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

Aude Bernheim\textsuperscript{1,2,3,4}, Alicia Calvo-Villamañán\textsuperscript{3}, Clovis Basier\textsuperscript{3}, Lun Cui\textsuperscript{3}, Eduardo P.C. Rocha\textsuperscript{1,2}, Marie Touchon\textsuperscript{1,2} & David Bikard\textsuperscript{3}

Type II CRISPR-Cas systems introduce double-strand breaks into DNA of invading genetic material and use DNA fragments to acquire novel spacers during adaptation. These breaks can be the substrate of several DNA repair pathways, paving the way for interactions. We report that non-homologous end-joining (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 using the NHEJ pathway from \textit{Bacillus subtilis} and the type II-A CRISPR-Cas systems from \textit{Streptococcus thermophilus} and \textit{Streptococcus pyogenes}. Our results suggest that the NHEJ system has no effect on CRISPR immunity. On the other hand, we provide evidence for the inhibition of NHEJ repair by the Csn2 protein. Our findings give insights on the complex interactions between CRISPR-Cas systems and repair mechanisms in bacteria, contributing to explain the scattered distribution of CRISPR-Cas systems in bacterial genome.
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. Yet, although most bacteria are infected by phages and other mobile genetic elements, Crispr-Cas systems are absent from the majority of bacterial genomes. The selective pressures and mechanisms that lead to the success of Crispr-Cas systems in some clades and not others remain poorly understood.

Crispr-Cas systems are classified into 6 types and 27 subtypes, according to the Cas proteins they carry. 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 array, three core genes (cas1, cas2, and cas9), and a small trans-activating Crispr RNA (crRNA) complementary to the Crispr repeat sequence. A fourth gene is involved in spacer acquisition, cas2 in the type II-A/II-B, and cas4 in type II-B systems. A third subtype, type II-C, only requires cas1, cas2, and cas9. All the Cas proteins of type II systems are necessary for spacer acquisition, but only Cas9 is necessary for interference. Cas9 is guided by small crRNA to introduce double-strand breaks (DSB) into target DNA. A short conserved sequence (2-5 bp) 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.

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, 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 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. In bacteria, the NHEJ system requires two core proteins: Ku and a ligase. Ligation is usually carried out by the ligD protein, but other ligases can be recruited by Ku when ligD is absent. The system is complemented by additional proteins in certain cases. Ku binds at the DSB and recruits the ligase to seal the break. 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. NHEJ repair can be mutagenic, leading to up to 50% error rates in certain bacteria.

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. Importantly, DNA repair pathways and Crispr-Cas systems are composed of proteins with structural similarities and interacting with the same substrates. For example, Cas4, a protein present in type I and type II-B systems shares structural and functional similarities with AddB, a component of the AddAB repair pathway and a functional homolog of RecBCD. In type II-A Crispr-Cas systems, Cas2 binds and slides along free DNA ends in the same manner as the Ku protein of the NHEJ system. 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 investigate it in bacteria. We hypothesize 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 test these hypotheses by assessing the patterns of co-occurrence of the two systems in bacterial genomes. This reveals one single case of co-occurrence of both systems among 5563 bacterial genomes, suggesting strong negative interaction. We then investigate experimentally mechanisms that could explain this 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. We could not measure any effect of the NHEJ system on type II-A Crispr-Cas interference and adaptation. On the other hand, our results suggest that the Csn2 protein inhibits NHEJ repair.

**Results**

**Negative association between NHEJ and type II-A Crispr-Cas.** We detected Crispr-Cas and NHEJ systems in 5563 fully sequenced bacterial genomes. 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. We sequenced 5563 fully sequenced bacterial genomes (Supplementary Data 1). The NHEJ system 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 Fig. 1 and Supplementary Table 1). Fimicutes and Proteobacteria were the only phyla with genomes encoding enough type II Crispr-Cas systems (respectively 209 and 101) and NHEJ (respectively 364 and 637), to perform robust statistical analyses (Supplementary Fig. 1). A possible confounding factor when studying the distribution of bacterial defense and DNA repair pathways is that their abundance co-varies with genome size. Accordingly, NHEJ systems were more frequent in larger genomes (P < 10^-4, χ² test on a logistic fit). In contrast, type II Crispr-Cas systems were only present in genomes smaller than 5 Mb (Supplementary Fig. 2). Hence, we focused our analysis on Fimicutes and Proteobacteria with genomes smaller than 5 Mb. They represent 56.5% of the total number of genomes. In this sample, the size of the genomes encoding NHEJ systems was independent of the presence of a type II crispr-Cas system (P = 0.99, Wilcoxon’s 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 Fimicutes (P < 10^-4, Fisher’s exact test), but not in Proteobacteria (P = 0.70, Fisher’s exact test) (Fig. 1b and Supplementary Fig. 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 Fimicutes encoded mostly type II-A systems (Fig. 1a). Type II-B systems were only detected in nine genomes and thus were not further analyzed. 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 freedom of the statistical analyses. To check whether systems are negatively associated while taking phylogeny into account, we built a tree of Fimicutes and tested if the binary traits (presence of both systems) evolved independently using BayesTraits. A strong negative association between NHEJ and type II-A Crispr-Cas systems was observed (Bayes factor (BF) = 9.7, Fig. 1c), 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, with no frameshifts or premature stop codons, and the adjacent CRISPR array encodes 44 spacers. We were also unable to detect in this Eggerthella genome anti-CRISPR proteins similar to the ones described in the literature38–40.

NHEJ has no measurable effect on 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 system10. 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 Note 1 and Supplementary Fig. 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). Streptococcus 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 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 (EOP) of phiNM4 when the NHEJ system

---

**Fig. 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.
is expressed. The CRISPR-Cas system provided a five order of magnitude reduction in the EOP of phage phiNM4 when compared with a spacer-less control, and no significant increase in the number of plaques was observed upon NHEJ induction (Fig. 2a). 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 eight mutant phages. All mutants had a point mutation in the PAM and none could not result from the unfaithful repair of Cas9 breaks through NHEJ, we sequenced the target position of eight 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 (ECOI), i.e., the proportion of cells that produce at least one functional phage particle, in the presence (pAB1) or absence (pE194) of the NHEJ system (n = 4, mean, NS double-sided t test P = 0.9991). The observed strain (sensitive to the phage). One would expect higher ECOI of functional phage particle after infection compared to the control (ECOI), i.e., the number of cells that produce at least one active gene.

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 (n = 3, mean). The EOP 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) was not supported by our experiments.

NHEJ has no measurable effect on type II-A CRISPR-Cas acquisition. Ku and Csn2 bind the same type of substrate—linear
double-stranded DNA—and might thus interfere antagonistically. To test if the NHEJ system affects spacer acquisition, we measured the cells’ ability to acquire new spacers in the presence of the NHEJ machinery. Streptococcus aureus cells carrying the NHEJ system (pAB1) and the type II-A CRISPR-Cas system (pRH87) were infected with phage PhiNM4 either with or without induction of the NHEJ system. 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 (Fig. 3a). 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 (analysis of variance, P = 0.16) (Fig. 3b).

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 in the rate of novel spacer acquisition between cells expressing the NHEJ machinery or the GFP (Wilcoxon’s test, P = 0.26) (Fig. 3c). 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. To test this hypothesis, we reproduced the experiment that led to the discovery of the NHEJ system in B. subtilis. When B. subtilis cells in stationary phase are irradiated by ionizing radiations, the DSB generated are repaired by the NHEJ system, as other repair systems cannot function in those specific conditions. Bacillus subtilis deleted for NHEJ do not survive irradiation as well as the wild type. If type II-A CRISPR-Cas systems limit NHEJ repair, cells bearing a type II-A CRISPR-Cas system are expected to show increased sensitivity to irradiation.

B. subtilis cells expressing the type II-A 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’s test, Fig. 4a). 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 an 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 pRH87 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’s test, Supplementary Fig. 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’s two-sided t test after validation of normality and homoscedasticity, Fig. 4b). 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...
Csn2 impacts NHEJ repair. Survival rates of irradiated *B. subtilis* cells. a–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 II-A CRISPR-Cas system (pRH87) or the control empty vector (pC194), and deleted for *ku* and *ligD* or not (*P = 0.0009, Wilcoxon’s). 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’s two-sided t test after validation of normality and homoscedasticity). c *B. subtilis* carrying the empty pC194 plasmid (Ø), expressing *csn2* from plasmid pAB56 or *cas9* from plasmid pDB114 (*P = 0.0048, Wilcoxon’s). 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 pAB88, *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 (gray bars) indicate the background of already circular DNA present in the sample before electroporation (*P = 0.006, two-sided t test)*.  

![Image](https://example.com/image.png)

**Fig. 4** Csn2 impacts NHEJ repair. Survival rates of irradiated *B. subtilis* cells. a–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 II-A CRISPR-Cas system (pRH87) or the control empty vector (pC194), and deleted for *ku* and *ligD* or not (*P = 0.0009, Wilcoxon’s). 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’s two-sided t test after validation of normality and homoscedasticity). c *B. subtilis* carrying the empty pC194 plasmid (Ø), expressing *csn2* from plasmid pAB56 or *cas9* from plasmid pDB114 (*P = 0.0048, Wilcoxon’s). 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 pAB88, *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 (gray bars) indicate the background of already circular DNA present in the sample before electroporation (*P = 0.006, 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 Fig. 4a). The *B. subtilis* Ku and LigD were able to circulate the plasmid DNA in *S. aureus*, but we obtained on average fivefold fewer colonies when Csn2 was co-expressed with Ku and LigD compared to the NHEJ system alone (Fig. 4d). In this assay the NHEJ system is strongly overexpressed compared to the natural expression of Ku and LigD in *B. subtilis* during stationary phase (Supplementary Fig. 6). Note that such overexpression was necessary to observe plasmid recircularization events in *S. aureus*. On the other hand, Csn2 was only slightly overexpressed compared to its expression level in *S. pyogenes* SF370. 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. subtilis*, 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. The efficiency of repair in these experimental setups was very low. Consistently, our results show that NHEJ repair cannot lead to a...
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. These DNA binding properties were noted in previous reports to be very similar to that of the Ku protein. 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 and this frequency might be much smaller in uncultivated bacteria. This is in striking contrast with other known to be transferred horizontally at a high rate suggesting that 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. In particular, the ability of Csn2 to block NHEJ repair could prove especially useful in genome editing experiments performed in Eukaryotes where NHEJ repair of Cas9-mediated breaks can compete with homology-directed repair and limit the efficiency of precise editing.

**Methods**

**Detection of repair systems and CRISPR-Cas systems.** NHEJ and type II CRISPR-Cas systems were detected using MacSyFinder (default parameters). The published models were used for the detection of type II CRISPR-Cas systems. To detect NHEJ, we retrieved protein profiles from TIGRFAM: Ku (PF02735) and ligD (TIGR02777, TIGR02778, TIGR02779). We built a MacSyFinder model where the presence of Ku was defined as mandatory and that of ligD as accessory (Supplementary Note 2). Other ligases can indeed be recruited by Ku with. With this method, 74% of genomes detected encoded both Ku and ligD, 26% encoded only Ku. We compared these results to a previous analysis using other methods. One only out of 113 genomes was discordant (we identified a NHEJ system in Sinorhizobium meliloti were none had been previously). We made 1000 ultra-fast bootstraps to evaluate node support (options −w 1000 −wbtl in IQ-TREE).

**Phylogenetic analyses.** We built persistent genomes for 245 Firmicutes genomes smaller than 5 Mb available in GenBank RefSeq (Dataset). 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. 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 strain 168 was used as pivot for each clade. Hits with <37% similarity in amino acid sequence and more than 20% difference in protein length were discarded. We made a persistent genome tree by concatenation of the multiple alignments of the persistent genes obtained with MAFFT v7.205 (with default options, PMID: 23329690) and BMEGE (with default options, PMID: 20626897). Missing genes were 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. This model gave the lowest 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−s 1000−wblt in IQ-TREE).

We applied BayesTraits v.2.071 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 BF which can be interpreted as follows: <2 weak evidence, >2 positive evidence, 5–10 strong evidence, and >10 very strong evidence. We made 1000 ultra-fast bootstraps to evaluate node support (options−s 1000−wblt in IQ-TREE).

We applied BayesTraits v.2.071 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 BF which can be interpreted as follows: <2 weak evidence, >2 positive evidence, 5–10 strong evidence, and >10 very strong evidence.

**Bacterial strains and growth conditions.** S. aureus strain RN4220 was grown in TSB (Tryptic Soy Broth) or TSA (Tryptic Soy Agar) at 37 °C. Whenever applicable, media were supplemented with chloramphenicol (10 μg ml−1), erythromycin (10 μg ml−1), tetracycline (100 ng ml−1), or spectinomycin (120 μg ml−1) to ensure the maintenance of pC194-derived, pE194-derived, pT181-derived, and pLZ-derived plasmid, respectively. Expression from pT promoters was induced by the addition of anhydrotetracycline (aTc) at 0.5 μg ml−1. Streptococcus thermophilus strain DGGC7710 was grown in LM17 at 37 °C. Whenever applicable, media were supplemented with chloramphenicol (3 μg ml−1) to ensure the maintenance of pNZ123-derived plasmids. Bacillus subtilis strain 168 was grown in LB or LB agar at 37 °C. Whenever applicable, media were supplemented with chloramphenicol (5 μg ml−1) or erythromycin (1 μg ml−1) to ensure the maintenance of pC194-derived plasmids and the integration of pMUTIN4-derived plasmids.

---

**Supplementary Information.**

We thank A. de Toledo for help with strain genetic construction and sequencing, J. Hommer and J. Suau for helpful discussion and comments on the manuscript. We acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness, project AGL2015-67604-C2-1-R (MINECO).
Plasmids and strains construction. The cloning strategies employed for each plasmid are summarized in Supplementary Table 2 and the primers used are listed in Supplementary Table 3. PCR fragments were assembled with Gibson assembly. DNA sequencing was performed at the Department of Life Sciences, University of Copenhagen, to confirm the sequence. Cell lines were cultured in 10% CO₂, 37 °C in TIBCO media (Corning) supplemented with 5% fetal bovine serum (FBS), 100 U/mL penicillin, and 100 μg/mL streptomycin.

NHEJ functionality assay. The plasmid pC194 was linearized by PCR using primers B329 and B330 (Supplementary Table 3). Strains with the plasmids carrying the NHEJ system were grown to optical density (OD) 0.5 and the NHEJ system was induced by adding a tetracycline (tC). Cells were grown to OD 0.8 and made competent by washing them three times in ice-cold water, supplemented with 10% glycerol for the last wash, and concentrated 100-fold. We transformed 200 μg of linearized pC194 in those electro-competent cells and added a tC to the recovery medium. Cells were plated on selective media and incubated overnight at 37 °C. We resuspended single colonies in lysis buffer with 15 mM MgCl₂ and lysostaphin and incubated them at 37 °C for 10 min, then 98 °C for 10 min. Following centrifugation (11 000 g), 1 μL of the supernatant was used as template for DreamTaqPCR amplification with primer A9, A10 (Supplementary Table 3). PCR products were then purified and sequenced.

CRISPR-Cas interference efficiency assay using phages. We used two types of assays to assess the impact of Ku and Ligase on CRISPR-Cas immunity. Phage titer assay: we used agar lawns supplemented with 5 mM CaCl₂ and inoculated with strains bearing the NHEJ system or not were poured on selective plates (with or without tC for induction in S. aureus). We spotted serial dilutions of PhN14M or PhN2972 on the lawns of S. aureus and S. thermophilus, respectively. Streptococcus aureus strain RN4220 carried the S. pyogenes CRISPR-Cas system on plasmid pDB114 or a derivative plasmid derived from pV-INTGTTAACAAATATACGTGGTA (pMD021). S. thermophilus strain DGCC7710 and a bacteriophage-insensitive strain DGCC7710' carrying spacer 5'-TTGTTAAAAAGACGACTA-GAGGTGATTACG-3' in the first position of the CRISPR-1 locus were used. EOP was determined after overnight incubation at 37 °C. Productive infection assays: cells were diluted 1:100 from overnight cultures in TSB supplemented with 5 mM CaCl₂ and the appropriate antibiotics and incubated at 37 °C. The NHEJ system was induced using a tC at OD600 0.2. After 30 min of incubation allowing the expression of the NHEJ system, we added plasmid PhN14M at an MOI (multiplicity of infection) of 1. Adsorption was allowed for 5 min at 37 °C. CaCl₂ and the appropriate antibiotics and incubated at 37 °C. The NHEJ system were grown to optical density (OD) 0.3 and the NHEJ systems expressing the NHEJ system were grown to OD 0.8, induced by adding aTc and further grown to OD 0.8. Cells were then washed twice with ice-cold water, once with 10% glycerol and resuspended in 10 mL of 3 mM CaCl₂ and the appropriate antibiotics and incubated at 37 °C. The cloning strategies employed for each strain are shown relative to expression in the wild-type strain (Ku in S. subtilis 168 or Can2 in S. pyogenes SF370).

DNA extraction. DNA was extracted from strains B. subtilis 168, B. subtilis 168 + pRH167, B. subtilis 168 + pAB16, S. pyogenes SF370, S. aureus RN4220 + pAB16, and S. aureus RN4220 + pAB1 + pRH167. Overnight cultures were spotted on 2% agarose gels. Overnight cultures were resuspended in 2 mL and incubated at 37 °C for 3 h. For strains with pAB1 or pAB2 plasmids, a tC (0.5 μg mL⁻¹) was added after 1 h of incubation. Four milliliters of RNA protect buffer (lysozyme 20 mg mL⁻¹). Streptococcus aureus cell pellets were resuspended in 200 μL of lysozymophospholipid solution (lysozyme 5 mg mL⁻¹). After 1 h incubation at 37 °C, 1 mL of Trizol was added, and regular Trizol reagent procedures for purifying the total RNA were followed.

RT-qPCR. All the RNA samples were treated with DNase (Turbio DNase free kit, Ambion), then all the RNA samples (1 μg for each sample) were reverse transcribed using the Transcriptor First strand cDNA synthesis Kit (Roche). The qPCR was performed using 1 μL of the reverse transcription reaction and the FastStart essential DNA green master mix (Roche) in a LightCycler 96 (Roche). Probes and PCR primers are listed in Supplementary Table 5. Relative gene expression was computed using the ∆∆CT method (2⁻ΔΔCT), where CqAB25 is the quantification cycle value for the 16S rRNA and CqTAR for the tested gene. Data are shown relative to expression in the wild-type strain (Ku in B. subtilis 168 or Can2 in S. pyogenes SF370).

Data availability. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. Data corresponding to 5536 complete genomes retrieved from NCBI RefSeq information can be found in Supplementary Data 1.

Received: 17 August 2017 Accepted: 22 November 2017
Published online: 12 December 2017

References

1. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561 (2005).
2. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA and provide additional tools for evolutionary studies. Microbiology 151, 653–663 (2005).
3. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening DNA sequences of class 2 CRISPR-Cas systems. J. Mol. Biol. 260, 740–752 (1996).
4. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in Streptococcus thermophilus. Science 315, 1799–1792 (2007).
5. Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736 (2015).
6. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. C. MacchiFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS ONE 9, e110726 (2014).
7. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169–182 (2017).
8. 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).
9. Delcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNAIII. Nature 471, 602–607 (2011).
10. Heier, R. et al. Cas9 specifies functional viral targets during CRISPR Cas adaptation. Nature 519, 199–202 (2015).
11. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in type II-A CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 356–361 (2015).
12. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).
13. Sparnaukas, R. et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nat. Microbiol. 3, 9275–9282 (2018).
14. Gasanis, G., Barrangou, R., Horvath, P. & Sikorski, R. Cas9 ribozyme. RNA 21, 2302–2308 (2015).
15. Cui, L. & Bikard, D. Cas9 cleavage in the chromosome of Escherichia coli. Nucleic Acids Res. 44, 4243–4251 (2016).
18. Shuman, S. & Glickman, M. S. Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.* 5, 852–861 (2007).

19. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).

20. Della, M. et al. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* 306, 683–685 (2004).

21. 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).

22. Bowater, R. & Doherty, A. J. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* 2, e8 (2006).

23. 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, 3312–3318 (2005).

24. Pitcher, R. S. et al. NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair* 6, 1271–1276 (2007).

25. 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).

26. Paris, U. et al. The NHEJ enzymes LigD and Ku participate in stationary-phase mutagenesis in *Pseudomonas putida*. *DNA Repair* 31, 11–18 (2015).

27. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510 (2015).

28. Arslan, Z. et al. Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Cas2. *Nucleic Acids Res.* 41, 6347–6359 (2013).

29. Zhang, J., Cornet, E. & Michel, B. Comparative and evolutionary analysis of bacterial homologous recombination systems. *PLoS Genet.* 1, e15 (2005).

30. Ellinger, P. et al. The crystal structure of the CRISPR-associated protein Cas2 from *Streptococcus agalactiae*. *J. Struct. Biol.* 178, 350–362 (2012).

31. Lee, K. H. et al. Identification, structural, and biochemical characterization of a group of large Cas2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins Struct. Funct. Bioinform.* 80, 2573–2582 (2012).

32. Nam, K. H. et al. The crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein revealed Ca2+-dependent double-stranded DNA binding activity. *J. Biol. Chem.* 286, 30759–30768 (2011).

33. Koo, Y., Jung, D. K. & Bae, E. Crystal structure of streptococcus pyogenes Cas2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS ONE* 7, e33401 (2012).

34. 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).

35. Silva, F. J., Latorre, A. & Moya, A. Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet.* 19, 172–176 (2003).

36. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible jump markov Chain Monte Carlo. *Ann. Math. 167*, 808–825 (2013).

37. Rauch, B. J. et al. Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* 168, 159–158 (2016).

38. Pawliuk, A. et al. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Microbiol.* 1, 16085 (2016).

39. Hynes, A. P. et al. An anti-CRISPR from a virulent streptococcus phage inhibits *Streptococcus pyogenes* Cas9. *Nat. Microbiol.* 2, 1374–1380 (2017).

40. Weller, G. R. et al. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 297, 1686–1689 (2002).

41. 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).

42. Burstein, D. et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defense systems. *Nat. Commun.* 7, 10613 (2016).

43. 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).

44. Bondy-Denomy, J. & Davidson, A. R. To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol.* 22, 218–25 (2014).

45. Jiang, W. et al. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS Genet.* 9, e1003844 (2013).

46. Rikard, D., Hatoum-Aslan, A., Muscida, D. & Marraffini, L. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12, 177–186 (2012).

47. Westra, E. R. et al. Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* 25, 1043–1049 (2015).

48. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71 (2010).

49. Nimwegen, E. Van Scaling laws in the functional content of genomes. *Trends Genet.* 19, 476–479 (2003).

50. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* 148, 1667–1686 (1998).

51. 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).

52. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. Introducing Markov Chain Monte Carlo. *Markov Chain Monte Carlo Pract.* 512, 1–19 (1996).

53. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345 (2009).

54. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* 3, e3647 (2008).