Evolution of the CRISPR-Cas9 defence system following a bacterial host shift

Authors: Thomas IPOUTCHA[1], Iason TSARMPOPOULOS[1], Géraldine GOURGUES[1], Vincent BABY[1], Paul DUBOS[1], Geoffrey E. HILL[2], Andrea DOWLING[3], 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] LaBRI, Univ. Bordeaux, F-33405 Talence, France
*Corresponding author: 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), Andrea DOWLING (0000-0003-0818-6092), 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).

Classification: Microbiology

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

The authors declare no conflict of interest.

Abstract

CRISPR-Cas systems are rapidly evolving bacterial defences targeting phages and mobile genetic elements. How these defences evolve in novel host environments remains, however, unknown. We address this question using *Mycoplasma gallisepticum*, a bacterial pathogen of poultry that jumped into a novel host ~30 yrs ago. In the 12 yrs following the jump, both active and inactive forms of MgCas9 were found in circulation, after which all isolates displayed inactive forms only. Isolates with active CRISPR-Cas harboured new sets of spacers, and a MgCas9 PI domain targeting a new protospacer adjacent motif (PAM), consistent with a change in the community of phages and mobile elements encountered within the novel host. The striking concordance in time between the rise of inactivated forms of CRISPR-Cas and the evolution of widespread host resistance, suggests that inactivation was subsequently necessary for adaptive bacterial responses. We highlight the need to consider both host and pathogen selection pressures on bacteria for understanding CRISPR systems evolution.

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems are effective defences found in most prokaryotes that protect them against the constant threat of phages and other sources of invading DNAs. While the flexibility and effectiveness of this defence system relies on the continuous acquisition of new spacers to the CRISPR array (1), its costly maintenance has, on the other hand, been invoked for its frequent inactivation and disappearance from prokaryote genomes (2,3). Rapid adaptations of spacer arrays, as well as gain/loss events of active CRISPR systems over the evolution of prokaryotes will, therefore, be shaped both by cost/benefit ratios (4) and by the complex ecological pressures experienced by prokaryotes. For pathogenic bacteria, the main ecological pressures arise from interactions with the host. Changes in host-driven selection pressures can occur over time as hosts evolve in response to pathogens but will be particularly dramatic in the event of host jumps. While such changing selection pressures, in particular following invasion of a novel host, are known to broadly impact bacterial genomes, how they will specifically shape bacterial defence systems against phages and other parasitic DNAs remains to be studied.

Instances in which bacterial pathogens have been documented to jump into and subsequently adapt to a new host are however rare (5), limiting our abilities to test the effect of any dramatic host environmental change on bacterial CRISPR systems. One such exception is the host shift of the bacterial pathogen, *Mycoplasma gallisepticum*, from poultry into a widely distributed North American passerine, the house finch (*Haemorhous mexicanus*). This novel host jump has been dated to the mid-1990s, with all subsequent *M. gallisepticum* lineages collected since in house finches found to have derived from this single event (6,7). *M. gallisepticum* is an avian pathogen belongs to the class *Mollicutes* (8), which are minimal bacteria characterized by a rapid evolution through drastic genome reduction from a Gram-positive ancestor. Such genetic loss is partially counter balanced by horizontal gene transfers (HGT) and exposure to mobile genetic elements acting as potential sources of genetic diversity and genome reorganization (9,10). Despite an evolutionary tendency to genome minimization, most mycoplasmas have retained various defence systems against phages and invading DNAs. Indeed, mycoplasmas are the second group of bacteria with the greatest number of Restriction-Modification systems (11) and are also the bacteria with the smallest genome where CRISPR systems have been described (12).

Changes in the CRISPR locus of *M. gallisepticum* have previously been described following the host shift into house finches (6,13). Indeed, a comparison of *M. gallisepticum* isolates collected from poultry and house finches evidences a complete change in the spacer set, with none being shared between both (6). Over time, house finch isolates also displayed a loss of spacers, as well as a degradation of the CRISPR locus. While these patterns suggest both a slowing down in the recruitment of new spacers and even an evolution towards this defence system becoming non-functional, further work is required to test whether such evolutionary trend has persisted over time in the new passerine host, as well as to functionally characterise other changes in the CRISPR-Cas system, and in particularly Cas9. Indeed, Cas9 endonuclease is the main actor of type-II CRISPR systems, both involved in the acquisition of new spacers and the double-strand DNA cleavage of target DNAs. In both processes, Cas9 selects functional spacers by recognizing their protospacer adjacent motif (PAM) sequence immediately downstream. Therefore, the change of spacer set that has accompanied *M. gallisepticum* host jump questions the evolution of MgCas9 PAM specificity as a possible explanation for this dramatic event.

Here, we conducted such a functional characterisation using a two-pronged approach. First, we performed *in silico* analyses on an extended set of *M. gallisepticum* isolates (79 collected from 1994 to 2015) to better understand the evolution of the CRISPR-Cas9 system in the novel finch host. We identified several evolutionary events leading to the remarkable divergence of spacer sets present in poultry and house finch arrays, as well as detected further evidence for a longer-term inactivation of the CRISPR-Cas system over time. We also identified a new mycoplasma ICE-like element and putative

phages, which may represent new targets of the CRISPR-Cas defence system in house finches. Second, we functionally characterized the CRISPR-Cas system of two *M. gallisepticum* strains infecting poultry (S6 strain) and house finch (CA06 strain) hosts. Using *in vitro* and *in vivo* approaches, we determined the PAM specificities of the MgCas9 and showed a correlation between a modification of the PAM signature, the evolution of the spacer repertoires, and the evolution of the PI domain of MgCas9. Such radical change in the CRISPR-Cas defence system of *M. gallisepticum* may have implications for the rapid adaptation of *M. gallisepticum* to its novel host (14,15).

Material & Methods

Oligonucleotides and plasmids.

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

Bacterial strains, culture conditions

M. gallisepticum S6 (Tax ID: 1006581) was cultivated in Hayflick modified medium (16) under a 5 % CO_2 atmosphere. Tetracycline 5 µg/mL was used in selective conditions. *M. gallisepticum* S6 strain with a modified PI domain of MgCas9 was obtained as described in (17).

***In vitro* determination of MgCas9 PAM specificity.** A 7N degenerated PAM plasmid library was provided by Dr. Gasiunas from Caszyme (18). Randomness of the 7 positions of the randomized PAM was validated as described in (19). *In vitro* cleavage assays were performed as explained in (18). Details of the protocol are provided in SI-1.

Transformation of *M. gallisepticum*

M. gallisepticum cells were grown for 36 h before transformation. At pH 6.2 to 6.5, aliquots of 10 mL were centrifuged during 15 min at 6000 x *g*, 10°C. Cells were then resuspended with 5 mL of HBSS 1X wash buffer (ThermoFisher, 14065056) and centrifuged 15 min at 6000 x *g*, 10°C. The pellet was resuspended in 250 µL of $CaCl_2$ 0.1 M and incubated for 30 min on ice. Cold $CaCl_2$-incubated cells were gently mixed with plasmid DNA (10 µg) and 10 µg of yeast tRNA (ThermoFisher, AM7119). Then, 2 mL of 40 % polyethylene glycol 6000 (PEG) (Sigma, 11130) dissolved in HBSS 1X buffer were added to the cells. After 2 min of incubation at room temperature, contact with the PEG was stopped by addition of 20 ml of HBSS 1X wash buffer. Mixture was centrifuged 15 min at 6000 x *g*, 10°C and the cells were resuspended in 1 mL of modified Hayflick medium pre-warmed at 37°C. After 2 hours of incubation at 37°C, cells were plated on Hayflick selective plates. After 10-15 days at 37°C with 5 % CO_2, colonies were counted, picked and re-suspended in 1 mL of Hayflick for three passages (one passage ~48 h).

Results

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

CRISPR-Cas sequences were extracted from previously published whole genome sequences of 19 poultry and 8 house finch strains (7,20–24), as well as from a further 79 newly-sequenced genomes of strains collected in house finches between 1994-2015 (Table S3). A single CRISPR locus was found in all *M. gallisepticum* isolates obtained from house finches (Figure 1A), with the exception of 6 isolates (1 collected in 2006, 3 in 2007, 1 in 2008 and 1 in 2015). In all 6 of these isolates, the CRISPR-Cas system was found to have been inactivated due to the loss of part of the locus harbouring the tracrRNA and the *cas* genes (i. e. *cas9, cas1, cas2* and *csn2*), leaving only the CRISPR array. An extended analysis of the genome region on both sides of the CRISPR locus further suggested this system being part of a larger defence island, also including a restriction-modification system, a toxin-antitoxin system and a

putative S8 serine peptidase (SI-2). At least for the 2 isolates with a completely assembled genome (strains collected in 2006 and 2008), the deletion was extended upstream of the CRISPR locus, including the putative S8 serine peptidase and other adjacent genes. The inactivation of the CRISPR-Cas system could also be predicted for a further subset of isolates based on mutations found in their MgCas9 gene. Indeed, we were able to predict three different truncated versions of MgCas9 (thereafter named V1, V2 and V3) (Figure 1B, table S3). The MgCas9 truncated form V1 form was observed only once in a 1998 house finch isolate; it is truncated at position 1498 in the REC domain as a result of a single base change that generates a TAG STOP codon instead of CAG Gln codon. The V2 form was detected in 7 isolates sampled in house finches in 2001, 2009 and 2015; it was predicted based on a frameshift at position 3445 of the CDS, leading to a protein truncated of the last 120 amino acids within the PAM Interacting domain (PI). Finally, the V3 form was identified in 29 isolates collected in house finches in 2003, as well as between 2011 to 2015; it is truncated at position 1018 of the CDS within the REC lobe (which exerts a key role for DNA cleavage) as a result of a single nucleotide mutation changing a GAA Glu codon into a TAA STOP codon.

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

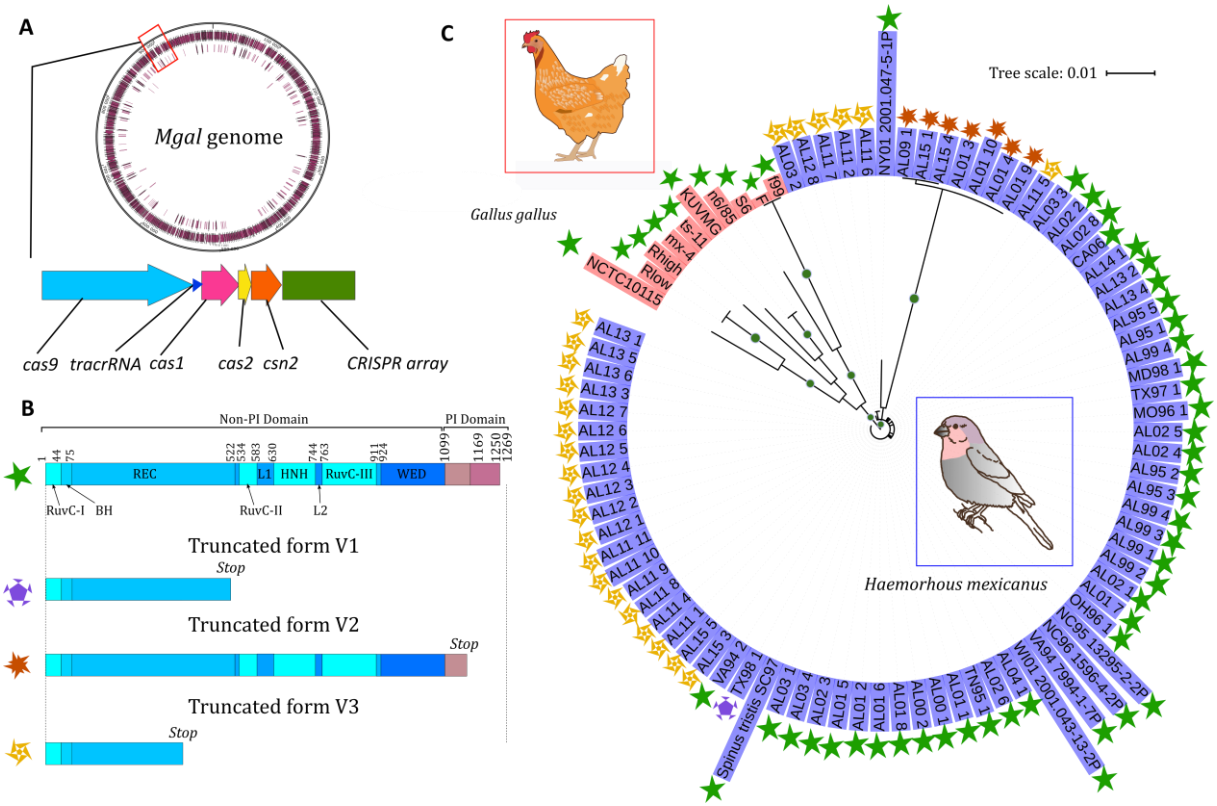


Figure 1. Presence and integrity of the CRISPR defence system in *M. gallisepticum* strains. A - Map of the *M. gallisepticum* genome highlighting the position of the CRISPR-Cas locus (red box) and its organization. **B** – Domains present in full-length and truncated forms of MgCas9 found in *M. gallisepticum*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

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

We compared the set of spacers inferred from the *M. gallisepticum* CRISPR locus across all our *M. gallisepticum* isolates (Figure S1, Table S4), and compiled a dataset of 490 non-redundant spacers (Table S4C). As a result, we identified a total of 438 unique spacers from 19 poultry strains and 55 unique spacers from 85 house finch isolates (Table S4C). Across the *M. gallisepticum* poultry strains, we observed a wide diversity of spacers, with no spacer found conserved across all strains (Figure S1). Indeed, 66% of spacers from the 19 poultry strains were present in only 1-2 genomes, while 2% were found in 9-10 genomes. In contrast, we found a lower diversity of spacers across the CRISPR arrays of house finch isolates. A quarter (24%) of spacers were shared across 80-90% of house finch isolates, while over half (58%) were present in at least 40% of them, and 21 out of 55 (38%) spacers were present in 10% or less of house finch isolates.

We found only 3 spacers that were shared across poultry and house finch isolates. Spacer SP_516* was identified in the poultry vaccinal strain ts-11, in 7 derivates obtained by chemical mutagenesis (K6372, K6356, K6212B, K6216D, K6208B, K5322C, K2966) and in all but 10 of the house finch isolates. Spacer SP_119* was found in 4 ts-11 derivate strains (K6369, K6208B, K5322C and K6222B) and in 3 house finch isolates for which the CRISPR-Cas9 system was predicted to be inactive (169125442_2001, 13_09_02_NC and 2015_J). Finally, spacer SP_325* was shared between the poultry strains R_{Low} and R_{High} and by 5 house finch isolates (NC95, NY01, WI01, NC06 and NC08). When considering the pattern

of spacer acquisition over time in house finches, we found that most new spacers (i.e. 46 out of 55; >80%) had been integrated before 2000 and were therefore present in the early years of *M. gallisepticum* emergence in house finches. Surprisingly, despite the inactivation of MgCas9 from 2007, we also found evidence for the acquisition of six new spacers between 2009 and 2015 (Table S4C).

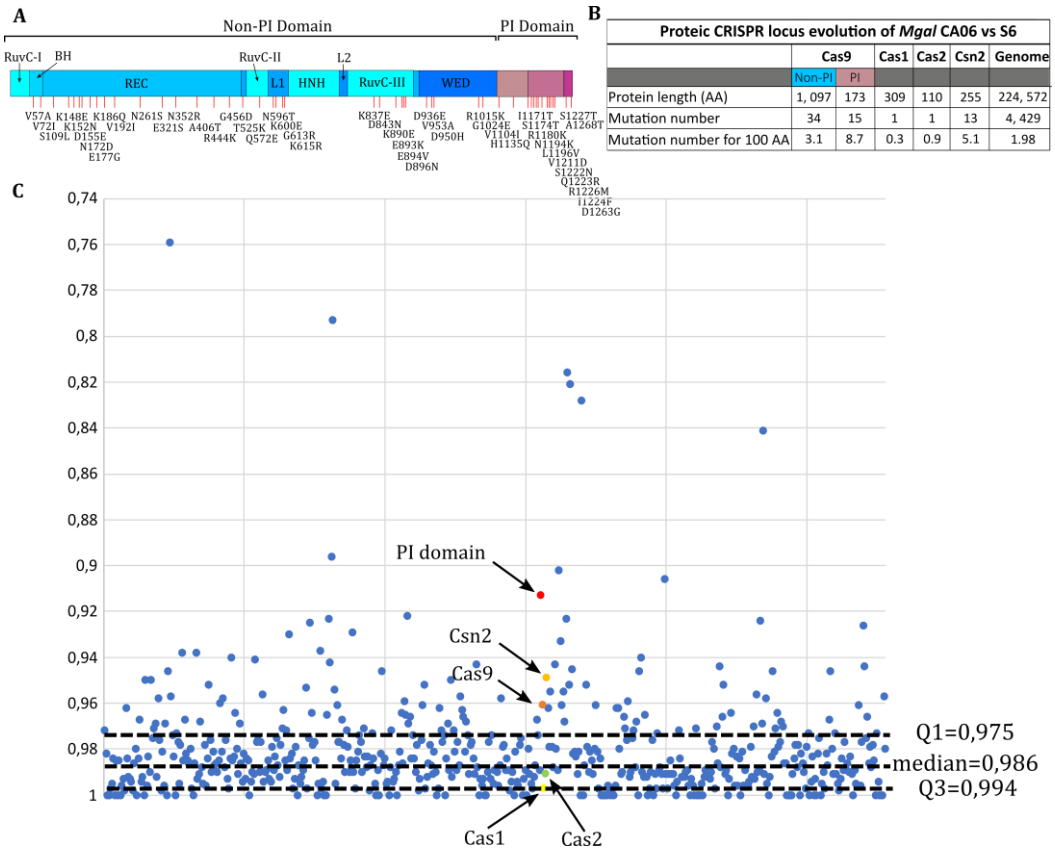
We investigated the potential origin of *M. gallisepticum* spacers using Blast queries against PHASTER, GenBank and IMG databases (Table S4D-S4G), as described in (25). Out of 490 spacers, 431 (88%) showed no hits in the explored databases. The PHASTER database returned hits on known bacterial phages for 34 spacers. GenBank returned hits on bacterial genomes (11 spacers) and phages (4 spacers). IMG returned hits on bacterial genomes for 8 spacers. In addition, two self-matching spacers were found in 2 strains of *M. gallisepticum* (ts-11 and NTCTC1015). Overall and across the 57 spacers with identified hits outside of *M. gallisepticum* genomes, we noted 6 spacers that were present in house finch isolates, and 51 in those from poultry. Out of the 19 spacers with hits on bacterial genomes, 10 returned non-CRISPR genomic regions of mycoplasma. Among them, 6 spacers from poultry strains (S6, KUVMG001, 6/85, ts-11 and ts-11 derivates) matched to the genome of *Mycoplasma imitans,* a duck pathogen phylogenetically related to *M. gallisepticum* (Table S4F); these potential protospacers are localized in a *M. imitans* chromosomal region of 14 kb that is not found in *M. gallisepticum* genomes, contrary to other surrounding loci. In this region, we identified several genes considered as essential for the mobilization and transfer of Mycoplasma Integrative and Conjugative Elements (MICE) (Figure S2) (26–28). This suggested those 6 spacers might be directed against a potential mobile element of 14 kb from bird mycoplasma, although no trace of this element has been detected in currently available *M. gallisepticum* genomes. Taken together, these results indicate that *M. gallisepticum* CRISPR-Cas9 systems are mostly directed against unknown targets and that predicted targets include most predominantly phages, as well as a newly identified ICE-like element of a bird mycoplasma.

Diversity and evolution of Cas proteins

We undertook further *in silico* analyses to document the intraspecific diversity of *M. gallisepticum* Cas proteins across isolates displaying a complete functional locus. To do so, we aligned the protein sequences of *M. gallisepticum* Cas9 (MgCas9) and constructed a phylogenetic tree (Figure 1C). The isolates were clearly distributed into two main branches, with a clustering of the poultry MgCas9 sequences in one branch and of the house finch ones in another. House finch MgCas9 sequences were found to be >99% identical to each other and the copy present in the CA06 isolate was selected (as in (18)) to be compared with a poultry MgCas9 from the S6 strain. This strain was chosen because it has not been mutagenized and it has been proven to be transformable, which was required for further experimental approaches. While being 96% identical in sequence, the highest proportion of amino acid variations was found within the PI domain, which is involved in the interaction with the PAM sequence (Figure 2A). Indeed, sequence variations of 8.7% and 3.1% were calculated for the PI domain and the other domains of the protein, respectively.

When considering other Cas proteins encoded within the CRISPR locus, we found that Csn2 sequences exhibited more variation (5.1%) than either Cas1 (0.3%) or Cas2 (0.9%) (Figure 2B). A proteome-wide comparison of *M. gallisepticum* CA06 and S6 isolates showed that Cas9 and Csn2 are part of the 12% most divergent proteins between the two isolates (Figure 2C, Table S4). Together, our results suggest that the jump of *M. gallisepticum* from poultry into house finches has been marked by a noticeable evolution of the MgCas9 PI domain and of Csn2 that are both key proteins of the adaptation stage of CRISPR-mediated immunity, known to be required for the acquisition of new spacers (29,30). Because the PI domain of Cas9 is involved in the PAM recognition, our results further suggest that the jump might be associated with a change in the PAM recognition specificity of MgCas9, a pattern which is

consistent with a change in pathogen pressure experienced by *M. gallisepticum* within the novel finch host.



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

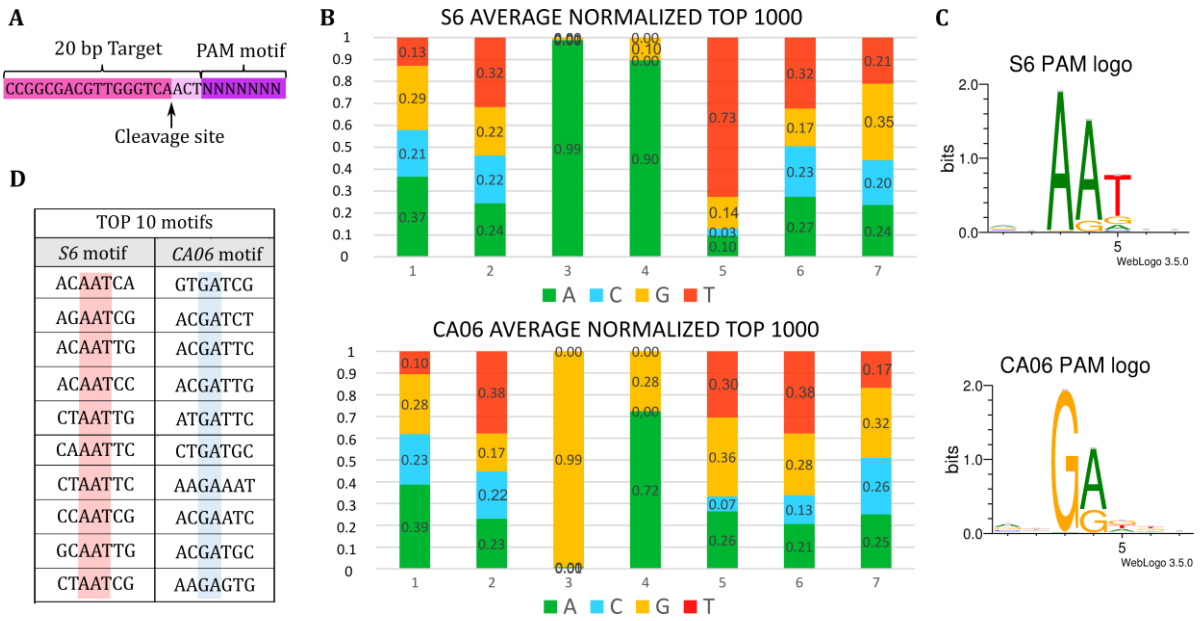
Figure 2. Evolution of the Cas proteins from *M. gallisepticum* poultry and house finch isolates. A - Representation of MgCas9 protein variable positions (red dash) between strains S6 (poultry) and CA06 (house finch) isolates. For each mutation, the corresponding amino-acid substitution is indicated.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

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

We used an *in vitro* assay previously developed by the Siksnys 's group (18,19) to compare the PAM recognition preference of MgCas9 between the poultry strain S6 and a house finch isolate CA06. *In vitro*-produced MgCas9/sgRNA ribonucleoparticles were used for the cleavage of a library of plasmids containing a target sequence and a PAM sequence randomized on 7 positions (Figure 3A, Figure S3, Table S6). After three independent *in vitro* cleavage of the plasmid library by MgCas9 S6 or MgCas9 CA06, molecules containing a recognized PAM sequence were captured using DNA ligation of adapters, enriched by PCR amplification, and submitted for deep sequencing. Cleavage was observed three bases upstream from the PAM sequence at more than 99.95% and results obtained for replicates of the two series of were nearly identical (Figures S4A and S4B). Out of the 16306 PAM sequences observed in the library, 10928 and 9709 PAM sequences were detected at least once after cleavage by MgCas9 S6 and MgCas9 CA06, respectively. The 1000 most frequent PAM sequences were selected for each MgCas9 and 74 PAM sequences were in common. Normalized frequencies were calculated, taking the bias of the original plasmid library into account (Table S6F); frequency matrices and sequence logos were produced (Figures 3B and 3C, Table S6G).

A comparison of the PAM recognition preference of poultry (S6) and the house finch (CA06) MgCas9 revealed that the PAM sequences they both recognise are 5 letters motifs with positions 3, 4 and 5 as the only ones to show a marked bias. Two lines of evidence indicate that the 5-positions PAM

sequences recognised by poultry S6 and house finch CA06 strains differ *in vitro*. First, a sequence logo comparison revealed that PAM preferences of MgCas9 proteins were clearly different between the two isolates. The main difference was observed at position 3 with 99% of sequences displaying an A in S6 and a G in CA06 (Figures 3B). Position 4 was preferentially A or G in both cases, with a bias towards A more marked in S6. In position 5, a T was present in 73% of sequences retrieved after cleavage by MgCas9 S6, compared to 30% by MgCas9 CA06, suggesting a notable difference at this position between the two isolates. Second, we found that the top 10 PAM sequences most frequently recognized differed between S6 and CA06 MgCas9 (Figure 3D).



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

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

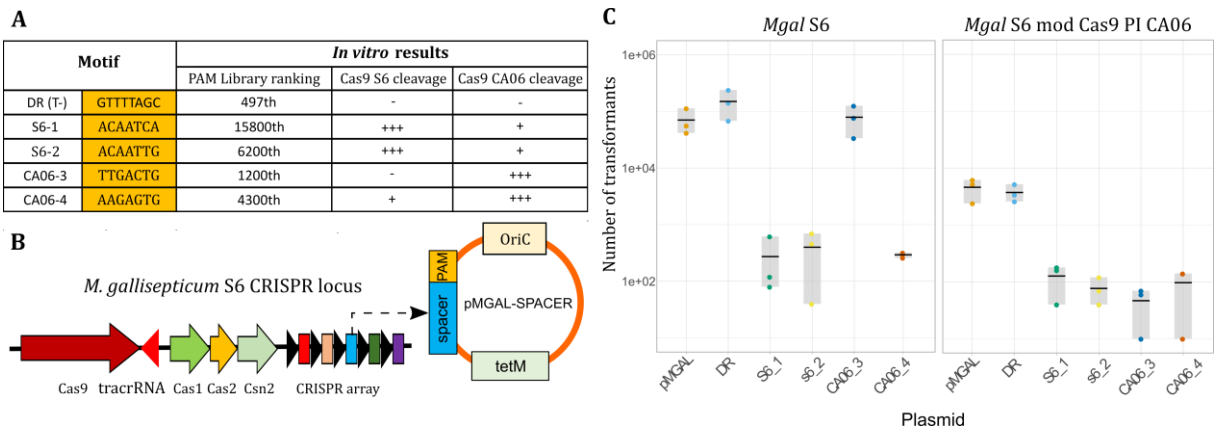
To confirm the difference of PAM recognition specificity, we selected two PAM sequences, specifically recognized by MgCas9 S6 (PAM 1, AT**AAA**AA) and MgCas9 CA06 (PAM 2, AA**GAG**AA), respectively. Three plasmids were constructed containing the target sequence followed by one of the two selected PAM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

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

We characterised the PAM specificity of poultry (S6) and house finch (CA06) MgCas9 using *in vitro* and *in vivo* approaches. Based on the previously obtained *in vitro* results, four PAM motifs found to be

differentially cut by poultry (S6) and house finch (CA06) MgCas9 were selected (Figure 4A). Motifs S6-1 and S6-2 were found to be recognized by MgCas9 S6 *in vitro*, but poorly by MgCas9 CA06, while motif CA06-3 was cleaved by MgCas9 CA06 but not MgCas9 S6, and CA06-4 was efficiently recognized by MgCas9 CA06 but only weakly by MgCas9 S6. To investigate whether these findings obtained *in vitro* are representative of processes occurring *in vivo*, we conducted interference assays on both the wild-type *M. gallisepticum* S6 strain and on a genetically engineered *M. gallisepticum* S6 (thereafter named *M. gallisepticum* S6 mod PI CA06) in which the MgCas9 PI domain was seamlessly replaced by the homologous domain of the MgCas9 CA06 (17). The *in vivo* cleavage assay was based on the transformation of *M. gallisepticum* with a replicative plasmid containing the third spacer naturally found in the CRISPR array of strain S6, followed by a selected PAM candidate (Figure 4B). A plasmid that contained the Direct Repeat of the CRISPR array as a PAM sequence was used as a negative control.

Transformations with plasmids containing PAM sequences expected to be cut by MgCas9 S6 (S6-1 and S6-2) led to 100-1000 fold less transformants compared to transformation with the control plasmid (Figure 4C). This confirmed that the CRISPR/Cas9 system is functional as a defense system in *M. gallisepticum* S6 and that PAM sequences recognized *in vitro* by MgCas9 S6 are also recognized *in vivo*. Similarly, plasmids harbouring PAMs frequently cleaved by MgCas9 CA06 *in vitro* were also cleaved by MgCas9 mod PI CA06 *in vivo*, demonstrating that the hybrid MgCas9 is fully active. Interestingly, in this interference assay, motifs that were poorly recognized *in vitro* (i.e. CA06-4 for MgCas9 S6; S6-1 and S6-2 for MgCas9 CA06), were efficiently cleaved *in vivo*. Finally, as we expected, we found that the PAM CA06-3 was not recognized by MgCas9 S6, but was efficiently cleaved by the hybrid MgCas9 that included a CA06 PI domain. This result confirms that the specificity of MgCas9 is mediated by the PI domain and suggests that mutations in the MgCas9 of house finch isolates has led to a change in PAM specificity (see SI-3b for further comparisons of *in vitro* and *in vivo* findings).



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	+	+++

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

Discussion

Using isolates from the bacterial pathogen, *M. gallisepticum*, that were collected in the original poultry host, as well as over time following a jump into a novel host species (the house finch), we investigated the impact of a dramatic change in the host environment on the evolution of the pathogen's CRISPR defence system. First, our results were consistent with previous findings on a restricted number of isolates sampled in house finches and which reported the gradual inactivation of *M. gallisepticum* Cas9 (MgCas9). Indeed, we found that, while all poultry strains displayed a predicted active CRISPR system, both active and inactive forms of MgCas9 were present in house finch isolates collected over 12 years following the host shift, after which they all displayed either an inactivated MgCas9 or a loss of their CRISPR locus. Second, of the 490 unique spacer sequences detected across poultry and house finch isolates, only three were shared across both avian hosts, with poultry isolates displaying a greater diversity of sequences than house finch ones. Third, we found that the protein sequences of MgCas9 of poultry and house finch isolates clustered into two distinct phylogenetic branches, with very little variation detected among house finch Mgcas9 protein sequences. Within MgCas9, the PAM-interacting (PI) domain (involved in the specific interaction between Cas9 and the protospacer adjacent motif (PAM) that follows the target DNA sequence) and the Csn2 protein sequences were found to be most variable across poultry and house finch isolates. Fourth, we found that MgCas9 recognises a 5-letter motif in the PAM sequence. The motif recognised nevertheless differed between poultry and house finch MgCas9 and, accordingly, we measured variation in cleavage efficiencies *in vitro*, with the poultry MgCas9 more effective at cleaving its target PAM sequences than the house finch ones, and vice versa for the house finch MgCas9. Finally, our results confirmed that differences in the PAM specificity of MgCas9 are driven by the differences in its PI domain. Indeed, differences in specificities were replicated across a wild-type poultry strain (S6) and a genetically engineered mutant of that strain in which the PI domain of MgCas9 was replaced by the homologous domain of a house finch isolate (CA06), with the latter more effective at cleaving PAMs from house finch isolates. Together, these results suggest that significant evolutionary changes took place in *M. gallisepticum*'s CRISPR defence system following the jump into the novel house finch host, with an initial pattern of acquisition and loss of CRISPR spacers, and of changes in PAM specificity driven by the MgCas9 PI domain, consistent with exposure to a new community of phages and Mobile Genetic Elements (MGEs). That such initial functional changes were followed by subsequent inactivation of this defence system, however, highlights the complexity of the selection pressures that it faces.

Our study on the CRISPR-Cas9 system of *M. gallisepticum* isolates collected in both poultry and house finch hosts provides support for previous work, while also broadening our understanding of the complexity of this system across vertebrate host environments. First, mycoplasma CRISPR-Cas9 systems have previously been classified as Type II *in silico*, which include the typical set of *cas*, *cas1*, *cas2*, *cas9* and *csn2* genes (12). In support, we found that the CRISPR-Cas systems of *M. gallisepticum* poultry S6 and house finch CA06 isolates are typical of the Type II systems. (Note that our finding contrasts with the conclusions from a previous study, in which the endogenous *M. gallisepticum* CRISPR-Cas system and an *oriC* plasmid carrying different CRISPR arrays were used to introduce targeted mutations into the bacterial chromosome (31,32); see SI-3a for further discussion). Second, our results are also consistent with a previous study that undertook the first functional investigation of the Cas9 endonuclease in a mycoplasma system as part of a larger initiative to characterize 79 Cas9 orthologs representative of 10 major clades in a Cas9 evolutionary tree (18). In this study, the house finch isolate CA06 was selected as representative mycoplasma Cas9 with a 5-positions motif NNGAD as its consensus PAM (18). Our results provide confirmation for MgCas9 CA06's consensus PAM. Interestingly, however, we identified a different PAM motif (NNAAT) in the MgCas9 of the poultry strain S6, which was found to exhibit an ultra-dominant A at position 3 (99%) and a marked trend (73%) towards a T at position 5. In addition, we found that this difference in the PAM recognition preference of MgCas9 was associated with differences in both the composition and diversity of the spacers

between poultry and house finch isolates, suggestive of a complete reset of the CRISPR array following the host shift.

Maintaining an active adaptive immunity system like CRISPR-Cas is expected to depend on the balance of costs and benefits of doing so (4,33,34). Obviously, an active CRISPR-Cas will be beneficial if it confers protection against phages and MGEs. However, there can also be maintenance costs if harbouring an active CRISPR-Cas decreases fitness. This can occur, for instance, when CRISPR-Cas gives rise to metabolic costs (35,36) or genetic conflicts (35,36); when it is less effective at preventing infections than other defence processes (37); when it prevents the acquisition of fitness-enhancing genes through horizontal gene transfer (4,38–43); or when there are risks of self-targeting spacers (44). The balance between the costs and benefits of maintaining an active CRISPR-Cas will therefore depend on CRISPR-Cas' contribution to the bacterium's global defence process, on its non-defence roles, and on the specific selective pressures exerted by different host environments.

That there were isolates in circulation with an active CRISPR-Cas system in the 12 years following the jump into house finches indicates that there were benefits to the maintenance of this defence mechanism during that time period. The reasons for such benefits are likely to be found in the distinct PAM specificity of MgCas9 in house finch isolates, alongside striking patterns of loss and acquisition of new spacers elements, as spacers that are no longer in use are expected to be lost, while new ones that confer resistance against novel pathogenic targets should be gained. Indeed, such marked differences from poultry isolates are highly suggestive of a change in the community of pathogens (phages and MGEs) faced by *M. gallisepticum* in the novel finch host. This would be consistent with previous evidence of a rapid evolution of PAM specificities reported to occur in response to an escape from detection by phages and MGEs displaying mutations in their PAM motifs (45–47). Whether *M. gallisepticum* experienced novel pathogen-driven selection in house finches, and the role that this might have played in facilitating or hindering its emergence remain, however, to be determined. Regardless, our results give rise to the hypothesis that the successful emergence of a bacterial pathogen in a new host species will not only depend on its ability to establish and spread in the new host, but also on its ability to cope with a new landscape of phages and other invading DNAs.

The costs of maintaining an active CRISPR-Cas system for house finch isolates seems, nevertheless, to have outweighed benefits over time. Isolates displaying an inactive CRISPR-Cas system were also found in circulation over the 12 years following the jump into house finches and, by 2007, all the isolates collected either displayed an inactivated MgCas9 or the complete loss of their *cas* genes locus, leaving only the CRISPR array (Figure 5). Across bacteria, the switching off of the CRISPR systems has been proposed as a mechanism by which to respond to selection (48). Indeed, in *Escherichia coli*, for example, CRISPR-Cas is thought to restrict adaptive change by limiting the acquisition of potentially beneficial genes through horizontal transfer, giving rise to a trade-off between defence and adaptive potential (49). As a result, we might expect the balance of costs and benefits of maintaining an active CRISPR-Cas system to have shifted for *M. gallisepticum* in the face of a significant change in selection in the novel host.

One major selective event faced by *M. gallisepticum* after the jump into house finches was the evolution of host resistance to infection (50–52). The jump gave rise to an epidemic that spread quickly and is thought to have killed millions (13,53), with high mortality rates arising because, in house finches, *M. gallisepticum* colonises the mucosal surfaces of the conjunctiva and upper respiratory tract, causing a severe conjunctivitis that can lead to death in the wild through blindness-induced starvation or predation (54,55). The intensity of this selective event led to the spread of resistance in exposed house finch populations from standing genetic variation and within only 12 years of epidemic outbreak (14,50). As a result, by 2007, the proportion of resistant individuals in exposed house finch populations had risen to ~80% (50). Such host resistance was, in turn, found to have significant selective consequences for *M. gallisepticum*. Indeed, resistance favoured the evolution of increasing pathogen virulence through antagonistic coevolution (14), with isolates causing ever greater host mortality and

more severe signs of conjunctivitis over time (15), as well as transmitting faster to uninfected sentinels (56). The concordance in the timing of the complete inactivation of CRISPR-Cas (i.e., from 2007) and the need for evolutionary changes in virulence in response widespread house finch resistance is striking. Further work is now required to test whether significant phenotypic and genetic changes have, indeed, taken place following the loss of a functioning CRISPR-Cas, as predicted if this defence system hindered the necessary adaptive responses.

In conclusion, our study of the CRISPR defence system of bacterial isolates collected in its original host, as well as in a novel host into which it jumped recently, reveals marked differences between both in term of: (1) the MgCas9 PAM recognition specificity, (2) the composition and diversity of the spacer repertoire, and (3) the gradual inactivation of CRISPR-Cas over time in the novel host. While the changes in MgCas9 PAM specificity and in the composition and diversity of spacers point towards a change in the landscape of phages and MGEs faced by the bacterium in the novel host, the subsequent inactivation of CRISPR-Cas is consistent with another important selection event. We hypothesize that the gradual inactivation of the CRISPR-Cas defence system, which took place over the first 12 years of the epidemic until this defence system became fully inactivated from 2007, occurred in response to the evolution of host resistance, which became widespread by 2007 and pushed for adaptive changes in the bacterium. Such adaptive changes would have necessitated the loss of a functional CRISPR defence system to take place. Together, our results highlight the need to consider not only a bacterium's pathogenic landscape, but also the host-driven selection pressures it experiences, for understanding the evolution of the CRISPR defence system.

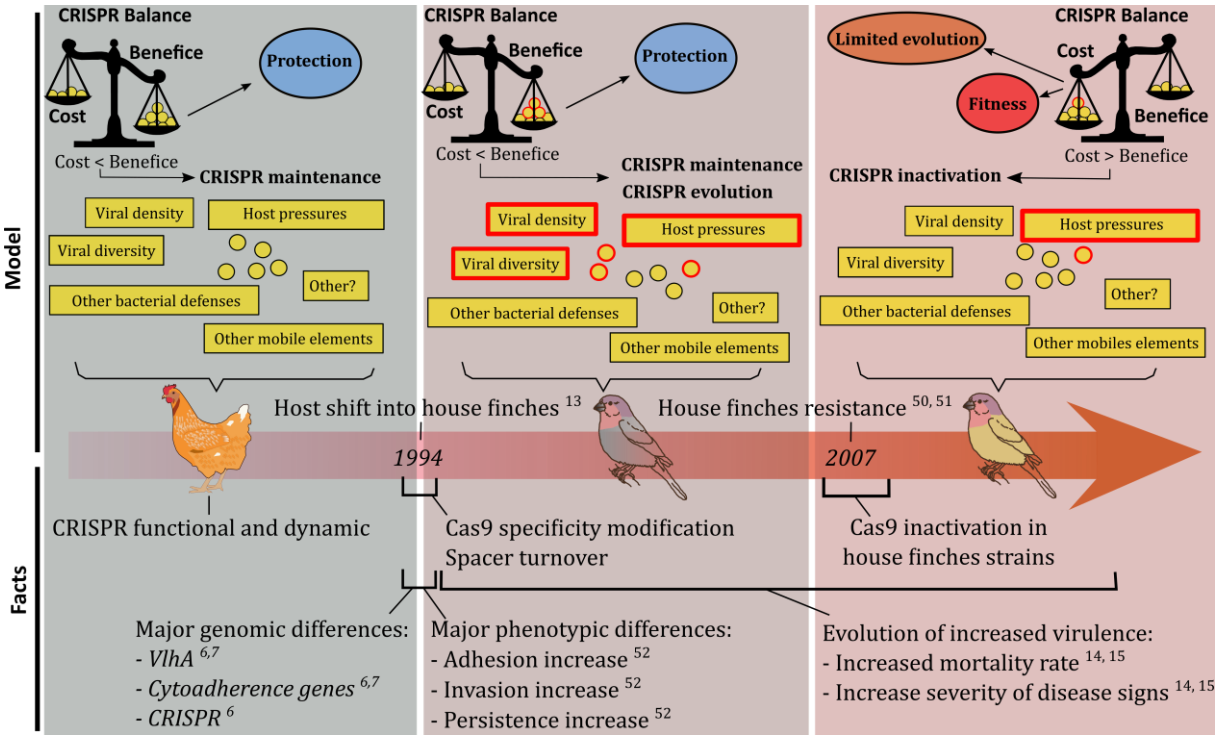


Figure 5. Hypothetical model of evolution of the CRISPR system in *M. gallisepticum* as it jumped from poultry and emerged into the novel house finch host. Phenotypic and genetic changes accompanying each step are obtained from the literature.

Acknowledgements

We are very thankful to Dr. Gasiunas from Caszyme for sharing the 7N PAM-degenerated pTZ57-7n library, the pET28ID116 plasmid encoding *M. gallisepticum* CA06 Cas9 and for appreciated advices on the protocol used for *in vitro* cleavage. We also want to thank Patrice Mora for helping in the constitution of the *M. gallisepticum* spacers database, Mammadou Sall for his contribution to the study of spacers and Alain Blanchard for much appreciated discussions on the work and manuscript.

References

1. Barrangou R, Fremaux C, Deveau H, Richardss M, Boyaval P, Moineau S, et al. CRISPR Provides Against Viruses in Prokaryotes. Science (80-). 2007;315(March):1709–12.
2. Koonin E V. CRISPR: a new principle of genome engineering linked to conceptual shifts in evolutionary biology. Biol Philos [Internet]. 2019;34(1). Available from: https://doi.org/10.1007/s10539-018-9658-7
3. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020;18(2):67–83.
4. Dimitriu T, Szczelkun MD, Westra ER. Evolutionary ecology of prokaryotic innate and adaptive immune systems and their interplay. Curr Biol. 2020;
5. Bonneaud C, Weinert LA, Kuijper B. Understanding the emergence of bacterial pathogens in novel hosts. Philos Trans R Soc B Biol Sci [Internet]. 2019 Sep 30;374(1782):20180328. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0328
6. Delaney NF, Balenger S, Bonneaud C, Marx CJ, Hill GE, Ferguson-Noel N, et al. Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. PLoS Genet. 2012;8(2).
7. Tulman ER, Liao X, Szczepanek SM, Ley DH, Kutish GF, Geary SJ. 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*. Microbiol (United Kingdom). 2012;158(8):2073–88.
8. May M, Balish MF, Blanchard A. The Order Mycoplasmatales. In: The Prokaryotes [Internet]. Springer Berlin Heidelberg; 2014. p. 515–50. Available from: http://link.springer.com/10.1007/978-3-642-30120-9_289
9. Sirand-Pugnet P, Citti C, Barré A, Blanchard A. Evolution of mollicutes: down a bumpy road with twists and turns. Res Microbiol [Internet]. 2007 Dec;158(10):754–66. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0923250807001751
10. Citti C, Dordet-Frisoni E, Nouvel LX, Kuo CH, Baranowski E. Horizontal Gene Transfers in Mycoplasmas (Mollicutes). Curr Issues Mol Biol [Internet]. 2018;29:3–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29648541
11. Oliveira PH, Touchon M, Rocha EPC. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. 2014;42(16):10618–32.
12. Ipoutcha T, Tsarmpopoulos I, Talenton V, Gaspin C, Moisan A, Walker CA, et al. Multiple Origins and Specific Evolution of CRISPR / Cas9 Systems in Minimal Bacteria (Mollicutes). Front Microbiol [Internet]. 2019;10(November):1–14. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02701/full
13. Dhondt AA, Tessaglia DL, Slothower RL. Epidemic From Conjunctivitis America in House Finches. J Wildl Dis. 1998;34(2):265–80.
14. Bonneaud C, Giraudeau M, Tardy L, Staley M, Hill GE, McGraw KJ. Rapid Antagonistic Coevolution in an Emerging Pathogen and Its Vertebrate Host. Curr Biol [Internet]. 2018;28(18):2978-2983.e5. Available from: https://doi.org/10.1016/j.cub.2018.07.003
15. Tardy L, Giraudeau M, Hill GE, McGraw KJ, Bonneaud C. Contrasting evolution of virulence and replication rate in an emerging bacterial pathogen. Proc Natl Acad Sci U S A. 2019;116(34):16927–32.
16. Freundt EA. CULTURE MEDIA FOR CLASSIC MYCOPLASMAS. In: Methods in Mycoplasmology. 1983.
17. Ipoutcha T, Gourgues G, Lartigue C, Blanchard A, Sirand-Pugnet P. Genome Engineering in Mycoplasma gallisepticum Using Exogenous Recombination Systems. ACS Synth Biol [Internet]. 2022 Feb 15;acssynbio.1c00541. Available from: https://pubs.acs.org/doi/10.1021/acssynbio.1c00541
18. Gasiunas G, Young JK, Karvelis T, Kazlauskas D, Urbaitis T, Jasnauskaite M, et al. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. Nat Commun. 2020;
19. Karvelis T, Young JK, Siksnys V. A pipeline for characterization of novel Cas9 orthologs [Internet]. 1st ed. Vol. 616, Methods in Enzymology. Elsevier Inc.; 2019. 219–240 p. Available from: http://dx.doi.org/10.1016/bs.mie.2018.10.021
20. Song Y, Kim T, La T-M, Lee H-J, Lee Y, Shin H, et al. Complete Genome Sequence of *Mycoplasma gallisepticum* Strain KUVMG001, an Isolate from South Korea. Microbiol Resour Announc [Internet]. 2021 May 13;10(19). Available from: http://www.ncbi.nlm.nih.gov/pubmed/33986095
21. Leigh SA, Evans JD, Branton SL. Complete Genome Sequences of Two *Mycoplasma gallisepticum* F-Strain Variants. Microbiol Resour Announc [Internet]. 2019 Aug 15;8(33). Available from: http://www.ncbi.nlm.nih.gov/pubmed/31416859
22. Szczepanek SM, Tulman ER, Gorton TS, Liao X, Lu Z, Zinski J, et al. Comparative genomic analyses of attenuated strains of Mycoplasma gallisepticum. Infect Immun. 2010;78(4):1760–71.
23. Papazisi L, Gorton TS, Kutish G, Markham PF, Browning GF, Nguyen DK, et al. The complete genome sequence of the avian pathogen Mycoplasma gallisepticum strain Rlow. Microbiology. 2003;
24. Leigh SA, Evans JD, Branton SL. Complete Genome Sequences of Two Vaccine Strains and One Field Isolate of *Mycoplasma gallisepticum*. Microbiol Resour Announc [Internet]. 2019 Dec 5;8(49). Available from: http://www.ncbi.nlm.nih.gov/pubmed/31806746
25. Vink JNA, Baijens JHL, Brouns SJJ. Comprehensive PAM prediction for CRISPR-Cas systems reveals evidence for spacer sharing, preferred strand targeting and conserved links with CRISPR repeats. bioRxiv [Internet]. 2021 Jan 1;2021.05.04.442622. Available from: http://biorxiv.org/content/early/2021/05/06/2021.05.04.442622.abstract
26. Calcutt MJ, Lewis MS, Wise KS. Molecular genetic analysis of ICEF, an integrative conjugal element that is present as a repetitive sequence in the chromosome of *Mycoplasma fermentans* PG18. J Bacteriol. 2002;184(24):6929–41.
27. Dordet Frisoni E, Marenda MS, Sagné E, Nouvel LX, Guérillot R, Glaser P, et al. ICEA of *Mycoplasma agalactiae*: A new family of self-

transmissible integrative elements that confers conjugative properties to the recipient strain. Mol Microbiol. 2013;89(6):1226–39.
28. Meygret A, Peuchant O, Dordet-Frisoni E, Sirand-Pugnet P, Citti C, Bébéar C, et al. High Prevalence of Integrative and Conjugative Elements Encoding Transcription Activator-Like Effector Repeats in Mycoplasma hominis. Front Microbiol [Internet]. 2019 Oct 18;10. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2019.02385/full
29. Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature. 2015;519(7542):199–202.
30. Wei Y, Terns RM, Terns MP. Cas9 function and host genome sampling in Type II-A CRISPR – Cas adaptation. Genes (Basel). 2015;356–61.
31. Klose SM, Wawegama N, Sansom FM, Marenda MS, Browning GF. Efficient disruption of the function of the mnuA nuclease gene using the endogenous CRISPR/Cas system in *Mycoplasma gallisepticum*. Vet Microbiol [Internet]. 2022 Jun;269:109436. Available from: http://www.ncbi.nlm.nih.gov/pubmed/35487018
32. Mahdizadeh S, Sansom FM, Lee S-W, Browning GF, Marenda MS. Targeted mutagenesis of *Mycoplasma gallisepticum* using its endogenous CRISPR/Cas system. Vet Microbiol. 2020 Nov;250:108868.
33. Koonin E V. CRISPR: a new principle of genome engineering linked to conceptual shifts in evolutionary biology. Biol Philos [Internet]. 2019 Feb 19;34(1):9. Available from: http://link.springer.com/10.1007/s10539-018-9658-7
34. Westra ER, Levin BR. It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements. Proc Natl Acad Sci U S A. 2020;117(45):27777–85.
35. Vale PF, Lafforgue G, Gatchitch F, Gardan R, Moineau S, Gandon S. Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. Proc R Soc B Biol Sci [Internet]. 2015 Aug 7;282(1812):20151270. Available from: https://royalsocietypublishing.org/doi/10.1098/rspb.2015.1270
36. Hall JPJ, Wright RCT, Harrison E, Muddiman KJ, Wood AJ, Paterson S, et al. Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation. de Visser JAGM, editor. PLOS Biol [Internet]. 2021 Oct 13;19(10):e3001225. Available from: https://dx.plos.org/10.1371/journal.pbio.3001225
37. Broniewski JM, Meaden S, Paterson S, Buckling A, Westra ER. The effect of phage genetic diversity on bacterial resistance evolution. ISME J [Internet]. 2020 Mar 2;14(3):828–36. Available from: http://www.nature.com/articles/s41396-019-0577-7
38. Pursey E, Dimitriu T, Paganelli FL, Westra ER, van Houte S. CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens. Philos Trans R Soc B Biol Sci [Internet]. 2022 Jan 17;377(1842). Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0464
39. Levin BR. Nasty Viruses, Costly Plasmids, Population Dynamics, and the Conditions for Establishing and Maintaining CRISPR-Mediated Adaptive Immunity in Bacteria. Guttman DS, editor. PLoS Genet [Internet]. 2010 Oct 28;6(10):e1001171. Available from: https://dx.plos.org/10.1371/journal.pgen.1001171
40. Wheatley RM, MacLean RC. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. ISME J [Internet]. 2021 May 21;15(5):1420–33. Available from: http://www.nature.com/articles/s41396-020-00860-3
41. Mackow NA, Shen J, Adnan M, Khan AS, Fries BC, Diago-Navarro E. CRISPR-Cas influences the acquisition of antibiotic resistance in *Klebsiella pneumoniae*. Kuo C-H, editor. PLoS One [Internet]. 2019 Nov 20;14(11):e0225131. Available from: https://dx.plos.org/10.1371/journal.pone.0225131
42. Palmer KL, Gilmore MS. Multidrug-Resistant Enterococci Lack CRISPR- cas. Losick R, editor. MBio [Internet]. 2010 Oct 29;1(4). Available from: https://journals.asm.org/doi/10.1128/mBio.00227-10
43. Alduhaidhawi AHM, AlHuchaimi SN, Al- Mayah TA, Al-Ouqaili MT, Alkafaas SS, Muthupandian S, et al. Prevalence of CRISPR-Cas Systems and Their Possible Association with Antibiotic Resistance in *Enterococcus faecalis* and *Enterococcus faecium* Collected from Hospital Wastewater. Infect Drug Resist [Internet]. 2022 Mar;Volume 15:1143–54. Available from: https://www.dovepress.com/prevalence-of-crispr-cas-systems-and-their-possible-association-with-a-peer-reviewed-fulltext-article-IDR
44. Wimmer F, Beisel CL. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. Front Microbiol. 2020;10(January):1–17.
45. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol. 2008;190(4):1390–400.
46. Paez-Espino D, Sharon I, Morovic W, Stahl B, Thomas BC, Barrangou R, et al. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. MBio. 2015;6(2):1–9.
47. Edraki A, Mir A, Ibraheim R, Gainetdinov I, Yoon Y, Song CQ, et al. A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. Mol Cell. 2019;73(4):714-726.e4.
48. Patterson AG, Yevstigneyeva MS, Fineran PC. Regulation of CRISPR–Cas adaptive immune systems. Curr Opin Microbiol [Internet]. 2017 Jun;37:1–7. Available from: http://dx.doi.org/10.1016/j.mib.2017.02.004
49. García-Martínez J, Maldonado RD, Guzmán NM, Mojica FJM. The CRISPR conundrum: evolve and maybe die, or survive and risk stagnation. Microb Cell [Internet]. 2018 Jun 4;5(6):262–8. Available from: http://microbialcell.com/researcharticles/the-crispr-conundrum-evolve-and-maybe-die-or-survive-and-risk-stagnation/
50. Bonneaud C, Balenger SL, Russell AF, Zhang J, Hill GE, Edwards S V. Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. Proc Natl Acad Sci U S A. 2011;108(19):7866–71.
51. Bonneaud C, Tardy L, Giraudeau M, Hill GE, McGraw KJ, Wilson AJ. Evolution of both host resistance and tolerance to an emerging bacterial pathogen. Evol Lett [Internet]. 2019 Oct;3(5):544–54. Available from: https://academic.oup.com/evlett/article/3/5/544/6697468
52. Dowling A, Hill GE, Bonneaud C. Multiple differences in pathogen- host cell interactions following a bacterial host shift. Sci Rep. 2020;1–12.
53. Nolan PM, Hill GE, Stoehr AM. Sex, size, and plumage redness predict house finch survival in an epidemic. Proc R Soc B Biol Sci. 1998;265(1400):961–5.
54. Adelman JS, Mayer C, Hawley DM. Infection reduces anti-predator behaviors in house finches. J Avian Biol [Internet]. 2017 Apr;48(4):519–28. Available from: https://onlinelibrary.wiley.com/doi/10.1111/jav.01058
55. Roberts SR, Nolan PM, Hill GE. Characterization of *mycoplasma gallisepticum* infection in captive house finches (Carpodacus mexicanus) in 1998. Avian Dis [Internet]. 2001;45(1):70–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11332501
56. Bonneaud C, Tardy L, Hill GE, McGraw KJ, Wilson AJ, Giraudeau M. Experimental evidence for stabilizing selection on virulence in a bacterial pathogen. Evol Lett [Internet]. 2020 Dec 11;4(6):491–501. Available from: https://academic.oup.com/evlett/article/4/6/491/6697513