32 nt, accounting for ~90% of acquired spacers (Fig. 4B), and the canonical GG
173 PAM (30, 36, 37, 46, 47) also accounted for ~90% of new protospacer PAMs (Fig. 4C and 4D).
174 Taken together, these results suggest that the chromosomal and plasmid Lens CRISPR-Cas
175 systems operate in a highly comparable manner during targeted spacer acquisition.
176
177 **The chromosomal Lens and plasmid Lens CRISPR-Cas systems can undergo cross-**
178 **priming**
179 Since the plasmid Lens CRISPR-Cas system and the chromosomal Lens CRISPR-Cas
180 system function in a very similar manner during targeted acquisition, we speculated that cross-
181 priming between the two systems could occur; that is, a targeted protospacer sequence for one
182 CRISPR-Cas system could initiate a primed acquisition response in the second CRISPR-Cas

183 system. In order to test this hypothesis, we analyzed spacer acquisition in the chromosomal
184 CRISPR array in populations transformed with the protospacer 50 plasmid (complementary to
185 the plasmid mid-array spacer) and analyzed spacer acquisition in the plasmid CRISPR array in
186 populations transformed with protospacer 23 plasmid (complementary to the chromosomal mid-
187 array spacer). We observed strikingly similar patterns of distribution for the new protospacers in
188 the two populations (Fig. 5), which were comparable with those seen in the previous targeted
189 acquisition experiments, indicating that the two CRISPR-Cas systems undergo a high degree of
190 cross-priming. There were some slight, but noticeable, differences in protospacer distribution on
191 the (+) strand on the 5' end of the priming protospacer for the chromosomal Lens primed,
192 plasmid Lens amplified sample. However, the peaks were not large enough for us to postulate
193 that they are "hotspot" regions of spacer acquisition, and we did not investigate them further.
194

195 **Bioinformatic analysis of *L. pneumophila* I-F CRISPR-Cas systems suggests cross-priming**
196 **can re-populate depleted CRISPR arrays**

197 We next aimed to further explore the implications of our observation that perfectly
198 targeted protospacer 1 plasmids result in populations enriched for spacer loss. While selecting for
199 maintenance of an efficiently targeted plasmid in the context of a wild-type CRISPR-Cas system
200 is a laboratory construct, such observations may have real-world implications as CRISPR-Cas
201 systems are known to acquire self-targeting spacers at a low, but detectable rate (17, 18, 28, 34,
202 36, 37). In such instances where a system accidentally acquires the ability to cleave its resident
203 genome, our data suggest that loss of one or more spacers (sometimes the entire array) might be
204 a mechanism by which to escape the dire consequences of such an event. Given our observations
205 that a protospacer 1 priming plasmid promoted spacer loss instead of spacer acquisition (Fig. 2),

206 and that cross-priming was occurring between the two Lens CRISPR-Cas systems (Fig. 5), we
207 bioinformatically tested the hypothesis that cross-priming between two related CRISPR-Cas
208 systems could be a way to re-populate a depleted CRISPR array.

209 In total, we analyzed five chromosome-based systems and three plasmid-based type I-F
210 CRISPR-Cas systems present in different *L. pneumophila* isolates, using data collected from our
211 previous study (39) and from Genbank (accessed September 2017). We evaluated three different
212 criteria in each CRISPR array: the repeat sequence, any mutations present in the last repeat, and
213 the number of spacers in the array (Table 2). Our analyses showed that 7/8 of the strains share
214 the same repeat sequence and the same mutated last repeat sequence, with the exception of a C to
215 T single nucleotide polymorphism present at position 12 in all repeats of the three plasmid-based
216 systems. Notably, the remaining chromosome-based system, in *L. pneumophila* str. Alcoy, has
217 no mutations in its last repeat. However, its repeat sequence is identical to the mutated last repeat
218 sequence present in the other chromosome-based systems. One intriguing interpretation of these
219 data is that Alcoy underwent a whole CRISPR array deletion through homologous recombination
220 between the first and last repeat sequences, leaving it with only the mutated last repeat. This
221 would have been followed by array replenishment, since the array contains 56 spacers, but no
222 mutations have emerged in the repetitive sequences, suggesting this was a relatively recent event.

223 The spacer sequences in the Alcoy array are unique and many of the spacer targets are
224 unknown. However, one spacer corresponds to a foreign plasmid element known as *Legionella*
225 mobile element-1 (LME-1), that was discovered as a common target for CRISPR-Cas in many *L.*
226 *pneumophila* strains (39). Together, our observations suggest that in strains with a depleted
227 CRISPR array, if a plasmid harboring a related CRISPR-Cas system was horizontally transferred
228 to the array-less strain, it could re-populate the CRISPR array through cross-priming when it

229 comes into contact with a widespread foe, such as LME-1. Subsequent loss of this plasmid
230 would leave little trace of such an event, other than a potential modification of the consensus
231 repeat sequence.
232
233 **Discussion**
234 We previously showed that type I-C CRISPR-Cas in *Legionella pneumophila* is highly
235 permissive, protects against a mobile genetic element, and is adaptive (34, 39). The patterns and
236 fidelity of primed spacer acquisition that we observed for *L. pneumophila* type I-C were
237 consistent with the previous observations of type I-F spacer acquisition in other bacterial species,
238 including *Pseudomonas aeruginosa* (36), *Escherichia coli* (36) and *Pectobacterium atrosepticum*
239 (30, 37). One strength of *L. pneumophila* as a model for studying CRISPR-Cas is the diversity of
240 system types present in this species and the frequent coexistence of multiple CRISPR-Cas
241 systems within the same isolate. We have bioinformatically identified eight distinct type I-F
242 systems in Legionella, and experimentally shown activity for 3 of them: *L. pneumophila* str.
243 Lens (plasmid and chromosome) and str. Mississauga-2006 (plasmid) (39). Each system
244 contains nearly identical *cas* genes but different spacer arrays. As we previously hypothesized
245 the diversification of type I-F arrays in *L. pneumophila* could emerge from extensive spacer
246 acquisition (39), we sought to directly test the adaptability of two of these arrays, both present in
247 *L. pneumophila* str. Lens.
248 The patterns of targeted acquisition observed in both the plasmid Lens and the
249 chromosomal Lens type I-F systems are remarkably similar to both primed and targeted
250 acquisition in other type I-F systems (30, 36, 37) (Figs. 3 and 4). Consistent with the similarity of
251 the *cas* genes, these two Lens systems undergo cross-priming, where the targeted sequence for

252 one system stimulates a primed acquisition response in the second system (Fig. 5). Regardless of
253 the source of priming, our data support the sliding model of primed acquisition, in which the
254 interference complex translocates away from the targeted sequence in a 3' to 5' manner, and
255 recruits Cas1 and Cas2 to capture a new spacer for array integration after recognizing an
256 appropriate PAM (30, 31, 37).

257 Our bioinformatic analyses of CRISPR arrays from type I-F systems in eight strains of *L.*
258 *pneumophila* showed that with the exception of a C to T polymorphism present at position 12 in
259 the three examined plasmid systems, the repetitive sequences are the same across all eight arrays
260 (Table 2). Additionally, 7/8 of the strains possessed a mutation in the last repeat of the array.
261 Based on these data, we hypothesize that the I-F system in *L. pneumophila* was horizontally
262 acquired from a plasmid and that this common ancestor has subsequently diverged based on the
263 spacer content and repeat sequences found in the varying arrays. Since the majority of the
264 examined arrays harbor mutations in the last repeat, it is plausible that genetic drift has occurred
265 since the acquisition of the I-F system to form the consensus repeat found in the remainder of the
266 array. This could be used to compare the timing of acquisition events within the array, as one
267 might expect other mutations to arise over time in the repeat sequences due to genetic drift.

268 Combining our bioinformatic analyses with our experimental data, we propose that *L.*
269 *pneumophila* str. Alcoy (which has a consensus repeat that matches the mutated last repeat of
270 other type I-F systems) underwent a mass spacer loss event followed by subsequent array
271 replenishment. We hypothesize that cross-priming between two CRISPR-Cas systems could be
272 yet another mechanism to not only protect against spacer loss, as spacers can be acquired at a
273 more frequent rate, but also to aid the system in quickly and efficiently replenishing an array that
274 has undergone a mass loss event.

275 Many of the I-F systems in *L. pneumophila* have different array lengths, ranging from 24
276 spacers to 74 spacers, with an average length of 54 spacers (Table 2). Toms and Barrangou
277 recently performed a global analysis of class I CRISPR arrays and found that the average array
278 length for type I-F systems was 33 spacers, with statistically significant differences between the
279 array lengths of different type I subtypes (48). Accordingly, if spacer acquisition is a driving
280 force in array divergence, it is likely coupled to spacer loss. Close examination of the
281 mechanisms driving spacer loss in these systems, combined with comparative genomics of
282 otherwise related strains, will be crucial to further testing the model of array diversification in *L.*
283 *pneumophila*.
284
285 **Methods and Materials**
286 **Bacterial strains, plasmids and oligos used**
287 The bacterial strains and plasmids used in this study are listed in supplementary table 1,
288 and the oligos used in this study are listed in supplementary table 2.
289 The priming plasmids were created by annealing oligos (see supplementary table 2) to
290 create the protospacer insert with the canonical GG PAM (30, 36, 37, 46, 47) and subsequently
291 ligating the insert into an ApaI/PstI-cut pMMB207 vector (49). The scrambled control plasmid
292 was created in the same manner, except it contained a 32-nt scrambled sequence in place of a
293 targeted protospacer sequence.
294
295 **Transformation efficiency assay and population pool generation**
296 The transformation efficiency assay was performed as we have previously described (39)
297 with some modifications. Briefly, overnight cultures of *L. pneumophila* str. Lens were grown in

298 ACES-buffered yeast extract (AYE) medium to an $OD_{600}$ of ~4.0 using two-day patches that
299 were grown on charcoal buffered ACES yeast extract (CYE) plates. Pellets from 4.0 $OD_{600}$ of
300 culture underwent three washing steps: twice with 1 mL of ice-cold ultrapure water and once
301 with 1 mL of ice-cold 10% glycerol. The pellet was then re-suspended in 200 uL of ice-cold 10%
302 glycerol and for every 50 uL of cell suspension, 100 ng of plasmid was added to the sample. The
303 solution was transferred to an ice-cold electroporation cuvette with a 2 mm gap and
304 electroporated with the following settings: 2500 kV, 600 Ω and 25 mF. After electroporation,
305 800 uL of AYE medium was added to each sample and the samples recovered for 3 hours at
306 37°C at 600 RPM in a shaking incubator. The samples were plated in a dilution series on CYE
307 plates supplemented with 5 mg $mL^{-1}$ of chloramphenicol and incubated at 37°C for 3 days. The
308 relative transformation efficiency for each targeted plasmid was calculated as a percentage of the
309 transformation efficiency obtained from the scrambled control plasmid. Three biological
310 replicates were performed for each transformation efficiency assay.
311 Population pools were generated by mixing together ≥ 50 colonies per population from
312 the CYE plates supplemented with 5 μg $mL^{-1}$ of chloramphenicol using AYE medium
313 supplemented with 5 μg $mL^{-1}$ of chloramphenicol. Population pools were made in triplicate for
314 each transformed plasmid.
315
316 **Serial passaging on an automated liquid handler**
317 The serial passaging of transformed *L. pneumophila* str. Lens populations was performed
318 as described previously (39). Briefly, overnight cultures of the populations pools in AYE
319 medium supplemented with 5 μg $mL^{-1}$ of chloramphenicol for plasmid maintenance were grown
320 to an $OD_{600}$ of ~2.0. The culture was then back diluted to an $OD_{600}$ of ~0.0625 and grown in a

321 flat-bottom 48-well plate (Greiner) in a shaking incubator at 37°C. A Freedom Evo 100 liquid
322 handler (Tecan) connected to an Infinite M200 Pro plate reader (Tecan) measured the optical
323 density of the plate every 20 minutes, until an $OD_{600}$ of ~2.0 was reached. The cultures were then
324 automatically back diluted to an $OD_{600}$ of ~0.0625 in the adjacent well to continue growth, and
325 the remaining culture was transferred to a 48-well plate that was kept at 4°C. In this manner,
326 each saved culture represented ~5 generations of growth. The passaging was done without
327 selection in AYE medium to allow for plasmid loss during passaging.
328
329 **Genomic DNA extraction, PCR and agarose gel screen**
330 Genomic DNA was extracted from the passaged cultures using the Machery-Nagel
331 Nucleospin Tissue kit as per the kit protocol. The extracted samples were used as a template in a
332 30-cycle PCR reaction with Econotaq Polymerase (Lucigen) to amplify the leader end of the
333 CRISPR array using primers listed in Table S2. The PCR products were then separated on a 3%
334 agarose gel to determine if spacer acquisition or spacer loss had occurred based on the presence
335 of an upper or lower band, respectively, relative to the control sample.
336
337 **Nextera library prep and Illumina sequencing**
338 The extracted genomic DNA was prepared for leader-end array sequencing by
339 performing a 20-cycle PCR using Kapa HiFi Polymerase (Kapa Biosystems) and the primers
340 listed in Table S2. The PCR products were purified using a Machery-Nagel Nucleospin Gel and
341 PCR Clean-up kit as per the manufacturer's instructions and normalized to 1 ng using Picogreen.
342 The DNA was then tagmented using the Nextera XT tagmentation kit as per the manufacturer's
343 instructions. The tagmented products were sequenced with a paired-end (2 x 150 bp) sequencing

344 run on an Illumina NextSeq platform at the Centre for the Analysis of Genome Evolution and
345 Function (CAGEF) at the University of Toronto.
346
347 **Bioinformatic analyses**
348 The bioinformatic analysis of the Illumina sequence data were performed as described
349 previously (34). Briefly, the raw paired-end reads were merged using FLASH (50), and any
350 unpaired reads were subsequently quality trimmed using Trimmomatic (51). These processed
351 reads were then combined and analyzed using a Perl script (available upon request) that
352 annotated existing spacers (S), newly acquired spacers (X), repetitive sequences (R) and the
353 downstream sequence (D). The newly acquired spacers were aligned to the priming plasmid, the
354 *L. pneumophila* str. Lens chromosome or the *L. pneumophila* str. Lens plasmid using BLASTN.
355 The results from the BLASTN alignment for the priming plasmid were then processed to obtain
356 the coverage per nucleotide, and plotted on the reference sequence using Circos (45). For the
357 PAM analyses, the flanking sequence of each new spacer was extracted and plotted using Web
358 Logo (52).
359 For the bioinformatics analyses of the *L. pneumophila* type I-F CRISPR arrays, the
360 repetitive sequence in *L. pneumophila* str. Lens was subjected to a BLAST search against other
361 *L. pneumophila* strains in Genbank (accessed September 2017). The hits were processed using
362 CRISPRFinder (53) to determine if there was a CRISPR system present in the strain and its type;
363 only eight strains with a type I-F system were examined, noting the repeat sequences, the number
364 of spacers present in each array and whether the system was on a chromosome or on a plasmid.
365 Mutations in the last sequence of each array were noted, as were any mutations between the
366 consensus repeat sequences of the different strains.

367

368 **Data Accessibility**

369 The raw Illumina reads have been deposited into the NCBI sequence read archive under
370 the BioProject PRJNA433194.

371

372 **Acknowledgements**

373 The authors thank Griffin Deecker (a volunteer high school student) for his assistance in
374 bioinformatically examining the diversity of I-F repeat sequences in *L. pneumophila*. We also
375 thank the Center for the Analysis of Genome Evolution and Function (CAGEF) at the University
376 of Toronto for performing Illumina sequencing. We thank members of the Ensminger laboratory
377 for their suggestions and careful reading of the manuscript, in particular Beth Nicholson and
378 Malene Urbanus. SRD is supported by a fellowship from the Department of Biochemistry,
379 University of Toronto. This work was supported by a Project Grant from the Canadian Institutes
380 of Health Research (PHT-148819), the Connaught Fund (NR-2015-16), and an infrastructure
381 grant from the Canada Foundation for Innovation and the Ontario Research Fund (30364) to
382 AWE.

383

384 **Figures**

385 **Figure 1 | The chromosomal and plasmid type I-F CRISPR-Cas systems in *L. pneumophila***
386 ***str.* Lens are active against plasmids containing protospacers**. *L. pneumophila* str. Lens was
387 transformed with plasmids containing targeted protospacer sequences matched to the first spacer
388 (sp1) or a mid-array spacer (sp23 or sp50, respectively) of the chromosomal Lens CRISPR-Cas
389 system **(A)** or the plasmid Lens CRISPR-Cas system **(B)**. After plating on selective media and
390 incubating for three days, transformation efficiencies were calculated as a percentage of the
391 transformation efficiency of a control plasmid with a scrambled targeted sequence. The average
392 for three biological replicates is shown where the error bars represent the standard error of the
393 mean.

394

395 **Figure 2 | Targeted spacer acquisition and spacer loss in the chromosomal Lens and**
396 **plasmid Lens CRISPR-Cas systems**. Spacer acquisition and loss were analyzed using a PCR
397 based screen where the leader-end of the CRISPR array for both the control samples and the
398 transformed samples was amplified with system-specific primers to differentiate between the
399 chromosomal Lens and the plasmid Lens arrays and visualized on an agarose gel. Products from
400 the transformed samples were compared to the control, which contained untransformed genomic
401 DNA. Bands representing spacer acquisition and loss are indicated.

402

403 **Figure 3 | Characterization of targeted acquisition in the plasmid Lens CRISPR-Cas**
404 **system.** Bacterial transformants with targeted plasmids were passaged for 20 generations without
405 antibiotic selection to enrich for spacer acquisition; the leader-end of the CRISPR array was
406 amplified and the amplicons were Illumina sequenced. Acquired protospacers were obtained

407 from the raw reads using an in-house bioinformatics pipeline and visualized with Circos. Unless
408 otherwise noted, all data is the average of three biological replicates. **A)** The distribution of
409 acquired protospacers mapped to the priming plasmid on the Circos plot reveals a strand bias in
410 targeted acquisition within the plasmid Lens CRISPR-Cas system. The height of the bars
411 indicates the number of spacers mapped to the position on the plasmid, up to 5% of total
412 acquired spacers. The priming protospacer sequence used on the targeted plasmid is denoted in
413 red and its PAM sequence is denoted in purple above the Circos plot. Quantification of the
414 Circos plot is shown in a quad plot, where (+) is the plus strand, (-) is the minus strand, (L) is the
415 left side of the plasmid (5' half) and (R) is the right side of the plasmid (3' half), relative to the
416 priming site. The red bar indicates the priming protospacer and the grey box indicates its PAM
417 location and targeted strand. **B)** Reversing the direction of the targeted sequence on the priming
418 plasmid creates a mirrored distribution bias in acquired protospacers. Labelling as in (A). **C)** The
419 distribution of spacer lengths acquired from the targeted plasmid (grey) compared to the wild-
420 type CRISPR-Cas array (black, n = 53). **D)** Quantification of the PAMs for the new protospacers
421 are shown in a stacked bar plot; the (+) strand PAMs are denoted by the hatched bars and the (-)
422 strand PAMs are denoted by the grey bars. **E)** Sequence logo analysis of PAMs from new
423 protospacers demonstrate a preference for the canonical GG PAM, with the protospacer denoted
424 in the grey box. The sequence logo represents the PAMs of all acquired protospacers for the
425 three biological replicates.
426
427 **Figure 4 | Characterization of targeted acquisition activity in the chromosomal Lens**
428 **CRISPR-Cas system.** Labelling and experimental set-up are as described for figure 3. **A)** The
429 distribution of acquired protospacers mapped to the priming plasmid on the Circos plot reveals a

430 strand bias in targeted acquisition within the chromosomal Lens CRISPR-Cas system. The
431 priming protospacer sequence and its PAM are shown above Circos plot, while the quantification
432 of the acquired protospacers is shown in the lower panel as a quad plot. **B)** The distribution of
433 spacer length acquired from the targeted plasmid compared to the wild-type CRISPR-Cas array
434 (n = 64). **C)** Quantification of the PAMs for the new protospacers in a stacked bar plot. **D)**
435 Sequence logo analysis of PAMs from new protospacers.
436
437 **Figure 5 | Cross-priming occurs between the chromosomal Lens and plasmid Lens**
438 **CRISPR-Cas systems.** Labelling and experimental set-up are as described for figure 3. The
439 distribution of acquired protospacers mapped to the priming plasmid reveals cross-priming
440 between the chromosomal Lens and plasmid Lens CRISPR-Cas systems. **(A)** shows
441 chromosomal Lens primed, plasmid Lens amplified while **(B)** shows plasmid Lens primed,
442 chromosomal Lens amplified.
443
444

445 **References**


446 1. **van Houte S**, **Buckling A**, **Westra ER**. 2016. Evolutionary ecology of prokaryotic
447 immune mechanisms. Microbiol Mol Biol Rev **80**:745–763.

448 2. **Barrangou R**, **Fremaux C**, **Deveau H**, **Richards M**, **Boyaval P**, **Moineau S**, **Romero**
449 **DA**, **Horvath P**. 2007. CRISPR provides acquired resistance against viruses in
450 prokaryotes. Science **315**:1709–1712.

451 3. **Brouns SJJ**, **Jore MM**, **Lundgren M**, **Westra ER**, **Slijkhuis RJH**, **Snijders APL**,
452 **Dickman MJ**, **Makarova KS**, **Koonin EV**, **van der Oost J**. 2008. Small CRISPR RNAs
453 guide antiviral defense in prokaryotes. Science **321**:960–964.

454 4. **Marraffini LA**, **Sontheimer EJ**. 2008. CRISPR interference limits horizontal gene
455 transfer in Staphylococci by targeting DNA. Science **322**:1843–1845.

456 5. **Makarova KS**, **Haft DH**, **Barrangou R**, **Brouns SJJ**, **Charpentier E**, **Horvath P**,
457 **Moineau S**, **Mojica FJM**, **Wolf YI**, **Yakunin AF**, **van der Oost J**, **Koonin EV**. 2011.
458 Evolution and classification of the CRISPR–Cas systems. Nature Reviews Microbiology
459 **9**:467–477.

460 6. **Makarova KS**, **Wolf YI**, **Alkhnbashi OS**, **Costa F**, **Shah SA**, **Saunders SJ**, **Barrangou**
461 **R**, **Brouns SJJ**, **Charpentier E**, **Haft DH**, **Horvath P**, **Moineau S**, **Mojica FJM**, **Terns**
462 **RM**, **Terns MP**, **White MF**, **Yakunin AF**, **Garrett RA**, **van der Oost J**, **Backofen R**,
463 **Koonin EV**. 2015. An updated evolutionary classification of CRISPR–Cas systems.
464 Nature Reviews Microbiology **13**:722–736.

465 7. **Koonin EV**, **Makarova KS**, **Zhang F**. 2017. Diversity, classification and evolution of
466 CRISPR-Cas systems. Current Opinion in Microbiology **37**:67–78.

467 8. **Ishino Y**, **Shinagawa H**, **Makino K**, **Amemura M**, **Nakata A**. 1987. Nucleotide
468 sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in
469 *Escherichia coli*, and identification of the gene product. Journal of Bacteriology
470 **169**:5429–5433.

471 9. **Mojica FJM**, **Díez-Villaseñor C**, **Soria E**, **Juez G**. 2000. Biological significance of a
472 family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria.
473 Molecular Microbiology **36**:244–246.

474 10. **Jansen R**, **Embden JDAV**, **Gaastra W**, **Schouls LM**. 2002. Identification of genes that
475 are associated with DNA repeats in prokaryotes. Molecular Microbiology **43**:1565–1575.

476 11. **Jackson SA**, **McKenzie RE**, **Fagerlund RD**, **Kieper SN**, **Fineran PC**, **Brouns SJJ**.
477 2017. CRISPR-Cas: Adapting to change. Science **356**:eaal5056–11.

478 12. **Hynes AP**, **Villion M**, **Moineau S**. 2014. Adaptation in bacterial CRISPR-Cas immunity
479 can be driven by defective phages. Nature Communications **5**:1–6.

480 13. **Haurwitz RE**, **Jinek M**, **Wiedenheft B**, **Zhou K**, **Doudna JA**. 2010. Sequence- and
481 structure-specific RNA processing by a CRISPR endonuclease. Science **329**:1355–1358.

482 14. **Jore MM**, **Lundgren M**, **van Duijn E**, **Bultema JB**, **Westra ER**, **Waghmare SP**,
483 **Wiedenheft B**, **Pul Ü**, **Wurm R**, **Wagner R**, **Beijer MR**, **Barendregt A**, **Zhou K**,
484 **Snijders APL**, **Dickman MJ**, **Doudna JA**, **Boekema EJ**, **Heck AJR**, **van der Oost J**,

485 **Brouns SJJ**. 2011. Structural basis for CRISPR RNA-guided DNA recognition by
486 Cascade. Nature Publishing Group **18**:529–536.

487 15. **Wiedenheft B**, **van Duijn E**, **Bultema JB**, **Waghmare SP**, **Zhou K**, **Barendregt A**,
488 **Westphal W**, **Heck AJR**, **Boekema EJ**, **Dickman MJ**, **Doudna JA**. 2011. RNA-guided
489 complex from a bacterial immune system enhances target recognition through seed
490 sequence interactions. Proc Natl Acad Sci USA **108**:10092–10097.

491 16. **Westra ER**, **van Erp PBG**, **Künne T**, **Wong SP**, **Staals RHJ**, **Seegers CLC**, **Bollen S**,
492 **Jore MM**, **Semenova E**, **Severinov KV**, **de Vos WM**, **Dame RT**, **de Vries R**, **Brouns**
493 **SJJ**, **van der Oost J**. 2012. CRISPR immunity relies on the consecutive binding and
494 degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular Cell
495 **46**:595–605.

496 17. **Yosef I**, **Goren MG**, **Qimron U**. 2012. Proteins and DNA elements essential for the
497 CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Research **40**:5569–5576.

498 18. **Datsenko KA**, **Pougach K**, **Tikhonov A**, **Wanner BL**, **Severinov KV**, **Semenova E**.
499 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial
500 immunity system. Nature Communications **3**:945–7.

501 19. **Levy A**, **Goren MG**, **Yosef I**, **Auster O**, **Manor M**, **Amitai G**, **Edgar R**, **Qimron U**,
502 **Sorek R**. 2015. CRISPR adaptation biases explain preference for acquisition of foreign
503 DNA. Nature **520**:505–510.

504 20. **Richter C**, **Gristwood T**, **Clulow JS**, **Fineran PC**. 2012. In vivo protein interactions and
505 complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas system.
506 PLoS ONE **7**:e49549.

507 21. **Nuñez JK**, **Kranzusch PJ**, **Noeske J**, **Wright AV**, **Davies CW**, **Doudna JA**. 2014.
508 Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive
509 immunity. Nature Publishing Group **21**:528–534.

510 22. **Wang J**, **Li J**, **Zhao H**, **Sheng G**, **Wang M**, **Yin M**, **Wang Y**. 2015. Structural and
511 mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell
512 **163**:840–853.

513 23. **Nuñez JK**, **Harrington LB**, **Kranzusch PJ**, **Engelman AN**, **Doudna JA**. 2015. Foreign
514 DNA capture during CRISPR–Cas adaptive immunity. Nature **527**:535–538.

515 24. **Rollie C**, **Schneider S**, **Brinkmann AS**, **Bolt EL**, **White MF**. 2015. Intrinsic sequence
516 specificity of the Cas1 integrase directs new spacer acquisition. eLife **4**:e08716.

517 25. **Deveau H**, **Barrangou R**, **Garneau JE**, **Labonte J**, **Fremaux C**, **Boyaval P**, **Romero**
518 **DA**, **Horvath P**, **Moineau S**. 2008. Phage response to CRISPR-encoded resistance in
519 *Streptococcus thermophilus*. Journal of Bacteriology **190**:1390–1400.

520 26. **Semenova E**, **Jore MM**, **Datsenko KA**, **Semenova A**, **Westra ER**, **Wanner B**, **van der**
521 **Oost J**, **Brouns SJJ**, **Severinov KV**. 2011. Interference by clustered regularly interspaced
522 short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad
523 Sci USA **108**:10098–10103.

524 27. **Swarts DC**, **Mosterd C**, **van Passel MWJ**, **Brouns SJJ**. 2012. CRISPR interference
525 directs strand specific spacer acquisition. PLoS ONE **7**:e35888–7.

526 28. **Savitskaya EE**, **Semenova E**, **Dedkov V**, **Metlitskaya A**, **Severinov KV**. 2013. High-
527 throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. RNA Biology
528 **10**:716–725.

529 29. **Fineran PC**, **Gerritzen MJH**, **Suarez-Diez M**, **Kunne T**, **Boekhorst J**, **van Hijum**
530 **SAFT**, **Staals RHJ**, **Brouns SJJ**. 2014. Degenerate target sites mediate rapid primed
531 CRISPR adaptation. Proc Natl Acad Sci USA **111**:E1629–E1638.

532 30. **Richter C**, **Dy RL**, **McKenzie RE**, **Watson BNJ**, **Taylor C**, **Chang JT**, **McNeil MB**,
533 **Staals RHJ**, **Fineran PC**. 2014. Priming in the Type I-F CRISPR-Cas system triggers
534 strand-independent spacer acquisition, bi-directionally from the primed protospacer.
535 Nucleic Acids Research **42**:8516–8526.

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

539 32. **Li M**, **Wang R**, **Xiang H**. 2014. *Haloarcula hispanica* CRISPR authenticates PAM of a
540 target sequence to prime discriminative adaptation. Nucleic Acids Research **42**:7226–
541 7235.

542 33. **Li M**, **Gong L**, **Zhao D**, **Zhou J**, **Xiang H**. 2017. The spacer size of I-B CRISPR is
543 modulated by the terminal sequence of the protospacer. Nucleic Acids Research **45**:4642–
544 4654.

545 34. **Rao C**, **Chin D**, **Ensminger AW**. 2017. Priming in a permissive type I-C CRISPR–Cas
546 system reveals distinct dynamics of spacer acquisition and loss. RNA **23**:1525–1538.

547 35. **Xue C**, **Seetharam AS**, **Musharova O**, **Severinov KV**, **Brouns SJJ**, **Severin AJ**,
548 **Sashital DG**. 2015. CRISPR interference and priming varies with individual spacer
549 sequences. Nucleic Acids Research **43**:10831–10847.

550 36. **Vorontsova D**, **Datsenko KA**, **Medvedeva S**, **Bondy-Denomy J**, **Savitskaya EE**,
551 **Pougach K**, **Logacheva MD**, **Wiedenheft B**, **Davidson AR**, **Severinov KV**, **Semenova**
552 **E**. 2015. Foreign DNA acquisition by the I-F CRISPR–Cas system requires all
553 components of the interference machinery. Nucleic Acids Research **43**:10848–10860.

554 37. **Staals RHJ**, **Jackson SA**, **Biswas A**, **Brouns SJJ**, **Brown CM**, **Fineran PC**. 2016.
555 Interference-driven spacer acquisition is dominant over naive and primed adaptation in a
556 native CRISPR-Cas system. Nature Communications **7**:1–13.

557 38. **D'Auria G**, **Jimenez-Hernandez N**, **Peris-Bondia F**, **Moya A**, **Latorre A**. 2010.
558 Legionella pneumophila pangenome reveals strain-specific virulence factors. BMC
559 Genomics **11**:181.

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

563 40. **Horvath P**, **Romero DA**, **Coute-Monvoisin AC**, **Richards M**, **Deveau H**, **Moineau S**,
564 **Boyaval P**, **Fremaux C**, **Barrangou R**. 2008. Diversity, activity, and evolution of
565 CRISPR loci in *Streptococcus thermophilus*. Journal of Bacteriology **190**:1401–1412.

566 41. **Tyson GW**, **Banfield JF**. 2008. Rapidly evolving CRISPRs implicated in acquired
567 resistance of microorganisms to viruses. Environ Microbiol **10**:200–207.

568 42. **Gudbergsdottir S**, **Deng L**, **Chen Z**, **Jensen JVK**, **Jensen LR**, **She Q**, **Garrett RA**.
569 2010. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when
570 challenged with vector-borne viral and plasmid genes and protospacers. Molecular
571 Microbiology **79**:35–49.

572 43. **Lopez-Sanchez M-J**, **Sauvage E**, **Da Cunha V**, **Clermont D**, **Ratsima Hariniaina E**,
573 **Gonzalez-Zorn B**, **Poyart C**, **Rosinski-Chupin I**, **Glaser P**. 2012. The highly dynamic
574 CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome.
575 Molecular Microbiology **85**:1057–1071.

576 44. **Jiang W**, **Maniv I**, **Arain F**, **Wang Y**, **Levin BR**, **Marraffini LA**. 2013. Dealing with
577 the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. PLoS
578 Genet **9**:e1003844–13.

579 45. **Krzywinski M**, **Schein J**, **Birol I**, **Connors J**, **Gascoyne R**, **Horsman D**, **Jones SJ**,
580 **Marra MA**. 2009. Circos: an information aesthetic for comparative genomics. Genome
581 Research **19**:1639–1645.

582 46. **Mojica FJM**, **Díez-Villaseñor C**, **García-Martínez J**, **Almendros C**. 2009. Short motif
583 sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology
584 **155**:733–740.

585 47. **Cady KC**, **Bondy-Denomy J**, **Heussler GE**, **Davidson AR**. 2012. The CRISPR/Cas
586 adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally
587 occurring and engineered phages. Journal of Bacteriology **194**:5728–5738.

588 48. **Toms A**, **Barrangou R**. 2017. On the global CRISPR array behavior in class I systems.
589 Biology Direct **12**:1–12.

590 49. **Solomon JM**, **Rupper A**, **Cardelli JA**, **Isberg RR**. 2000. Intracellular growth of
591 *Legionella pneumophila* in *Dictyostelium discoideum*, a system for genetic analysis of
592 host-pathogen interactions. Infect Immun **68**:2939–2947.

593 50. **Magoc T**, **Salzberg SL**. 2011. FLASH: fast length adjustment of short reads to improve
594 genome assemblies. Bioinformatics **27**:2957–2963.

595 51. **Bolger AM**, **Lohse M**, **Usadel B**. 2014. Trimmomatic: a flexible trimmer for Illumina
596 sequence data. Bioinformatics **30**:2114–2120.

597 52. **Crooks GE**. 2004. WebLogo: a sequence logo generator. Genome Research **14**:1188–
598 1190.

599 53. **Grissa I**, **Vergnaud G**, **Pourcel C**. 2007. CRISPRFinder: a web tool to identify clustered
600 regularly interspaced short palindromic repeats. Nucleic Acids Research **35**:W52–W57.

601

**TABLE 1**

The number of acquired spacers for the plasmid Lens and chromosomal Lens CRISPR-Cas systems and their respective targets.

| System | Number of acquired spacers[a] | % map to priming plasmid[a] | % map to chromosomal Lens genome[a] | % map to plasmid Lens genome[a] | % unknown[a] |
|---|---|---|---|---|---|
| Plasmid Lens | 78,753 | 98.38 | 0.05 | 0.01 | 1.57 |
| Chromosomal Lens | 86,789 | 98.66 | 0.01 | 0.00 | 1.33 |

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

**TABLE 2**

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

| Strain | System location | # of spacers in the array | Consensus Repeat (5'-3')[a] | Mutated last repeat (5'-3')[a] | Reference[b] |
|---|---|---|---|---|---|
| Lens | Chromosome | 64 | GTTCACTGCCGCACAGGCAGCTTAGAAA | GTTCACTGCCGCACAGGCAGCTTAGAA**G** | Rao et al. 2016 |
| FFI104 | Chromosome | 55 | GTTCACTGCCGCACAGGCAGCTTAGAAA | GTTCACTGCCGCACAGGCAGCTTAGAA**G** | GenBank (accession CP016872.1) |
| FFI105 | Chromosome | 55 | GTTCACTGCCGCACAGGCAGCTTAGAAA | GTTCACTGCCGCACAGGCAGCTTAGAA**G** | GenBank (accession CP016873.1) |
| FFI337 | Chromosome | 55 | GTTCACTGCCGCACAGGCAGCTTAGAAA | GTTCACTGCCGCACAGGCAGCTTAGAA**G** | GenBank (accession CP016876.1) |
| Alcoy | Chromosome | 56 | GTTCACTGCCGCACAGGCAGCTTAGAA**G** | -- | Rao et al. 2016 |
| Lens | Plasmid | 53 | GTTCACTGCCG**T**ACAGGCAGCTTAGAAA | GTTCACTGCCG**T**ACAGGCAGCTTAGAA**G** | Rao et al. 2016 |
| Missisauga-2006 | Plasmid | 74 | GTTCACTGCCG**T**ACAGGCAGCTTAGAAA | GTTCACTGCCG**T**ACAGGCAGCTTAGAA**G** | Rao et al. 2016 |
| C8_S | Plasmid | 24 | GTTCACTGCCG**T**ACAGGCAGCTTAGAAA | GTTCACTGCCG**T**ACAGGCAGCTTAGAA**G** | GenBank (accession CP015940.1) |

[a] Mutations in the last repeat relative to the consensus repeat, and in the consensus repeat between the chromosome based systems and the plasmid systems, are denoted in red

[b] The GenBank files were accessed September 2017.

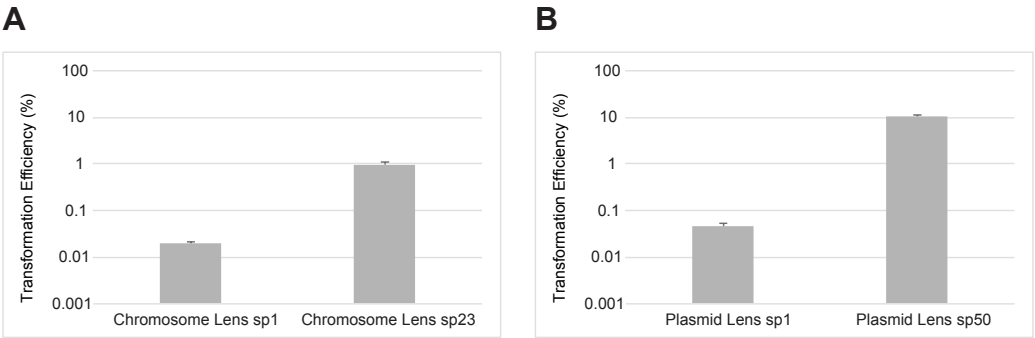


**Figure 1 | The chromosomal and plasmid type I-F CRISPR-Cas systems in *L. pneumophila* str. Lens are active against plasmids containing protospacers.** *L. pneumophila* str. Lens was transformed with plasmids containing targeted protospacer sequences matched to the first spacer (sp1) or a mid-array spacer (sp23 or sp50, respectively) of the chromosomal Lens CRISPR-Cas system **(A)** or the plasmid Lens CRISPR-Cas system **(B)**. After plating on selective media and incubating for three days, transformation efficiencies were calculated as a percentage of the transformation efficiency of a control plasmid with a scrambled targeted sequence. The average for three biological replicates is shown where the error bars represent the standard error of the mean..

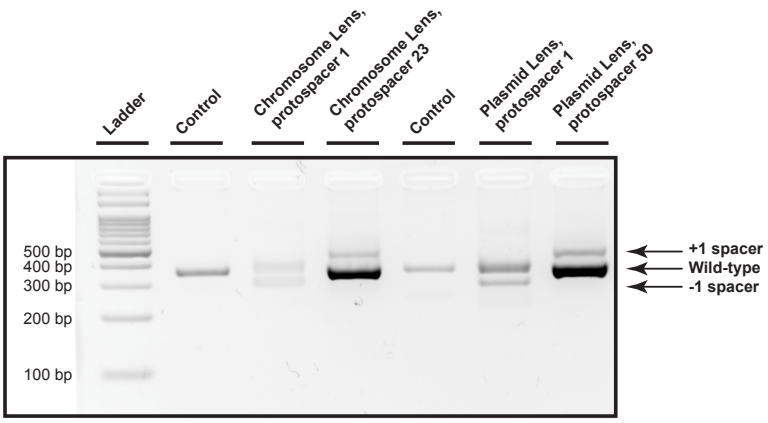


**Figure 2 | Targeted spacer acquisition and spacer loss in the chromosomal Lens and plasmid Lens CRISPR-Cas systems.** Spacer acquisition and loss were analyzed using a PCR based screen where the leader-end of the CRISPR array for both the control samples and the transformed samples was amplified with 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.

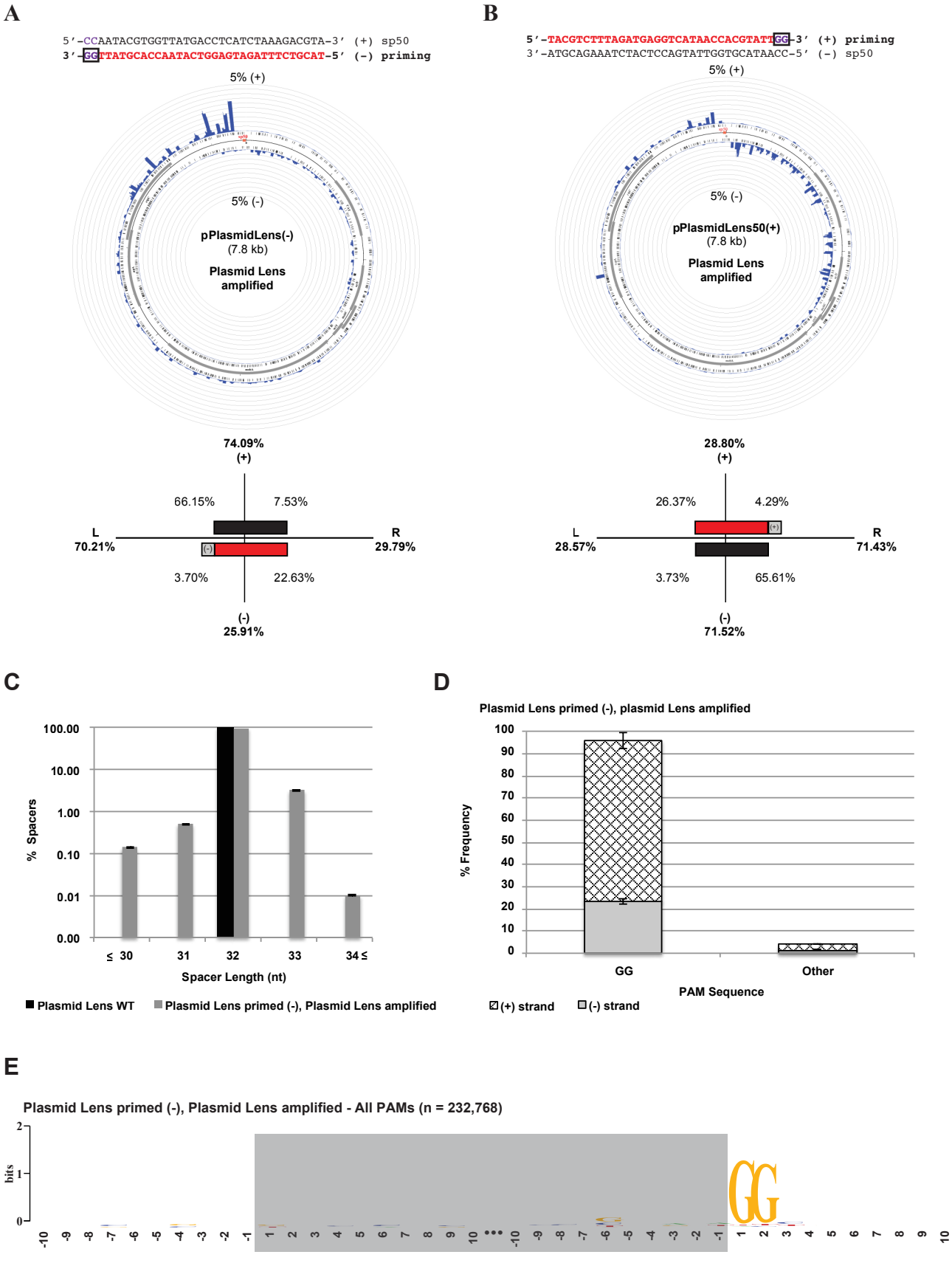


**Figure 3 | Characterization of targeted acquisition in the plasmid Lens CRISPR-Cas system.** (Legend on following page).

**Figure 3 | Characterization of targeted acquisition in the plasmid Lens CRISPR-Cas system.** 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. Acquired protospacers were obtained from the raw reads using an in-house bioinformatics pipeline and visualized with Circos. Unless otherwise noted, all data is the average of three biological replicates. **A)** The distribution of acquired protospacers mapped to the priming plasmid on the Circos plot reveals a strand bias in targeted acquisition within the plasmid Lens 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. The priming protospacer sequence used on the targeted plasmid is denoted in red and its PAM sequence is denoted in purple above the Circos plot. Quantification of the Circos plot is shown in a quad plot, where (+) is the plus strand, (-) is the minus strand, (L) is the left side of the plasmid (5' half) and (R) is the right side of the plasmid (3' half), relative to the priming site. The red bar indicates the priming protospacer and the grey box indicates its PAM location and targeted strand. **B)** Reversing the direction of the targeted sequence on the priming plasmid creates a mirrored distribution bias in acquired protospacers. Labelling as in (A). **C)** The distribution of spacer lengths acquired from the targeted plasmid (grey) compared to the wild-type CRISPR-Cas array (black, n = 53). **D)** Quantification of the PAMs for the new protospacers are shown in a stacked bar plot; the (+) strand PAMs are denoted by the hatched bars and the (-) strand PAMs are denoted by the grey bars. **E)** Sequence logo analysis of PAMs from new protospacers demonstrate a preference for the canonical GG PAM, with the protospacer denoted in the grey box. The sequence logo represents the PAMs of all acquired protospacers for the three biological replicates.

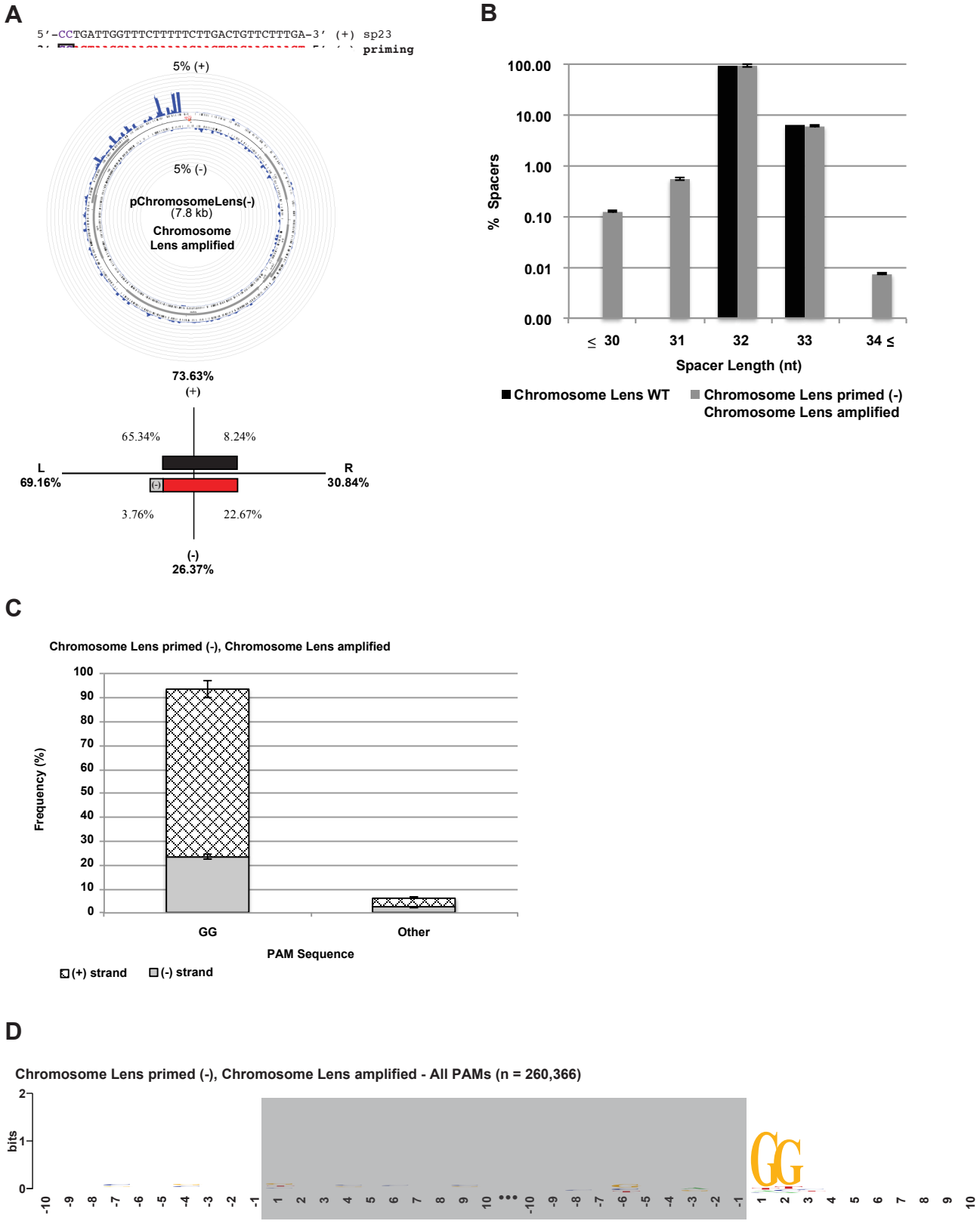


**Figure 4 | Characterization of targeted acquisition activity in the chromosomal Lens CRISPR-Cas system.** (Legend on following page).

**Figure 4 | Characterization of targeted acquisition activity in the chromosomal Lens CRISPR-Cas system.** Labelling and experimental set-up are as described for figure 3. **A)** The distribution of acquired protospacers mapped to the priming plasmid on the Circos plot reveals a strand bias in targeted acquisition within the chromosomal Lens CRISPR-Cas system. The priming protospacer sequence and its PAM are shown above Circos plot, while the quantification of the acquired protospacers is shown in the lower panel as a quad plot. **B)** The distribution of spacer length acquired from the targeted plasmid compared to the wild-type CRISPR-Cas array (n = 64). **C)** Quantification of the PAMs for the new protospacers in a stacked bar plot. **D)** Sequence logo analysis of PAMs from new protospacers.

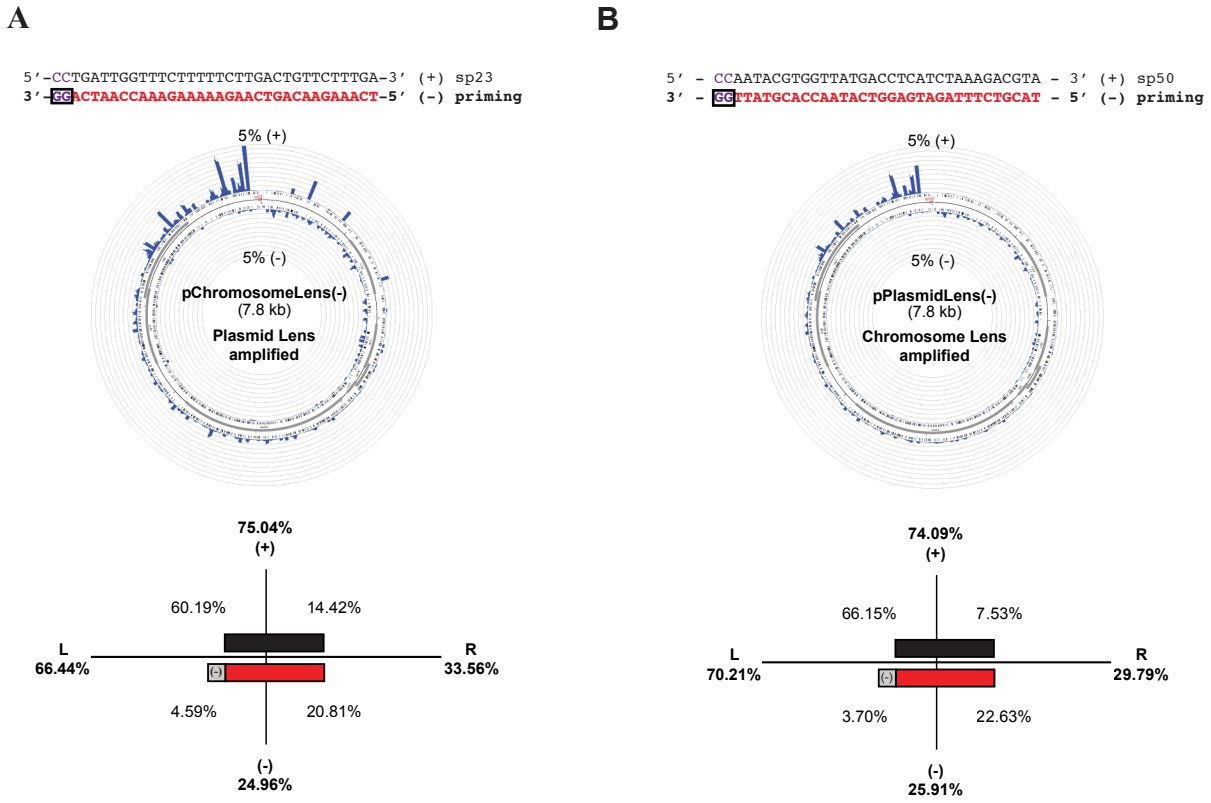


**Figure 5 | Cross-priming occurs between the chromosomal Lens and plasmid Lens CRISPR-Cas systems.** Labelling and experimental set-up are as described for figure 3. The distribution of acquired protospacers mapped to the priming plasmid reveals cross-priming between the chromosomal Lens and plasmid Lens CRISPR-Cas systems. **(A)** shows chromosomal Lens primed, plasmid Lens amplified while **(B)** shows plasmid Lens primed, chromosomal Lens amplified**.**