Evolution of the CRISPR-Cas9 defence system in *Mycoplasma gallisepticum* following colonization of a novel bird host

Authors: Thomas IPOUTCHA[1]*, Iason TSARMPOPOULOS[1], Géraldine GOURGUES[1], Vincent BABY[1], Paul DUBOS[1], Geoffrey E. HILL[2], Yonathan ARFI[1], Carole LARTIGUE[1], Patricia THEBAULT[4], Camille BONNEAUD[3], Pascal SIRAND-PUGNET[1]*

[1] Univ. Bordeaux, INRAE, UMR BFP, F-33882, Villenave d'Ornon, France
[2] Department of Biological Sciences, Auburn University, Auburn, Alabama, 36849-5414, United States of America
[3] Centre for Ecology and Conservation, Tremough Campus, University of Exeter, Penryn TR10 9FE, United Kingdom
[4] Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France
*Corresponding authors: Thomas Ipoutcha, ipoutcha.thomas@orange.fr; Pascal Sirand-Pugnet, pascal.sirand-pugnet@inrae.fr

ORCID identifier: Thomas IPOUTCHA (0000-0001-6266-9381), Vincent BABY (0000-0002-4938-1487), Geoffrey E. HILL (0000-0001-8864-6495), Yonathan ARFI (0000-0002-6064-7899), Carole LARTIGUE (0000-0001-5550-7579), Patricia THEBAULT (0000-0003-2276-4573), Camille BONNEAUD (0000-0003-2248-3288) and Pascal SIRAND-PUGNET (0000-0003-2613-0762).

Data summary: The authors confirm all supporting data and protocols have been provided within the article or through supplementary data files available in the online version of this article. GenBank accession numbers of all publicly available *M. gallisepticum* genomes are listed in Table S3. Sequences of the CRISPR locus of other strains are also provided in Table S3.

Impact statement: Mycoplasma are minimal bacteria involved in many diseases affecting humans and a wide diversity of animals. In this paper, we report the evolution of the Type II CRISPR-Cas system of the bird pathogen, *Mycoplasma gallisepticum*, following an host jump from its original poultry host into its novel house finch host in the early 90's. Instances in which bacterial pathogens have been documented to jump into and subsequently adapt to a new host are rare, and the well documented case of *M. gallisepticum* is a unique model to evaluate the effect of any dramatic host environmental change on bacterial CRISPR-Cas defence systems. First, we performed in silico analyses on an extended set of 98 *M. gallisepticum* genomes to better understand the evolution of

34 the CRISPR-Cas9 system in the novel finch host. We documented several evolutionary events leading
35 to the drastic divergence of spacer sets present in poultry and house finch arrays, as well as the
36 progressive inactivation of the CRISPR-Cas system after 12 years in the novel finch host. Second,
37 using in vitro and in vivo assays, we demonstrated that the evolution of the MgCas9 PI domain,
38 involved in the protospacer adjacent motif (PAM) recognition has led to a major change in the
39 defence system, with a modification of the recognized PAM in the novel host. Such radical change in
40 the CRISPR-Cas defence system of *M. gallisepticum* may have implications for the its rapid adaptation
41 to its novel host. Together, our results highlight the need to consider not only the host-driven
42 selection pressures a bacterium experiences, but also the complex interplay between phages and
43 defence systems for better understanding the key factors driving the emergence of a pathogenic
44 bacterium in a novel host.

45 Abstract

46 CRISPR-Cas systems are bacterial defences that target bacteriophages and mobile genetic elements.
47 How these defences evolve in novel host environments remains, however, unknown. We studied the
48 evolution of the CRISPR-Cas system in *Mycoplasma gallisepticum*, a bacterial pathogen of poultry
49 that jumped into a passerine host ~30 years ago. Over the decade following the host shift, all isolates
50 displayed a functional CRISPR-Cas system were found not only to harbour completely new sets of
51 spacers, but the DNA protospacer adjacent motif (PAM) recognised by the main effector MgCas9 was
52 also different. These changes in CRISPR-Cas diversity and specificity are consistent with a change in
53 the community of phages and mobile elements infecting *M. gallisepticum* as it colonised the novel
54 host. In the years following the host shift, we also detected a gradual rise in isolates displaying non-
55 functional MgCas9. After 12 years, all circulating isolates harboured inactive forms only. This loss of
56 CRISPR-Cas function comes at a time when the passerine host is known to have evolved widespread
57 resistance, which in turn drove the evolution of increasing *M. gallisepticum* virulence through
58 antagonistic coevolution. Such striking concordance in the rise of inactivated forms of CRISPR-Cas

and the evolution of host resistance suggests that inactivation of the CRISPR-Cas system was necessary for enabling adaptive bacterial responses to host-driven selection. We highlight the need to consider both host and pathogen selection pressures on bacteria for understanding the evolution of CRISPR-Cas systems and the key factors driving the emergence of a pathogenic bacterium in a novel host.

Keywords: CRISPR-Cas, mycoplasma, bird pathogen, host shift, evolution

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems are defences found in most prokaryotes that effectively protect cells against the constant threat of phages and other sources of invading DNAs. The flexibility and effectiveness of this defence system relies on the continuous acquisition of new spacers to the CRISPR array (Barrangou et al., 2007), which entails substantial maintenance costs. As a consequence, inactivation and disappearance of CRISPR systems from prokaryote genomes is predicted to occur when benefits of this defence system are relaxed (Koonin, 2019; Makarova et al., 2020). Ultimately, gain or loss events of active CRISPR systems will be shaped both by cost/benefit ratios (Dimitriu et al., 2020) and by the complex ecological pressures experienced by prokaryotes. For pathogenic bacteria, the strongest ecological pressures arise from interactions with hosts. Changes in host-driven selection pressures may occur, for instance, as hosts evolve in response to pathogens, but should be particularly dramatic when a prokaryotic parasite colonizes a novel host. While invasion of a novel host is known to broadly impact bacterial genomes, how such host shifts shape bacterial defence systems against phages and other parasitic DNAs remains largely unstudied.

82 Instances in which bacterial pathogens have been documented to jump into and subsequently adapt
83 to a new host species are rare (Bonneaud et al. 2019), limiting opportunities to test the effect of host
84 environmental change on bacterial CRISPR systems. One model in which the adaptation of a
85 bacterium to a novel host can be studied is the *Mycoplasma gallisepticum* / house finch
86 (*Haemorhous mexicanus*) system. *M. gallisepticum* is a bacterial pathogen that is a common in
87 domestic chickens (*Gallus gallus*), a gallinaceous bird. It jumped into a very distantly related and
88 widely distributed North American passerine, the house finch. This novel host shift was documented
89 in 1994 by wildlife disease experts (Ley et al., 1996) and a mid-1990s host-shift date has been
90 confirmed by genome analyses of *M. gallisepticum* isolates collected from chickens and finches. *M.*
91 *gallisepticum* lineages subsequently collected across the range of the house finches indicate that all
92 *M. gallisepticum* circulating in house finch populations are derived from this single host-shift event
93 (Delaney et al., 2012; Tulman et al., 2012).

94 *M. gallisepticum* is an avian pathogen that belongs to the class *Mollicutes* (May et al., 2014), a group
95 of minimal bacteria characterized by a rapid evolution through drastic genome reduction from a
96 Gram-positive ancestor. Such genetic losses are partially counter-balanced by horizontal gene
97 transfers (HGT) and exposure to Mobile Genetic Elements (MGEs) acting as potential sources of
98 genetic diversity and genome reorganization (Citti et al., 2018; Sirand-Pugnet et al., 2007). Despite an
99 evolutionary tendency toward genome minimization, most mycoplasmas have retained various
100 defence systems against phages and invading DNAs. Indeed, mycoplasmas are the second group of
101 bacteria with the greatest number of Restriction-Modification systems (Oliveira et al., 2014), and are
102 also the free-living bacteria with the smallest genome in which CRISPR-Cas systems have been
103 described (Ipoutcha et al., 2019).

104 Changes in the CRISPR locus of *M. gallisepticum* have previously been reported following the host
105 shift into house finches (Delaney et al., 2012; Dhondt et al., 1998). Indeed, a comparison of *M.*
106 *gallisepticum* isolates obtained from ancestral poultry and novel house finch hosts showed a

107 complete change in the spacer set, with no elements being shared between poultry and house finch
108 strains (Delaney et al., 2012). House finch isolates also displayed a loss of spacers over time, as well
109 as a degradation of their CRISPR locus. While these patterns suggest both a slowing down in the
110 recruitment of new spacers and even an evolution towards this defence system becoming non-
111 functional, further work is required to test whether these evolutionary changes have persisted over
112 time in the new passerine host. Making sense of observed changes in the house finch CRISPR also
113 requires functionally characterising other changes that have occurred in the CRISPR-Cas system, and
114 particularly in *M. gallisepticum* Cas9 (MgCas9). Previous research has shown that Cas9 endonuclease
115 is the main effector of type-II CRISPR systems, being involved in both the acquisition of new spacers
116 and the double-strand DNA cleavage of target DNAs (Heler et al., 2015; Wei et al., 2015). In both
117 processes, Cas9 selects functional spacers by recognizing their protospacer adjacent motif (PAM)
118 sequence immediately downstream. The change of spacer set that has accompanied *M. gallisepticum*
119 host jump therefore calls into question the evolution of MgCas9 PAM specificity as a possible driver.

120 To better understand the changes in the *M. gallisepticum* CRISPR system that took place during the
121 colonization of the novel avian host, we conducted a functional characterisation using a two-pronged
122 approach. First, we performed *in silico* analyses on an extended set of *M. gallisepticum* isolates (79
123 collected from 1994 to 2015) to better understand evolutionary changes in CRISPR-Cas9. We
124 identified several events leading to the remarkable divergence of spacer sets in poultry versus house
125 finch arrays. We also detected further evidence for an ongoing inactivation of the CRISPR-Cas system
126 in the house finch host over time. Second, we functionally characterized the CRISPR-Cas system of
127 two *M. gallisepticum* strains infecting poultry (S6 strain) and house finch (CA06 strain) hosts. Using *in*
128 *vitro* and *in vivo* approaches, we determined the PAM specificities of the MgCas9 and showed a
129 correlation between a modification of the PAM signature, the evolution of the spacer repertoires,
130 and the evolution of the PI domain of MgCas9. Such radical change in the CRISPR-Cas defence system
131 of *M. gallisepticum* may have implications for the rapid adaptation of *M. gallisepticum* to its novel
132 host (Bonneaud et al., 2018; Tardy et al., 2019).

133

134 ## Results

135 ### Comparison of the CRISPR-Cas locus across *M. gallisepticum* isolates

136 CRISPR-Cas sequences were extracted from previously published whole genome sequences of 19
137 poultry and 8 house finch strains (Leigh et al., 2019a, 2019b; Papazisi et al., 2003; Song et al., 2021;
138 Szczepanek et al., 2010; Tulman et al., 2012), as well as from a further 79 newly-sequenced genomes
139 of strains collected in house finches between 1994-2015 (Table S3). A single CRISPR locus was found
140 in all *M. gallisepticum* isolates obtained from house finches (Figure 1A), with the exception of 6
141 isolates (1 collected in 2006, 3 in 2007, 1 in 2008 and 1 in 2015). In all 6 of these isolates, the CRISPR-
142 Cas system was found to have been inactivated due to the loss of part of the locus harbouring the
143 tracrRNA and the *cas* genes (i. e. *cas9, cas1, cas2* and *csn2*), leaving only the CRISPR array. An
144 extended analysis of the genome region on both sides of the CRISPR locus further suggested this
145 system being part of a larger defence island, also including a restriction-modification system, a toxin-
146 antitoxin system and a putative S8 serine peptidase (SI-2). For at least the 2 isolates with a
147 completely assembled genome (strains collected in 2006 and 2008), the deletion was extended
148 upstream of the CRISPR locus, including the putative S8 serine peptidase and other adjacent genes.
149 The inactivation of the CRISPR-Cas system could also be predicted for a further subset of isolates
150 based on several mutations found in their MgCas9 gene. Indeed, we were able to predict three
151 different truncated versions of MgCas9 (thereafter named V1, V2 and V3) (Figure 1B, table S3). The
152 MgCas9 truncated form V1 form was observed only once in a 1998 house finch isolate; it is truncated
153 at position 1498 in the REC domain as a result of a single base change that generates a TAG STOP
154 codon instead of CAG Gln codon. The V2 form was detected in 7 isolates sampled in house finches in
155 2001, 2009 and 2015; it was predicted based on a frameshift at position 3445 of the CDS, leading to a
156 protein truncated of the last 120 amino acids within the PAM Interacting domain (PI). Finally, the V3
157 form was identified in 29 isolates collected in house finches in 2003, as well as between 2011 to

158 2015; it is truncated at position 1018 of the CDS within the REC lobe (which exerts a key role for DNA
159 cleavage) as a result of a single nucleotide mutation changing a GAA Glu codon into a TAA STOP
160 codon.

161 The pattern of occurrence of active and inactive MgCas9 in house finch isolates suggests an overlap
162 of both forms over a relatively short period and followed by the complete inactivation of MgCas9.
163 Indeed, all the house finch isolates for which we could predict a complete MgCas9 were collected
164 between 1995 and 2006, whereas all the isolates collected after 2007 displayed either an inactivated
165 MgCas9 or a partial deletion of the CRISPR locus. Isolates with active and inactive MgCas9 were
166 therefore only found to co-occuring in house finches between 2001 and 2006. This gradual pattern of
167 inactivation contrasts with what we see in poultry, as all strains collected in chicken and turkeys >60
168 years (~1958-2019) are predicted to harbour an active CRISPR system and did not display any
169 inactivating mutations in MgCas9.

170 **Analysis of spacers and identification of *M. gallisepticum* CRISPR targets**

171 We compared the set of spacers inferred from the *M. gallisepticum* CRISPR locus across all studied
172 *M. gallisepticum* isolates (Figure S1, Table S4), and compiled a dataset of 490 non-redundant spacers
173 (Table S4C). As a result, we identified a total of 438 unique spacers from 19 poultry strains and 55
174 unique spacers from 85 house finch isolates (Table S4C). Across the *M. gallisepticum* poultry strains,
175 we observed a wide diversity of spacers, with no spacer found conserved across all strains (Figure
176 S1). Indeed, 66% of spacers from the 19 poultry strains were present in only 1-2 (0-10%) genomes,
177 while 2% were found in 9-10 (40-50%) genomes. In contrast, we found a lower diversity of spacers
178 across the CRISPR arrays of house finch isolates. A quarter (24%) of spacers found in house finches
179 were shared across more than 80% of isolates, while over half (58%) were present in at least 40% of
180 them, and 21 out of 55 (38%) spacers were present in 10% or less of house finch isolates.

181 We found only 3 spacers that were shared across poultry and house finch isolates (Table S4C). Spacer
182 SP_516* was identified in the poultry vaccinal strain ts-11, in 7 derivates obtained by chemical

183 mutagenesis (K6372, K6356, K6212B, K6216D, K6208B, K5322C, K2966) and in all but 10 of the house
184 finch isolates. Spacer SP_119* was found in 4 ts-11 derivate strains (K6369, K6208B, K5322C and
185 K6222B) and in 3 house finch isolates for which the CRISPR-Cas9 system was predicted to be inactive
186 (169125442_2001, 13_09_02_NC and 2015_J). Finally, spacer SP_325* was shared between the
187 poultry strains R_{Low} and R_{High} and by 5 house finch isolates (NC95, NY01, WI01, NC06 and NC08).
188 When considering the pattern of spacer acquisition over time in house finches, we found that most
189 new spacers (i.e. 46 out of 55; >80%) had been integrated before 2000 and were therefore present in
190 the early years of *M. gallisepticum* emergence in house finches. Surprisingly, despite the inactivation
191 of MgCas9 from 2007, we also found evidence for the acquisition of six new spacers between 2009
192 and 2015 (Table S4C).

193 We investigated the potential origin of *M. gallisepticum* spacers using Blast queries against PHASTER,
194 GenBank and IMG databases (Tables S4D-S4G), as described in (Vink et al., 2021). Out of 490 spacers,
195 431 (88%) showed no hits in the explored databases. The PHASTER database returned hits on known
196 bacterial phages for 34 spacers. GenBank returned hits on bacterial genomes (11 spacers) and
197 phages (4 spacers). IMG returned hits on bacterial genomes for 8 spacers. In addition, two self-
198 matching spacers were found in 2 strains of *M. gallisepticum* (ts-11 and NTCTC1015). Overall and
199 across the 57 spacers with identified hits outside *M. gallisepticum* genomes, we noted 6 spacers that
200 were present in house finch isolates, and 51 in those from poultry. Out of the 19 spacers with hits on
201 bacterial genomes, 10 returned non-CRISPR genomic regions of mycoplasma. Among them, 6 spacers
202 from poultry strains (S6, KUVMG001, 6/85, ts-11 and ts-11 derivates) matched to the genome of
203 *Mycoplasma imitans,* a duck pathogen phylogenetically related to *M. gallisepticum* (Table S4F); these
204 potential protospacers are localized in a *M. imitans* chromosomal region of 14 kb that is not found in
205 the *M. gallisepticum* genomes used in this study, contrary to other surrounding loci. In this region,
206 we identified several genes considered as essential for the mobilization and transfer of Mycoplasma
207 Integrative and Conjugative Elements (MICE) (Figure S2) (Calcutt et al., 2002; Dordet Frisoni et al.,
208 2013; Meygret et al., 2019). This suggested those 6 spacers might be directed against a potential

209 mobile element of 14 kb from bird mycoplasma, although no trace of this element has been detected
210 in the *M. gallisepticum* genomes studied here. Taken together, these results indicate that *M.*
211 *gallisepticum* CRISPR-Cas9 systems are mostly directed against unknown targets and that predicted
212 targets include most predominantly phages, as well as a newly identified ICE-like element of a bird
213 mycoplasma.

214 **Diversity and evolution of Cas proteins**

215 We undertook further *in silico* analyses to document the intraspecific diversity of *M. gallisepticum*
216 Cas proteins across isolates displaying a complete functional locus. To do so, we aligned the protein
217 sequences of MgCas9 and constructed a phylogenetic tree (Figure 1C). The isolates were clearly
218 distributed into two main branches, with a clustering of the poultry MgCas9 sequences in one branch
219 and of the house finch ones in another. House finch MgCas9 sequences were found to be >99%
220 identical to each other and the copy present in the CA06 isolate was selected (as in (Gasiunas et al.,
221 2020)) to be compared with a poultry MgCas9 from the S6 strain. This strain was chosen because it
222 has not been mutagenized and it has been proven to be transformable, which was required for
223 further experimental approaches. While being 96% identical in sequence, the highest proportion of
224 amino acid variations was found within the PI domain, which is involved in the interaction with the
225 PAM sequence (Figure 2A). Indeed, sequence variations of 8.7% and 3.1% were calculated for the PI
226 domain and the other domains of the protein, respectively.

227 When considering other Cas proteins encoded within the CRISPR locus, we found that Csn2
228 sequences exhibited more variation (5.1%) than either Cas1 (0.3%) or Cas2 (0.9%) (Figure 2B). A
229 proteome-wide comparison of *M. gallisepticum* CA06 and S6 isolates showed that Cas9 and Csn2 are
230 part of the 12% most divergent proteins between the two isolates (Figure 2C, Table S4). Together,
231 our results suggest that the jump of *M. gallisepticum* from poultry into house finches has been
232 marked by a noticeable evolution of the MgCas9 PI domain and of Csn2, both of which are key
233 proteins in the ability of CRISPR-mediated immunity to acquire new spacers (Heler et al., 2015; Wei

234 et al., 2015). Since the PI domain of Cas9 is involved in PAM recognition, our results further suggest
235 that the jump might have been associated with a change in the PAM recognition specificity of
236 MgCas9, a pattern which is consistent with a change in pathogen pressure experienced by *M.*
237 *gallisepticum* within the novel finch host.

238 ***In vitro* PAM specificity differs between poultry (S6) and house finch (CA06) MgCas9**

239 We used an *in vitro* assay previously developed (Gasiunas et al., 2020; Karvelis et al., 2019) to
240 compare the PAM recognition preference of MgCas9 between the poultry strain S6 and the house
241 finch isolate CA06. *In vitro*-produced MgCas9/sgRNA ribonucleoparticles were used for the cleavage
242 of a library of plasmids containing a target sequence and a PAM sequence randomized on 7 positions
243 (Figure 3A, Figure S3, Table S6). After three independent *in vitro* cleavage of the plasmid library by
244 MgCas9 S6 or MgCas9 CA06, molecules containing a recognized PAM sequence were captured using
245 DNA ligation of adapters, enriched by PCR amplification, and submitted for deep sequencing.
246 Cleavage was observed three bases upstream from the PAM sequence at >99.95% and results
247 obtained for replicates of the two series of were nearly identical (Figures S4A and S4B). Out of the
248 16,306 PAM sequences contained in the library, 10,928 and 9,709 PAM sequences were detected at
249 least once after cleavage by MgCas9 S6 and MgCas9 CA06, respectively. The 1,000 most frequent
250 PAM sequences were selected for each MgCas9 including 74 PAM sequences that were common to
251 both MgCas9 S6 and MgCas9 CA06 top1,000 PAM sequences. Normalized frequencies were
252 calculated, taking the bias of the original plasmid library into account (Table S6F); frequency matrices
253 and sequence logos were produced (Figures 3B and 3C, Table S6G).

254 A comparison of the PAM recognition preference of MgCas9 S6 and MgCas9 CA06 revealed that the
255 PAM sequences recognised by both correspond to a 5-letter motif, with a marked bias observed at
256 positions 3, 4 and 5 only. Indeed, two lines of evidence indicate that the 5-position PAM sequences
257 recognised by MgCas9 from poultry S6 and house finch CA06 strains differ *in vitro*. First, a sequence
258 logo comparison revealed that PAM preferences of MgCas9 proteins were clearly different between

259 the two isolates. The main difference was observed at position 3 with 99% of sequences displaying
260 an A in MgCas9 S6 and a G in MgCas9 CA06 (Figures 3B). Position 4 was preferentially A or G in both
261 cases, with a bias towards A more marked in MgCas9 S6. In position 5, a T was present in 73% of
262 sequences retrieved after cleavage by MgCas9 S6, compared to 30% by MgCas9 CA06, suggesting a
263 notable difference at this position between the two isolates. Second, we found that the top 10 PAM
264 sequences most frequently recognized differed between S6 and CA06 MgCas9 (Figure 3D).

265 To confirm the difference of PAM recognition specificity, we selected two PAM sequences,
266 specifically recognized by MgCas9 S6 (Motif 1, AT**AAA**AA) and MgCas9 CA06 (Motif 2, AA**GAG**AA),
267 respectively. Three plasmids were constructed containing the target sequence followed by one of the
268 two selected PAM sequences or by the 7 first nucleotides of Direct Repeat (DR) sequence
269 interspacing spacers in *M. gallisepticum* CRISPR arrays (Motif 3, GT**TTT**AG) (Figure S5A). After *in vitro*
270 cleavage of these plasmids by *in vitro*-produced MgCas9/sgRNA complexes, we observed that
271 MgCas9 S6 efficiently cleaved Motif 1 but not Motif 2, whereas MgCas9 CA06 preferentially cleaved
272 the Motif 2 (Figure S5B). By contrast, the control Motif 3 remained uncleaved after incubation with
273 one MgCas9 or the other. This confirmed the difference of PAM specificity of MgCas9 originating
274 from poultry and house finch strains, *in vitro*. Together, our results show that poultry and house finch
275 MgCas9 recognise different 5-letter motifs in the PAM sequences and display different cleavage
276 specificities of PAM sequences as a result.

277 ***In vivo* PAM specificity of MgCas9 is driven by PI domain evolution**

278 We characterised the PAM specificity of poultry (S6) and house finch (CA06) MgCas9 using *in vitro*
279 and *in vivo* approaches. Based on the results previously obtained *in vitro*, we selected four PAM
280 motifs found to be differentially cut by poultry (S6) and house finch (CA06) MgCas9 (Figure 4A).
281 Motifs S6-1 and S6-2 were found to be recognized by MgCas9 S6 *in vitro*, but poorly by MgCas9 CA06,
282 while motif CA06-3 was cleaved by MgCas9 CA06 but not MgCas9 S6, an d CA06-4 was
283 efficiently recognized by MgCas9 CA06 but only weakly by MgCas9 S6. To investigate whether these

284 findings obtained *in vitro* are representative of processes occurring *in vivo*, we conducted
285 interference assays on both the wild-type *M. gallisepticum* S6 strain and on a genetically engineered
286 *M. gallisepticum* S6 (thereafter named *M. gallisepticum* S6 mod PI CA06) in which the MgCas9 PI
287 domain was seamlessly replaced by the homologous domain of the MgCas9 CA06 (Ipoutcha et al.,
288 2022). The *in vivo* cleavage assay was based on the transformation of *M. gallisepticum* with a
289 replicative plasmid containing the third spacer naturally found in the CRISPR array of strain S6,
290 followed by a selected PAM candidate (Figure 4B). A plasmid that contained the Direct Repeat of the
291 CRISPR array as a PAM sequence was used as a negative control.

292 Transformations with plasmids containing PAM sequences expected to be cut by MgCas9 S6 (S6-1
293 and S6-2) led to 100-1,000 fold less transformants compared to transformation with the control
294 plasmid (Figure 4C). This confirmed that the CRISPR-Cas9 system is functional as a defense system in
295 *M. gallisepticum* S6 and that PAM sequences recognized *in vitro* by MgCas9 S6 are also recognized *in*
296 *vivo*. Similarly, plasmids harbouring PAMs frequently cleaved by MgCas9 CA06 *in vitro* were also
297 cleaved by MgCas9 mod PI CA06 *in vivo*, demonstrating that the hybrid MgCas9 is fully active.
298 Interestingly, in this interference assay, motifs that were poorly recognized *in vitro* (i.e. CA06-4 for
299 MgCas9 S6; S6-1 and S6-2 for MgCas9 CA06), were efficiently cleaved *in vivo* (see SI-3 for further
300 comparisons of *in vitro* and *in vivo* findings). Finally, as we expected, we found that the PAM CA06-3
301 was not recognized by MgCas9 S6, but was efficiently cleaved by the hybrid MgCas9 that included a
302 CA06 PI domain. This result confirms that the specificity of MgCas9 is mediated by the PI domain and
303 suggests that mutations in the MgCas9 of house finch isolates has led to a change in PAM specificity.

304

305 Discussion

306 We studied the genomes of the bacterial pathogen, *M. gallisepticum*, that were collected in the
307 original poultry host and from house finches periodically over a 21-year span following a host shift.

308 Our focus was investigating the impact of a dramatic change in the host environment on the
309 evolution of the pathogen's CRISPR-Cas defence system. First, our results were consistent with a
310 previous observation, made from a small number of isolates, that there has been a gradual
311 inactivation of MgCas9 in the house finch host since colonization. Indeed, we found that, while all
312 poultry strains displayed a an active CRISPR-Cas system, both active and inactive forms of MgCas9
313 were present in house finch isolates collected in the initial 12 years following the host shift. After the
314 twelfth year following colonization, however, all *M. gallisepticum* collected from house finches
315 displayed either an inactivated MgCas9 or a partial loss of their CRISPR locus (*cas* genes). Second, of
316 the 490 unique spacer sequences detected across poultry and house finch isolates, only three were
317 shared across both avian hosts, with poultry isolates displaying a greater diversity of sequences than
318 house finch isolates. Third, we found that the protein sequences of MgCas9 of poultry and house
319 finch isolates clustered into two distinct phylogenetic branches, with very little variation detected
320 among house finch MgCas9 protein sequences. Within MgCas9, the PAM-interacting (PI) domain
321 (involved in the specific interaction between Cas9 and the protospacer adjacent motif (PAM) that
322 follows the target DNA sequence) and the Csn2 protein sequences were found to be most variable
323 across poultry and house finch isolates. Fourth, we found that MgCas9 recognises a 5-letter motif in
324 the PAM sequence. The motif recognised differed between poultry and house finch MgCas9 and,
325 accordingly, we measured variation in cleavage efficiencies *in vitro*. The poultry MgCas9 was more
326 effective at cleaving its target PAM sequences than the house finch MgCas9, while the house finch
327 MgCas9 was most effective at cleaving its target PAM. Finally, our results confirmed that differences
328 in the PAM specificity of MgCas9 are driven by differences in the PI domain between poultry and
329 house finch strains. Indeed, differences in specificities were replicated across a wild-type poultry
330 strain (S6) and a genetically engineered mutant of that strain in which the PI domain of MgCas9 was
331 replaced by the homologous domain of a house finch isolate (CA06), with the latter more effective at
332 cleaving PAMs from house finch isolates. Together, these results suggest that significant evolutionary
333 changes took place in *M. gallisepticum*'s CRISPR-Cas defence system following the jump into the

334 novel house finch host. We documented an initial pattern of acquisition and loss of CRISPR spacers,
335 and of changes in PAM specificity driven by the MgCas9 PI domain. This observation is consistent
336 with an evolutionary response to exposure to a new community of phages and MGEs. That such
337 initial functional changes were followed by the subsequent inactivation of this defence system,
338 however, highlights the complexity of the selection pressures that it faced.

339 Our study on the CRISPR-Cas9 system of *M. gallisepticum* isolates collected in both poultry and house
340 finch hosts provides support for previous work, while also broadening our understanding of the
341 complexity of this system across vertebrate host environments. First, mycoplasma CRISPR-Cas9
342 systems have previously been classified as Type II *in silico*, which include the typical set of *cas* genes,
343 *cas1*, *cas2*, *cas9* and *csn2* (Ipoutcha et al., 2019). Accordingly, we found that the CRISPR-Cas systems
344 of *M. gallisepticum* poultry S6 and house finch CA06 isolates are typical of the Type II systems.
345 Second, our results are also consistent with a previous study that undertook the first functional
346 investigation of the Cas9 endonuclease in a mycoplasma system as part of a larger initiative to
347 characterize 79 Cas9 orthologs representative of 10 major clades in a Cas9 evolutionary tree
348 (Gasiunas et al., 2020). In this study, the house finch isolate CA06 was selected as representative
349 mycoplasma Cas9 with a 5-positions motif NNGAD as its consensus PAM (Gasiunas et al., 2020). Our
350 results provide confirmation for MgCas9 CA06's consensus PAM. Interestingly, however, we
351 identified a different PAM motif (NNAAT) in the MgCas9 of the poultry strain S6, which was found to
352 exhibit an ultra-dominant A at position 3 (99%) and a marked trend (73%) towards a T at position 5.
353 In addition, we found that this difference in the PAM recognition preference of MgCas9 was
354 associated with differences in both the composition and diversity of the spacers between poultry and
355 house finch isolates, suggestive of a complete reset of the CRISPR array following the host shift.

356 An active adaptive immunity system like CRISPR-Cas is expected to be maintained only when the
357 benefits of such a system outweigh the costs of sustaining it (Dimitriu et al., 2020; Koonin, 2019;
358 Westra & Levin, 2020). The main benefit of an active CRISPR-Cas system is the protection that it

359 confers against phages and MGEs. For example, in the case of *M. gallisepticum*, the presence in
360 poultry strains of 6 spacers matching to the genome of the duck pathogen *M. imitans* suggests that
361 avian mycoplasma bacteria are under selection to actively protect themselves against invasion by
362 circulating MICEs. However, maintaining an active CRISPR-Cas undoubtedly incurs fitness costs. Such
363 costs can arise, for instance, when CRISPR-Cas generates metabolic costs (Hall et al., 2021; Vale et al.,
364 2015) and genetic conflicts (Hall et al., 2021; Vale et al., 2015) or when it is less effective at
365 preventing infections than other defence processes (Stern et al. 2010; Broniewski et al. 2020; Watson
366 et al. 2024).CRISPR-Cas can also create a risk of self-targeting spacers (Stern et al., 2010; Wimmer &
367 Beisel, 2020), and can prevent the acquisition of fitness-enhancing genes through horizontal gene
368 transfer (HGT) (Alduhaidhawi et al., 2022; Bikard et al., 2012; Dimitriu et al., 2020; Kogay et al., 2024;
369 Levin, 2010; Mackow et al., 2019; Palmer & Gilmore, 2010; Pursey et al., 2022; Wheatley & MacLean,
370 2021). The balance between the costs and benefits of maintaining an active CRISPR-Cas will therefore
371 depend on CRISPR-Cas' contribution to the bacterium's global defence process, on its non-defence
372 roles, and on the specific selective pressures exerted by different host environments.

373 That there were isolates in circulation with an active (albeit reconfigured) CRISPR-Cas system in the
374 12 years following the jump into house finches indicates that there were benefits to the maintenance
375 of this defence mechanism during that time period (Figure 5). The retention of an active CRISPR-Cas
376 system is particularly striking when one considers the function of the system was maintained even
377 with a large number of mutations (49 amino-acid changes when comparing MgCas9 S6 and MgCAs9
378 CA06) that accumulated in the house finch *cas9* gene following the jump. The reasons for such
379 benefits are likely to be found in the distinct PAM specificity of MgCas9 in house finch isolates,
380 alongside salient patterns of loss of 'poultry-type' spacers and acquisition of new spacers elements.
381 Indeed, spacers that are no longer in use are expected to be lost, while new ones that confer
382 resistance against novel pathogenic targets should be gained. *M. gallisepticum* isolates from house
383 finches display dramatic changes both in spacer set and MgCas9 PAM specificity, which suggest that
384 the bacterium was faced with a massively different community of phages and invasive MGEs in the

385 novel host. Indeed, such a finding is consistent with previous evidence of a rapid evolution of PAM
386 specificities reported to occur in response to an escape from detection by phages and MGEs
387 displaying mutations in their PAM motifs (Deveau et al., 2008; Edraki et al., 2019; Paez-Espino et al.,
388 2015). Whether *M. gallisepticum* experienced novel pathogen-driven selection in house finches, and
389 the role that this might have played in facilitating or hindering its emergence remain to be
390 determined. Regardless, our results give rise to the hypothesis that the successful emergence of a
391 bacterial pathogen in a new host species will not only depend on its ability to establish and spread in
392 the new host, but also on its ability to cope with a new landscape of phages and other invading
393 DNAs.

394 As *M. gallisepticum* adapted to its new host and novel sets of phages and invasive MGEs, the costs of
395 maintaining an active CRISPR-Cas system seem to have outweighed the benefits of such a system.
396 Isolates displaying an inactive CRISPR-Cas system were also found in circulation over the 12 years
397 following the jump into house finches and, by 2007, all the isolates collected either displayed an
398 inactivated MgCas9 or the complete loss of their *cas* genes locus, leaving only the CRISPR array
399 (Figure 5). Indeed, *cas9* inactivation evolved at least three times over that period from different point
400 mutations leading to stop codons. Across bacteria, the switching off of the CRISPR-Cas systems has
401 been proposed as a mechanism by which to respond to selection (Patterson et al., 2017). In
402 *Escherichia coli*, for example, CRISPR-Cas is thought to restrict adaptive change by limiting the
403 acquisition of potentially beneficial genes through horizontal transfer, giving rise to a trade-off
404 between defence and adaptive potential (García-Martínez et al., 2018). As a result, we might expect
405 the balance of costs and benefits of maintaining an active CRISPR-Cas system to have shifted for *M.*
406 *gallisepticum* in the face of a significant change in selection in the novel host.

407 One major selective event faced by *M. gallisepticum* after the jump into house finches was the
408 evolution of host resistance to infection (Bonneaud et al. 2011; Bonneaud et al. 2019; Dowling et al.
409 2020). The jump gave rise to an epidemic that spread quickly and is thought to have killed millions of

410 house finches (Dhondt et al., 1998; Nolan et al., 1998). The high mortality rates of *M. gallisepticum* in
411 house finches was a result of *M. gallisepticum* localizing in the mucosal surfaces of the conjunctiva
412 and upper respiratory tract, causing a severe conjunctivitis that can lead to death in the wild through
413 blindness-induced starvation or predation (Adelman et al., 2017; Roberts et al., 2001). The intensity
414 of this selective event led to the spread of resistance in house finch populations from standing
415 genetic variation and within only 12 years of epidemic outbreak (Bonneaud et al., 2011, 2018). As a
416 result, by 2007, the proportion of resistant individuals in exposed house finch populations had risen
417 to ~80% (Bonneaud et al., 2011). Such host resistance was, in turn, found to have significant selective
418 consequences for *M. gallisepticum*. Indeed, resistance favoured the evolution of increasing pathogen
419 virulence through antagonistic coevolution (Bonneaud et al., 2018), with isolates causing ever greater
420 host mortality and more severe signs of conjunctivitis over time (Tardy et al., 2019), as well as
421 transmitting faster to uninfected sentinels (Bonneaud et al., 2020). The concordance in the timing of
422 the complete inactivation of CRISPR-Cas (i.e., from 2007) and the need for evolutionary changes in
423 virulence in response to widespread house finch resistance is striking. Further work is now required
424 to test whether significant phenotypic and genetic changes have, indeed, taken place following the
425 loss of a functioning CRISPR-Cas, as predicted if this defence system hindered the necessary adaptive
426 responses.

427 In conclusion, our study of the CRISPR-Cas defence system of bacterial isolates collected from the
428 original host, as well as from a novel host into which it jumped recently, reveals marked differences
429 between both in term of: (1) the MgCas9 PAM recognition specificity, (2) the composition and
430 diversity of the spacer repertoire, and (3) the gradual inactivation of CRISPR-Cas over time in the
431 novel host. While the changes in MgCas9 PAM specificity and in the composition and diversity of
432 spacers indicates a change in the landscape of phages and MGEs faced by the bacterium in the novel
433 host, the subsequent inactivation of CRISPR-Cas is consistent with another important selection event.
434 We hypothesize that the gradual inactivation of the CRISPR-Cas defence system, which took place
435 over the first 12 years of the epidemic until this defence system became fully inactivated from 2007,

436 occurred in response to the evolution of host resistance, which became widespread by 2007 and
437 pushed for adaptive changes in the bacterium. Such adaptive changes would have necessitated the
438 loss of a functional CRISPR-Cas defence system to take place. Together, our results highlight the need
439 to consider not only the host-driven selection pressures a bacterium experiences, but also the
440 complex interplay between phages and defence systems for better understanding the key factors
441 driving the emergence of a pathogenic bacterium in a novel host.

442

443 ## Material & Methods

444 **Oligonucleotides and plasmids.**

445 All oligonucleotides used in this study were supplied by Eurogentec and are described in Table S1. All
446 plasmids constructed and used in this study are listed in Table S2. Detailed protocols for plasmid
447 construction and other methods are provided in SI-1.

448 **Bacterial strains, culture conditions**

449 *M. gallisepticum S6* (Tax ID: 1006581) was cultivated in Hayflick modified medium (Freundt, 1983)
450 under a 5 % CO_2 atmosphere. Tetracycline 5 µg/mL was used in selective conditions. *M. gallisepticum*
451 S6 strain with a modified PI domain of MgCas9 was obtained as described in (Ipoutcha et al., 2022).

452 ***In vitro* determination of MgCas9 PAM specificity.** A 7N degenerated PAM plasmid library was
453 provided by Dr. Gasiunas from Caszyme (Gasiunas et al., 2020). Randomness of the 7 positions of the
454 randomized PAM was validated as described in (Karvelis et al., 2019). *In vitro* cleavage assays were
455 performed as explained in (Gasiunas et al., 2020). Details of the protocol are provided in SI-1.

456 **Transformation of *M. gallisepticum***

457 *M. gallisepticum* cells were grown for 36 h before transformation. At pH 6.2 to 6.5, aliquots of 10 mL
458 were centrifuged during 15 min at 6000 x *g*, 10°C. Cells were then resuspended with 5 mL of HBSS 1X

459 wash buffer (ThermoFisher, 14065056) and centrifuged 15 min at 6000 x *g*, 10°C. The pellet was
460 resuspended in 250 µL of $CaCl_2$ 0.1 M and incubated for 30 min on ice. Cold $CaCl_2$-incubated cells
461 were gently mixed with plasmid DNA (10 µg) and 10 µg of yeast tRNA (ThermoFisher, AM7119).
462 Then, 2 mL of 40 % polyethylene glycol 6000 (PEG) (Sigma, 11130) dissolved in HBSS 1X buffer were
463 added to the cells. After 2 min of incubation at room temperature, contact with the PEG was stopped
464 by addition of 20 ml of HBSS 1X wash buffer. The mixture was centrifuged 15 min at 6000 x *g*, 10°C
465 and the cells were resuspended in 1 mL of modified Hayflick medium pre-warmed at 37°C. After 2
466 hours of incubation at 37°C, cells were plated on Hayflick selective plates. After 10-15 days at 37°C
467 with 5 % CO_2, colonies were counted, picked and re-suspended in 1 mL of Hayflick with selection for
468 three passages (one passage ~48 h).

469 ## Acknowledgements

470 We are very thankful to Dr. Gasiunas from Caszyme for sharing the 7N PAM-degenerated pTZ57-7n
471 library, the pET28ID116 plasmid encoding *M. gallisepticum* CA06 Cas9 and for appreciated advices on
472 the protocol used for *in vitro* cleavage. We thank M. Staley for growing and shipping the house finch
473 isolates and Andrea J. Dowling for growing the isolates, extracting their DNA and organising the
474 sequencing of the genomes. We also want to thank Patrice Mora for helping in the constitution of
475 the *M. gallisepticum* spacers database, Mammadou Sall for his contribution to the study of spacers
476 and Alain Blanchard for much appreciated discussions on the work and manuscript.

477 ## Funding

478 This research did not receive any specific grant from funding agencies in the public, commercial, or
479 not-for-profit sectors.

480 ## Author contributions

Thomas Ipoutcha: Conceptualization, methodology, validation, formal analysis, investigation, writing - original draft, writing - review & editing

Iason Tsarmpopoulos: methodology, investigation

Géraldine Gourgues: investigation

Vincent Baby: software, formal analysis, investigation

Paul Dubos: software

Geoffrey E. Hill: review & editing

Yonathan Arfi: Conceptualization, writing - review & editing, supervision

Carole Lartigue: Conceptualization, methodology, writing - review & editing, supervision

Patricia Thébault: methodology, software, formal analysis, investigation, data curation

Camille Bonneaud: Conceptualization, writing - original draft, writing - review & editing

Pascal Sirand-Pugnet: Conceptualization, methodology, investigation, data curation, writing - original draft, writing - review & editing, supervision

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Adelman, J. S., Mayer, C., & Hawley, D. M. (2017). Infection reduces anti-predator behaviors in house finches. *Journal of Avian Biology*, *48*(4), 519–528. https://doi.org/10.1111/jav.01058

Alduhaidhawi, A. H. M., AlHuchaimi, S. N., Al- Mayah, T. A., Al-Ouqaili, M. T., Alkafaas, S. S., Muthupandian, S., & Saki, M. (2022). Prevalence of CRISPR-Cas Systems and Their Possible Association with Antibiotic Resistance in *Enterococcus faecalis* and *Enterococcus faecium* Collected from Hospital Wastewater. *Infection and Drug Resistance*, *Volume 15*, 1143–1154. https://doi.org/10.2147/IDR.S358248

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, *315*(5819), 1709–1712. https://doi.org/10.1126/science.1138140

Bikard, D., Hatoum-Aslan, A., Mucida, D., & Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host & Microbe*, *12*(2), 177–186. https://doi.org/10.1016/j.chom.2012.06.003

Bonneaud, C., Balenger, S. L., Russell, A. F., Zhang, J., Hill, G. E., & Edwards, S. V. (2011). Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7866–7871. https://doi.org/10.1073/pnas.1018580108

514 Bonneaud, C., Giraudeau, M., Tardy, L., Staley, M., Hill, G. E., & McGraw, K. J. (2018). Rapid
515 Antagonistic Coevolution in an Emerging Pathogen and Its Vertebrate Host. *Current Biology*,
516 *28*(18), 2978-2983.e5. https://doi.org/10.1016/j.cub.2018.07.003

517 Bonneaud, C., Tardy, L., Giraudeau, M., Hill, G. E., McGraw, K. J., & Wilson, A. J. (2019). Evolution of
518 both host resistance and tolerance to an emerging bacterial pathogen. *Evolution Letters*, *3*(5),
519 544–554. https://doi.org/10.1002/evl3.133

520 Bonneaud, C., Tardy, L., Hill, G. E., McGraw, K. J., Wilson, A. J., & Giraudeau, M. (2020). Experimental
521 evidence for stabilizing selection on virulence in a bacterial pathogen. *Evolution Letters*, *4*(6),
522 491–501. https://doi.org/10.1002/evl3.203

523 Bonneaud, C., Weinert, L. A., & Kuijper, B. (2019). Understanding the emergence of bacterial
524 pathogens in novel hosts. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
525 *374*(1782), 20180328. https://doi.org/10.1098/rstb.2018.0328

526 Broniewski, J. M., Meaden, S., Paterson, S., Buckling, A., & Westra, E. R. (2020). The effect of phage
527 genetic diversity on bacterial resistance evolution. *The ISME Journal*, *14*(3), 828–836.
528 https://doi.org/10.1038/s41396-019-0577-7

529 Calcutt, M. J., Lewis, M. S., & Wise, K. S. (2002). Molecular genetic analysis of ICEF, an integrative
530 conjugal element that is present as a repetitive sequence in the chromosome of *Mycoplasma*
531 *fermentans* PG18. *Journal of Bacteriology*, *184*(24), 6929–6941.
532 https://doi.org/10.1128/JB.184.24.6929-6941.2002

533 Citti, C., Dordet-Frisoni, E., Nouvel, L. X., Kuo, C. H., & Baranowski, E. (2018). Horizontal Gene
534 Transfers in Mycoplasmas (Mollicutes). *Current Issues in Molecular Biology*, *29*, 3–22.
535 https://doi.org/10.21775/cimb.029.003

536 Delaney, N. F., Balenger, S., Bonneaud, C., Marx, C. J., Hill, G. E., Ferguson-Noel, N., Tsai, P., Rodrigo,
537 A., & Edwards, S. V. (2012). Ultrafast evolution and loss of CRISPRs following a host shift in a
538 novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genetics*, *8*(2).
539 https://doi.org/10.1371/journal.pgen.1002511

540 Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A.,
541 Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in
542 *Streptococcus thermophilus*. *Journal of Bacteriology*, *190*(4), 1390–1400.
543 https://doi.org/10.1128/JB.01412-07

544 Dhondt, A. A., Tessaglia, D. L., & Slothower, R. L. (1998). Epidemic From Conjunctivitis America in
545 House Finches. *Journal of Wildlife Diseases*, *34*(2), 265–280.

546 Dimitriu, T., Szczelkun, M. D., & Westra, E. R. (2020). Evolutionary ecology of prokaryotic innate and
547 adaptive immune systems and their interplay. *Current Biology*.
548 https://doi.org/10.1016/j.cub.2020.08.028.Evolutionary

549 Dordet Frisoni, E., Marenda, M. S., Sagné, E., Nouvel, L. X., Guérillot, R., Glaser, P., Blanchard, A.,
550 Tardy, F., Sirand-Pugnet, P., Baranowski, E., & Citti, C. (2013). ICEA of *Mycoplasma agalactiae*: A
551 new family of self-transmissible integrative elements that confers conjugative properties to the
552 recipient strain. *Molecular Microbiology*, *89*(6), 1226–1239.
553 https://doi.org/10.1111/mmi.12341

554 Dowling, A., Hill, G. E., & Bonneaud, C. (2020). Multiple differences in pathogen- host cell interactions
555 following a bacterial host shift. *Scientific Reports*, 1–12. https://doi.org/10.1038/s41598-020-
556 63714-0

557 Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C. Q., Cao, Y., Gallant, J., Xue, W.,
558 Rivera-Pérez, J. A., & Sontheimer, E. J. (2019). A Compact, High-Accuracy Cas9 with a
559 Dinucleotide PAM for In Vivo Genome Editing. *Molecular Cell*, *73*(4), 714-726.e4.
560 https://doi.org/10.1016/j.molcel.2018.12.003

561 Freundt, E. A. (1983). CULTURE MEDIA FOR CLASSIC MYCOPLASMAS. In *Methods in Mycoplasmology*.
562 https://doi.org/10.1016/b978-0-12-583801-6.50029-9

563 García-Martínez, J., Maldonado, R. D., Guzmán, N. M., & Mojica, F. J. M. (2018). The CRISPR
564 conundrum: evolve and maybe die, or survive and risk stagnation. *Microbial Cell*, *5*(6), 262–268.
565 https://doi.org/10.15698/mic2018.06.634

566 Gasiunas, G., Young, J. K., Karvelis, T., Kazlauskas, D., Urbaitis, T., Jasnauskaite, M., Grusyte, M. M.,
567 Paulraj, S., Wang, P., Hou, Z., Dooley, S. K., Cigan, M., Alarcon, C., Chilcoat, N. D., Bigelyte, G.,
568 Curcuru, J. L., Mabuchi, M., Sun, Z., Fuchs, R. T., … Siksnys, V. (2020). A catalogue of
569 biochemically diverse CRISPR-Cas9 orthologs. *Nature Communications*.
570 https://doi.org/10.1038/s41467-020-19344-1

571 Hall, J. P. J., Wright, R. C. T., Harrison, E., Muddiman, K. J., Wood, A. J., Paterson, S., & Brockhurst, M.
572 A. (2021). Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by
573 compensatory mutation. *PLOS Biology*, *19*(10), e3001225.
574 https://doi.org/10.1371/journal.pbio.3001225

575 Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., & Marraffini, L. A. (2015).
576 Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature*, *519*(7542), 199–
577 202. https://doi.org/10.1038/nature14245

578 Ipoutcha, T., Gourgues, G., Lartigue, C., Blanchard, A., & Sirand-Pugnet, P. (2022). Genome
579 Engineering in Mycoplasma gallisepticum Using Exogenous Recombination Systems. *ACS*
580 *Synthetic Biology*, acssynbio.1c00541. https://doi.org/10.1021/acssynbio.1c00541

581 Ipoutcha, T., Tsarmpopoulos, I., Talenton, V., Gaspin, C., Moisan, A., Walker, C. A., Brownlie, J.,
582 Blanchard, A., Thebault, P., & Sirand-pugnet, P. (2019). Multiple Origins and Specific Evolution
583 of CRISPR / Cas9 Systems in Minimal Bacteria (Mollicutes). *Frontiers in Microbiology*,
584 *10*(November), 1–14. https://doi.org/10.3389/fmicb.2019.02701

585 Karvelis, T., Young, J. K., & Siksnys, V. (2019). A pipeline for characterization of novel Cas9 orthologs.
586 In *Methods in Enzymology* (1st ed., Vol. 616). Elsevier Inc.
587 https://doi.org/10.1016/bs.mie.2018.10.021

588 Kogay, R., Wolf, Y. I., & Koonin, E. V. (2024). Defence systems and horizontal gene transfer in
589 bacteria. *Environmental Microbiology*, *26*(4), e16630. https://doi.org/10.1111/1462-
590 2920.16630

591 Koonin, E. V. (2019). CRISPR: a new principle of genome engineering linked to conceptual shifts in
592 evolutionary biology. *Biology & Philosophy*, *34*(1), 9. https://doi.org/10.1007/s10539-018-9658-
593 7

Leigh, S. A., Evans, J. D., & Branton, S. L. (2019a). Complete Genome Sequences of Two *Mycoplasma gallisepticum* F-Strain Variants. *Microbiology Resource Announcements*, *8*(33). https://doi.org/10.1128/MRA.00485-19

Leigh, S. A., Evans, J. D., & Branton, S. L. (2019b). Complete Genome Sequences of Two Vaccine Strains and One Field Isolate of *Mycoplasma gallisepticum*. *Microbiology Resource Announcements*, *8*(49). https://doi.org/10.1128/MRA.01237-19

Levin, B. R. (2010). Nasty Viruses, Costly Plasmids, Population Dynamics, and the Conditions for Establishing and Maintaining CRISPR-Mediated Adaptive Immunity in Bacteria. *PLoS Genetics*, *6*(10), e1001171. https://doi.org/10.1371/journal.pgen.1001171

Ley, D. H., Berkhoff, J. E., & McLaren, J. M. (1996). Mycoplasma gallisepticum isolated from house finches (Carpodacus mexicanus) with conjunctivitis. *Avian Diseases*, *40*(2), 480–483.

Mackow, N. A., Shen, J., Adnan, M., Khan, A. S., Fries, B. C., & Diago-Navarro, E. (2019). CRISPR-Cas influences the acquisition of antibiotic resistance in *Klebsiella pneumoniae*. *PLOS ONE*, *14*(11), e0225131. https://doi.org/10.1371/journal.pone.0225131

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x

May, M., Balish, M. F., & Blanchard, A. (2014). The Order Mycoplasmatales. In *The Prokaryotes* (pp. 515–550). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30120-9_289

Meygret, A., Peuchant, O., Dordet-Frisoni, E., Sirand-Pugnet, P., Citti, C., Bébéar, C., Béven, L., & Pereyre, S. (2019). High Prevalence of Integrative and Conjugative Elements Encoding Transcription Activator-Like Effector Repeats in Mycoplasma hominis. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02385

Nolan, P. M., Hill, G. E., & Stoehr, A. M. (1998). Sex, size, and plumage redness predict house finch survival in an epidemic. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1400), 961–965. https://doi.org/10.1098/rspb.1998.0384

Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research*, *42*(16), 10618–10632. https://doi.org/10.1093/nar/gku734

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B. C., Barrangou, R., & Banfielda, J. F. (2015). CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio*, *6*(2), 1–9. https://doi.org/10.1128/mBio.00262-15

Palmer, K. L., & Gilmore, M. S. (2010). Multidrug-Resistant Enterococci Lack CRISPR- cas. *MBio*, *1*(4). https://doi.org/10.1128/mBio.00227-10

Papazisi, L., Gorton, T. S., Kutish, G., Markham, P. F., Browning, G. F., Nguyen, D. K., Swartzell, S., Madan, A., Mahairas, G., & Geary, S. J. (2003). The complete genome sequence of the avian pathogen Mycoplasma gallisepticum strain Rlow. *Microbiology*. https://doi.org/10.1099/mic.0.26427-0

Patterson, A. G., Yevstigneyeva, M. S., & Fineran, P. C. (2017). Regulation of CRISPR–Cas adaptive immune systems. *Current Opinion in Microbiology*, *37*, 1–7. https://doi.org/10.1016/j.mib.2017.02.004

Pursey, E., Dimitriu, T., Paganelli, F. L., Westra, E. R., & van Houte, S. (2022). CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1842). https://doi.org/10.1098/rstb.2020.0464

Roberts, S. R., Nolan, P. M., & Hill, G. E. (2001). Characterization of *mycoplasma gallisepticum* infection in captive house finches (Carpodacus mexicanus) in 1998. *Avian Diseases*, *45*(1), 70–75.

Sirand-Pugnet, P., Citti, C., Barré, A., & Blanchard, A. (2007). Evolution of mollicutes: down a bumpy road with twists and turns. *Research in Microbiology*, *158*(10), 754–766. https://doi.org/10.1016/j.resmic.2007.09.007

Song, Y., Kim, T., La, T.-M., Lee, H.-J., Lee, Y., Shin, H., Ahn, G., Hur, W., Song, S.-U., Lee, J.-B., Park, S.-Y., Choi, I.-S., & Lee, S.-W. (2021). Complete Genome Sequence of *Mycoplasma gallisepticum* Strain KUVMG001, an Isolate from South Korea. *Microbiology Resource Announcements*, *10*(19). https://doi.org/10.1128/MRA.00331-21

Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in Genetics: TIG*, *26*(8), 335–340. https://doi.org/10.1016/j.tig.2010.05.008

Szczepanek, S. M., Tulman, E. R., Gorton, T. S., Liao, X., Lu, Z., Zinski, J., Aziz, F., Frasca, S., Kutish, G. F., & Geary, S. J. (2010). Comparative genomic analyses of attenuated strains of Mycoplasma gallisepticum. *Infection and Immunity*, *78*(4), 1760–1771. https://doi.org/10.1128/IAI.01172-09

Tardy, L., Giraudeau, M., Hill, G. E., McGraw, K. J., & Bonneaud, C. (2019). Contrasting evolution of virulence and replication rate in an emerging bacterial pathogen. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 16927–16932. https://doi.org/10.1073/pnas.1901556116

Tulman, E. R., Liao, X., Szczepanek, S. M., Ley, D. H., Kutish, G. F., & Geary, S. J. (2012). Extensive variation in surface lipoprotein gene content and genomic changes associated with virulence during evolution of a novel North American house finch epizootic strain of *Mycoplasma gallisepticum*. *Microbiology (United Kingdom)*, *158*(8), 2073–2088. https://doi.org/10.1099/mic.0.058560-0

Vale, P. F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., & Gandon, S. (2015). Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151270. https://doi.org/10.1098/rspb.2015.1270

Vink, J. N. A., Baijens, J. H. L., & Brouns, S. J. J. (2021). Comprehensive PAM prediction for CRISPR-Cas systems reveals evidence for spacer sharing, preferred strand targeting and conserved links with CRISPR repeats. *BioRxiv*, 2021.05.04.442622. https://doi.org/10.1101/2021.05.04.442622

Watson, B. N. J., Capria, L., Alseth, E. O., Pons, B. J., Biswas, A., Lenzi, L., Buckling, A., van Houte, S., Westra, E. R., & Meaden, S. (2024). CRISPR-Cas in Pseudomonas aeruginosa provides transient population-level immunity against high phage exposures. *The ISME Journal*, *18*(1). https://doi.org/10.1093/ismejo/wrad039

675 Wei, Y., Terns, R. M., & Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A
676 CRISPR – Cas adaptation. *Genes*, 356–361. https://doi.org/10.1101/gad.257550.114.

677 Westra, E. R., & Levin, B. R. (2020). It is unclear how important CRISPR-Cas systems are for protecting
678 natural populations of bacteria against infections by mobile genetic elements. *Proceedings of*
679 *the National Academy of Sciences of the United States of America*, *117*(45), 27777–27785.
680 https://doi.org/10.1073/pnas.1915966117

681 Wheatley, R. M., & MacLean, R. C. (2021). CRISPR-Cas systems restrict horizontal gene transfer in
682 *Pseudomonas aeruginosa*. *The ISME Journal*, *15*(5), 1420–1433.
683 https://doi.org/10.1038/s41396-020-00860-3

684 Wimmer, F., & Beisel, C. L. (2020). CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers.
685 *Frontiers in Microbiology*, *10*(January), 1–17. https://doi.org/10.3389/fmicb.2019.03078

686

687 ## Figure captions

688 **Figure 1. Presence and integrity of the CRISPR defence system in *M. gallisepticum* strains. A -** Map
689 of the *M. gallisepticum* genome highlighting the position of the CRISPR-Cas locus (red box) and its
690 organization. **B** – Domains present in full-length and truncated forms of MgCas9 found in *M.*
691 *gallisepticum* strains. Green star, strains with a complete CRISPR locus and a complete MgCas9 ;
692 other motifs, inactivated CRISPR system with truncated forms V1, V2 and V3 of MgCas9. **C -**
693 Phylogenetic tree of MgCas9. Strains isolated from poultry and house finch hosts are highlighted in
694 red and purple, respectively. MgCas9 forms are indicated for all strains. Green dots on branches
695 indicate bootstrap values >95%.

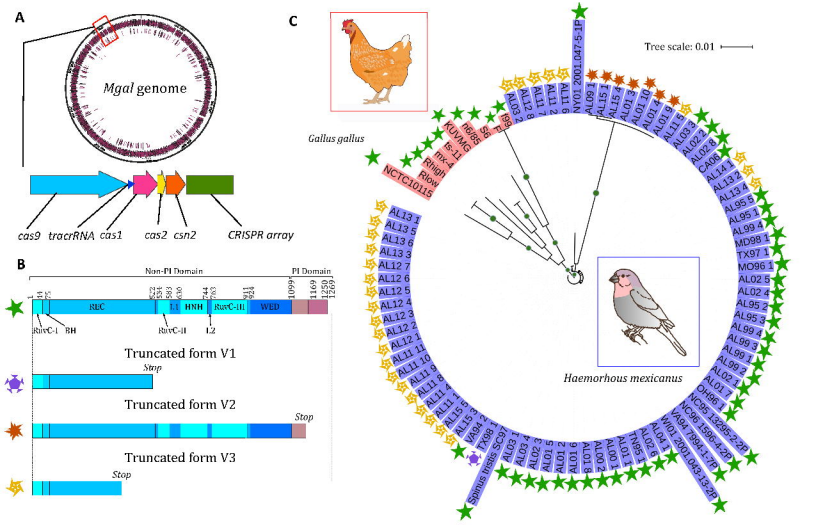
696 **Figure 2**. **Evolution of the Cas proteins from *M. gallisepticum* poultry and house finch isolates. A** -
697 Representation of MgCas9 protein variable positions (red dash) between strains S6 (poultry) and
698 CA06 (house finch) isolates. For each mutation, the corresponding amino-acid substitution is
699 indicated. **B**- Protein sequence comparisons between CRISPR effectors of S6 and CA06 isolates. **C** -
700 Protein identity plot. Percentages of identity are represented in the graph for each protein found in
701 *M. gallisepticum* S6 genome and CA06 genome from PATRIC analysis.

702 **Figure 3. *In vitro* characterization of PAM sequences recognized by MgCas9 from poultry (S6) and**
703 **house finch (CA06) strains. A**- Position of the observed cleavage site for the two MgCas9. **B**- Position
704 Frequency Matrix (PFM) observed after cleavage of a 7-positions PAM degenerated library by
705 MgCas9 from S6 and CA06 *M. gallisepticum* strains. Frequencies were averaged on three replicates
706 and corrected considering the bias in the distribution of PAM sequences in the initial plasmid library.
707 Adenine are represented in green, Cytosine in blue, Guanine in yellow and Thymine in red.
708 Frequencies of each base at each position are indicated on the graph. **C**- Web logos of consensus
709 PAM sequences recognized by MgCas9 S6 or CA06. **D**- Top 10 of the best motifs recognized *in vitro* by
710 MgCas9 S6 and CA06.

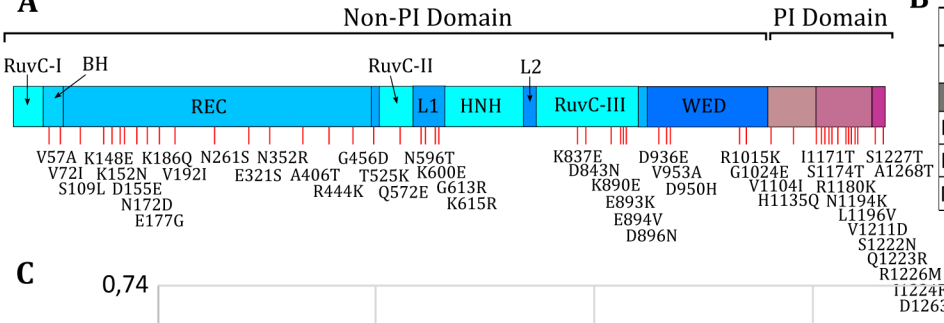
711 **Figure 4. Evaluation of MgCas9 PAM specificity *in vivo* using a plasmid interference assay. A**- PAM
712 sequences for *in vivo* interference assays were selected from *in vitro* results. DR is the direct repeat
713 sequence naturally present in the *M. gallisepticum* CRISPR array; this motif cannot be recognized by
714 MgCas9 and is used as a positive control of transformation and negative control for cleavage. Four
715 other PAM sequences where chosen, that are differentially cut *in vitro* by MgCas9 S6 and CA06. PAM
716 sequences frequently retrieved after cleavage are noted "+++" whereas lowly and not recognized
717 sequences are indicated by "+" and "-", respectively. **B**- Plasmid design for interference assays in *M.*
718 *gallisepticum* S6 and *M. gallisepticum* S6 mod PI CA06. Plasmid pMgal contains *M. gallisepticum oriC*,
719 a *tet(M)* resistance marker and the third spacer which is present in the S6 CRISPR array, followed by
720 selected PAM sequences. **C**- *In vivo* result of transformation assays for the two *M. gallisepticum*
721 strains. The numbers of transformants obtained after three transformation replicates are
722 represented as a dot plot.

723 **Figure 5. Hypothetical model of evolution of the CRISPR-Cas system in *M. gallisepticum* as it**
724 **jumped from poultry and emerged into the novel house finch host.** Phenotypic and genetic changes
725 accompanying each step are obtained from the literature.

726


A
Mgal genome
cas9
tracrRNA
cas1
cas2
csn2
CRISPR array
B
Non-PI Domain
PI Domain
RuvC-I
RH
RuvC-II
L2
RuvC-III
WED
Truncated form V1
Stop
Truncated form V2
Stop
Truncated form V3
Stop
C
Tree scale: 0.01
Gallus gallus
Haemorhous mexicanus
Spinus tristis

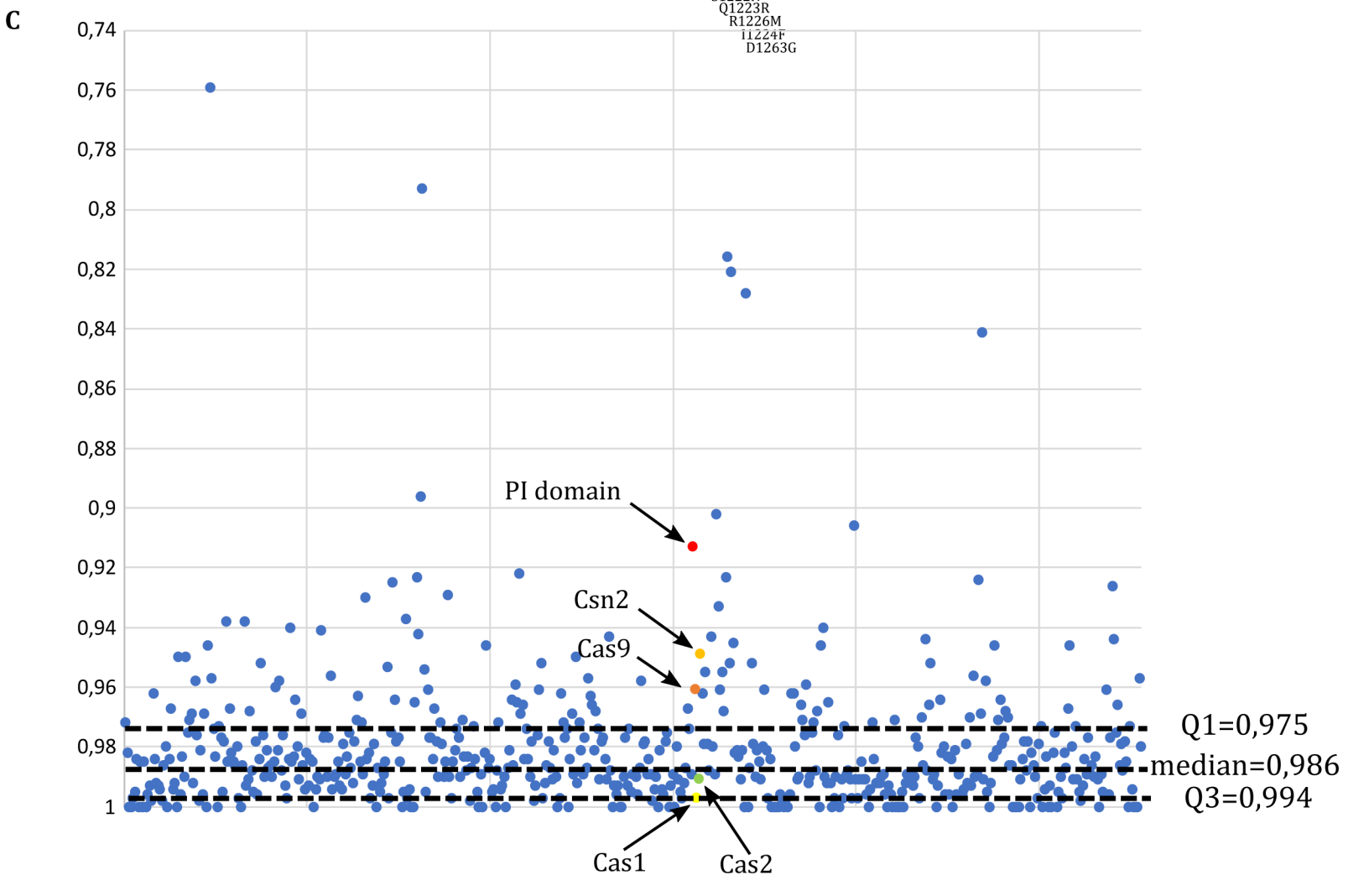
A



B

Proteic CRISPR locus evolution of *Mgal* CA06 vs S6						
	Cas9		Cas1	Cas2	Csn2	Genome
	Non-PI	PI				
Protein length (AA)	1, 097	173	309	110	255	224, 572
Mutation number	34	15	1	1	13	4, 429
Mutation number for 100 AA	3.1	8.7	0.3	0.9	5.1	1.98

C

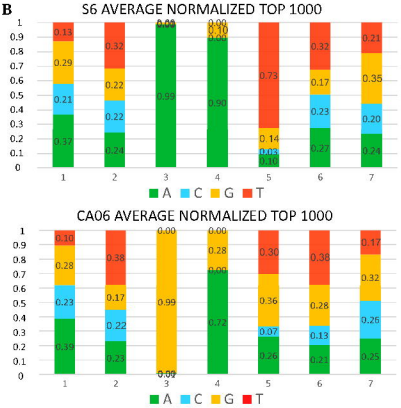


A

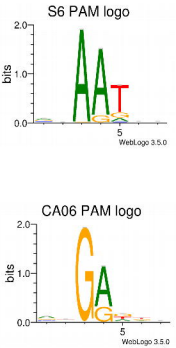
D

TOP 10 motifs	
S6 motif	*CA06* motif
ACAATCA	GTGATCG
AGAATCG	ACGATCT
ACAATTG	ACGATTC
ACAATCC	ACGATTG
CTAATTG	ATGATTC
CAAATTC	CTGATGC
CTAATTC	AAGAAAT
CCAATCG	ACGAATC
GCAATTG	ACGATGC
CTAATCG	AAGAGTG

B



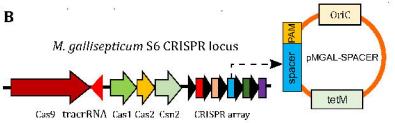
C



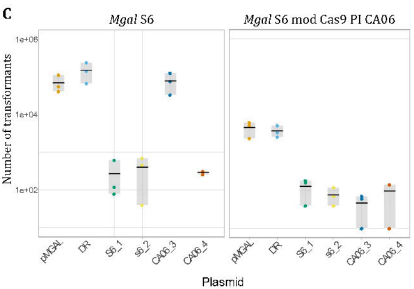
A

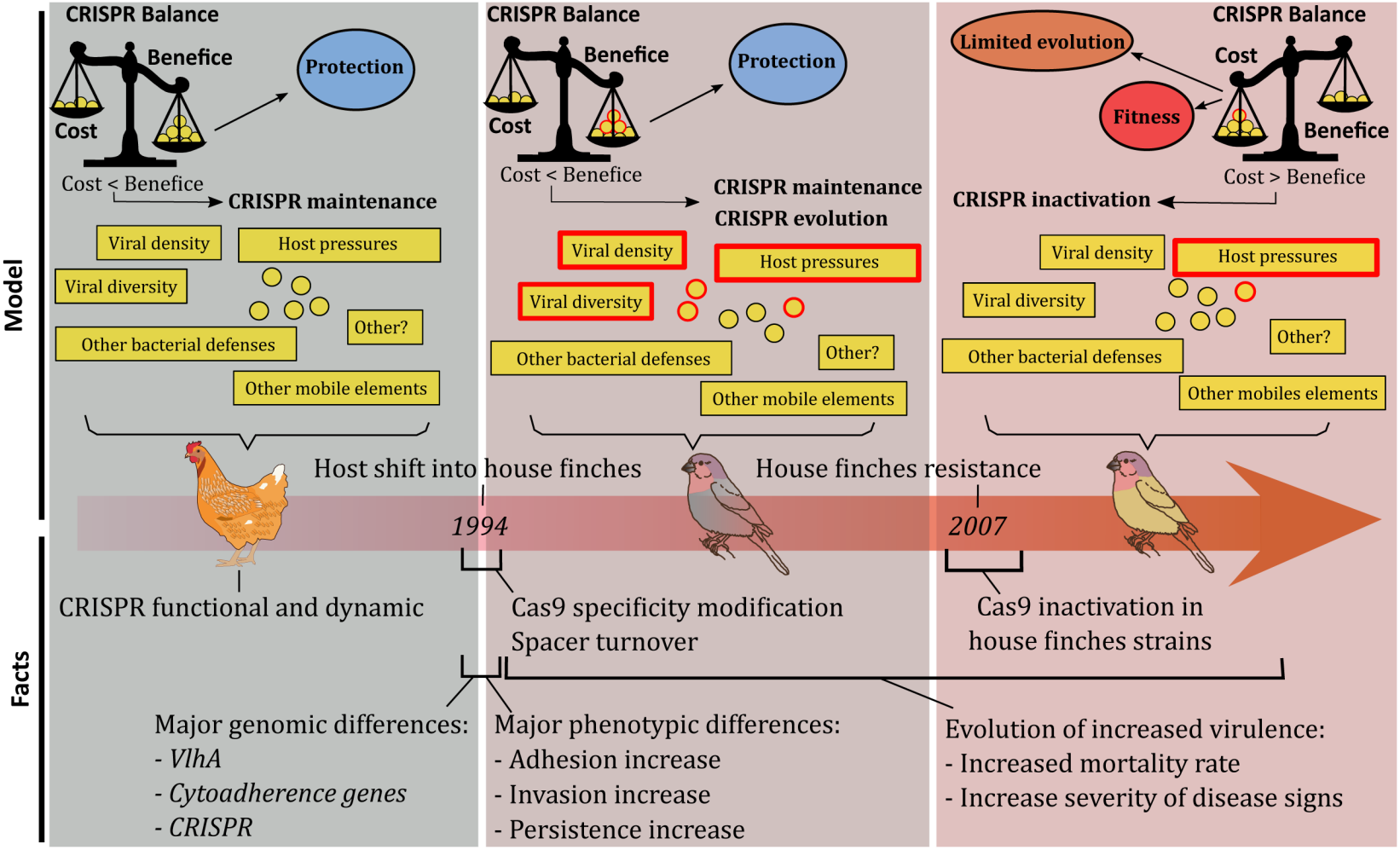
Motif		*In vitro* results		
		PAM Library ranking	Cas9 S6 cleavage	Cas9 CA06 cleavage
DR (T-)	GTTTTAGC	497th	-	-
S6-1	ACAATCA	15800th	+++	+
S6-2	ACAATTG	6200th	+++	+
CA06-3	TTGACTG	1200th	-	+++
CA06-4	AAGAGTG	4300th	+	+++

B



C




Model
Facts
CRISPR Balance
Benefice
Cost
Protection
Cost < Benefice
CRISPR maintenance
Viral density
Host pressures
Viral diversity
Other?
Other bacterial defenses
Other mobile elements
CRISPR Balance
Benefice
Cost
Protection
Cost < Benefice
CRISPR maintenance
CRISPR evolution
Viral density
Host pressures
Viral diversity
Other?
Other bacterial defenses
Other mobile elements
CRISPR Balance
Limited evolution
Cost
Fitness
Benefice
Cost > Benefice
CRISPR inactivation
Viral density
Host pressures
Viral diversity
Other?
Other bacterial defenses
Other mobiles elements
Host shift into house finches
House finches resistance
1994
2007
CRISPR functional and dynamic
Cas9 specificity modification
Spacer turnover
Cas9 inactivation in
house finches strains
Major genomic differences:
- VlhA
- Cytoadherence genes
- CRISPR
Major phenotypic differences:
- Adhesion increase
- Invasion increase
- Persistence increase
Evolution of increased virulence:
- Increased mortality rate
- Increase severity of disease signs