Cas9 specifies functional viral targets during CRISPR–Cas adaptation

Robert Heler1*, Poulami Samai1*, Joshua W. Modell1, Catherine Weiner1, Gregory W. Goldberg1, David Bikard1,2 & Luciano A. Marraffini1

Clustered regularly interspaced short palindromic repeat (CRISPR) loci and their associated (Cas) proteins provide adaptive immunity against viral infection in prokaryotes. Upon infection, short phage sequences known as spacers integrate between CRISPR repeats and are transcribed into small RNA molecules that guide the Cas9 nuclease to the viral targets (protospacers). *Streptococcus pyogenes* Cas9 cleavage of the viral genome requires the presence of a 5′–NGG–3′ protospacer adjacent motif (PAM) sequence immediately downstream of the viral target. It is not known whether and how viral sequences flanked by the correct PAM are chosen as new spacers. Here we show that Cas9 selects functional spacers by recognizing their PAM during spacer acquisition. The replacement of cas9 with alleles that lack the PAM recognition motif or recognize an NNGNG PAM eliminated or changed PAM specificity during spacer acquisition, respectively. Cas9 associates with other proteins of the acquisition machinery (Cas1, Cas2 and Csn2), presumably to provide PAM-specificity to this process. These results establish a new function for Cas9 in the genesis of prokaryotic immunological memory.

Cas9 is required for spacer acquisition

To investigate the mechanisms of recognition of PAM-adjacent protospacers during spacer acquisition, we cloned the type II-A CRISPR–Cas locus of *S. pyogenes* (Fig. 1a) into the staphylococcal vector pC194 (ref. 16) and introduced the resulting plasmid (pWJ40 (ref 17)) into *Staphylococcus aureus* RN4220 (ref. 18), a strain lacking CRISPR–Cas loci. We chose this experimental system because it facilitates the genetic manipulation of the *S. pyogenes* CRISPR–Cas system. We first tested the ability of the cells to mount adaptive CRISPR immunity by infecting

---

1Laboratory of Bacteriology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. 2Synthetic Biology Group, Institut Pasteur, 28 Rue du Dr. Roux, 75015 Paris, France.

*These authors contributed equally to this work.
them with the staphylococcal phage φNM4γ4, a lytic variant of φNM4 (ref. 19) (see Methods for a description of φNM4γ4 isolation). Plate-based assays performed by mixing bacteria and phage in top agar allowed the selection of phage-resistant colonies that were checked by PCR to look for the expansion of the CRISPR array (Extended Data Fig. 1a). On average 50% of the colonies acquired one or more spacers (8/13, 5/11 and 7/16 in three independent experiments), whereas the rest of the resistant colonies survived phage infection by a non-CRISPR mechanism, most likely including phage receptor mutations (Extended Data Fig. 2a). To maximize the capture of new spacer sequences, we performed the same assay in liquid and recovered surviving bacteria at the end of the phage challenge. These were analysed by PCR of the CRISPR array and the amplification products of expanded loci were subjected to Illumina MiSeq sequencing to determine the extent of spacer acquisition. Analysis of 2.96-million reads detected protospacers adjacent to 2,083 out of 2,687 NGG sequences present in the viral genome, although with variation in the frequency of acquisition of each sequence (Extended Data Fig. 1b). The data revealed a prominent selection of spacers matching protospacers with downstream NGG PAM sequences (99.97%, Extended Data Fig. 1c). The acquisition of new spacers by cells in liquid culture proved to be simple and highly efficient, providing the possibility to look at millions of new spacers in a single step. It was therefore used in the rest of our studies.

To determine the genetic requirements for spacer acquisition, we made individual deletions of cas1, cas2 or csn2 and challenged the mutant strains with phage φNM4γ4. Spacer acquisition was decreased to levels below our limit of detection in each of these mutants (Fig. 1b), corroborating previous experiments. Therefore, although Cas1, Cas2 and Csn2 are dispensable for anti-phage immunity in the presence of a pre-existing spacer (Extended Data Fig. 2b, c), they are required for spacer acquisition. To determine whether these genes are also sufficient for this process, we overexpressed cas1, cas2 and csn2 in the absence of cas9 using a tetracycline-inducible promoter in plasmid pH223 and looked for the integration of new spacers in the absence of phage infection using a highly sensitive PCR assay (Extended Data Fig. 3). We were unable to detect new spacers even in the presence of the inducer (Fig. 1c). However, the addition of a second plasmid expressing tracrRNA (see below) and Cas9 from their native promoters (Extended Data Fig. 3) enabled spacer acquisition only in the presence of the inducer, with all the new spacers matching chromosomal or plasmid sequences (Fig. 1c and Extended Data Table 1). Although it is most likely that the acquisition of such spacers causes cell death or plasmid curing, respectively, the acquisition event can still be detected in liquid culture using our highly sensitive PCR assay (Extended Data Fig. 3b, c). The tracrRNA (Fig. 1a) is a small RNA bound by Cas9 that is required for crRNA processing and Cas9 nuclelease activity. We wondered if Cas9 involvement in spacer acquisition also required the presence of the tracrRNA. Deletion of the tracrRNA prevented spacer acquisition in the absence of phage infection (Fig. 1c), suggesting that apo-Cas9 is not sufficient to promote spacer acquisition and that association with its cofactor is also required. Altogether these data indicate that Cas1, Cas2 and Csn2 are necessary but not sufficient for the incorporation of new spacers and that a tracrRNA–Cas9 complex is also required. This is in contrast to the type I-E CRISPR–Cas system of E. coli, in which overexpression of Cas1 and Cas2 alone is sufficient for spacer acquisition. It is important to note that the CRISPR array used in this assay consists of a single repeat, without pre-existing spacers (Extended Data Fig. 3). Therefore the Cas9 requirement is not a consequence of the phenomenon known as ‘primed’ spacer acquisition. This refers to an increase in the frequency of spacer acquisition observed in type I CRISPR–Cas systems that relies on the presence of a pre-existing spacer with a partial match to the phage genome as well as the full targeting complex (Cascade) on the presence of a pre-existing spacer with a partial match to the phage genome.  

Cas9 specifies the PAM of newly acquired spacers  

Given this newfound requirement in the CRISPR adaptation process and the well-established PAM recognition function of Cas9 during the surveillance and destruction of viral target sequences, we hypothesized that this nuclease could participate in the selection of PAM sequences during spacer acquisition. To test this we exchanged the cas9 genes of S. pyogenes (Sp) and S. thermophilus (St) CRISPR–Cas systems to create two chimaeric CRISPR loci: tracrRNA<sup>Sp</sup>–cas9<sup>Sp</sup>–cas1<sup>Sp</sup>–cas2<sup>Sp</sup>–csn2<sup>Sp</sup> and tracrRNA<sup>St</sup>–cas9<sup>St</sup>–cas1<sup>St</sup>–cas2<sup>St</sup>–csn2<sup>St</sup> (Fig. 2a). We chose the type II-A CRISPR–Cas system of S. thermophilus (also known as CRISPR3 (ref. 23)) because it is an orthologue of the S. pyogenes system. While the PAM sequence for the Sp CRISPR–Cas system is NGG, the PAM sequence for the St system is NGGN<sup>24</sup> (Fig. 2b and Extended Data Table 1). We infected each naïve strain with phage φNM4γ4, sequenced the newly acquired spacers, and obtained the PAM of the matching protospacers using WebLogo. We found that each chimaeric system acquired spacers with PAMs that correlated with the cas9, but not the tracrRNA, cas1, cas2 or csn2, allele present (Fig. 2b and Extended Data Table 1). To rule out the possibility that non-functional spacers are negatively selected during phage infection, that is, they are acquired randomly but only those cells containing spacers with a correct PAM for Cas9 cleavage provide immunity and allow cell survival, we sequenced the PAMs of spacers acquired in the absence of phage infection (Figs 1c and 2c). Either Cas9<sup>Sp</sup> or Cas9<sup>St</sup> were produced in cells overexpressing Cas1<sup>Sp</sup>, Cas2<sup>Sp</sup> and Csn2<sup>Sp</sup>. In this experiment, as explained earlier, spacers matching chromosomal or plasmid sequences were acquired. The PCR products containing new spacers were cloned into a commercial vector from which they were sequenced (Extended Data Table 1). Expression of Cas9<sup>Sp</sup> led to the incorporation of spacers matching protospacers with an NGG PAM sequence, whereas the expression of Cas9<sup>St</sup> in the same cells shifted the composition of the PAM to NGGN<sup>25</sup> (Fig. 2d). These results demonstrate that Cas9 determines PAM sequences during CRISPR adaptation to ensure the acquisition of functional spacers.

Cas9 associates with Cas1, Cas2 and Csn2  

In type I CRISPR–Cas systems, Cas1 and Cas2 form a complex and the dsDNA nuclease activity of Cas1 has been implicated in the initial cleavage of the invading viral DNA to generate a new spacer<sup>26</sup>. The genetic analyses presented above suggest that in the type II S. pyogenes CRISPR–Cas system, the PAM-binding function of Cas9 observed in vitro<sup>27</sup> could specify a PAM-adjacent site of cleavage for Cas1, or other members of the spacer acquisition machinery. This would guarantee that newly acquired spacers have the correct PAM needed for Cas9 activity later in this immune pathway. This hypothesis predicts an interaction between Cas9 and Cas1, Cas2 and/or Csn2. To test this we expressed the type II Cas operon in E. coli, using a histidyl tagged version of Cas9, and looked for other proteins that co-purify. We observed an abundant co-purifying protein with an apparent molecular weight close to 33 kDa, the expected size of Cas1 (Extended Data Fig. 4a). Mass spectrometry confirmed the identity of both of these proteins, as well as the presence of Cas2 and Csn2 co-purifying with Cas9 (Extended Data Table 2). This result suggested the formation of a Cas9–Cas1–Cas2–Csn2 complex and therefore we explored other purification strategies to unequivocally determine its existence. We were able to isolate a Cas9–Cas1–Cas2–Csn2 complex when the histidyl tag was added to Csn2 (Fig. 3a, b). The identity of the purified proteins was confirmed by mass spectrometry (Extended Data Table 3). This demonstrates a biochemical link between the Cas9 nuclease and the other Cas proteins that function exclusively to acquire new spacers, supporting the role of Cas9 as a PAM specificity factor in the adaptation phase of CRISPR immunity.

Cas9 PAM-binding motif is needed for spacer selection  

Within this complex the PAM-binding domain of Cas9 would specify a functional spacer (one adjacent to a correct PAM) and the nuclease activity of Cas1 and/or Cas9 would cleave the invading DNA to extract the spacer sequence. To test this model, we performed adaptation studies in the absence of phage selection as described in Extended Data Fig. 3 but using different combinations of wild-type Cas1, Cas1(E220A) (catalytically dead or dCas1<sup>26</sup>), wild-type Cas9, Cas9(D10A,H840A)
(catalytically dead or dCas9 (ref. 6)) and Cas9(R1333Q,R1335Q) (abbreviated here as Cas9\(^{R\text{PAM}}\), containing mutations in the PAM-binding motif that substantially reduces binding to target DNA sequences with NGG PAMs \textit{in vitro} \(^2\)). We observed that the nuclease activity of Cas1 is necessary for spacer acquisition (Fig. 3c). In contrast, the nuclease and PAM-binding function of Cas9 are dispensable for this process. Next we determined the PAM of the acquired spacers in the presence of mutated Cas9 (Fig. 3d). We found that whereas spacers acquired in the presence of dCas9 displayed correct PAMs, those acquired in the presence of Cas9\(^{R\text{PAM}}\) matched DNA regions without a conserved flanking sequence, that is, without a PAM sequence. Cells containing the catalytically dead Cas9(D10A,H847A) from \textit{S. thermophilus} acquired spacers with NGGNG PAMs (Extended Data Fig. 5). These results indicate that Cas1 and Cas9 are part of a complex dedicated to spacer acquisition which requires Cas1 nuclease activity and Cas9 PAM-binding properties for the selection of new spacer sequences.

**Discussion**

The selection of new spacers with a correct PAM is fundamental for the survival of the infected host during CRISPR–Cas immunity. In the simplest scenario there is no active selection of PAM-flanked protospacers; any spacer sequence can be acquired but only those with the correct PAM allow Cas9 cleavage of the invader and survival. Bacteria that acquire spacers with ineffective flanking sequences are killed by the virus and as a consequence PAM-flanking spacers are enriched in the population. Here we show that even in the absence of phage selection, the type II CRISPR–Cas system acquires new spacers with correct PAMs, a result that rules out the possibility of random spacer selection with subsequent selection for functional spacers. How are PAM-flanked protospacers selected during type II CRISPR–Cas immunity? One possibility is that the proteins exclusively dedicated to spacer acquisition perform the PAM-selection function. The inability of cells overexpressing only \textit{cas1}, \textit{cas2} and \textit{csn2} to expand the CRISPR array strongly suggest that none of the proteins encoded by these genes can recognize and select correct PAMs. Another possibility is that the known PAM-recognition function of Cas9 (refs 15, 27), essential for destroying the invading virus, could also be used during spacer acquisition to recognize PAM-flanking viral sequences. Experiments showing that the \textit{cas9} allele, but not the \textit{cas1} or \textit{cas2} or \textit{csn2} alleles, determines the PAM sequence of the newly acquired spacers, demonstrated that this scenario is probably correct. Regarding the molecular mechanism by which Cas9 participates in CRISPR adaptation, our experiments show that Cas9 forms a stable complex with Cas1, Cas2 and Csn2 that presumably participates in the selection of new spacers. The nuclease activity of Cas1, but not of Cas9, is required for spacer acquisition. The tracrRNA is also required, suggesting that the apo–Cas9 structure\(^2\), very different from holo–Cas9 (ref. 15), does not have the correct conformation to participate in spacer acquisition. The key residues involved in Cas9 PAM recognition are not required for spacer acquisition, but they are necessary for the incorporation of new spacers with the correct PAM sequence. This suggests that the reported non-specific DNA binding
property of Cas9 (refs 6, 7) is sufficient for spacer acquisition, but not for the selection of functional spacers. There are currently two models for the incorporation of new spacers into the CRISPR array, one where the future spacer sequence is cut from the invading viral DNA, the ‘cut and paste’ model, and another where this sequence is copied from the viral genome, the ‘copy and paste’ model26. In the context of the first model, our data suggests that, at a low frequency that may reflect the dynamics of spacer acquisition, Cas9 cleaves the invading genome to extract a new spacer sequence. However, on its own, Cas1 nuclelease activity is non-specific26. Therefore we propose that through the formation of the Cas9–Cas1–Cas2–Cas32 complex, Cas9 binding to PAM-adjacent sequences provides specificity to Cas1 endonuclease activity. In the copy and paste model, Cas1 nuclelease activity is most likely necessary for downstream events, such as the cleavage of the repeat sequence that precedes spacer insertion, and Cas9 is required to ‘mark’ sequences adjacent to GG motifs to be copied into the CRISPR array. In any case, following as yet unknown processing and integration events, the selected DNA becomes a new functional spacer, that is, its match-ability to the CRISPR array to recognize it as a new spacer in the CRISPR array are still unknown. All genes of the type II-A CRISPR–Cas locus (tracrRNA, cas9, cas1, cas2 and casm2) are required for spacer acquisition, therefore most likely all the members of the Cas9–Cas1–Cas2–Cas32 complex participate in the process. Future work will address this and other aspects of the mechanisms of spacer integration in different CRISPR–Cas systems. The present work reveals a new function for Cas9 in CRISPR immunity. This nuclelease is fundamental for both the execution of immunity, participating in the surveillance and destruction of infectious target viruses, and the generation of immunological memory, selecting the viral sequences that allow adaptation and resistance to viral predators. 

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 October 2014; accepted 20 January 2015.
Published online 18 February 2015.

1. Barrangou, R. & Marrisfni, L. A. CRISPR–Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell 54, 234–244 (2014).
2. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712 (2007).
3. Delkhcheva, E. et al. CRISPR-RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607 (2011).
4. Makarova, K. S. et al. Evolution and classification of the CRISPR–Cas systems. Nature Rev. Microbiol. 9, 467–477 (2011).
5. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71 (2010).
6. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).
7. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–67 (2014).
8. Gasman, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-cRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl Acad. Sci. USA 109, E2579–E2586 (2012).
9. Szczelkun, M. D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl Acad. Sci. USA 111, 9798–9803 (2014).
10. Díez-Villaseñor, C., Guzman, N. M., Almendros, C., Garcia-Martinez, J. & Mojica, F. J. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR–Cas1–E variants of Escherichia coli. RNA Biol. 10, 792–802 (2013).
11. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40, 5569–5576 (2012).
12. Datsenko, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945 (2012).
13. Nun˜ez, J. K. et al. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nature Struct. Mol. Biol. 21, 528–534 (2014).
14. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marrisfni, L. A. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. Nature Biotechnol. 31, 233–239 (2013).
15. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569–573 (2014).
16. Hornouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pCF194, a plasmid that specifies inducible chloramphenicol resistance. J. Bacteriol. 150, 815–829 (1982).
17. Goldberg, G. W., Jiang, W., Bikard, D. & Marrisfni, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR–Cas targeting. Nature 514, 633–637 (2014).
18. Kreiswirth, B. N. et al. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. Nature 305, 709–712 (1983).
19. Bae, T., Baba, T., Hiramatod, K. & Schneewind, O. Prophages of Staphylococcus aureus Newman and their contribution to virulence. Mol. Microbiol. 62, 1035–1047 (2006).
20. Sizova, V. et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 39, 9275–9282 (2011).
21. Li, M., Wang, R., Zhao, D. & Xiang, H. Adaptation of the Haliarcaul hispanica CRISPR–Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res. 42, 2483–2492 (2014).
22. Richter, C. et al. Priming in the Type I-F CRISPR–Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. 42, 8516–8526 (2014).
23. Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. 190, 1401–1412 (2008).
24. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR–Cas systems. Nucleic Acids Res. 42, 2577–2590 (2014).
25. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. Weblogo: a sequence logo generator. Genome Res. 14, 1188–1190 (2004).
26. Wiedenhett, B. et al. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 21, 904–912 (2014).
27. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 343, 1247997 (2014).
28. Arslan, Z., Hermans, V., Wurm, R., Wagner, R. & Pel, U. Detection and characterization of spacer integration intermediates in type I CRISPR–Cas systems. Nucleic Acids Res. 43, 7884–7893 (2015).

Acknowledgements We thank members of the laboratory for critical discussion of the experiments and their results. A. Zaytsev for help with the deep sequencing data analysis and A. Sherlock for help with plasmid construction. R.H. is the recipient of a Howard Hughes International Student Research Fellowship. P.S. is supported by a Heilmsley Postdoctoral Fellowship for Basic and Translational Research on Disorders of the Digestive System at The Rockefeller University. J.W.M. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. D.B. is supported by a Harvey L. Karp Discovery Award and the Bettencourt Schuller Foundation. L.A.M. is supported by the Rita Allen Scholars Program, an Irma T. Hirschl Award, a Sinsheimer Foundation Award and a NIH Director’s New Innovator Award (1DP2AI104556-01).

Author Contributions R.H., P.S., D.B. and L.A.M. conceived the study and designed experiments. R.H. and P.S. executed the experimental work with help from C.W. J.W.M. set up the experimental system to detect spacer acquisition in the absence of phage infection. G.W.G. isolated and characterized phage eNM4-4 and constructed the pgp32 plasmid. D.B. analysed MiSeq data. L.A.M. wrote the paper with the help of the rest of the authors.

Author Information The full sequence of eNM4-4 has been deposited in GenBank under accession number KF299285. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (david.bikard@gasteur.fr) or L.A.M. (marrisfni@rockefeller.edu).
**METHODS**

**Bacterial strains and growth conditions.** Cultivation of *S. aureus* RN4220 (ref. 18), was carried out in brain-heart infusion (BHI) or heart infusion (HI) medium (BD) at 37 °C. Whenever applicable, media were supplemented with chloramphenicol at 10 μg ml⁻¹ or erythromycin at 5 μg ml⁻¹ to ensure plasmid maintenance, respectively.

**On-plate spacer acquisition assay.** To detect individual colonies on a plate, cells from overnight cultures were mixed with plate at a m.o.i. value of 1 in top agar containing appropriate antibiotic and 5 mM CaCl₂. The mixture was poured on BHI plates with antibiotic and incubated at 37 °C overnight. Subsequently, colonies that survived phage infection were restreaked on fresh BHI plates in order to remove contaminating virus and dead cells. Plates were incubated at 37 °C overnight. To check for spacer acquisition, individual colonies were resuspended in lysis buffer (250 mM KCl, 5 mM MgCl₂, 50 mM Tris-HCl at pH 9.0, 0.5% Triton X-100), treated with 50 ng ml⁻¹ lysothphin and incubated at 37 °C for 5 min, then 98 °C for 5 min. Following centrifugation (16,000g), a sample of the supernatant was used as template for TopTaq PCR amplification with primers L400 and H050. The PCR reactions were analysed on 2% agarose gels (Fig. 1a).

**In-liquid spacer acquisition assay.** Overnight cultures launched from single colonies were diluted 1:1,000 into a fresh 10-mI culture of BHI containing appropriate antibiotic and 5 mM CaCl₂. When the cultures reached OD₆₀₀nm of 0.4, depending on the strain, they were either infected with phage MOI value of 1 (Fig. 1b) or induced with 1 μg ml⁻¹ anhydrotetracycline (Fig. 1c). After 1 h, plasmids carrying the CRISPR systems were extracted using a slightly modified QIAprep Spin Miniprep Kit protocol: the pelletted bacterial cells were resuspended in 250 μl buffer P1 containing 50 ng ml⁻¹ lysothphin and incubated at 37 °C for 1 h, followed by the standard QIAprep protocol. 100 ng of plasmid DNA was used to amplify the CRISPR locus using Phusion DNA Polymerase (New England Biolabs) with the following primer mix: 3 parts JW8 and 1 part each of JW3, JW4 and JW5 (Extended Data Table 4). The following cycling conditions were used: (1) 98 °C for 10 s; (2) (for 30 times) 98 °C for 10 s, 64 °C for 20 s, 72 °C for 30 s (3) 72 °C for 5 min. The PCR reactions were analysed on 2% agarose gels. To sequence individual spacers, the adapted bands were extracted, gel-purified and cloned via Zero Blunt TOPO PCR Cloning Kit (Invitrogen). CRISPR loci of individual clones were checked for expansion of the arrays by PCR using the primers listed above and sent for sequencing.

**Phage adsorption assay.** The phage adsorption assay was performed as described previously with minor modifications. Cells were grown in BHI and 10 mM CaCl₂ to a D₆₀₀nm (OD₆₀₀nm) of 0.4. The phage solution was prepared at 10⁶ plaque-forming units (p.f.u.) ml⁻¹ of this was added to 900 μl of cells. The mixture was incubated for 10 min at 37 °C to allow adsorption of the phage to the cellular membrane. The mixture was centrifuged for 1 min at 16,000g and the number of phage particles left in the supernatant was determined by phage titre assay.

**Phage titre assay.** Serial dilutions of the phage stock were prepared in triplicate and spotted on fresh top agar lawns of RN4220 in HI agar supplemented with the appropriate antibiotic and 5 mM CaCl₂. Plates were incubated at 37 °C overnight (Extended Data Fig. 2).

**High-throughput sequencing.** Plasmid DNA was extracted from adapted cultures using the in-liquid spacer acquisition assay described above. 100 ng of plasmid DNA was used as template for Phusion PCR to amplify the CRISPR locus with primers H182 and H183 (Extended Data Table 4). Following gel extraction and purification of the adapted bands, samples were subject to Illumina MiSeq sequencing.

**Plasmid construction.** Construction of pW140 was described elsewhere. For the construction of pc194-derived and pe194-derived plasmids, cloning was performed using chemically competent *S. aureus* cells, as described previously. The Δcas1 (pRH059), Δcas2 (pRH061) and Δcas2 (pRH063) mutants were constructed by one-piece Gibson assembly from pW140 using the primers H16–H17 and H18–H19, respectively (Extended Data Table 4). Plasmid pRH087 containing the wild type cas genes of *S. pyogenes* was obtained by inserting the first spacer of *S. pyogenes* (annealed primers H049 and H050 containing compatible BsaI overhangs) in pDB184 using BsaI cloning. BsaI cloning was also used to construct pRH097 and pRH233 by inserting a nM4Δ targetting spacer (annealed primers H029 and H030) into pDB114 and pDB184, respectively. Plasmid pRH200 harbours the wild-type CRISPR3 system from *S. thermophilus* LMD-9 amplified with H168 and H169 from genomic DNA. The fragment was inserted on pRH200 using Gibson assembly. pRH213 was constructed by replacing Cas9 on pRH087 with Cas9 on pRH200 using the primers pairs H223–H232 and H231–H234, respectively, pRH214 was constructed by replacing Cas9 on pRH200 with Cas9 from pRH087 using the primers H227–H230 and H228–H229, respectively. PGG32 was created by reducing the CRISPR locus using Phusion DNA Polymerase (New England Biolabs) with the following primer mix: 3 parts JW8 and 1 part each of JW3, JW4 and JW5 (Extended Data Table 4). The following cycling conditions were used: (1) 98 °C for 10 s; (2) (for 30 times) 98 °C for 10 s, 64 °C for 20 s, 72 °C for 30 s (3) 72 °C for 5 min. The PCR reactions were analysed on 2% agarose gels. To sequence individual spacers, the adapted bands were extracted, gel-purified and cloned via Zero Blunt TOPO PCR Cloning Kit (Invitrogen). CRISPR loci of individual clones were checked for expansion of the arrays by PCR using the primers listed above and sent for sequencing.

**Protein purification of Cas9.** PM806 (wild-type Cas9) plasmid was obtained from Addgene. The proteins were purified as described before with minor modifications as follows. The proteins were expressed in *E. coli* BL21 Rosetta (DE3) codon plus cells (EMD Millipore). Cultures (2 litres) were grown at 37 °C in Terrific Broth medium containing 50 μg ml⁻¹ kanamycin and 34 μg ml⁻¹ chloramphenicol until the D₆₀₀nm reached 0.6. The cultures were supplemented with 0.2 mM isopropyl-1-thio-β-D-galactopyranoside and incubation was continued for 16 h at 16 °C with constant shaking. The cells were collected by centrifugation and the pellets stored at −80 °C. All subsequent steps were performed at 4 °C. Thawed bacteria were resuspended in 30 ml of buffer A (50 mM Tris–HCl pH 7.5, 500 mM NaCl, 200 mM LiSO₄, 10% sucrose, 15 mM imidazole) supplemented with complete EDTA free protease inhibitor tablet (Roche). Triton X-100 and lysozyme were added to final concentrations of 0.1% and 0.1 mg ml⁻¹, respectively. After 30 min, the lysate was sonicated to reduce viscosity. Insoluble material was removed by centrifugation for 1 h at 16,200g in a Beckman JA-3050 rotor. The soluble extract was bound in batch to mixed for 1 h with 5 ml of Ni⁺²-Nitrotriacetic-acid-garose resin (Qiagen) that had been pre-equilibrated with buffer A. The resin was recovered by centrifugation, and then washed extensively with buffer A. The bound protein was eluted step-wise with aliquots of IMAC buffer (50 mM Tris–HCl pH 7.5, 250 mM NaCl, 10% glucose) eluting increasing concentrations of imidazole. The 25 mM imidazole eluate containing the His₆-MBP tag was pooled and dialysed. The His₆-MBP affinity tag was removed by cleavage with TEV protease during overnight dialysis against 20 mM Tris–HCl pH 7.5, 150 mM KCl, 1 mM TCEP and 10% glycerol. The tagless Cas9 protein was separated from the fusion tag by using a 5 ml SP Sepharose Hitrap column (GE Life Sciences). The protein was further purified by size exclusion chromatography using a Superdex 200 10/300 GL in 20 mM Tris HCl pH 7.5, 150 mM KCl, 1 mM TCEP, and 5% glycerol. The elution peak from the size exclusion was aliquoted, frozen and kept at −80 °C.

**Protein purification of Cas1.** Plasmid pKW01 (wild-type Cas1) was constructed by using the single piece Gibson protocol for pW140 as a template for polymerase chain reactions (PCRs) to clone Cas1 into pET28b-His6Sm3 using the primers PS192 and PS193 (Extended Data Table 4). Full sequencing of cloned DNA fragment confirmed perfect matches to the original sequence. The pKW01 plasmid was transformed into *E. coli* BL21 (DE3) Rosetta 2 cells (EMD Millipore). Cultures were grown and protein was purified by Ni-affinity chromatography step, as mentioned before in Cas9 purification. The 200 mM imidazole elutes containing the His₆-Sm3 tagged Cas1 polypeptide was pooled together. The His₆-Sm3 affinity tag was removed by cleavage with SUMO protease during overnight dialysis against 50 mM Tris–HCl pH 7.5, 250 mM NaCl, 20 mM imidazole and 10% glycerol. The tagless Cas1 protein was further purified by size exclusion chromatography using a Superdex 200 10/300 GL in 20 mM Tris HCl pH 7.5, 500 mM KCl, 1 mM TCEP, and 5% glycerol. The elution peak from the size exclusion was aliquoted, frozen and kept at −80 °C.
Protein purification of Cas2. The sequence encoding Cas2 was PCR amplified with primers PS334 and PS335 from pWJ40 and inserted into a pET-His6 MBP TEV cloning vector (Addgene Plasmid number 29656) using ligation independent cloning (LIC). Sequencing of the resultant plasmid (pPS059) confirmed the matches to the wild-type sequence. The protein was expressed and purified following the same procedure as that for Cas9.

Protein purification of Csn2. Plasmid pPS060 was constructed by through amplification of pWJ40 as a template for polymerase chain reactions (PCRs) to clone Csn2 into pET28b-His10Smt3 using the primers PS336 and PS337. Full sequencing of cloned DNA fragment confirmed perfect matches to the original sequence. Csn2 was expressed and purified following the same method as that of Cas1. Previously Csn2 was shown to form a tetramer\(^34\). Protein concentrations for all the purifications were determined by using the Bradford dye reagent with BSA as the standard.

Protein purification of Cas9–Cas1–Cas2–Csn2 complex. pKW07 (His\(_{10}\)-Cas9–Cas1–Cas2–Csn2) was constructed by amplification of pWJ40 with primers PS199/PS202 and pET16b (Novagen) with primers PS200/PS203, followed by Gibson assembly of the fragments. Full sequencing of cloned DNA fragment was done to confirm perfect matches to the original sequence. The proteins were expressed in E. coli BL21 Rosetta 2(DE3) codon plus cells (EMD Millipore). Cultures were grown and protein was purified by Ni-affinity chromatography step, as mentioned before in Cas9 purification with minor modifications. The 200 mM imidazole eluates were dialysed overnight against 20 mM Tris-HCl pH 7.5, 150 mM KCl, 1 mM TCEP and 10% glycerol and subjected to mass spectrometry for the identification of the co-purifying proteins. pKW06 (Cas9–Cas1–Cas2–Csn2–His\(_{6}\)) was constructed by amplification of pWJ40 with primers PS204/PS205 and pET23a (Novagen) with primers PS206/PS207 (Extended Data Table 4), followed by Gibson assembly of the fragments. Full sequencing of cloned DNA fragment was done to confirm perfect matches to the original sequence. The proteins were expressed in E. coli BL21 Rosetta 2(DE3) codon plus cells (EMD Millipore). Cultures were grown and protein was purified by Ni-affinity chromatography step, as mentioned before in Cas9 purification with minor modifications. The 200 mM imidazole eluates were dialysed overnight against 20 mM Tris-HCl pH 7.5, 150 mM KCl, 1 mM TCEP and 10% glycerol. The proteins were further purified using a 5 ml SP Sepharose HiTrap column (GE Life Sciences), eluting with a linear gradient of 150 mM–1 M KCl.

Sample size. No statistical methods were used to predetermine sample size.

29. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibiotics. J. Bacteriol. 150, 804–814 (1982).
30. Duplessis, M. & Moïneau, S. Identification of a genetic determinant responsible for host specificity in Streptococcus thermophilus bacteriophages. Mol. Microbiol. 41, 325–336 (2001).
31. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6, 343–345 (2009).
32. Bikard, D. et al. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. Nature Biotechnol. 32, 1146–1150 (2014).
33. Moore, S. D. & Prevelige, P. E., Jr. A P22 scaffold protein mutation increases the robustness of head assembly in the presence of excess portal protein. J. Virol. 76, 10245–10255 (2002).
34. 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).
Extended Data Figure 1 | The *S. pyogenes* type II CRISPR–Cas system displays a strong bias for the acquisition of spacers matching viral protospacers with NGG PAMs.  

**a**, Analysis of bacteriophage-insensitive mutant colonies using PCR and agarose gel electrophoresis, representative of five technical replicates. Bacteria and phage were mixed in top agar and incubated overnight. DNA was isolated from individual colonies resistant to phage infection and used as template for a PCR reaction with primers (arrows) H182 and H183 (Extended Data Table 4), which amplify the 5′ end of the *S. pyogenes* CRISPR array. The size of the PCR band indicates the number of new spacers (shown at the top of the gel). Cells without additional spacers resist infection by a CRISPR-independent mechanism, presumably envelope resistance. 

**b**, Analysis of acquired spacers during phage infection of a population of bacteria carrying the *S. pyogenes* type II CRISPR–Cas system. Liquid cultures of bacteria were infected with phage, surviving cells were collected at the end of the infection, DNA extracted and used as template for a PCR reaction as described above. Amplification products were separated by agarose gel electrophoresis and the DNA of the bands corresponding to products with additional spacers was extracted and sent for Mi-Seq next-generation sequencing. Reads corresponding to newly acquired spacers were plotted according to their position in the phage φNM44 genome (x axis) and their abundance (y axis). Each dot represents a unique spacer sequence; blue and red dots indicate a corresponding protospacer with an NGG or non-NGG PAM. Top and bottom plots indicate protospacers in the top and bottom strands of the φNM44 DNA. The map as well as the different functions of the phage genes are indicated in between the plots. The raw data used to make this graph is in the Source Data File. 

**c**, Weblogo showing the conservation of the 5′ flanking sequences of 10,000 protospacers randomly selected from the experiment shown in **b**. Absolute conservation of the NGG PAM was observed.
Extended Data Figure 2 | *cas1*, *cas2* and *csn2* are not required for the execution of immunity. 

**a.** Analysis of bacteriophage-resistant mutants that do not acquire a new spacer. Three colonies that survived phage infection in our in-plate adaptation assay (Extended Data Fig. 1) were subjected to phage adsorption assay. Briefly, surviving colonies as well as the wild-type *S. aureus* RN4220 control were grown in liquid and mixed with bacteriophage. After a brief incubation, cells were pelleted by centrifugation and the phages present in the supernatant (unable to bind and infect cells) were counted on a lawn of sensitive cells. The number of plaque-forming units (p.f.u.) of a control experiment in the absence of host cells were used to determine the 100% free-phage, or 0% adsorption value. No plaques were observed in the control experiment using wild-type cells and this value was used to set the 100% adsorption limit. The three CRISPR-independent, bacteriophage-resistant mutants displayed a marked defect in phage adsorption (about 50%), indicating that most likely they carry envelope resistance mutations. Error bars: mean ± s.d. (*n* = 3).

**b.** *cas1*, *cas2* and *csn2* are not required for the execution of immunity using previously acquired spacers. Position within the phage φNM4γ4 genome of the type II CRISPR–Cas target used in this experiment. The protospacer sequence is in the bottom strand (shown in 3′–5′ direction) and flanked by a TGG PAM (in green).

**c.** Comparison of immunity provided by a type II CRISPR–Cas system programmed to target the sequence shown in panel a in the presence (wild-type, wt) or absence (Δ*cas1*, Δ*cas2*, Δ*csn2*) of *cas1*, *cas2* and *csn2*. Immunity is measured as the p.f.u. of a φNM4γ4 phage lysate spotted on top agar lawns of *S. aureus* RN420 cells containing no CRISPR system (–), a wild-type *S. pyogenes* CRISPR–Cas type II system (wt, pH233), or the same CRISPR–Cas systems with a deletion of *cas1*, *cas2* and *csn2* genes (Δ*cas1*, Δ*cas2*, Δ*csn2*, pH079). Error bars: mean ± s.d. (*n* = 3).
Extended Data Figure 3 | Generation of an experimental system for the overexpression of cas1, cas2 and csn2 and the detection of spacer acquisition in the absence of phage infection. a, Plasmids used in the spacer acquisition experiments presented in Figs 1c and 2c, d. pRH223 contains cas1, cas2 and csn2 from S. pyogenes under a tetracycline-inducible promoter. Cells containing this plasmid only acquired spacers when a second plasmid expressing cas9 was introduced, pRH240 or pRH241, containing the tracrRNA gene, the leader and first repeat from the S. pyogenes type II CRISPR–Cas system as well as cas9 from S. pyogenes (cas9Sp) or S. thermophilus (cas9St), respectively. The leader is a short, AT-rich sequence immediately upstream of the first repeat that contains the promoter for the transcription of the CRISPR array. b, Highly sensitive PCR assay to enrich for amplification products of adapted CRISPR loci. Arrows indicate primer annealing position and direction. The forward primer (JW8) anneals on the leader. For the reverse primer, a cocktail of JW3, JW4 and JW5 was used. The three reverse primers anneal on the repeat and differ only in their 3’-end nucleotide that never matches the last nucleotide of the leader (red arrowhead). Because this nucleotide is critical for the annealing of the primers, loci that acquire spacers ending in A, C or T are preferentially amplified over unadapted loci. c, To quantify the sensitivity of this technique, we mixed pGG32 (one repeat, unadapted) with pRH087 (repeat-spacer-repeat, adapted) in known ratios. The amplification of adapted plasmid was detected even when it represented 0.01% (10^-2) of the total plasmid template, representative of three technical replicates. This highly sensitive PCR assay is not required to detect acquisition during phage infection, as in this case adapted cells survive and are enriched within the population, making their detection much easier.
Extended Data Figure 4 | Purification of a Cas9–Cas1–Cas2–Csn2 complexes. a, The cas9–cas1–cas2–csn2 operon of S. pyogenes SF370 was cloned into the pET16b vector (generating pKW07) to add an N-terminal histidyl tag to Cas9 and express all proteins in E. coli. Purification was performed using Ni-NTA affinity chromatography. SDS–PAGE followed by Coomassie staining of the purified proteins revealed a co-purifying protein that was identified as Cas1 by mass spectrometry, in a result representative of five technical replicates. Mass spectrometry identification of all the eluted proteins co-purifying with Cas9 is shown in Extended Data Table 2. b, The cas9–cas1–cas2–csn2 operon of S. pyogenes SF370 was cloned into the pET23a vector (generating pKW06) to add a C-terminal histidyl tag to Csn2 and express all proteins in E. coli. Purification was performed using Ni-NTA affinity chromatography followed by ion exchange chromatography. The elution fractions that constituted the peak containing the complex (Fig. 3a) were separated by SDS–PAGE and visualized by Coomassie staining, representative of three technical replicates.
Extended Data Figure 5 | dCas9<sup>St</sup> can also support spacer acquisition. A plasmid derived from pRH241 containing mutations in the active site of <i>S. thermophilus</i> Cas9 (D10A, H847A; dCas9<sup>St</sup>) was used to characterize spacer acquisition in the absence of phage infection. Upon overexpression of Cas1, Cas2 and Csn2 using anydrotetracycline (aTc), we were able to detect spacer acquisition. Sequencing of spacers and alignment of the protospacer flanking sequences demonstrated the selection of an NGGNG PAM. The image is representative of three technical replicates.
Extended Data Figure 6 | A model for the selection of PAM-flanking spacers by Cas9. After injection of the phage DNA, an adaptation complex formed by Cas9, Cas1, Cas2 and Csn2 uses the Cas9 PAM binding domain to specify functional protospacers, that is, those followed by the correct PAM. It is not known how the protospacer sequence is extracted from the viral DNA to become a spacer. In the ‘cut and paste’ model, a nuclease, possibly Cas1, cuts the viral DNA to generate the spacer. In the ‘copy and paste’ model the protospacer sequence is copied first. Once loaded with the selected protospacer sequence, this complex promotes the integration of this sequence into the CRISPR array, thus becoming a new spacer. Previous studies demonstrated that Cas1 dimerizes and interacts with Cas2 (ref. 13); Csn2 has been determined to form a tetramer.34
### Extended Data Table 1 | Sequences of the spacers as obtained to sequence the logos in this study

| Figure | Spacer | Sequence | PAM | Target |
|--------|--------|----------|-----|--------|
| **2d** | 1 | gctaactgtagggctcttcattgtaag | tGGaG | phage |
| 2 | tattttttgtagattttactttactttagtgtaag | aGGaG | chromosome |
| 3 | ttttaagctattcattttaaaaggtcatat | gGGca | phage |
| 4 | tgaagaaatttttatacattgattattcaccaac | aGGca | phage |
| 5 | ttaggagatacttttactttccattacttaaatca | gGGAg | phage |
| 6 | tcggactgttagggtacgcgaagggcaaaa | gGGtG | chromosome |
| 7 | arrtttgcattttagaataatttttcttcttcaacaccatctatggc | caaaa | pRH223/243 |
| 8 | tcttatcttgataataagggtaactattgc | tGGtc | chromosome |
| 9 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 10 | gactactaatgacttgcactgcttcgatgaagaaatgattcaagaa | tGGta | chromosome |
| **3d** | 1 | aagtcgaacttcataatcatcgctttcgg | aGGtG | chromosome |
| 2 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 3 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 4 | actttttctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 5 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 6 | tcttatcttgataataagggtaactattgc | tGGtc | chromosome |
| 7 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 8 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 9 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 10 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 11 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 12 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 13 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |

**Fig. 5**

| Ext. Data | 1 | cgtaactgtagggctcttcattgtaag | tGGaG | phage |
| 2 | tattttttgtagattttactttactttagtgtaag | aGGaG | chromosome |
| 3 | ttttaagctattcattttaaaaggtcatat | gGGca | phage |
| 4 | tgaagaaatttttatacattgattattcaccaac | aGGca | phage |
| 5 | ttaggagatacttttactttccattacttaaatca | gGGAg | phage |
| 6 | tcggactgttagggtacgcgaagggcaaaa | gGGtG | chromosome |
| 7 | arrtttgcattttagaataatttttcttcttcaacaccatctatggc | caaaa | pRH223/243 |
| 8 | tcttatcttgataataagggtaactattgc | tGGtc | chromosome |
| 9 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 10 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 11 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 12 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 13 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 14 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 15 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 16 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 17 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 18 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 19 | atgagatgaggcgataaaagaacgtcgcta | tgtttgccctccaaatatgaaaacatggcc | chromosome |
| 20 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |
| 21 | ccattctctggttttctttgataataagggtaactattgc | tGGtc | chromosome |
| 22 | aagtttgggagctcattatcggctttttaac | tGGcG | chromosome |

©2015 Macmillan Publishers Limited. All rights reserved
| Accession | Protein                                      | % Coverage | Unique Peptides | Total peak area |
|-----------|----------------------------------------------|------------|----------------|-----------------|
| P77398    | Bifunctional polymyxin resistance protein ArnA (arnA) | 85.76      | 43             | 8.2×10⁸         |
| P60422    | 50S ribosomal protein L2 (rplB)               | 67.40      | 24             | 1.9×10⁸         |
| P17169    | Glucosamine--fructose-6-phosphate aminotransferase (glmS) | 79.31      | 38             | 1.8×10⁸         |
| P0AA43    | Ribosomal small subunit pseudouridine synthase A (rsuA) | 85.71      | 17             | 8.9×10⁸         |
| P0A9K9    | FKBP-type peptidyl-prolyl cis-trans isomerase (slyD) | 68.88      | 7              | 3.7×10⁹         |
| P0ACJ8    | Catabolite gene activator (crp)              | 82.86      | 18             | 5.4×10⁹         |
| P45395    | Arabinose 5-phosphate isomerase (kdsD)       | 73.17      | 21             | 1.2×10⁹         |
| P0A6F5    | 60 kDa chaperonin (groL)                     | 83.94      | 38             | 2.8×10⁹         |
| P0A9A9    | Ferric uptake regulation protein (fur)       | 78.38      | 8              | 1.2×10⁹         |
| P08622    | Chaperone protein DnaJ (dnaJ)                | 72.07      | 19             | 1.4×10⁹         |
| P00393    | NADH dehydrogenase (ndh)                    | 59.22      | 16             | 3.6×10⁸         |
## Extended Data Table 3 | Mass spectrometry analysis of protein bands from the purified Cas9–Cas1–Cas2–Csn2 complex

| Protein | % Coverage | Unique Peptides | Total peak area |
|---------|------------|-----------------|-----------------|
| Cas1    | 67.82      | 26              | $3.4 \times 10^8$ |
| Cas2    | 90.27      | 13              | $1.2 \times 10^9$ |
| Cas9    | 68.49      | 111             | $4.1 \times 10^8$ |
| Csn2    | 82.27      | 19              | $4.1 \times 10^3$ |
## Extended Data Table 4 | Oligonucleotides used in this study

| Primer | Sequence |
|--------|----------|
| B337   | gacgctattgtgcagttgctaaacctcttgattgatttct |