Inhibition of NHEJ repair by type II-A CRISPR-Cas systems

Aude Bernheim[1,2,3,4], Alicia Calvo Villamanan[3], Clovis Basier[3], Eduardo PC Rocha[1,2], Marie Touchon[1,2,*], David Bikard[3,*]

[1] Microbial Evolutionary Genomics, Institut Pasteur, 25-28 rue Dr Roux, Paris, 75015, France
[2] CNRS, UMR3525, 25-28 rue Dr. Roux, Paris, 75015, France
[3] Synthetic Biology Group, Institut Pasteur, 25-28 rue Dr. Roux, Paris, 75015, France
[4] AgroParisTech, F-75005 Paris, France

* co-senior authors

Abstract

CRISPR-Cas systems introduce double strand breaks into DNA of invading genetic material and use DNA fragments to acquire novel spacers during adaptation. Double strand breaks are the substrate of several bacterial DNA repair pathways, paving the way for interactions between them and CRISPR-Cas systems. Here, we hypothesized that non-homologous end joining (NHEJ) interferes with type II CRISPR-Cas systems. We tested this idea by studying the patterns of co-occurrence of the two systems in bacterial genomes. We found that NHEJ and type II-A CRISPR-Cas systems only co-occur once among 5563 fully sequenced prokaryotic genomes. We investigated experimentally the possible molecular interactions causing this negative association using the NHEJ pathway from *Bacillus subtilis* and the type II-A CRISPR-Cas systems from *Streptococcus thermophilus* and *Streptococcus pyogenes*. Our results suggest that the NHEJ system has no effect on type II-A CRISPR-Cas interference and adaptation. On the other hand, we provide evidence for the inhibition of NHEJ repair by the Csn2 protein from type II-A CRISPR-Cas system. Our findings give insights on the complex interactions between CRISPR-Cas systems and repair mechanisms in bacteria and contribute to explain the scattered distribution of CRISPR-Cas systems in bacterial genomes.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays and their associated (Cas) proteins confer Bacteria and Archaea adaptive immunity against phages and other exogenous mobile genetic elements [1,2]. Yet, even if most bacteria are infected by phages and other mobile genetic elements, CRISPR-Cas systems are absent from the majority of bacterial genomes[3,4]. The selective pressures and mechanisms that lead to the success of CRISPR-Cas systems in some clades and not others remains poorly understood.

CRISPR-Cas systems are classified in six types and twenty-seven subtypes, according to the Cas proteins they carry [3,5]. The recent development of CRISPR-Cas9-based genetic engineering technologies has made type II CRISPR-Cas systems the focus of many investigations. Type II systems include the CRISPR repeat-spacer array, three core genes (*cas1*, *cas2* and *cas9*), and a small trans-activating CRISPR RNA (tracrRNA) complementary to the CRISPR repeat sequence [6,7]. A fourth gene is involved in spacer acquisition, *csn2* in the type II-A [8–11], and *cas4* in type II-B systems [6]. A third subtype, type II-C, only requires *cas1*, *cas2* and *cas9* [3,6]. All the Cas proteins of type II systems are necessary for spacer acquisition [12,13], but only Cas9 is necessary for interference [14,15]. The Cas9 protein is guided by small CRISPR RNA (crRNA) to introduce double strand breaks (DSB) into target DNA[14,16]. A short conserved sequence (2-5bp) adjacent to the protospacer known as the PAM (protospacer Adjacent Motif) is essential to distinguish foreign from self DNA and can be different for CRISPR-Cas systems of the same type [17,18].

In bacteria, DSB can be repaired either by Homologous Recombination (HR) or by Non-Homologous End Joining (NHEJ). These mechanisms could thus affect the efficiency of CRISPR-Cas interference by repairing the breaks. Type II CRISPR-Cas systems introduce DSB at the same position in all copies of the target DNA molecule [19], and the concomitant lack of an intact DNA template should preclude the repair of these DSB by HR. However, NHEJ repairs DSB without requiring template DNA[20] and could mend DSB generated by Cas9. In Eukaryotic cells, breaks introduced by Cas9 can efficiently be repaired by NHEJ, a strategy now widely used to introduce indel mutations [21]. In bacteria, the NHEJ system requires two core proteins: Ku and a ligase [22]. Ligation is usually carried out by the LigD protein, but other ligases can be recruited by Ku when LigD is absent [20]. The system is complemented by additional proteins in certain cases [23]. Ku binds at the DSB and recruits the ligase to seal the break [24,25]. NHEJ offers a mean to repair DSB when only a single copy of the genome is available, such as after sporulation or during stationary phase [26,27]. NHEJ repair can be mutagenic [28], leading to up to 50% error rates in certain bacteria [25].

DNA repair pathways could also affect the acquisition of novel spacers by CRISPR-Cas systems because they modulate the availability of DSB and/or compete with the Cas machinery for the DNA substrate. Conversely, the action of Cas proteins at DSB could hinder DNA repair pathways. It was shown that novel spacers of type I CRISPR-Cas systems can be acquired after

DSB from RecBCD degradation products [29]. Importantly DNA repair pathways and CRISPR-Cas systems are composed of proteins with structural similarities and interacting with the same substrates [9]. For example, Cas4, a protein present in type I and type II-B systems shares structural and functional similarities with AddB[9,30] , a component of the AddAB repair pathway and a functional homolog of RecBCD[31]. In type II-A CRISPR-Cas systems, Csn2 binds and slides along free DNA ends in the same manner as the Ku protein of the NHEJ system [9]. Csn2 has been shown to be mandatory to acquire new spacers [8,12,13]. If Cas proteins and proteins involved in DNA repair mechanisms recognize the same substrate, a competition might arise leading to antagonistic interactions between the two processes.

The interaction between the NHEJ system and Cas9 is at the heart of the CRISPR-Cas based genetic engineering technologies, and we now search to understand it in bacteria. We hypothesized that the NHEJ system could interfere with the activities of type II CRISPR-Cas systems by repairing DSB generated by Cas9 during interference or by competing with Cas proteins for the same substrate during adaptation. Alternatively, type II CRISPR-Cas systems could interfere with NHEJ during repair. We tested these hypotheses by assessing the patterns of co-occurrence of the two systems in bacterial genomes. This revealed one single case of co-occurrence of both systems among 5563 bacterial genomes, suggesting strong negative epistasis. We then studied experimentally the causes of this negative interaction, by introducing the NHEJ system from *B. subtilis* and/or the CRISPR-Cas system from *S. pyogenes* in *B. subtilis*, *S. thermophilus* and *S. aureus*.

Results

Negative association between NHEJ system and type II-A CRISPR-Cas systems

We detected CRISPR-Cas and NHEJ systems in 5563 fully sequenced bacterial genomes (Supplementary Table 1). The NHEJ pathway was present in 24.7% and the type II CRISPR-Cas system in 6.9% of the genomes, and these systems were very unevenly distributed among bacterial phyla (Supplementary Figure 1 and Supplementary Table 2). Firmicutes and Proteobacteria were the only phyla with genomes encoding enough type II CRISPR-Cas systems (resp. 209 and 101) and NHEJ (resp. 364 and 637), to perform robust statistical analyses (Supplementary Figure 1). A possible confounding factor when studying the distribution of bacterial defense and DNA repair pathways is that there abundance co-vary with genome size [32,33]. Accordingly, NHEJ systems were more frequent in larger genomes ($P<10^{-4}$, χ^2 test on a logistic fit). In contrast, type II CRISPR-Cas systems were only present in genomes smaller than 5Mb (Supplementary Figure 2). Hence, we focused our analysis on Firmicutes and Proteobacteria with genomes smaller than 5Mb. They represent 56.5% of the total number of genomes. In this sample, the size of the genomes with the NHEJ system was independent of the presence of a type II CRISPR-Cas system (P=0.99, Wilcoxon test).

We analyzed the patterns of co-occurrence of NHEJ and CRISPR-Cas systems to test if they were independently distributed. We observed that NHEJ and type II systems were negatively associated in Firmicutes ($P<10^{-4}$, Fisher Exact Test), but not in Proteobacteria (P=0.70, Fisher Exact Test) (Figure 1.b and Supplementary Figure 3). Note however that different subtypes of type II CRISPR-Cas systems are distributed differently in these two phyla. Proteobacteria encoded many type II-C and no type II-A systems, whereas Firmicutes encoded mostly type II-A systems (Figure 1.a). Type II-B systems were only detected in 9 genomes and will not be analyzed any further. To test if different subtypes could have different interactions with NHEJ systems, we looked at them separately. When studying co-occurrences of genes, it is important to consider that genomes are linked by a common evolutionary history, which decreases the degrees of the freedom of the statistical analyses. To check whether systems are negatively associated while taking phylogeny into account, we built a tree of Firmicutes and tested if the binary traits (presence of both systems) evolved independently using BayesTraits[34]. A strong negative association between NHEJ and type II-A CRISPR-Cas systems was observed (Bayes Factor BF=9.7, Figure 1.c), while no associations between NHEJ and type II-C CRISPR-Cas systems was detected. Only one genome among the 5563 encodes both NHEJ and type II-A: the actinobacteria *Eggerthella sp. YY7918*. In this genome, both NHEJ and type II-A systems seem intact, since the *cas* operon contains all four genes, lacking frameshifts or premature stop codons, and the adjacent CRISPR array encodes 44 spacers.

Taken together, these results show a strong negative association between NHEJ and type II-A CRISPR-Cas systems that is independent of the phylogenetic structure of the data. This negative

association suggests the existence of a negative interaction between these systems in the bacterial cell. We devised three hypotheses to explain this negative association: 1) NHEJ impairs type II-A CRISPR-Cas interference, 2) NHEJ impairs type II-A CRISPR-Cas adaptation, *i.e.*, the ability of the system to acquire new spacers 3) type II-A CRISPR-Cas impairs NHEJ.

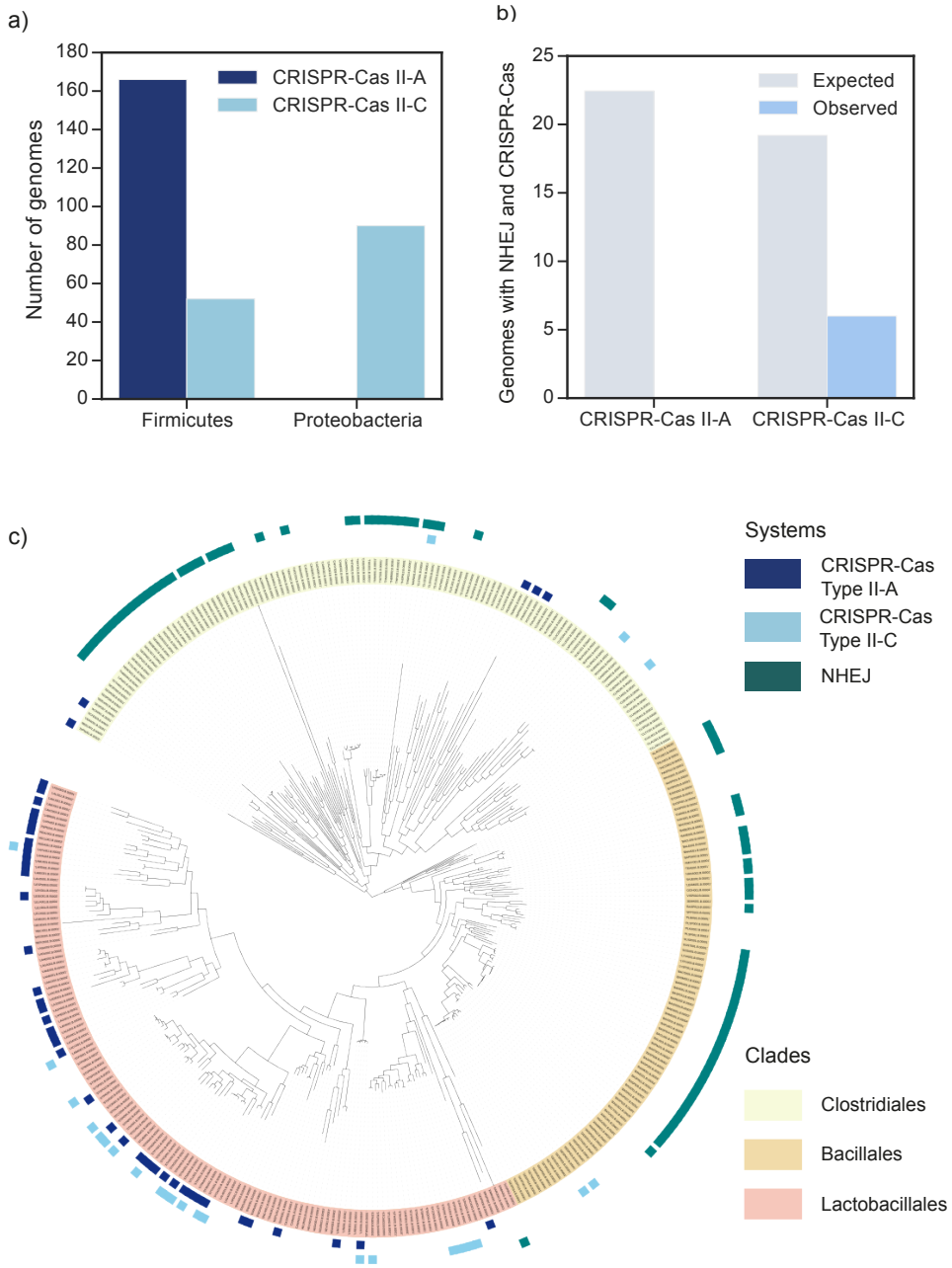

Figure 1: Negative association between NHEJ and type II-A CRISPR-Cas systems.
a, Distribution of the subtypes II-A and II-C in Proteobacteria and Firmicutes genomes. **b,** Associations between NHEJ and subtypes II-A and II-C CRISPR-Cas systems. Expected values correspond to the number of co-occurrences that would be obtained if the systems were randomly distributed. **c,** Presence of NHEJ and type II CRISPR-Cas systems in Firmicutes. A system is annotated as present in a given species when more than half of the genomes available for this species encode the system.

NHEJ system does not impact type II-A CRISPR-Cas interference

We first tested if the *B. subtilis* NHEJ system could affect type II-A CRISPR-Cas interference, using the previously described *S. aureus* model system[12]. The *ku* and *ligD* genes were cloned under the control of a Ptet promoter (plasmid pAB1) into *S. aureus* RN4220 cells. This system was able to circularize linearized plasmids after electroporation, showing it is functional (Supplementary Text 1, Supplementary Figure 4). The type II-A CRISPR-Cas system from *S. pyogenes* was introduced on plasmid pDB114 and programmed with a single spacer targeting phage phiNM4 (pMD021). *S. aureus* cells carrying both systems were then challenged in phage infection assays. A NHEJ system might facilitate phage escape from CRISPR-Cas by promoting the introduction of mutations at the target site through unfaithful repair, or by efficiently and faithful repairing DSB generated by Cas9, making CRISPR immunity inefficient.

First, the unfaithful repair of Cas9 breaks could lead to the formation of indels that would block further cleavages. The generation of such mutant phages should lead to a higher efficiency of plaquing of phiNM4 when the NHEJ system is expressed. The CRISPR-Cas system provided a five order of magnitude reduction in the E.O.P. of phage phiNM4 when compared with a spacer-less control, and no significant increase in the number of plaques was observed upon NHEJ induction (Figure 2.a). To confirm that the small number of plaques obtained could not result from the unfaithful repair of Cas9 breaks through NHEJ, we sequenced the target position of 8 mutant phages. All mutants had a point mutation in the PAM and none presented an indel.

Second, the faithful repair of Cas9 breaks could lead to a cycle of repair and cleavage that would allow the production of functional phage particles. In this case it might not be possible to observe plaque formation as the competition between NHEJ and CRISPR interference might lower burst sizes. To test this hypothesis, we measured the efficiency of center of infection (E.C.O.I), *i.e.*, the number of cells that produce at least one functional phage particle after infection compared to the control strain (sensitive to the phage). One would expect higher E.C.O.I of phiNM4 when cells express the NHEJ system. The observed E.C.O.I was ~10^{-2} regardless whether the NHEJ system was induced or not (Figure 2.b).

We further tested whether NHEJ could reduce CRISPR-Cas9 immunity against plasmids. To this end, we cloned the PhiNM4 target sequence used above on plasmid pAB2 and transformed this plasmid in strains carrying the NHEJ system or not. While a control target-less plasmid could be efficiently introduced in the cells, no clones were recovered after transformation of pAB2 regardless of the presence of the NHEJ machinery. This shows that the CRISPR-Cas system efficiently blocks plasmid transformation and that the NHEJ system did not measurably reduce the efficiency of CRISPR immunity, nor introduced mutations in the target plasmid at a detectable rate (Figure 2.c).

To confirm these results in a bacterium that naturally carries a type II-A CRISPR-Cas system, we measured interference against phage Phi2972 in *S. thermophilus*, in the presence or absence of the NHEJ system from *B. subtilis*. Genes *ku* and *ligD* were cloned under the control of a constitutive promoter on plasmid pNZ123 and introduced in a derivative of strain DGCC7710 whose CRISPR1 locus carries a spacer targeting phage Phi2972. The resistance provided by the CRISPR-Cas system was as strong in the presence of the NHEJ system as in the presence of a control GFP carried by the same plasmid (Figure 2.d). All in all, our results do not support the hypothesis that NHEJ affects type II-A CRISPR-Cas interference.

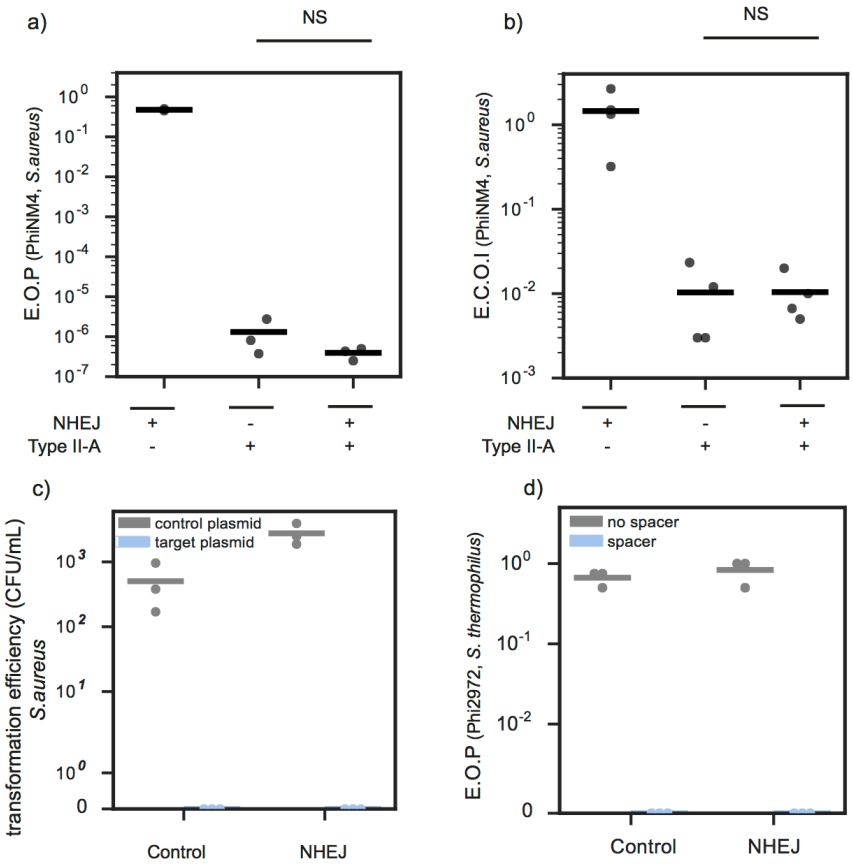

Figure 2: NHEJ system has no effect on type II-A CRISPR-Cas interference

a, Resistance to phage phiNM4 provided by the *S. pyogenes* CRISPR-Cas9 system in *S. aureus* in the presence (pAB1) or absence (pE194) of the NHEJ system from *B. subtilis* (n=3, mean, NS double sided t-test P=0.9999). **b,** Efficiency of Center Of Infection (E.C.O.I), *i.e.*, the proportion of cells that produce at least one functional phage particle, in the presence (pAB1) of absence (pE194) of the NHEJ system (n=4, mean, NS double sided t-test P=0.9998). **c,** Transformation efficiency of plasmid pT181 either empty or carrying a target sequence (pAB2) in *S. aureus* RN4220 cells expressing the CRISPR-Cas system from plasmid pMD021 in the presence (pAB1) of absence (pE194) of the NHEJ system from *B. subtilis* (n=3, mean). **d,** E.O.P. of phage Phi2972 on a bacteriophage insensitive mutant of *S. thermophilus* DGCC7710 carrying a spacer against Phi2972. Cells express either the *B. subtilis* NHEJ system from plasmid pAB66 or a control GFP from plasmid pAB69. (n=3, mean).

B. subtilis* NHEJ machinery does not prevent spacer acquisition in *S. aureus* and in *S. thermophilus

Ku and Csn2 bind the same type of substrate - linear double stranded DNA [9]- and might thus interfere antagonistically. To test if the NHEJ system affects spacer acquisition, we measured the cells' ability to acquire new spacers in presence of the NHEJ machinery. *S. aureus* cells carrying the NHEJ system and the type II-A CRISPR-Cas system were infected by phage PhiNM4 either with or without induction of the NHEJ system[12]. In this experiment, cells can escape phage infection either by capturing a novel spacer or by using other mechanisms of defense. Survivors were screened by PCR to check for acquisition of novel spacers and measure adaptation rate (Figure 3.a). No effect of the NHEJ system on the adaptation rate was observed. As a control the expression of Ku alone, ligD alone or GFP were also observed to have no effect (ANOVA, P=0,16) (Figure 3.b).

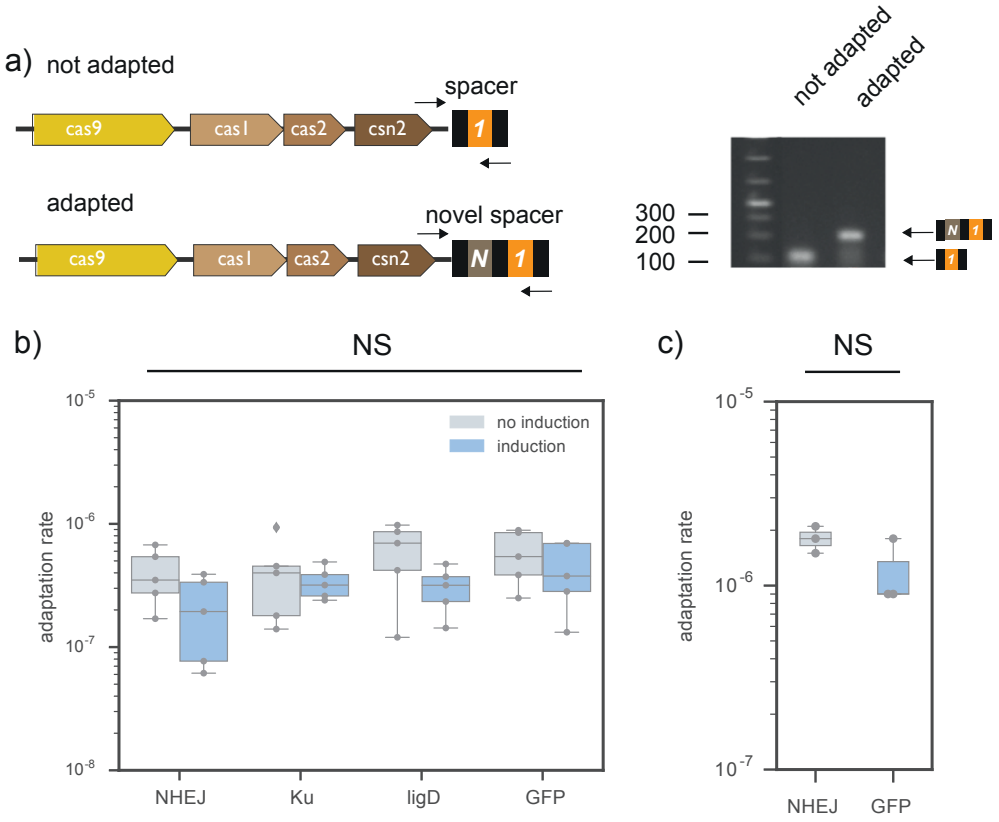

Figure 3 : NHEJ system does not impact adaptation of type II-A CRISPR-Cas system.

a, *S. aureus strain* RN4220 carrying the type II-A CRISPR-Cas system on plasmid pRH87 was challenged with phage phiNM4. Spacer acquisition was assessed by PCR on isolated colonies that survived the infection (oligonucleotides are depicted as black arrows). **b** Adaptation rate measured in the presence of NHEJ, ku, ligD or GFP carried by plasmids pAB23, pAB24, pAB25 and pAB62 respectively (n=5, ANOVA, NS P=0.5674). **c,** Adaptation rate of *S. thermophilus* DGCC7710 against phage Phi2972 when expressing the *B. subtilis* NHEJ system from plasmid pAB66 or a control GFP from plasmid pAB69 (n=3, two sided t test, NS P=0.91).

To corroborate these results, a similar experiment was performed in *S. thermophilus*. Cells carrying the *B. subtilis* NHEJ system or a control GFP on a plasmid were infected with phage Phi2972. We observed no difference the rate of novel spacer acquisition between cells expressing the NHEJ machinery or the GFP (Wilcoxon test, P=0.26) (Figure 3.c). Altogether these results indicate that NHEJ has no effect on the acquisition of novel spacers by a type II-A CRISPR-Cas system.

Csn2 inhibits NHEJ repair

As Csn2 binds to the same substrate as Ku, it could interfere with NHEJ repair[9,10,35]. To test this hypothesis, we reproduced the experiment that led to the discovery of the NHEJ system in *B. subtilis* [36]. When *B. subtilis* cells in stationary phase are irradiated by ionizing radiations (IR), the DSB generated are repaired by the NHEJ system, as other repair systems cannot function in those specific conditions. *B. subtilis* deleted for NHEJ do not survive irradiation as well as the wild-type. If type IIA CRISPR-Cas systems limit NHEJ repair, cells bearing a type IIA CRISPR-Cas system are expected to show increased sensitivity to irradiation.

B. subtilis cells expressing the type IIA CRISPR-Cas system from plasmid pRH087 were more sensitive to irradiation than cells carrying a control empty vector and showed the same level of sensitivity as the Δku-ligD mutant ($P<10^{-4}$, Wilcoxon, Figure 4.a). If the increased sensitivity provided by the CRISPR-Cas system is due to an impairment of NHEJ repair, then we expect to observe no cumulative effects when the NHEJ system is deleted and the CRISPR-Cas system added. Indeed, cells deleted for the NHEJ system and carrying the type II-A CRISPR-Cas system have the same survival as the ones deleted for the NHEJ system, pointing towards a interaction between the two systems. Another prediction that results from this hypothesis is that the CRISPR-Cas system should have no effect on the sensitivity to irradiation in species that lack a NHEJ system. To test this, we performed irradiation experiments on *S. aureus* cells carrying plasmid pRH087 or the control pC194. The presence or absence of the CRISPR-Cas system did not have an effect on survival in *S. aureus* (P=0.5, Wilcoxon, Supplementary Figure 5). Taken together, these results support the hypothesis that the type II-A CRISPR-Cas system impairs the NHEJ system.

To understand if a specific protein was responsible for this phenotype, we deleted or mutated individual *cas* genes from plasmid pRH87 and performed the same assay. While the effect size is small, the only mutant that significantly rescued *B. subtilis* cells upon irradiation was the delta *csn2* mutant (P=0.02, Student two sided t-test after validation of normality and homoscedasticity, Figure 4.b). When expressed alone, Csn2 was able to decrease survival of irradiated cells to the same level as the whole CRISPR-Cas system, while no effect could be observed with an empty vector or Cas9 alone ($P<10^{-4}$, Wilcoxon, Figure 4.c).

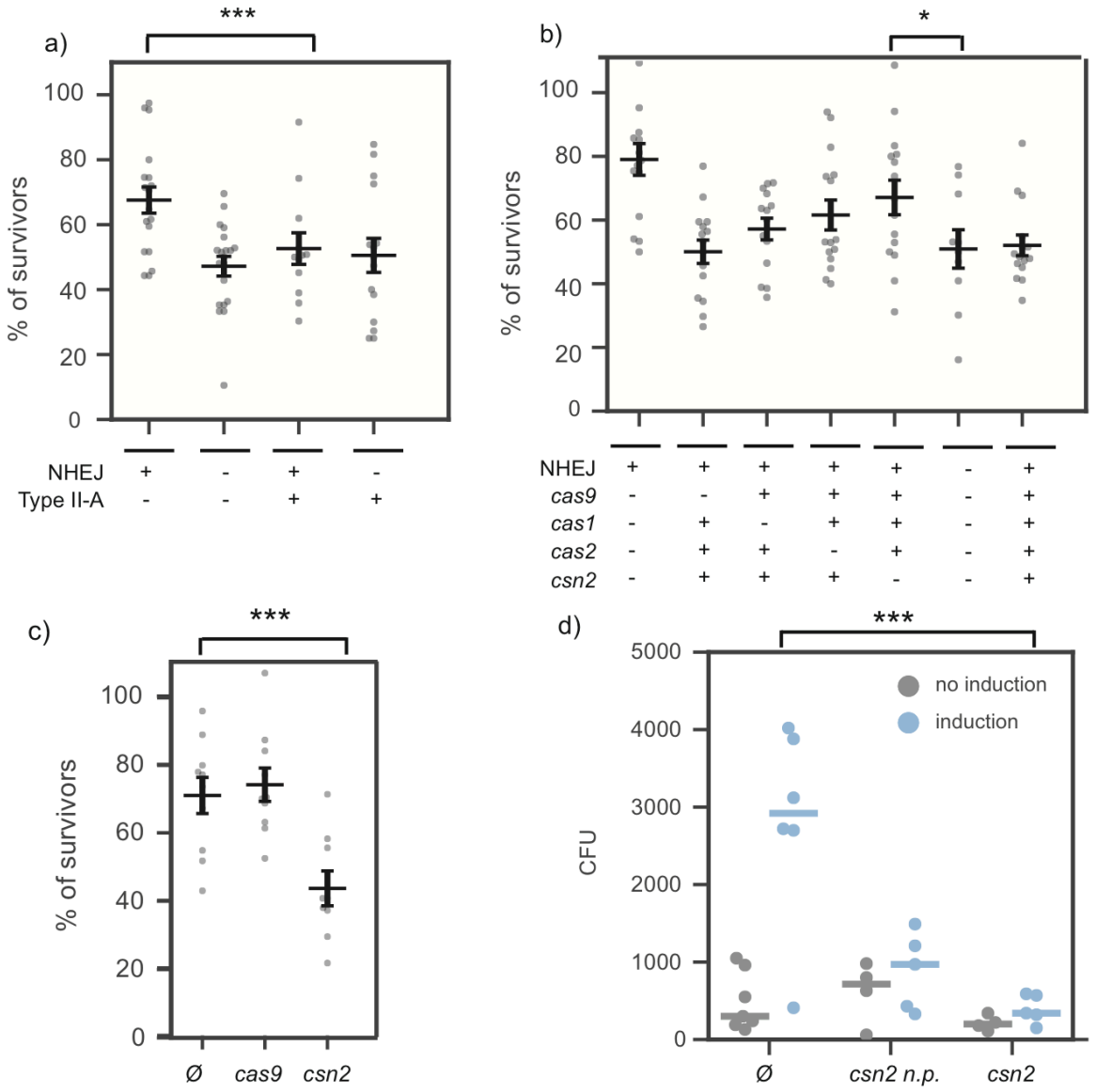

Figure 4: Type II-A CRISPR-Cas system impact NHEJ repair in *B. subtilis*

Survival rates of irradiated *B. subtilis* cells (a,b,c). Individual replicates (points) and average (horizontal bars) are shown. Error bars correspond to the standard error of the mean (s.e.m.). **a,** cells carrying the type IIA CRISPR-Cas system (pRH87) or the control empty vector (pC194), and deleted for *ku* and *ligD* or not (P=0.0009, Wilcoxon). **b,** *B. subtilis* carrying the CRISPR-Cas system with the dCas9 mutations (pRH121) or deleted for *csn2* (pRH63), *cas1* (pRH61), or *cas2* (pRH62)(P=0.02, Student two sided t-test). **c,** *B. subtilis* carrying the empty pC194 plasmid (Ø), expressing *csn2* from plasmid pAB56 or *cas9* from plasmid pDB114 (P=0.0048, Wilcoxon). **d,** A linearized plasmid providing resistance to chloramphenicol (pC194) was electroporated into *S. aureus* RN4220 cells carrying the NHEJ system either alone (plasmid pAB1, Ø) or with *csn2* cloned downstream of *ligD* (plasmid pAB81, csn2) or under the control of its natural promoter (plasmid pAB82, csn2 n.p.). The number of CFUs obtained with or without induction of the NHEJ system using aTc are reported. The number of CFU obtained without induction (grey bars) indicate the background of already circular DNA present in the sample before electroporation (P=0.0060, two sided t-test).

To obtain more direct evidence that Csn2 blocks NHEJ repair, we investigated its ability to inhibit the recircularization of linear plasmid DNA upon electroporation into *S. aureus*. The *csn2* gene was added to plasmid pAB1 which encodes Ku and LigD, either under the control of a Ptet promoter (pAB82), or under the control of the *cas* operon promoter (pAB81). We then electroporated a linearized plasmid providing resistance to chloramphenicol (pC194) into cells expressing the NHEJ system or both NHEJ and Csn2 (protocol presented in Supplementary Figure 4.a). The *B. subtilis* Ku and LigD were able to circularize the plasmid DNA in *S. aureus*, but we obtained on average 5-fold fewer colonies when Csn2 was co-expressed with Ku and LigD compared to the NHEJ system alone (Figure 4.d). Altogether, these results show that Csn2 hinders NHEJ repair.

Discussion

We found that with the exception of a single case, NHEJ and type II-A CRISPR-Cas systems do not co-occur in fully sequenced bacterial genomes available to date. A possible incompatibility between NHEJ and type II-A CRISPR-Cas systems was investigated in a variety of experimental systems encompassing *S. aureus*, *B. subtillis* and *S. thermophilus*. Our results indicate that NHEJ does not affect CRISPR immunity against phages and plasmids, nor the capture of novel spacers. Previous studies showed that NHEJ repair pathways are able to repair Cas9-mediated DNA breaks in various bacterial species[19,37]. In agreement with our results, the efficiency of repair in these experimental setups was very low and thus cannot lead to a meaningful reduction in phage infectivity or plasmid transfer. Our results rather show that the Csn2 protein from type II-A CRISPR-Cas systems is able to inhibit NHEJ repair (Figure 5).

The strong avoidance of co-occurrences between NHEJ and type II-A systems was not observed with type II-C systems. This is consistent with the fact that type II-C systems lack Csn2. Csn2 is a multimeric toroidal protein that can bind double stranded DNA ends and slide inward through rotation-coupled translocation [9]. These DNA binding properties were noted in previous reports to be very similar to that of the Ku protein [9]. When present in the same cell, these two proteins will likely compete for the same substrate. We suggest that the binding of Csn2 at DNA-ends could block access to Ku or inhibit its function preventing efficient repair by the NHEJ machinery.

CRISPR-Cas systems are present in 47% of fully sequenced bacterial genomes[3] and this frequency might be much smaller in uncultivated bacteria [38]. This is in striking contrast with other defense systems, such as R-M systems, present on average at two copies per genome[39]. CRISPR-Cas systems are known to be transferred horizontally at a high rate[40], suggesting that they should spread in the bacterial world very rapidly if they were always advantageous. This brings to the fore the intriguing question of what is preventing further CRISPR rise in bacteria. Hypothesis that have been put forward include the cost of autoimmunity, the cost of limiting horizontal gene transfer, and the cost of inducible defenses [41–44]. Our results suggest another (non-mutually exclusive) reason: negative epistasis between the genetic background of a bacteria and a CRISPR-Cas system acquired by horizontal transfer can lead to a decreased fitness. In the present case, the type II-A CRISPR-Cas system affects the efficiency of NHEJ repair, thereby decreasing the fitness gain associated with the acquisition of the system. We therefore propose that NHEJ is a barrier to the establishment of this type of CRISPR-Cas systems in bacteria.

We have observed an intriguing tendency of type II CRISPR-Cas systems to be absent from the largest genomes. DNA repair mechanisms are more frequent in larger genomes, presumably as a result of the presence of more abundant accessory functions[45], and to maintain constant genomic mutation rates[46]. If these larger genomes endure stronger selection for the presence of NHEJ,

then incoming type II-A CRISPR-Cas systems will not be maintained in the genome. In agreement with the hypothesis of a trade-off between the two functions, nearly all of the largest genomes of Firmicutes encode NHEJ systems.

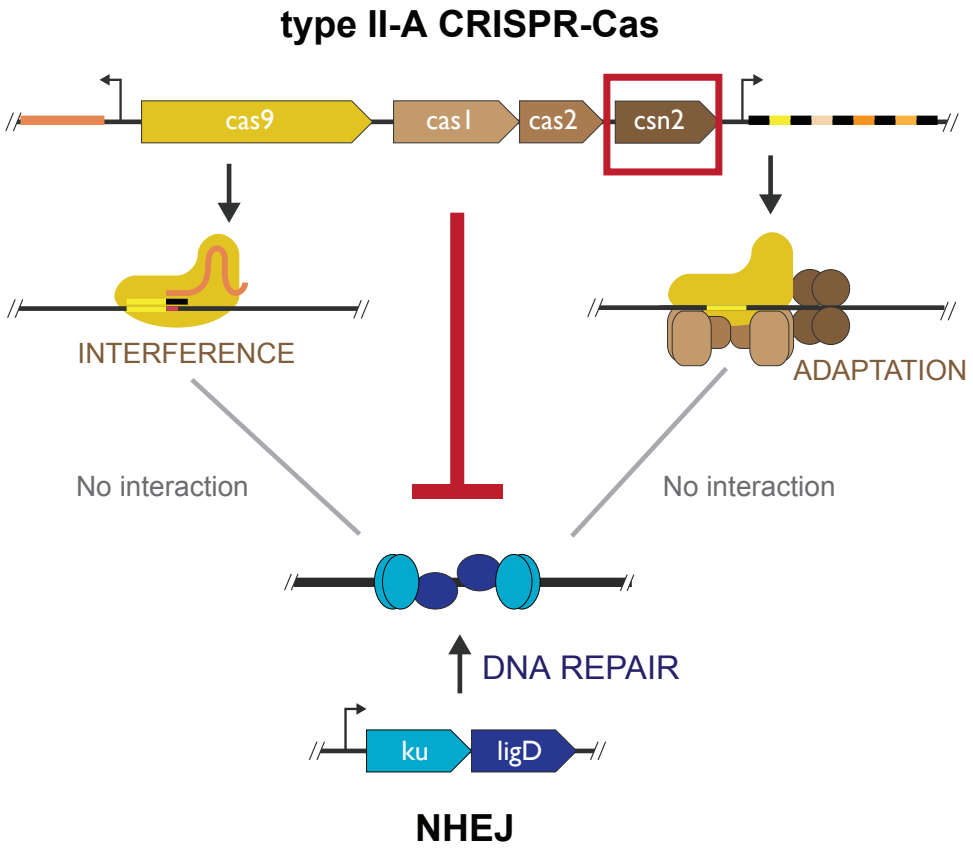

Figure 5: Graphical summary of the results. Three possible modes of negative interactions between type II-A CRISPR-Cas systems and NHEJ systems were tested: NHEJ could block CRISPR interference, NHEJ could block CRIPSR adaptation or CRISPR could block NHEJ repair. The last hypothesis was shown to be correct and Csn2 to be responsible for the inhibition of NHEJ repair.

Sorek and colleagues previously reported a positive effect of recBCD function on type I-E CRISPR spacer acquisition[47]. Since CRISPR-Cas systems acts by cutting DNA, epistatic interactions between these systems and DNA repair pathways might be numerous. These interactions are not only relevant to the evolution of bacterial genomes, but are also at the core of CRISPR genome editing technologies which rely on the repair of DNA breaks generated by Cas nucleases. Further studies will likely uncover novel aspects of DNA repair pathways' interaction with CRISPR-Cas systems.

Materials and Methods

Detection of repair systems and CRISPR-Cas systems

NHEJ and type II CRISPR-Cas systems were detected using MacSyFinder (default parameters) and the corresponding models for these systems (Abby et al. 2014) [4]. To detect NHEJ, we retrieved protein profiles from TIGRFAM: Ku (PF02735), ligD (TIGR02777, TIGR02778, TIGR02779). We build a MacSyFinder model for this system (Supplementary Text 2), where the presence of Ku was defined as mandatory and that of LigD as accessory (since other ligases can be recruited by Ku [20]) and both genes should be within 5 genes of distance. With this method, 74% of the detected systems encoded Ku and LigD; 26% encoded only Ku. We compared these results to a previous analysis using other methods [31]. Only one out of 113 genomes was discordant (we identified a NHEJ system in *Sinorhizobium meliloti* were none had been found previously)[31].

Genome dataset

We analyzed 5563 complete genomes retrieved from NCBI RefSeq *(ftp://ftp.ncbi.nih.gov/genomes/,* last accessed in November 2016) representing 2437 species of Bacteria.

Phylogenetic analyses

We built persistent genomes for 245 Firmicutes genomes smaller than 5 Mb available in GenBank RefSeq (Dataset), A list of orthologs was identified as reciprocal best hits using end-gap free global alignment, between the proteome of a pivot and each of the other strain's proteomes. *Bacillus subtilis* str.168 was used as pivot for each clade. Hits with less than 37% similarity in amino acid sequence and more than 20% difference in protein length were discarded. The persistent genome of each clade was defined as the intersection of pairwise lists of orthologs that were present in at least 90% of the genomes. We made a persistent genomes tree from concatenate of the multiple alignments of the persistent genes obtained with MAFFT v.7.205 (with default options, PMID: 23329690) and BMGE (with default options, PMID: 20626897). Missing genes have been replaced by stretches of "-" in each multiple alignment. The tree was computed with IQ-TREE multicore v.1.5.4 under the LG+R10 model[48]. This model was the one minimizing the Bayesian Information Criterion (BIC) among all models available (option -m TEST in IQ-TREE). We made 1000 ultra fast bootstraps to evaluate node support (options –bb 1000 –wbtl in IQ-TREE).

We applied BayesTraits v.2.0[34] to test the correlations among pairs of traits that adopt a finite number of discrete states. We ran two models (Independent and Dependent) in MCMC mode (priorAll exp 10) and computed the Bayes Factor BF which can be interpreted as follow : <2 weak evidence, >2 positive evidence, 5-10 strong evidence, >10 very strong evidence[49] .

Bacterial strains and growth conditions.

S. aureus strain RN4220 was grown in TSB or TSA at 37 °C. Whenever applicable, media were supplemented with chloramphenicol at 10ug/ml or erythromycin at 10ug/ml or tetracycline at 100ng/mL or spectinomycine at 120ug/ml to ensure pC194-derived, pE194-derived, pT181, pLZ-derived plasmid maintenance, respectively and anhydrotetracycline (aTc) at 0.5ug/mL to ensure induction.

S. thermophilus strain DGCC7710 was grown in LM17 at 37 °C. Whenever applicable, media were supplemented with chloramphenicol at 5ug/ml to ensure pNZ123-derived plasmid maintenance.

B. subtilis strain 168 was grown in LB or LB agar at 37 °C. Whenever applicable, media were supplemented with chloramphenicol at 5ug/ml or erythromycin at 1ug/ml to ensure pC194-derived plasmid maintenance or pMUTIN4-derived selection.

Plasmids and strains construction

We constructed the pC194-derived and pE194-derived plasmids using electrocompetent *S. aureus* cells and pNZ123-derived, pLZ12-derived and pMUTIN4-derived plasmids using electrocompetent *E. coli* cells. Cloning strategies for each plasmid is summarized in Supplementary Table 3 and primers used are listed in Supplementary Table 4. PCR fragments were assembled using Gibson assembly except when mentioned otherwise. pAB2, pAB17, pAB18, pAB56 were obtained with a single PCR and recircularized by ligation. pMD021 was assemble by Golden Gate.

CRISPR-Cas interference efficiency assay using phages

We used two types of assays to assess the impact of the NHEJ system on efficiency of CRISPR-Cas immunity. **Phage titre assay**. Top agar lawns supplemented with 5mM CaCl2 of strains bearing the NHEJ system or not were poured on selective plates (with aTc for induction in *S. aureus*). We spotted serial dilutions of PhiNM4 or Phi2972 on the lawns of *S. aureus* and *S. thermophilus* respectively. *S. aureus* strain RN4220 carried the *S. pyogenes* CRISPR-Cas system on plasmid pDB114 or a derivative with spacer `5′-AAAATGTTTTAACACCTATTAACGTAGTAT-3′` (pMD021). *S. thermophilus* strain DGCC7710 and a bacteriophage insensitive mutant of strain DGCC7710 carrying spacer `5′-TGTTAAAAGAAGCACTAGAGGTGATTTACG-3′` in the first position of the CRISPR-1 locus were used. E.O.P was determined after overnight incubation at 37°C. **Productive infection assays**. Cells were diluted 1:100 from overnight cultures in TSB supplemented with 5mM CaCl2 and appropriate antibiotics, grown to OD 0.2 at 37°C. The NHEJ system was induced by aTc. After 30 minutes of incubation allowing the expression of the NHEJ system, we added phage PhiNM4 at M.O.I (Multiplicity of Infection) of 1. Adsorption was allowed for 5 minutes at 37°C with shaking. Cells were then put on ice and washed twice

with ice cold TSB. We then diluted and spotted them on top agar lawns of RN4220 supplemented with CaCl2. E.C.O.I was determined after overnight incubation at 37°C.

CRISPR-Cas interference efficiency assay using plasmids

Cells carrying a type II-A CRISPR-Cas systems (pRH87) and the NHEJ system (pAB1) or the empty vector as a control (pE194) were made electro-competent as follow: cells were grown until OD 0.4, induced by adding aTc and further grown to OD 0.8. Cells were washed twice with ice-cold water, once with 10% glycerol and then resuspended in 1/100 of their volume in 10% glycerol. 100 ng of plasmid pT181 or pAB2 were electroporated in 50ul of electrocompetent cells (2500V, 25µF, 100Ω and 2mm cuvettes). Cells were then incubated in 1ml TSB for one hour at 37°C and plated on tetracycline only. Transformation efficiency was assessed after overnight incubation at 37°C.

Adaptation assays

The spacer acquisition assay was described elsewhere [12]. We mixed cells from overnight cultures (induced and non induced when necessary) with phages (M.O.I value of 1) in top agar supplemented with 5mM CaCl2 and poured them on plates containing appropriate antibiotics and supplemented with aTc when necessary, followed by overnight incubation at 37°C. For *S. aureus,* single colonies were resuspended in lysis buffer (250mM KCl, 5mM $MgCl_2$, 50mM Tris-HCl at pH 9.0, 0.5% Triton X-100) supplemented with 20ng/mL lysostaphin and incubated at 37°C for 10 min, then 98°C for 10 min. Following centrifugation (11 000g), 1ul of the supernatant was used as template for DreamTaq PCR amplification with primers AB23 and AB24. We provide a list of 15 acquired spacers in supplementary table 5. For *S. thermophilus*, single colonies were resuspended in 10ul of water, 1ul of which was used as template for DreamTaqPCR amplification with primers AB103 and AB104. The PCR reactions were analyzed on 2% agarose gels. Adaptation rates were computed as the estimated number of clones that acquired a spacer divided by the estimated number of cells in the initial population.

Irradiation assay

The NHEJ repair assay was described elsewhere [36]. 100ul of overnight cultures of *B. subtilis* strains were irradiated at 100 Gy (RS Xstrahl, 42 minutes, 250kV, 12mA, 30cm from focal point). We plated 1:10 000 dilution on appropriate antibiotics. CFU was determined after overnight incubation at 37°C. Survival rate was determined as the ratio of the CFUs obtained for irradiated cells over CFUs obtained for non-irradiated cells.

References

1. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).
2. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–12 (2007).
3. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13,** 722–736 (2015).
4. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. C. MacSyFinder: A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems. *PLoS One* **9,** e110726 (2014).
5. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* (2017).
6. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42,** 6091–6105 (2014).
7. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).
8. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–12 (2007).
9. Arslan, Z. *et al.* Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Res.* **41,** 6347–6359 (2013).
10. Ellinger, P. *et al.* The crystal structure of the CRISPR-associated protein Csn2 from Streptococcus agalactiae. *J. Struct. Biol.* **178,** 350–362 (2012).
11. Lee, K. H. *et al.* Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins Struct. Funct. Bioinforma.* **80,** 2573–2582 (2012).
12. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519,** 199–202 (2015).
13. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in Type II-A CRISPR – Cas adaptation. *Genes Dev.* **29,** 356–361 (2015).
14. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–21 (2012).
15. Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR / Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39,** 9275–9282 (2011).
16. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E2579-86 (2012).
17. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1390–1400 (2008).
18. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).
19. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of Escherichia

coli. *Nucleic Acids Res.* **44,** 4243–4251 (2016).
20. Shuman, S. & Glickman, M. S. Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.* **5,** 852–61 (2007).
21. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
22. Della, M. *et al.* Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science.* **306,** 683–685 (2004).
23. Aniukwu, J., Glickman, M. S. & Shuman, S. The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev.* **22,** 512–527 (2008).
24. Bowater, R. & Doherty, A. J. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* **2,** e8 (2006).
25. Gong, C. *et al.* Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat. Struct. Mol. Biol.* **12,** 304–312 (2005).
26. Pitcher, R. S. *et al.* NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair* . **6,** 1271–1276 (2007).
27. Moeller, R. *et al.* Role of DNA repair by nonhomologous-end joining in Bacillus subtilis spore resistance to extreme dryness, mono- and polychromatic UV, and ionizing radiation. *J. Bacteriol.* **189,** 3306–3311 (2007).
28. Paris, Ü. *et al.* NHEJ enzymes LigD and Ku participate in stationary-phase mutagenesis in Pseudomonas putida. *DNA Repair.* **31,** 11–18 (2015).
29. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 505–510 (2015).
30. Zhang, J., Kasciukovic, T. & White, M. F. The CRISPR Associated Protein Cas4 Is a 5' to 3' DNA Exonuclease with an Iron-Sulfur Cluster. *PLoS One* **7,** e47232 (2012).
31. Rocha, E. P. C., Cornet, E. & Michel, B. Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet.* **1,** e15 (2005).
32. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42,** 10618–10631 (2014).
33. Silva, F. J., Latorre, A. & Moya, A. Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet.* **19,** 172–176 (2003).
34. Pagel, M. & Meade, A. Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. *Am. Nat.* **167,** 808–825 (2013).
35. Koo, Y., Jung, D. K. & Bae, E. Crystal structure of streptococcus pyogenes Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS One* **7,** e33401 (2012).
36. Weller, G. R. *et al.* Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* **297,** 1686–1689 (2002).
37. Su, T. *et al.* A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. *Sci. Rep.* **6,** 37895 (2016).
38. Burstein, D. *et al.* Major bacterial lineages are essentially devoid of CRISPR-Cas viral defense systems. *Nat. Commun.* **7,** 10613 (2016).
39. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42,**

10618–10632 (2014).
40. Godde, J. S. & Bickerton, A. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* **62,** 718–29 (2006).
41. Bondy-Denomy, J. & Davidson, A. R. To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol.* **22,** 218–25 (2014).
42. Jiang, W. *et al.* Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* **9,** e1003844 (2013).
43. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L. a. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12,** 177–86 (2012).
44. Westra, E. R. *et al.* Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. *Curr. Biol.* **25,** 1043–1049 (2015).
45. Nimwegen, E. van. Scaling laws in the functional content of genomes. *Trends Genet.* **19,** 476–479 (2003).
46. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148,** 1667–1686 (1998).
47. Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520,** 505–510 (2015).
48. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32,** 268–274 (2015).
49. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. Introducing Markov Chain Monte Carlo. *Markov Chain Monte Carlo in Practice* 512 (1996).

Acknowledgements
We thank Philippe Horvath for providing *S. thermophillus* strain DGCC7710 and phage Phi2972 and Sylvain Moineau for providing plasmid pNZ123. We also want to thank them both for their very helpful advice. We thank Matt Deyell for plasmid pMD021. A.B. is a member of the "Ecole Doctorale Frontière du vivant (FdV)". Funding: This work was supported by the European Research Council (ERC) under the Europe Union's Horizon 2020 research and innovation programme (grant agreement No [677823]); the French Government's Investissement d'Avenir program; Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' [ANR-10-LABX-62-IBEID]; the Pasteur-Weizmann consortium.

Author contributions.
A.B., E.R., M.T. and D.B. designed the research. A.B., A.C.V., C.B. performed the experiments. A.B., E.R., M.T. and D.B wrote the manuscript.

Competing financial interests.
The authors declare no competing financial interests.