Primed acquisition and microhomology-mediated end-joining cooperate to confer specific CRISPR immunity against invasive genetic elements

Xiaojie Wang^{1#}, Bo Wu^{2#}, Zhufeng Zhang¹, Tao Liu¹, Yingjun Li¹, Guoquan Hu², Mingxiong He^{2*}, and Nan Peng^{1, 2*}

¹ State Key Laboratory of Agricultural Microbiology, College of Life Science and Technology, Huazhong Agricultural University, Wuhan, 430070, Hubei, P.R. China

² Key Laboratory of Development and Application of Rural Renewable Energy (Ministry of Agriculture), Biomass Energy Technology Research Centre, Biogas Institute of Ministry of Agriculture, Chengdu, 610041, Sichuan, P. R. China

[#]These authors contributed equally to this work.

^{*}To whom correspondence should be addressed

Tel: +86 27 8728 1267, Fax: +86 27 8728 0670, email: nanp@mail.hzau.edu.cn; or tel & fax: +86 28 8524 2281, email: hemingxiong@caas.cn

Abstract

CRISPR-Cas systems provide archaea and bacteria with adaptive immunity against invasive genetic elements through acquisition of invader-derived spacers. *De novo* spacer acquisition generally adapts spacers from both invaders and hosts with a bias towards invaders, while primed acquisition shows higher specificity to adapt spacers from invaders. Here, in *Zymomonas mobilis* subtype I-F system, primed acquisition showed much higher efficiency than *de novo* acquisition in this system. However, both routes recognised a large proportion of protospacers with the less conserved 5'-end CC PAM. Moreover, primed acquisition showed a preference towards protospacers located at the opposite strand and 3' direction of the priming protospacer sites, differing from the canonical subtype I-F system. Further, self-spacers were adapted at a higher frequency during *de novo* spacer acquisition, probably leading to self-interference. Importantly, this species employed microhomology-mediated end-joining (MMEJ) for repair of host DNA breaks guided by self-targeted spacers, and overexpression of a host NAD⁺-dependent Ligase-A significantly increased the repair efficiency. In summary, our findings demonstrate that *Z. mobilis* uses primed acquisition for higher specific uptake of invader DNA and employs MMEJ to repair host DNA breaks guided by self-targeted spacers, showing specific immunity against invasive genetic elements.

CRISPR-Cas systems, consisting of CRISPR arrays containing identical repeats separated by unique spacers and associated *cas* genes, defend Bacteria and Archaea against invasive plasmids and viruses ¹⁻³. CRISPR-Cas adaptive immunity occurs in three stages: acquisition of *de novo* spacers, crRNA biogenesis, and nucleic acid targeting and cleavage ^{2,4}. CRISPR-Cas systems acquire DNA fragments of invading nucleic acids (prespacers) and integrate them into CRISPR arrays as spacers, thus forming heritable immunological memory ⁵. DNA motifs in invasive genetic elements, or in CRISPR loci, play crucial roles in the spacer-acquisition process. For example, the protospacer-adjacent motif (PAM) associated with protospacers ⁶⁻¹⁰ and the leader-repeat junction ^{10,11} have been found to direct spacer acquisition in different CRISPR-Cas systems. Recently, it was found that CRISPR spacer acquisition requires genome-stability proteins ^{12,13} as well as *trans*- and *cis*-acting factors that maintain integration specificity ^{14,15}. In a *de novo* spacer acquisition assay, *E. coli* subtype I-E Cas1 and Cas2 have been found to be required ^{10,16-18}. Furthermore, the target DNA interference complex is required in the primed spacer acquisition process, which is triggered by a pre-existing spacer matching the target DNA ^{16,17,19-21}.

Discrimination of self and non-self DNA is crucial for immunity. CRISPR-Cas systems have developed diverse mechanisms to avoid autoimmunity, including a

PAM requirement for target interference^{22,23}. However, because *de novo* spacer acquisition adapts new spacers from host genomic DNA^{9,24}, avoiding host DNA sampling at the CRISPR spacer acquisition stage is more important. *E. coli* has developed a strategy to limit prespacer supply from the host through *Chi* site-assisted reduction of DNA resection in its genome²⁵. Further, primed adaptation allows cells to effectively counter viruses and plasmids that escape CRISPR interference^{16,26-28}, leading to highly efficient acquisition of new *cis*-located spacers of the protospacer of the invading genetic elements with mutations^{7,21,29,30}. Primed adaptation requires functional Cas1, Cas2, and Cas3 proteins, suggesting a functional link between CRISPR interference and primed adaptation^{20,31,32}.

However, adaptation of host self DNA cannot be avoided. In *de novo* spacer acquisition process, 32%, 7%, 16%, and 22.8% or 1.8% (for induction or non-induction of Cas1, Cas2 expression, respectively) of new spacers have been derived from host genomic DNA in *S. thermophilus* subtype II-A²⁴, *S. islandicus* subtype I-A⁹, *Pectobacterium atrosepticum* subtype I-F³³, and *E. coli* subtype I-E²⁵ systems. Even in primed spacer acquisition processes, 0.01–0.03% of new spacers have been derived from host genomic DNA in *E. coli* subtype I-E³⁰ and *P. atrosepticum* subtype I-F³³ systems, suggesting that host cells might employ different mechanisms to avoid self-spacer guided interference. Here, we studied CRISPR spacer acquisition and interference of the *Zymomonas mobilis* subtype I-F system, to demonstrate that *Z. mobilis* employs primed spacer acquisition for specific uptake of invader DNA and utilizes the MMEJ system to maintain genome stability after self DNA interference.

Results

Identification of leaders and PAM variants for primed spacer acquisition

Z. mobilis ZM4 encodes a subtype I-F CRISPR-Cas system, including three CRISPR arrays and an adaptation and a DNA interference modules (Fig. 1a). Here, we investigated the effect of leader sequences and PAM variants on primed acquisition in *Z. mobilis*. Plasmids constructed with “NN-S1” cassettes were transformed into *Z. mobilis* ZM4mrr (Fig. 1b). The NC or CN PAM (N = A, T or G) of the priming protospacer (PPS) triggered strong spacer acquisition at CRISPR locus 1 (C1), while no new spacer was detected in the negative control sample carrying the empty vector pZM15Asp. However, CC PAM, which conferred strong DNA interference (Fig. S1)³⁴, triggered weak acquisition (Fig. 1c), due to a single nucleotide mutation at CC PAM (CC to CT) or protospacer sequence, or a deletion of the 3′ portion of the PPS on the plasmid. In contrast, other NN combinations could not trigger strong primed spacer acquisition (Fig. 1c). Further, CRISPR locus 2 and 3 showed similar acquisition patterns to locus 1, indicating that the leader sequences were all functional (Fig. 1c). However, acquisition efficiency at locus 2 and 3 were weaker than locus 1

(Fig. 2d), demonstrating the leader sequence variations had an effect on primed spacer acquisition (Fig. S2). Moreover, by extending the culture time, expanded bands on the gel representing new spacers became more apparent than the parental bands (Fig. 1d), and TT, GT, and TG PAM sequences triggered primed acquisition at 5 (Fig. 1d), 10, and 15 days (Fig. S3), respectively, suggesting NT or TN PAMs, in addition to CN or NC PAMs, had the ability to promote priming.

Self-discrimination and PAM recognition in *de novo* and primed spacer acquisition

Cells carrying pEZ15Asp, pAC-S1, and pTT-S1 (Fig. 2a) were cultured for 5 days to analyse acquisition efficiency. Plasmid pEZ15Asp didn't induce detectable acquisition, even with the extended cultivation time or the presence/absence of antibiotics (Fig. 2b). The pCC-S1 construct showed strong primed acquisition in $\pm$ antibiotic medium with the extended culture (Fig. 2b). The pTT-S1 plasmid only induced moderate acquisition in antibiotic-free medium (Fig. 2b). The expanded bands of day 1 antibiotic-free cultures were purified and analysed by high-throughput sequencing. The efficiencies of *de novo* spacer acquisition and pTT-S1 induced primed acquisition were $\sim 0.04\%$ and $\sim 4.70\%$ of the pAC-S1 induced primed acquisition (Fig. 2c). Most of the new spacers ($>99\%$) were adapted from plasmid DNA in pAC-S1 and pTT-S1 primed acquisition, while only $\sim 80\%$ of new spacers were adapted from plasmid DNA in pEZ15Asp-induced *de novo* spacer acquisition (Fig. 2c), indicating more efficient self-discrimination during primed acquisition than *de novo* acquisition in the *Z. mobilis* subtype I-F system.

Further, all constructs recognised protospacers with a high proportion of non-canonical CC PAM. Approximately 50% of protospacers identified during *de novo* acquisition had NC/CN PAM sequences, while $\sim 20\%$ from both primed acquisition experiments had NC or CN PAMs (Fig. 2d), which is discordant with other studies on subtype I-F systems reporting that primed protospacers strictly required canonical PAM sequences^{33,35}. Protospacers with NC or CN PAMs could probably be recognised at the canonical CC PAM but were acquired by sliding with -2, -1, +1, or +2 nucleotides related to CC PAM (Fig. 2e and f), confirming that acquisition of these protospacer was initiated at a canonical CC PAM but not a random PAM.

Identification of seed sequence for DNA interference and primed acquisition

Systematic transversion substitution was introduced at each nucleotide of the protospacer in the pCC-S1 construct (Fig. 3a). Single-nucleotide mutations at nucleotides 1–5 and 7 (partially) rescued transformation efficiency (Fig. 3b), indicating that these nucleotides acted as the seed sequence for CRISPR interference. Single colonies of these transformants were isolated, and sequencing results revealed

no additional mutations at the PAM or protospacer sequences. However, designed mutations at this region triggered moderate primed spacer acquisition (Fig. 3c). Mutations at nucleotides 6 and 8–16 had no effect on CRISPR interference (Fig. 3b). From isolated mutant colonies, plasmids in some cells showed additional mutations at the PAM or protospacer sequences, resulting in weak primed acquisition, while the intact plasmid in construct M14 induced strong primed acquisition (Fig. 3c).

Further, mutations at the region from 17–29 nucleotides were found to partially rescue transformation efficiency (Fig. 3b). Most single colonies of these transformants showed additional mutations at the protospacer sequence on the plasmid, and weak primed acquisition was found (Fig. 3c). Nucleotides 30–32 were not essential for interference (Fig. 3b). However, only two single colonies were isolated for M30, and the intact plasmid in these two colonies showed the strongest primed acquisition efficiency (Fig. 3c).

Priming protospacer determines location and strand biases of new acquisitions

New spacers were visualised by mapping their cognate protospacer locations on the plasmids (Fig. 4). Protospacer mapping was remarkably consistent across the replicates for the colonies carrying pEZ15Asp, pAC+ and pAC- plasmids (Fig. S4). In the *de novo* acquisition experiment (colonies carrying pEZ15Asp), protospacer hotspots were identified near the origin for plasmid replication in *Z. mobilis*, however, another hotspot was identified at the negative strand on the antibiotic-resistance gene in two replicates (Fig. S4a). Primed spacer acquisitions showed distinct patterns for protospacer distributions related to the PPSs (Fig. 4a and b). A significant protospacer peak was found on the same strand of PPS around the priming sites (Fig. 4a and b). On the scale of whole plasmid DNA, primed spacer acquisition showed preference for the paired strand related to the location of the PPSs (Fig. 4c and d) and showed preference for the 3' regions of both DNA strands related to the location the PPS (Fig. 4c and e).

Microhomology-mediated end-joining of self-targeted DNA breaks

To study the host cell to escape targeting guided by adapted self-spacers, we constructed self-interference plasmids carrying mini-CRISPR cassette encoding spacers matching a non-essential gene (*Zmo0346*) and an putative essential folylpolyglutamate synthase gene (*Zmo0582*) (Fig. 5a). Transformation efficiency of pCC-S1 was very low, since this plasmid was targeted by the host CRISPR complex guided by the CRISPR locus 1-derived spacer (Fig. 5b). Similarly, transformation efficiency of the pS0582 plasmid, which encoded a spacer matching a protospacer on the putative essential gene *Zmo0582*, was low, indicating that the *Z. mobilis* CRISPR-

Cas system cleaved its genomic DNA, leading to cell death without (or with inefficient) DNA damage repair (Fig. 5b). Surprisingly, transformation efficiency of the pS0346 plasmid, encoding a spacer matching a protospacer on the non-essential gene *Zmo0346*, was much higher than that of pS0582 (Fig. 5b). This result suggested CRISPR RNA-guided DNA double strand break probably was repaired by some mechanism with very high efficiency.

To explain the phenomenon, dozens of single colonies transformed with pS0346 and pS0582 plasmids were selected, and the target gene loci were PCR-amplified for sequencing. Different deletion mutations or point mutations were identified at the *Zmo0346* gene (Fig. 5c) and *Zmo0582* gene loci (Fig. 5d). For the essential gene *Zmo0582*, most mutations occurred at the CC PAM or the seed sequence of the protospacer (Fig. 5d), leading to host escape of self-interference. For the non-essential *Zmo0346* gene, deletion regions covering the protospacer were identified, leading to escape of CRISPR interference (Fig. 5c). Deletions at the *Zmo0346* gene locus were different in length (63-531 bp); however, short direct repeats, 6–10 bp in length, flanking the deletion regions were identified (Fig. 5c). Similar result was also found for the experiments designed to target other genes. For example, self-targeting at *Zmo1062* gene resulted in ~75% randomly selected colonies was repaired with deletion mutations covering the target site (Fig. S5). We also sequenced the plasmids from the single colonies transformed with the pCC-S1 plasmid and found point mutations at the PAM sequence (20%) or deletion mutations at the protospacer sequence (~10%). Analysis of the deletion mutations on the plasmid found a 5-bp direct repeat adjacent to the deletion region (Fig. 5e), indicating that CRISPR interference at both genomic and plasmid loci induced microhomology-dependent end-joining (MMEJ), resulting in deletion mutations.

We further studied whether the replicative NAD⁺-dependent Ligase-A (LigA) was involved in MMEJ. Overexpression *ligA* gene on the plasmid reduced the transformation efficiency (Fig. S6), suggesting cytotoxicity. Further, the *ligA* gene was overexpressed on the pS0346 self-interference plasmid (pS0346OEligA), and its end-joining efficiency was assessed. Transformants carrying the control plasmid pEZ15Asp showed normal (large) colonies on the antibiotic plates; however, transformants carrying the self-interference plasmid pS0346 showed many small colonies and few normal colonies (Fig. 5f). PCR analysis indicated that ~46% of the random selected colonies transformed with pS0346 showed MMEJ repair at the *Zmo0346* locus (Fig. 5g). All transformants carrying the pS0346OEligA plasmid formed large colonies on the plates, similar to the control plasmid pEZ15Asp (Fig. 5f). PCR analysis indicated that ~60% of the randomly selected colonies overexpressing LigA showed MMEJ repair at the *Zmo0346* locus. However, only ~10% of the normal colonies carrying pS0346 interference plasmid were MMEJ

repaired (Fig. 5g), indicating that *ligA* overexpression increased efficiency of MMEJ repair of the self-targeted DNA break.

Discussion

Diversity of primed acquisition in different CRISPR subtypes

Primed spacer acquisition has been studied in different CRISPR-Cas subtypes [7,16,20,21,36](#). In *Z. mobilis* subtype I-F system, more than 50% of PAM mutations promoted primed spacer acquisition (Fig. 1C and Fig. S2), similar to other systems, such as *E. coli* subtype I-E [37](#) and *H. hispanica* subtype I-B [7](#). Priming sampled protospacers mostly from the targeted plasmid in *Z. mobilis* subtype I-F system, consistent with previous studies which reported that priming reinforces immunity against invasive genetic elements [38](#). In nearly all of the studied primed acquisition systems, for example, *E. coli* subtype I-E [16](#), *H. hispanica* subtype I-B [17](#), *P. atrosepticum* subtype I-F [33](#) and *L. pneumophila* subtype I-C [36](#), primed acquisition is strictly dependent on recognition of a canonical PAM. It was found that >90% of the identified protospacers harboured a canonical PAM sequence; however, in the *Z. mobilis* subtype I-F system, only ~77% of identified protospacers had a canonical CC PAM (Fig. 2d). Further analysis of nearby sequences of protospacers with less canonical PAM, a ± 2 nt regions showed identity with CC PAM, indicating a slipped PAM recognition pattern as reported previously [33,36](#), but occurring at a higher frequency. These slipped-acquired spacers might guide the CRISPR surveillance complex to recognise the NC or CN PAM (Fig. 2e), therefore triggering a second-round priming.

Priming acquired new spacers close to the PPS, and the protospacer distribution of most subtypes was strand- and direction-biased [38](#). Previous studies demonstrated that subtypes I-B, I-C, and I-E showed biases towards the acquisition of new spacers from the same strand as the PPS [17,30,36](#); however, the *P. atrosepticum* subtype I-F system acquired spacers almost equally from both strands [21,33](#). Here, priming in *Z. mobilis* subtype I-F system was biased towards acquired new spacers from the paired strand of the PPS on the whole plasmid DNA level (Fig. 4d), opposite to other systems. However, priming highly preferred the same strand around the PPS regions in *Z. mobilis* subtype I-F system (Fig. 4a and b). Regarding directional bias, the *P. atrosepticum* subtype I-F system showed a priming bias towards the 5' direction [33](#). In contrast, in *Z. mobilis* subtype I-F system, priming was biased towards the 3' direction of the PPS site (Fig. 4e), similar to the I-B, I-C, and I-E systems [38](#). Asymmetric protospacer distribution was generally found in our priming experiments (Fig. 5b–d), similar to the *P. atrosepticum* subtype I-F system [33](#). Protospacers that mapped near the PPS site showed clustering. However, a high density of protospacers

at the far 3' region of PPS was detected (Fig. 4b and c), opposite of the I-F system ^{21,33}, but similar to I-B, I-C, and II-A systems ³⁸.

Primed acquisition and MMEJ cooperate to avoid self-targeting

In this study, we found primed spacer acquisition was much more robust than *de novo* acquisition (Fig. 2b), showing greater specificity to invasive DNA (Fig. 2c). However, the pre-existing spacers required *de novo* acquisition for “first step” integration (Fig. 6). While, similar to other systems, *de novo* acquisition adapted a large proportion of spacers from the host genomic DNA in our system (Fig. 2c), possibly resulting in self-targeting (Fig. 6). In *Z. mobilis*, transformation of self-targeting plasmids still resulted in a relatively high transformation efficiency (Fig. 5b). Further analysis of the target locus of transformants revealed that the MMEJ pathway was involved in repair of the host DNA DSBs, rescuing the targeted host cells (Fig. 5c and Fig. 6). The MMEJ pathway is an error-prone repair process, resulting in deletion of target regions covering the protospacer site (Fig. 5c) ³⁹ and finally avoiding self-targeting. Therefore, *Z. mobilis* developed primed acquisition and robust DNA repair systems to promote specific CRISPR immunity against invasive genetic elements and avoid self-targeting (Fig. 6).

MMEJ also repair the targeted plasmid DNA (Fig. 5b). Although very few single colonies were isolated, ~10% of these single colonies were found carrying a short deletion (27 bp) at the protospacer and its flanking region (Fig. 5e). In the *de novo* acquisition, protospacers on the invasive genetic elements formed a hotspot at the replication origin (Fig. 4a) ²⁵, while most protospacers from the host genomic DNA were located on non-essential genes (Table S3). Therefore, MMEJ-mediated deletion of protospacers flanking the Zmo origin would probably hinder plasmid replication, while deletion of non-essential genes was nonlethal for the cells.

Error-prone repair systems contributing to avoid CRISPR self-targeting could be widespread in bacteria. For example, the non-homologous end-joining (NHEJ) pathway is co-present with the CRISPR-Cas system in some bacterial species ^{40,41}. However, the NHEJ system has no effect on CRISPR immunity when both NHEJ genes and CRISPR-Cas subtype II-A system were co-expressed in a heterogeneous host ⁴⁰. Furthermore, *Mycobacterium smegmatis* encodes subtype III-A and NHEJ systems; its NHEJ system only works for heterogenous CRISPR-FnCpf1 cleaved host DNA, not for the heterogenous CRISPR-TdCas9_m or CRISPR-NmCas9 systems ⁴¹. Another end-joining repair system, MMEJ, found in *E. coli* has been experimentally demonstrated to repair DNA DSBs ^{39,42}. Here, the repair efficiency for DSBs by MMEJ system in *Z. mobilis* was significantly higher than that in *E. coli* ($\sim 3.0 \times 10^{-1}$ vs $\sim 3.0 \times 10^{-5}$, Table S2) ^{39,43}. In summary, our results demonstrated the function of the MMEJ system to cooperate with the CRISPR-Cas system to repair CRISPR self-

targeted host DNA, therefore promoting immune specificity against invasive genetic elements.

MATERIALS AND METHODS

Strains, growth, and transformation of *Z. mobilis*

Z. mobilis strains, including ZM4, ZM4mrr⁴⁴, and ZM4mrr carrying different plasmids, were cultured in RM medium (20 g/L glucose, 10 g/L yeast extract, and 2 g/L K₂HPO₄) at 37°C. *Z. mobilis* ZM4mrr lacking an endonuclease gene was used as the genetic host in this study. Plasmid DNA, modified in the *E. coli* Trans10 strain, was transformed into *Z. mobilis* ZM4mrr by electroporation, and transformants were selected on RM medium agar plates with 100 µg/mL spectamycin. Single colonies from the plates were cultured in RM liquid medium for cell growth. *E. coli* Trans10 cells used for DNA cloning were cultured at 37°C in Luria–Bertani medium, and spectamycin was added to the culture to a final concentration of 100 µg/mL, if required.

Construction of plasmids

The PAM variable challenging (or priming) plasmids were constructed by cloning the “NN-S1” (NN = any nucleotide, S1 = the first spacer of CRISPR locus 1 of *Z. mobilis* ZM4) oligonucleotides into the shuttle vector pEZ15Asp at EcoRI/XbaI restriction sites, resulting in 16 (4²) challenging plasmids. The seed sequence variable priming plasmids were constructed by introduction of a transversion substitution at each nucleotide of the “Spacer 1” of the plasmid (pCC-S1) carrying the “CC-Spacer 1” cassette. The interference plasmids carrying the expression cassette of “Repeat-Spacer-Repeat” against genomic loci, Zmo0346 and Zmo0582 genes, were constructed by cloning the 32-bp sequences following a CC PAM on these gene coding sequences into the pEZ15Asp vector under control of the *casI* gene promoter (*PcasI*), resulting in interference plasmid pS0346 and pS0582. Zmo0364 (*ligA* gene) overexpression plasmids were constructed by cloning the Zmo0364 open reading frame under control of the Zmo0367 promoter⁴⁵ into pEZ15Asp or the interference plasmid pS0346.

PCR amplification of the leader-proximal CRISPR regions

Z. mobilis strains harbouring an empty vector (pEZ15Asp) or the challenging (priming) plasmids were cultured in 10 mL of RM medium at 37°C without shaking for one day or passaged daily for 15 days by transferring 100 µL cell culture to 10 mL fresh medium with or without antibiotics. Samples of each culture (0.1 mL) were taken at a desired time point, and total DNA from these cells was used as the PCR template. The leader-proximal regions of three CRISPR loci or only CRISPR locus 1

were amplified by PCR using Taq polymerase and the forward/reverse primer sets (Table S1) for locus 1, 2, or 3, respectively. PCR products were separated on a 1.5% agarose gel and visualised by ethidium bromide staining.

High-throughput sequencing and bioinformatics analysis

Three single colonies of each transformant were selected from the plate and inoculated in RM medium for growth. Genomic DNA was extracted from each enlarged culture of these single colonies and used for amplification of the leader-proximal region of CRISPR locus 1. Sequencing was carried out as described previously⁴⁶. Using the BLASTN program against pEZ15Asp plasmid or pEZ15Asp carrying PPS or the *Z. mobilis* ZM4 genome sequence⁴⁷, the protospacer sequences were identified for new integrated spacers. The 2-bp sequence at the 5'-end of each protospacer was considered as the PAM region. Perl scripts were run to analyse the protospacers. All high-throughput sequencing reads have been deposited into the SRA database with the accession number PRJNA579098. The protospacer locations of the newly acquired spacers in the CRISPR locus from *de novo* acquisition or primed acquisition were mapped on pEZ15Asp and the priming plasmids pAC-SS1(+) and pAC-S1(-) (pAC+ and pAC- for short), respectively, using a sliding 150 nt binning window, as described previously³³.

References

- 1 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).
- 2 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:10.1126/science.1159689 (2008).
- 3 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).
- 4 Westra, E. R. *et al.* The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet* **46**, 311-339, doi:10.1146/annurev-genet-110711-155447 (2012).
- 5 van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M. & Brouns, S. J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34**, 401-407, doi:10.1016/j.tibs.2009.05.002 (2009).
- 6 Yosef, I. *et al.* DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. *Proc Natl Acad Sci U S A* **110**, 14396-14401, doi:10.1073/pnas.1300108110 (2013).
- 7 Li, M., Wang, R. & Xiang, H. Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res* **42**, 7226-7235, doi:10.1093/nar/gku389 (2014).
- 8 Erdmann, S. & Garrett, R. A. Selective and hyperactive uptake of foreign DNA by

366 adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* **85**,
367 1044-1056, doi:10.1111/j.1365-2958.2012.08171.x (2012).

368 9 Liu, T. *et al.* Transcriptional regulator-mediated activation of adaptation genes triggers
369 CRISPR de novo spacer acquisition. *Nucleic Acids Res* **43**, 1044-1055,
370 doi:10.1093/nar/gku1383 (2015).

371 10 Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the
372 CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* **40**, 5569-5576,
373 doi:10.1093/nar/gks216 (2012).

374 11 Wei, Y., Chesne, M. T., Terns, R. M. & Terns, M. P. Sequences spanning the leader-repeat
375 junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic*
376 *Acids Res* **43**, 1749-1758, doi:10.1093/nar/gku1407 (2015).

377 12 Ivancic-Bace, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability
378 proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic*
379 *Acids Res* **43**, 10821-10830, doi:10.1093/nar/gkv1213 (2015).

380 13 Leon-Sobrinho, C., Kot, W. P. & Garrett, R. A. Transcriptome changes in STSV2-infected
381 Sulfolobus islandicus REY15A undergoing continuous CRISPR spacer acquisition. *Mol*
382 *Microbiol* **99**, 719-728, doi:10.1111/mmi.13263 (2016).

383 14 Nunez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR
384 immunological memory requires a host factor for specificity. *Mol Cell* **62**, 824-833,
385 doi:10.1016/j.molcel.2016.04.027 (2016).

386 15 McGinn, J. & Marraffini, L. A. CRISPR-Cas systems optimize their immune response by
387 specifying the site of spacer integration. *Mol Cell* **64**, 616-623,
388 doi:10.1016/j.molcel.2016.08.038 (2016).

389 16 Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas
390 adaptive bacterial immunity system. *Nat Commun* **3**, 945, doi:10.1038/ncomms1937
391 (2012).

392 17 Li, M., Wang, R., Zhao, D. & Xiang, H. Adaptation of the Haloarcula hispanica CRISPR-
393 Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res* **42**,
394 2483-2492, doi:10.1093/nar/gkt1154 (2014).

395 18 Arslan, Z., Hermanns, V., Wurm, R., Wagner, R. & Pul, U. Detection and characterization
396 of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res*
397 **42**, 7884-7893, doi:10.1093/nar/gku510 (2014).

398 19 Diez-Villasenor, C., Guzman, N. M., Almendros, C., Garcia-Martinez, J. & Mojica, F. J.
399 CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition
400 specificities among CRISPR-Cas I-E variants of Escherichia coli. *RNA Biol* **10**, 792-802,
401 doi:10.4161/rna.24023 (2013).

402 20 Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation.
403 *Proc Natl Acad Sci U S A* **111**, E1629-1638, doi:10.1073/pnas.1400071111 (2014).

404 21 Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent
405 spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res* **42**,
406 8516-8526, doi:10.1093/nar/gku527 (2014).

407 22 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif
408 sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*
409 **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).

410 23 Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus*
411 *thermophilus*. *J Bacteriol* **190**, 1390-1400, doi:10.1128/jb.01412-07 (2008).

412 24 Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in Type II-
413 A CRISPR-Cas adaptation. *Genes Dev* **29**, 356-361, doi:10.1101/gad.257550.114 (2015).

414 25 Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign
415 DNA. *Nature* **520**, 505-510, doi:10.1038/nature14302 (2015).

416 26 Fineran, P. C. & Charpentier, E. Memory of viral infections by CRISPR-Cas adaptive
417 immune systems: acquisition of new information. *Virology* **434**, 202-209,
418 doi:10.1016/j.virol.2012.10.003 (2012).

419 27 Swarts, D. C., Mosterd, C., van Passel, M. W. & Brouns, S. J. CRISPR interference
420 directs strand specific spacer acquisition. *PLoS One* **7**, e35888,
421 doi:10.1371/journal.pone.0035888 (2012).

422 28 Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* **356**,
423 doi:10.1126/science.aal5056 (2017).

424 29 Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic
425 repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**,
426 10098-10103, doi:10.1073/pnas.1104144108 (2011).

427 30 Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. & Severinov, K. High-
428 throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol* **10**,
429 716-725, doi:10.4161/rna.24325 (2013).

430 31 Kunne, T. *et al.* Cas3-derived target DNA degradation fragments fuel primed CRISPR
431 adaptation. *Mol Cell* **63**, 852-864, doi:10.1016/j.molcel.2016.07.011 (2016).

432 32 Dillard, K. E. *et al.* Assembly and translocation of a CRISPR-Cas primed acquisition
433 complex. *Cell* **175**, 934-946.e915, doi:10.1016/j.cell.2018.09.039 (2018).

434 33 Staals, R. H. *et al.* Interference-driven spacer acquisition is dominant over naive and
435 primed adaptation in a native CRISPR-Cas system. *Nat Commun* **7**, 12853,
436 doi:10.1038/ncomms12853 (2016).

437 34 Zheng, Y. *et al.* Characterization and repurposing of the endogenous Type I-F CRISPR-
438 Cas system of *Zymomonas mobilis* for genome engineering. *Nucleic Acids Res*,
439 doi:10.1093/nar/gkz940 (2019).

440 35 Vorontsova, D. *et al.* Foreign DNA acquisition by the I-F CRISPR-Cas system requires all
441 components of the interference machinery. *Nucleic Acids Res* **43**, 10848-10860,
442 doi:10.1093/nar/gkv1261 (2015).

443 36 Rao, C., Chin, D. & Ensminger, A. W. Priming in a permissive type I-C CRISPR-Cas
444 system reveals distinct dynamics of spacer acquisition and loss. *RNA* **23**, 1525-1538,
445 doi:10.1261/rna.062083.117 (2017).

446 37 Xue, C. *et al.* CRISPR interference and priming varies with individual spacer sequences.
447 *Nucleic Acids Res* **43**, 10831-10847, doi:10.1093/nar/gkv1259 (2015).

448 38 Nicholson, T. J. *et al.* Bioinformatic evidence of widespread priming in type I and II
449 CRISPR-Cas systems. *RNA Biol* **16**, 566-576, doi:10.1080/15476286.2018.1509662
450 (2019).

451 39 Chayot, R., Montagne, B., Mazel, D. & Ricchetti, M. An end-joining repair mechanism in
452 *Escherichia coli*. *Proc Natl Acad Sci U S A* **107**, 2141-2146,
453 doi:10.1073/pnas.0906355107 (2010).

- 454 40 Bernheim, A. *et al.* Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in
455 bacteria. *Nat Commun* **8**, 2094, doi:10.1038/s41467-017-02350-1 (2017).
- 456 41 Sun, B. *et al.* A CRISPR-Cpf1-assisted non-homologous end joining genome editing
457 system of Mycobacterium smegmatis. *Biotechnology journal* **13**, e1700588,
458 doi:10.1002/biot.201700588 (2018).
- 459 42 Bhattacharyya, S. *et al.* Phage Mu Gam protein promotes NHEJ in concert with
460 Escherichia coli ligase. *Proc Natl Acad Sci U S A* **115**, E11614-e11622,
461 doi:10.1073/pnas.1816606115 (2018).
- 462 43 Huang, C. *et al.* CRISPR-Cas9-assisted native end-joining editing offers a simple strategy
463 for efficient genetic engineering in Escherichia coli. *Appl Microbiol Biotechnol* **103**,
464 8497-8509, doi:10.1007/s00253-019-10104-w (2019).
- 465 44 Kerr, A. L., Jeon, Y. J., Svenson, C. J., Rogers, P. L. & Neilan, B. A. DNA restriction-
466 modification systems in the ethanologen, Zymomonas mobilis ZM4. *Appl Microbiol*
467 *Biotechnol* **89**, 761-769, doi:10.1007/s00253-010-2936-1 (2011).
- 468 45 Yang, Y. *et al.* Prediction and characterization of promoters and ribosomal binding sites of
469 Zymomonas mobilis in system biology era. *Biotechnol Biofuels* **12**, 52,
470 doi:10.1186/s13068-019-1399-6 (2019).
- 471 46 Zhang, Z., Pan, S., Liu, T., Li, Y. & Peng, N. Cas4 Nucleases Can Effect Specific
472 Integration of CRISPR Spacers. *J Bacteriol* **201**, e00747-00718, doi:10.1128/JB.00747-
473 18 (2019).
- 474 47 Yang, S. *et al.* Complete genome sequence and the expression pattern of plasmids of the
475 model ethanologen Zymomonas mobilis ZM4 and its xylose-utilizing derivatives 8b and
476 2032. *Biotechnol Biofuels* **11**, 125, doi:10.1186/s13068-018-1116-x (2018).

477

478 **Acknowledgements**

479 This work was supported by the National Natural Science Foundation of China (No.
480 31671291 to N.P., No. 31570055 to M.H. and 31801035 to Y.L.), and the
481 Fundamental Research Funds for the Central Universities (No. 2662019PY028 to
482 N.P.). Funding for open access charge: National Natural Science Foundation of China
483 (No. 31671291).

484

485 **Author contributions**

486 X. W. and B. W. conducted the experiments, Z. Z., T. L. and Y. L. analyzed the
487 acquisition data, M. H., G.H. and N. P. designed the experiments and analyzed the
488 data. N.P. wrote the manuscript.

489

Figure legends

Figure 1. Identification of leader and PAM variants for primed spacer

acquisition in *Z. mobilis* ZM4. **a**, organisation of CRISPR-Cas subtype I-F module in *Z. mobilis* ZM4. Three CRISPR arrays (C1, C2, and C3) were indicated. **b**, Schematic of the primed spacer acquisition assay. The plasmids carrying a protospacer matching spacer 1 of CRISPR locus 1 adjacent to different PAMs (NN, N = A, T, C or G) were transformed into *Z. mobilis* ZM4mrr. Three single colonies of each transformants were selected and cultured for 1 day in RM medium. The *cas* genes are indicated in grey arrows, repeats are indicated in black squares, and spacers are indicated by diamonds. **c**, Detection of spacer acquisition at three CRISPR loci (C1, C2, and C3) on 1.5% agarose gels. Each colony of transformant was cultured in liquid medium for 1 day and used for amplification of the leader proximal regions. The parental and expanded bands are indicated. These data represent three independent spacer acquisition analyses for each construct. **d**, cells used for amplification of the leader proximal region of CRISPR locus 1 were cultured in the same medium without antibiotics for 5days. For the extended culture, 1% cells were transferred into fresh medium daily. Red arrows indicate newly expanded bands identified after extended culture. Each gel is a representative of three repeated experiments using three independent single colonies for each construct.

Figure 2. Comparison of *de novo* and primed spacer acquisition.

a, Schematic of *de novo* and primed spacer acquisition assay. The empty plasmid (pEZ15Asp) or plasmid carrying protospacers matching a pre-existing spacer 1 of CRISPR locus 1 adjacent to two atypical PAM sequences (AC and TT) were transformed into *Z. mobilis* ZM4mrr. **b**, Detection of spacer acquisition at CRISPR locus 1 of *Z. mobilis* transformants cultured in medium with or without antibiotics on 1.5% agarose gels. Each gel is a representative of three repeated experiments using three independent single colonies for each construct. **c**, The PCR-amplified leader-proximal region containing newly adapted spacers was analysed by high-throughput sequencing, and the data are summarised. Data were collected from three independent colonies of each transformant (two independent colonies for pEZ15Asp transformant). **d**, analysis of the PAM sequences of protospacers identified in the high-throughput sequencing. **e**, schematic representation of slipping, which results in capture of spacers that map to non-canonical PAMs. Spacer sequence is indicated by black letters, and 5'-end CC PAM is indicated by (specify symbol/identifier here). **f**, frequency of slipping events observed in each dataset. Others indicated that 5'-end of protospacer was >2 nt from the CC PAM.

Figure 3. Seed sequence of the protospacer for DNA interference and primed spacer acquisition.

a, Protospacer sequence matching spacer 1 of CRISPR locus 1 with a 5'-CC PAM was cloned into pEZ15Asp vector. A single transversion substitution was introduced at each nucleotide. **b**, each plasmid mutant was transformed into *Z. mobilis* ZM4mrr, and transformation efficiencies related to the pEZ15Asp plasmid are shown. **c**, Detection of spacer acquisition at CRISPR locus 1

on 1.5% agarose gels. Each colony of transformants was cultured in liquid medium for 1 day and used for amplification of the leader proximal regions. The parental and expanded bands are indicated. These data represent three independent spacer acquisition analyses for each construct.

Figure 4. Protospacer distribution reveals biased spacer incorporation. The protospacer locations were mapped on **a**, pAC⁺, and **b**, pAC⁻ plasmids using a sliding 150 nt binning window. Protospacers on the plus and minus strand are indicated by blue and orange, respectively. PAM distributions in a sliding 150 nt binning window are indicated by black lines. The position of priming protospacer (PPS) in pAC⁺ and pAC⁻ plasmids is indicated with a red dash line and arrows. Plasmid-encoded genes and cis acting elements are indicated by arrows. **c**, Proportion of protospacers distribution at the 5' and 3' of the PPS on each strand. PPS is indicated by a red dash line, and the direction is indicated by a red arrow. Blue and orange squares indicate the plus and minus strands, respectively. **d**, proportion of protospacers present on each strand. Blue and orange squares indicate the plus and minus strands, respectively. **e**, Proportion of protospacers distribution at the 5' and 3' of the PPS on both strands. Data shown represent mean $\pm$ SD for the different three (for pAC⁺) or two (for pAC⁻) independent transformants of each plasmid.

Figure 5. Microhomology-mediated end-joining repair of self-targeted DNA. **a**, schematic of self-targeting assay. Self-targeting plasmids carrying "Repeat-Spacer-Repeat" expression cassettes, in which the spacer matches to the Zmo0346 or Zmo0582 gene, were transformed into *Z. mobilis* ZM4mrr. The challenging plasmid pCC_PS1 in Figure 1 was used as the control. **b**, transformation efficiency of the empty plasmid, self-targeting plasmids, and challenging plasmid. **c**, **d**, and **e**, locations of mutations at the genomic Zmo0346, Zmo0582 genes, and protospacer on the plasmid in the randomly selected single colonies carrying pS0346, pS0582, or pCC-S1 plasmids. Protospacer location is indicated by a blue bar for each gene. Grey bars indicate sequenced regions, and white bars indicate deletion regions. Direct repeat sequences flanking the deletion regions are indicated upon the bars. For the Zmo0582 gene, PAM is indicated by an arrow, and a point mutation within the protospacer is indicated by a red bar. Deletion or mutation sites, related to the gene start codon for **c** or the first nucleotide of the protospacer for **d** and **e**, are indicated at the right of each graph. **f**, single colonies of *Z. mobilis* transformed with pEZ15Asp, pS0346, and pS0346OEligA plasmids on the plates. Normal and tiny colonies were indicated by arrows. MMEJ repair rates at Zmo0346 gene locus of randomly selected single colonies (**g**), or randomly selected normal colonies (**h**) carrying pS0346 or pS0346OEligA plasmids. The significance of the MMEJ efficiency was determined using a *t*-test; $p < 0.01$ **.

Figure 6. Proposed model for CRISPR immunity against invasive genetic elements in *Z. mobilis* subtype I-F system. *De novo* spacer acquisition samples spacers from both invasive genetic elements and host DNA. Non-self DNA sampling

578 leads to CRISPR interference of invasive genetic elements or escape of CRISPR
579 interference. Escape triggers primed acquisition, finally leading to interference and
580 clearance of invasive (non-self) genetic elements. Host self DNA sampling leads to
581 interference of self DNA, resulting in cell death. However, MMEJ repairs the self-
582 spacer-guided DNA DSBs, resulting in cell survival and genomic alterations. The
583 width of the arrows indicates the probabilities. Red arrows indicate threats to the host,
584 and green arrows indicate immunity against the invasive genetic elements.
585

a

b

c

d

a **b** **c**

	pEZ15Asp	pAC-PS1	pTT-PS1
New spacer reads/ total reads (%)	$0.75 \pm 0.12 (\times 10^{-2})$	17.44 ± 1.29	0.82 ± 0.10
New spacers from genome/total new (%)	19.28 ± 2.54	0.02 ± 0.01	0.24 ± 0.03
New spacers from plasmid/total new (%)	80.72 ± 18.77	99.98 ± 7.42	99.76 ± 0.03

d **e**

5' -3' spacer

NNCCNNN... -2

NNCCNNN... -1

NNCCNNN... 0

NNCCNNN... +1

NNCCNNN... +2

f

a

b

c

d

e

g

h

f