*Published online 24 October 2019* *Nucleic Acids Research, 2019, Vol. 47, No. 21* ***11461–11475***
*doi: 10.1093/nar/gkz940*

# Characterization and repurposing of the endogenous Type I-F CRISPR–Cas system of *Zymomonas mobilis* for genome engineering

**Yanli Zheng, Jiamei Han, Baiyang Wang, Xiaoyun Hu, Runxia Li, Wei Shen, Xiangdong Ma, Lixin Ma, Li Yi*, Shihui Yang* and Wenfang Peng***

State Key Laboratory of Biocatalysis and Enzyme Engineering, Hubei Engineering Research Center for Bio-enzyme Catalysis, Environmental Microbial Technology Center of Hubei Province, Hubei Collaborative Innovation Center for Green Transformation of Bio-resources, School of Life Sciences, Hubei University, Wuhan 430062, P.R. China

Received August 13, 2019; Revised October 05, 2019; Editorial Decision October 07, 2019; Accepted October 09, 2019

## ABSTRACT

**Application of CRISPR-based technologies in non-model microorganisms is currently very limited. Here, we reported efficient genome engineering of an important industrial microorganism, *Zymomonas mobilis*, by repurposing the endogenous Type I-F CRISPR–Cas system upon its functional characterization. This toolkit included a series of genome engineering plasmids, each carrying an artificial self-targeting CRISPR and a donor DNA for the recovery of recombinants. Through this toolkit, various genome engineering purposes were efficiently achieved, including knockout of *ZMO0038* (100% efficiency), *cas2/3* (100%), and a genomic fragment of >10 kb (50%), replacement of *cas2/3* with *mCherry* gene (100%), *in situ* nucleotide substitution (100%) and His-tagging of *ZMO0038* (100%), and multiplex gene deletion (18.75%) upon optimal donor size determination. Additionally, the Type I-F system was further applied for CRISPRi upon Cas2/3 depletion, which has been demonstrated to successfully silence the chromosomally integrated *mCherry* gene with its fluorescence intensity reduced by up to 88%. Moreover, we demonstrated that genome engineering efficiency could be improved under a restriction–modification (R–M) deficient background, suggesting the perturbance of genome editing by other co-existing DNA targeting modules such as the R–M system. This study might shed light on exploiting and improving CRISPR–Cas systems in other microorganisms for genome editing and metabolic engineering practices.**

## INTRODUCTION

Nearly 50% of bacteria and 90% of archaea equip themselves with clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) adaptive immune systems to defend against invading genetic elements (1). Six main types (Type I to VI) have been defined for various CRISPR–Cas systems, which basically employ two classes of effector complexes to achieve interference. Class 1 systems involving Type I, III and IV encode multi-subunit effector complexes, and Class 2 systems involving Type II, V and VI have single-Cas machineries (2). Despite of the diversity, the CRISPR–Cas immunity is exclusively CRISPR RNA (crRNA)-based and Cas-driven, functioning in three distinct molecular steps: the integration of short DNA stretches (spacers) into the CRISPR array in a polarized manner (spacer adaptation), the processing of the CRISPR transcript into mature crRNAs (crRNA biogenesis), and the execution of the crRNA-guided target DNA/RNA destruction (target interference) (3).

Class I systems naturally occur in more than 90% of sequenced genomes of bacteria and archaea (2). Among them, the Type I systems, containing seven subtypes, *i.e.* I-A through I-F plus I-U are the most abundant, diverse and widespread in nature (2,4), and have been extensively studied in recent years. These systems encode the Cascade (CRISPR-associated complex for antiviral defense) complex to bind to a *bona fide* target (protospacer) upon PAM (protospacer adjacent motif) recognition, and subsequently recruit the featuring Cas3 nuclease–helicase to execute DNA cleavage (5,6). Uniquely, the trans-acting nuclease-helicase in Type I-F systems is encoded as a Cas2–Cas3 fusion, *i.e.* Cas2/3, whose activity is repressed by Cas1 and can be rescued by the target bound Type I-F Cascade (7). PAMs for Type I systems are found directly upstream of

*To whom correspondence should be addressed. Tel: +86 27 88661237; Email: wenfang@hubu.edu.cn
Correspondence may also be addressed to Shihui Yang. Tel: +86 15607121038; Email: Shihui.Yang@hubu.edu.cn
Correspondence may also be addressed to Li Yi. Tel: +86 27 88661237; Email: liyi@hubu.edu.cn

the target sequences in foreign DNAs (2,8) and a 5′-CC-3′ motif was previously proven to be a functional PAM for a Type I-F system (9).

Given the simplicity and programmability of Class 2 systems, CRISPR–Cas9 and CRISPR–Cas12a, for example, were widely developed as genome engineering tools in both eukaryotes and prokaryotes, including human cells, *Saccharomyces cerevisiae* and *Escherichia coli* (10). These systems have also been modified for applications beyond genome editing (11). However, Class 2 systems are relatively rare in nature, being found in fewer than 10% of sequenced prokaryotic genomes (2). Moreover, for the majority of prokaryotes, the exogenous Class 2 machineries are difficult to be exploited due to possibly their large size and severe toxicity to host cells (12), which has largely limited their applications in many alternative production hosts, especially some industrially important species, such as *Zymomonas mobilis*.

*Zymomonas mobilis* is a facultative anaerobic ethanologen with many attractive physiological attributes. For instance, *Z. mobilis* is generally regarded as safe (GRAS), and capable of tolerating a high ethanol concentration up to 16% (v/v) and a broad pH range (3.5–7.5) (13). It has evolved specifically to fit the high sugar and ethanol environment with unique physiological features such as a unique hopanoid membrane structure, a truncated tricarboxylic acid cycle (TCA) pathway, and an efficient Entner-Doudoroff (ED) pathway (14). In order to fully take the advantages of *Z. mobilis*' capabilities in biorefinery (15) to fit the growing global demand for alternative sustainable biofuels and biochemicals, a series of genetic tools, including shutter vectors and transformation methods, have been explored (16). Although the CRISPR–Cas9-based toolkit has been used for plasmid clearance in *Z. mobilis*, there has not been any follow-up report on its application (17), indicative of uncertain factors for its further use in this species. Therefore, more readily useful high-throughput genetic engineering toolkits are needed for performing effective genome editing in *Z. mobilis*.

In recent years, several endogenous Type I CRISPR–Cas systems, including I-A (18), I-B (19–21), and I-E (22), have been harnessed as an alternative strategy to study prokaryotic engineering in archaea and bacteria. For example, the native Type I-B system of *Clostridium pasteurianum* facilitated efficient genome editing of around 4 folds more than that of the heterologously expressed Cas9 system (20). Very recently, the Type I-E system of *Lactobacillus crispatus* has been exploited for genome editing including gene knockout, knockin and point mutation (22), which further highlighted the usefulness of native Type I CRISPR–Cas system for *in situ* genome modification. In addition, the Type I-B of *Haloferax volcanii* (23) and Type I-E of *Escherichia coli* (24,25) systems were harnessed for efficient repression of transcription upon Cas3 depletion. However, to the best of our knowledge, very few native CRISPR-based toolkits were developed for non-model industrial microorganisms. Furthermore, to date there has not been any report on the exploitation of the Type I-F CRISPR–Cas, a close cousin of Type I-E, for genome manipulation application yet, although Type I-F is among the most well-studied CRISPR–Cas systems.

In this study, we characterized the DNA interference capability of the Type I-F CRISPR–Cas system in *Z. mobilis* ZM4. Upon the characterization, an efficient native Type I-F CRISPR-based genome editing toolkit was established to fulfil diverse genome engineering purposes, including gene deletion and replacement (100% efficiency), *in situ* modifications (100%), large fragment (of >10 kb, ca. 5‰ of the genome) deletion (50%), and simultaneous multiple gene editing (18.75%). Efficient gene repression by programming the Type I-F system was also achieved. This work thus provided a versatile and powerful genetic manipulation toolkit for the development and further improvement of *Z. mobilis* as an ideal chassis for biorefinery and synthetic biology studies.

## MATERIALS AND METHODS

### Strains, growth conditions and electroporation transformation of *Z. mobilis*

*Zymomonas mobilis* ZM4 and derivatives constructed in this work were listed in Supplementary Table S1. *Z. mobilis* strains were grown at 30°C in an RMG2 medium (20 g/L glucose, 10 g/L yeast extract, 2 g/L $KH_2PO_4$). If required, spectinomycin was supplemented to a final concentration of 200 μg/mL for *Z. mobilis* and 50 μg/mL for *E. coli*. *Z. mobilis* competent cells were prepared as previously described (16) and transformed with plasmids by electroporation using Bio-Rad Gene Pulser (0.1-cm gap cuvettes, 1.6 kV, 200 Ω, 25 μF) (Bio-Rad, Hercules, CA, USA) following the method developed for *Z. mobilis* (26). Following the electroporation, the cells were incubated in an RMG2 medium for 3 h at 30°C prior to plating.

### Construction of plasmids

Interference plasmids were initially constructed with two spacers, Spacer7 of CRISPR2 (C2S7) and Spacer4 of CRISPR3 (C3S4). Oligonucleotides were designed to bear the entire sequences of the selected spacers following a 5′-CCC-3′ PAM or the last three nucleotides of the repeat (5′-AAA-3′). Each set of oligonucleotides was annealed by first being heated to 95°C for 5 min and subsequently cooled down gradually to room temperature, followed by double digestion using XbaI and EcoRI. These DNA fragments were ligated with the XbaI and EcoRI linearized vector (pEZ15Asp, an *E. coli–Z. mobilis* shuttle vector (16)), yielding the interference plasmids (pInt plasmids) and the corresponding reference plasmids (pRef plasmids). All plasmids were listed in Supplementary Table S1.

Artificial CRISPR expression plasmids were constructed based on the pEZ15Asp vector in this study. A DNA block consisting of the leader sequence of the chromosomal CRISPR2 as a promoter and two CRISPR repeats spaced by two BsaI restriction sequences in opposite orientation was firstly synthesized from GenScript (Nanjing, China), and used as a template for PCR amplification with the primer pair of L2R-XbaI-F/l2R-EcoRI-R (Supplementary Table S2). Then, the PCR product was digested with XbaI and EcoRI and subsequently inserted into the pEZ15Asp vector, generating the base vector pL2R. Digestion of pL2R with BsaI generated a linearized plasmid

having protruding repeat sequences of 4 nt at both ends. Double-stranded spacer DNAs were prepared by annealing two spacer oligonucleotides through being heated to 95°C for 5 min followed by cooling down gradually to room temperature. The spacer fragments were designed to carry 4 nt protruding ends complementary to those in the linearized pL2R.

Therefore, plasmids each bearing an artificial mini-CRISPR with a self-targeting spacer were generated by individually ligating the spacer inserts with the linearized pL2R vector. Subsequently, donor DNA fragments each containing a mutant allele of a target gene were generated by splicing and overlap extension PCR (SOE-PCR) (27) and individually cloned into their cognate pST plasmids through the T5 exonuclease-dependent DNA assembly (TEDA) method (28), yielding the all-in-one genome engineering plasmids listed in Supplementary Table S1. By repeating the actions, the pGME-Cs was yielded. In addition, DNA of the mCherry expression cassette was amplified from a previously constructed plasmid (29) and used for constructing the donor DNA of the pRep-*cas2/3* plasmid (Supplementary Table S1).

All oligonucleotides were synthesized from GenScript (Nanjing, China) and listed in Supplementary Table S2. Restriction enzymes and T5 exonuclease were purchased from New England Biolabs (Beijing) Ltd (Beijing, China).

### Construction and screening of mutants

The plasmids used for genome engineering were individually introduced into *Z. mobilis* cells. Electroporated cells were spread on RGM agar plates containing spectinomycin at a final concentration of 200 μg/mL (RGMSp) and incubated at 30°C until colonies were observed. Mutant candidates were screened by colony PCR using primers listed in Supplementary Table S2. The resulting PCR products were analysed by agarose gel electrophoresis and confirmed by Sanger sequencing (GenScript, Nanjing, China).

### Curing of genome engineering plasmids

Cells of genome engineering plasmid transformants were grown up in an RMG2 broth with the supplement of spectinomycin. Then, the cells were spread on an RMG agar plate with spectinomycin for selection. Single colonies were picked and individually suspended in 10 μL of sterilized $ddH_2O$, and 1 μL of each cell suspension was spotted on an RMG agar plate with or without spectinomycin. Cells from the same suspension grew up on the plate without spectinomycin but not on that with spectinomycin were regarded as those lost the genome engineering plasmid.

### Transcriptional analysis of the CRISPR–Cas of *Z. mobilis*

Fastq files of sRNA-Seq and RNA-seq were checked using FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), data passing the quality control were imported into CLC Genomics Workbench, and the reads were trimmed for nucleotides with quality score less than 30. sRNA-Seq data (GSE57773) and three RNA-seq datasets of total RNA without rRNA depletion from previous studies and unpublished inhouse datasets were also used to map four extended CRISPR clusters using CLC Genomics Workbench. The mapped reads were then manually checked for potential crRNA processing, and figures for the coverage of mapped reads to these four clusters were then exported from CLC Genomics Workbench. In addition, RPKM values of 109 microarray and 75 RNA-Seq datasets from previous studies and unpublished inhouse datasets were composed and imported into JMP Genomics (SAS Inc., USA), which were then $log_2$-transformed before statistical analysis as previously reported (30,31) to compare the expression of genes encoding Cas proteins as well as the mean and median values of overall gene expressions in different conditions. The GEO accession numbers for sRNA study and transcriptomic studies in *Z. mobilis* that have been deposited into NCBI are GSE57773 (sRNA), GSE108890, GSE63540, GSE57553, GSE49620, GSE39558, GSE37848, GSE25443, GSE21165, GSE18106 and GSE10302.

### FACS analysis

The protocol used for FACS analysis was modified slightly based on a previous study (32). Briefly, cells were washed with phosphate buffered saline (PBS) twice and then resuspended into PBS to a concentration of $10^7$ cells/mL. Cells were analyzed by flow cytometry using Beckman CytoFLEX FCM (Beckman Coulter, Inc., USA) with the phosphate buffered saline as the sheath fluid. The cells fluorescence of mCherry were excited with the 561 nm and detected with PC5.5 (33).

### Western blotting

*Z. mobilis* His-ZMO0038 and His-ZMO0038M strains were cultured in RMG2 medium. Cells were then harvested by centrifugation when $OD_{600\ nm}$ reached 0.4, followed by suspension in 50 mM phosphate buffer for sonication lysis. The crude protein samples were fractionated on 12% SDS-PAGE and transferred onto a nylon membrane using the Semi-Dry Electrophoretic Transfer Cell System (Bio-Rad, Hercules, CA, USA). The membrane was incubated with a hybridization buffer including an antibody against His-tag peptide (GenScript, Nanjing, China). The antibody bound to the His-tagged proteins was then recognized by a secondary antibody (Goat Anti-Mouse IgG, GenScript, Nanjing, China). The results were then visualized by chemiluminescent detection using the clarity Western ECL substrate (Bio-Rad, Hercules, CA, USA) and analyzed using the Amersham Imager 600 device (GE Healthcare, Seoul, South Korea).

## RESULTS

### *In silico* analysis of the Type I-F CRISPR–Cas system of *Z. mobilis* ZM4

Based on the complete genome sequence of *Z. mobilis* ZM4 (34–36), a Type I-F CRISPR–Cas system was identified, consisting of four CRISPR arrays (CRISPR1-4) and a *cas* gene cluster (Figure 1A). CRISPR1 is solely located on one strand and CRISPR2-4 are on the other

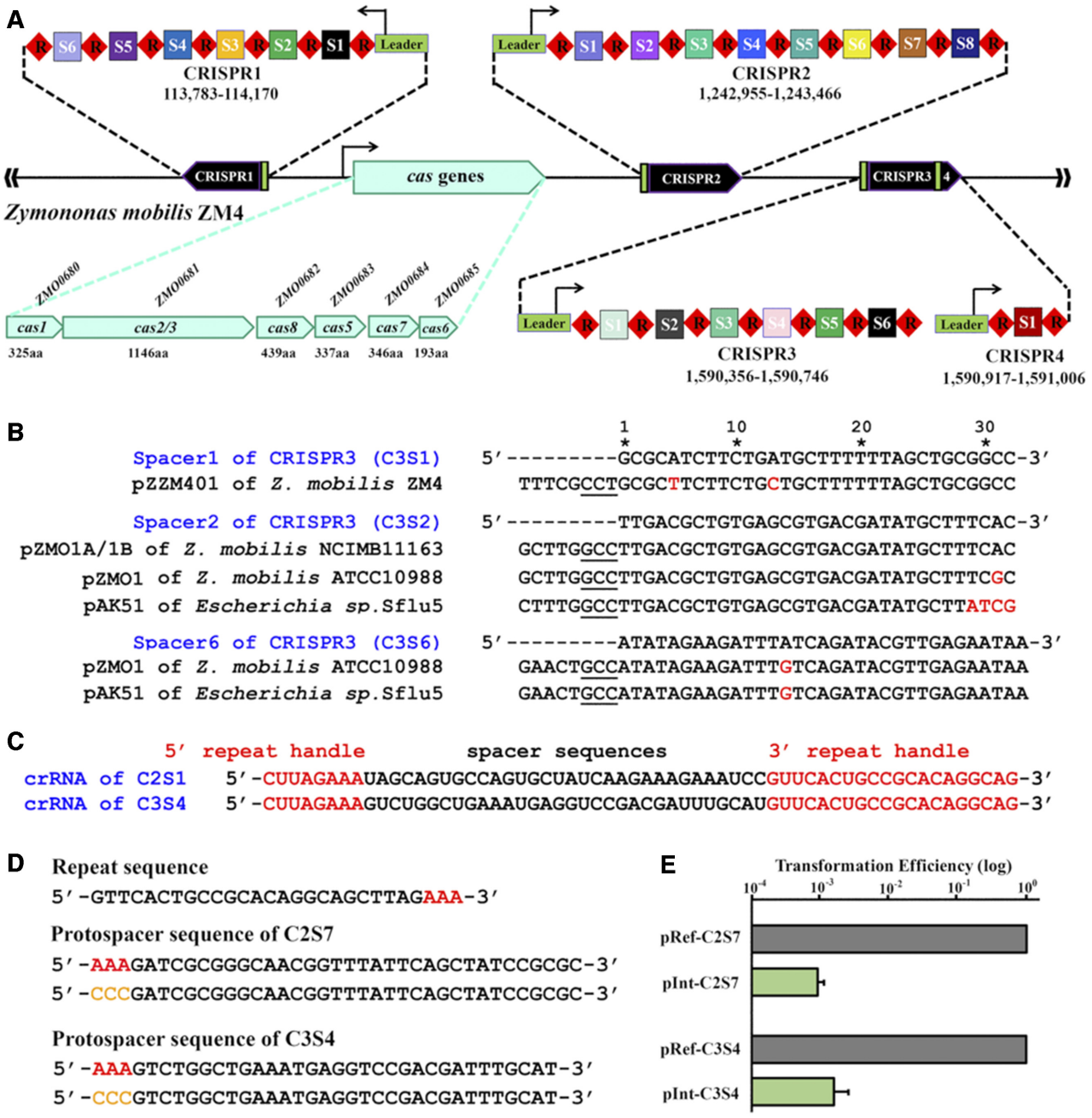



**Figure 1.** *Z. mobilis* ZM4 CRISPR and *cas* loci and demonstration of the Type I-F DNA interference activity. (A) Schematic of CRISPR arrays and a *cas* locus in *Z. mobilis* ZM4. (B) *In silico* identification of putative protospacer matches. Mismatches between protospacer and spacer are indicated with red colour. Potential PAM motifs are underlined. (C) Identification of mature crRNA reads matching Spacer1 of CRISPR2 (C2S1) and C3S4 from transcriptomic data. Repeat-derived handles at both ends are shown in red fonts. (D) Representatives of the repeat sequence, the C2S7 and C3S4 spacers, and the construction of protospacers used in interference plasmids and their corresponding reference plasmids. (E) Plasmid interference assay. Two interference plasmids, pInt-C2S7 and pIntC3S4, were used to assess the DNA interference activity of the Type I-F CRISPR-Cas system. Transformation efficiencies of each interference plasmid are expressed as relative values to the efficiencies of their corresponding reference plasmids, the latter of which are set to 1.0. Three replications were performed for each DNA sample.

with CRISPR4 being immediately preceded by CRISPR3, which contain 6, 8, 6 and 1 spacers of 32–34 nt, respectively, with two-thirds (14/21) are in length of 32 nt (Figure 1A). The sequence of most of the repeats is 5′-GTTC ACTGCCGCA**C**AGGCAGCTTAGAAA-3′, whereas four repeats in CRISPR1 have sequence of 5′-GTTCACTGC CGCA**T**AGGCAGCTTAGAAA-3′. These sequences are identical to the functional repeat sequences of two previously well studied Type I-F CRISPR–Cas systems of *Pectobacterium atrosepticum* SCRI1043 (37) and *Pseudomonas aeruginosa* UCBPP-PA14 (38), respectively (Supplementary Table S3). Furthermore, the terminal repeats in arrays of CRISPR1, 2 and 4 have mutations within the last eight nucleotides (Supplementary Table S3). The *cas* genes are organized as an operon of *cas1–cas2/3–cas8*(*csy1*)*–cas5*(*csy2*)*–cas7*(*csy3*)*–cas6*(*csy4*), being architecturally the same as the model Type I-F systems (37,38). Each Cas protein has a homolog in *P. aeruginosa* UCBPP-PA14, exhibiting 35–67% amino acid sequence identity. Type I-F system can be identified in almost all sequenced genomes of *Z. mobilis* species using the online CRISPRFinder tools (39,40).

In order to find out sequences matching spacers in attacking DNAs, such as plasmid and phage/prophage, we analyzed all 21 spacers in the four CRISPR loci using the online CRISPRTarget tool (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html) with the default parameters (41). Sequences of three spacers showed significant matches to putative protospacers present in plasmids of *Zymomonas* or *Escherichia* species (Figure 1B). The three spacers occur exclusively in the CRISPR3 locus, designated C3S1, C3S2 and C3S6 for Spacer1, 2 and 6 of CRISPR3, respectively, exhibiting high sequence identity to their corresponding putative protospacers, ranging from 87.5% to 100%. The 32-nt C3S2 sequence perfectly matched sequences carried by two plasmids of *Z. mobilis* NCIMB11163, pZMO1A (NC_019198) and pZMO1B (NC_019210.1), while had a single mismatch at position 31 to the sequence in the pZMO1 (NC_011363.1) plasmid of *Z. mobilis* ATCC10988. Identical protospacer sequences of the 33-nt C3S6 were also detected in both plasmids pZMO1 and pAK51, carrying one mismatch to C3S6 at position 14. As for the 32-nt C3S1, a putative protospacer was seen in a plasmid of *Z. mobilis* ZM4, pZZM401 (NC_013784.1), with two mismatching nucleotides at positions 5 and 13 in the 5′-part sequence. Since in Type I systems, PAM sequences are found directly upstream of the protospacer (2), the motifs immediately preceding the detected protospacers, *i.e.* 5′-GCC-3′ (upstream of C3S2 and C3S6) and 5′-CCT-3′ (upstream of C3S1) (Figure 1B), are PAM candidates.

In addition to analyzing the Type I-F CRISPR–Cas architecture on DNA level, we performed transcriptome analysis to reveal the expression of both the CRISPR loci and the *cas* operon. We found two small RNA molecules containing Spacer1 of CRISPR2 (C2S1) and Spacer4 of CRISPR3 (C3S4) in the transcriptomic data. They were identified to be Type I effector-associated mature crRNAs produced by Cas6 homologues (42) as both of which possess the typical feature: an intact spacer sequence with an 8-nt repeat-derived handle and a 20-nt repeat sequence (the reminder) located at the 5′ and 3′ termini, respectively (Figure 1C). These findings indicated the effective expression of Cas6 for processing the CRISPR transcripts into mature crRNAs (43). Since the *cas6* gene is located at the farthest end of the *cas* operon (Figure 1A), its expression could reflect the expression of all other *cas* genes. These results were further confirmed using the datasets from previous microarray and RNA-Seq studies (Supplementary Figure S1). Under different strain background and growth conditions, the Type I-F *cas* genes were constitutively expressed, among which the expression levels of *cas1* and *cas2/3* were slightly lower than the mean and median values of all genes, while those of the *csy* genes (*cas5–cas8*) were significantly above the mean and median values of all genes (Supplementary Figure S1).

### Demonstration of the DNA targeting activity of the Type I-F system of *Z. mobilis*

To test the DNA targeting capability of the Type I-F system of *Z. mobilis* ZM4, we employed a plasmid interference assay since it could efficiently express the protospacer-matching crRNA and be eliminated from the cells afterwards (44). Two interference plasmids were generated with one containing C2S7 protospacer (pInt-C2S7) and the other containing C3S4 protospacer (C3S4) using a functional 5′-CCC-3′ PAM sequence (9) based on the *E. coli–Z. mobilis* shuttle vector pEZ15Asp (16). For the corresponding reference plasmids (pRef-C2S7 and pRef-C3S4), the PAM sequence was replaced with the last 3 nt of the repeat to mimic the genetic structure of CRSIPR arrays in the genome (Figure 1D). Upon introducing the plasmid constructs into *Z. mobilis* cells, hundreds-fold lower transformation rates were obtained with pInt-C2S7 or pInt-C3S4 than those with reference plasmids (Figure 1E), suggesting that the interference plasmids triggered much stronger defense responses than those of reference plasmids. The effectiveness of C3S4 and C2S7 further confirmed the transcription and processing of the CRISPR loci, and revealed that a spacer with a length of 34 bp could work as well as the predominant 32-bp ones. Moreover, these results confirmed that a 5′-CCC-3′ motif is a functional PAM sequence and more importantly demonstrated the DNA targeting activity of the Type I-F CRISPR–Cas system, which had the potential to be hijacked for genome engineering in all *Z. mobilis* species.

### Determination of PAMs required for the Type I-F CRISPR immunity in *Z. mobilis*

Although the 5′-CCC-3′ PAM sequence was confirmed to be functional, it remained unclear whether there are other functional PAMs for the Type I-F system. This is a critical issue for proper on-targeting design and off-targeting evaluation, especially for experiments with more than one sequences being simultaneously targeted. Hence, we set out to identify and characterize all potential PAMs for the Type I-F CRISPR–Cas system of *Z. mobilis*.

Plasmid interference assay was employed again to systematically examine the putative PAM sequences. DNA fragments were generated bearing C3S4 preceded by trinucleotide combinations and were individually cloned to the

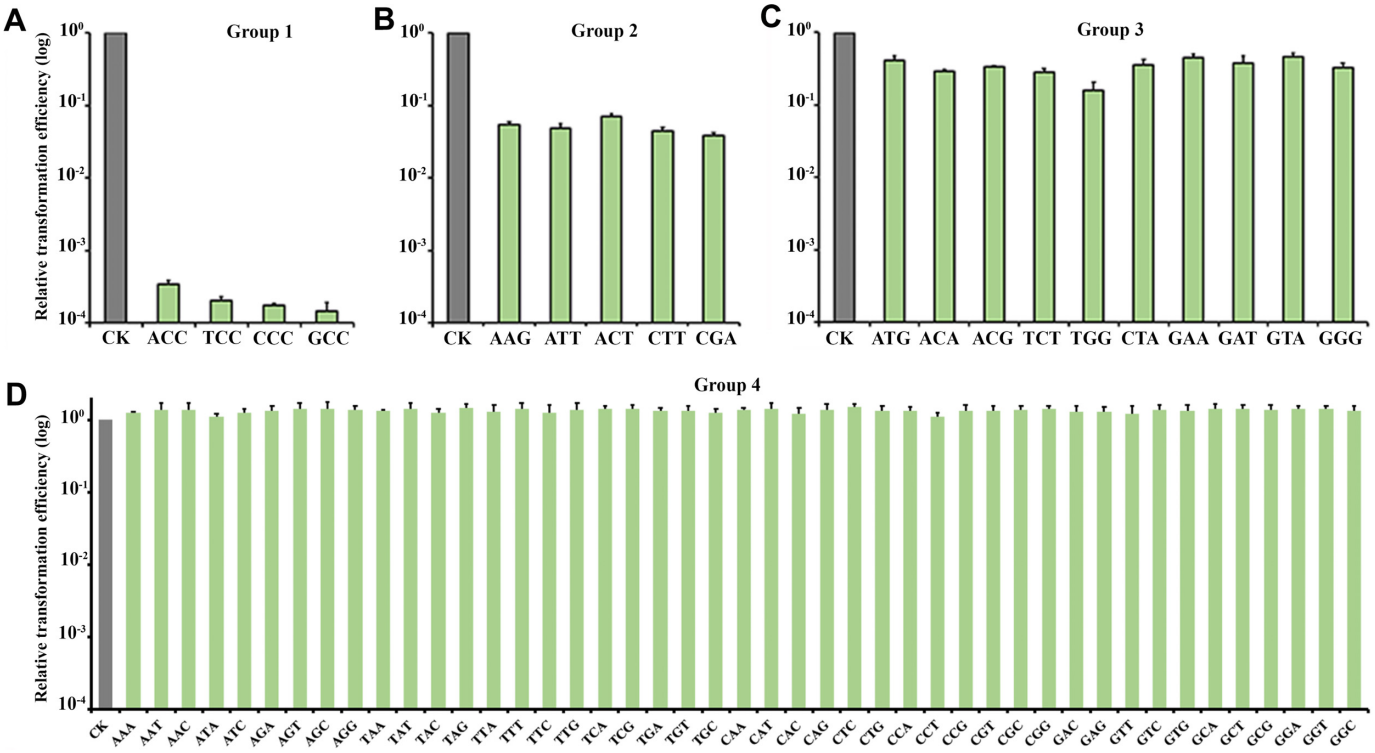



**Figure 2.** Effects of PAMs on triggering target attack by the Type I-F CRISPR–Cas system of *Z. mobilis*. Plasmid interference assay was performed as depicted in Figure 1E. Relative transformation efficiencies of each interference plasmid are normalized by setting the transformation efficiency of vector pEZ15Asp (CK) to be 1.0. PAMs resulting in obvious drop of transformation efficiency are presented and classified into 3 groups upon the descent degree: more than >1,000-fold in Group 1 (A), ca. 10-fold in Group 2 (B), and <4-fold in Group 3 (C), respectively. The trinucleotides in the rest plasmids with a similar level of transformation efficiencies as that of the vector pEZ15Asp are listed as Group 4 unfunctional PAMs (D). Three replications were performed for each DNA sample.

pEZ15Asp vector. Together with 5'-CCC-3' in pInt-C3S4 and 5'-AAA-3' in pRef-C3S4 constructed previously, a total of 64 different candidate PAM sequences were evaluated. These plasmid constructs, as well as the vector pEZ15Asp as a reference, were then individually introduced into the host cells.

Transformation rates significantly varied with plasmids containing different PAM sequences adjacent to C3S4, which were presented in Figure 2 as relative transformation efficiencies to that of the vector pEZ15Asp (set to 1.0). These PAM sequences can be artificially categorized into four groups. In Group 1, plasmids contained any of the 5′-NCC-3′ sequences as a PAM got extremely lowered transformation rates (more than 1,000-fold decrease), indicating that very strong defense reaction was triggered (Figure 2A). This is consistent with previous bioinformatic prediction (8) and experimental observations (9). In Group 2, plasmids with 5′-AAG-3′, 5′-ATT-3′, 5′-CTT-3′, 5′-CGA-3′ or 5′-ACT-3′, exhibited an around 10-fold reduction in transformation rates (Figure 2B). In Group 3, plasmids with ten other 3-nt motifs, *i.e.* 5′-ATG-3′, 5′-ACA-3′, 5′-ACG-3′, 5′-TCT-3′, 5′-TGG-3′, 5′-CTA-3′, 5′-GAA-3′, 5′-GAT-3′, 5′-GTA-3′ and 5′-GGG-3′, also had effect on triggering CRISPR targeting but showed less than 4-fold drop in transformation rates (Figure 2C). The trinucleotides in the rest plasmids (Group 4) were not considered as functional PAM sequences because transformation with plasmids carrying any of them yielded transformation rates at a similar level with that of the vector pEZ15Asp (Figure 2D). The presence of the motifs containing same last two nucleotides in different groups, *e.g.* 5′-ATT-3′ and 5′-CTT-3′ in Group 2 whereas 5′-TTT-3′ and 5′-GTT-3′ in Group 4, indicated that the first nucleotide in this case also contributed to target recognition by Cascade and therefore PAMs in Type I-F function as trinucleotides, alike those determined for other Type I systems (44–46). Reasonably, the dinucleotide 5′-CC-3′ (actually 5′-NCC-3′) could work functionally (9).

In short, we confirmed that the 5′-NCC-3′ sequences in Group 1 are functional PAMs, while those in Groups 2 and 3 also triggered CRISPR targeting on the protospacer albeit the targeting effect was much weaker than that of 5′-NCC-3′. These data may provide us with important indications for on-targeting design in genome engineering and fine-tuning of gene expression in CRISPR interference study in the future.

### Establishment and optimization of the Type I-F CRISPR-based genome engineering toolkit for *Z. mobilis*

Given the fact that the endogenous Type I-F CRISPR–Cas system of *Z. mobilis* ZM4 exhibited strong interference activity against the protospacer-bearing plasmids, we redirected its DNA cleavage activity to a PAM-flanking sequence on the chromosome for self-targeting and subsequent genome engineering. The *cas2/3* (*ZMO0681,* 3,441 bp) and *ZMO0038* (603 bp), two non-essential genes for cell

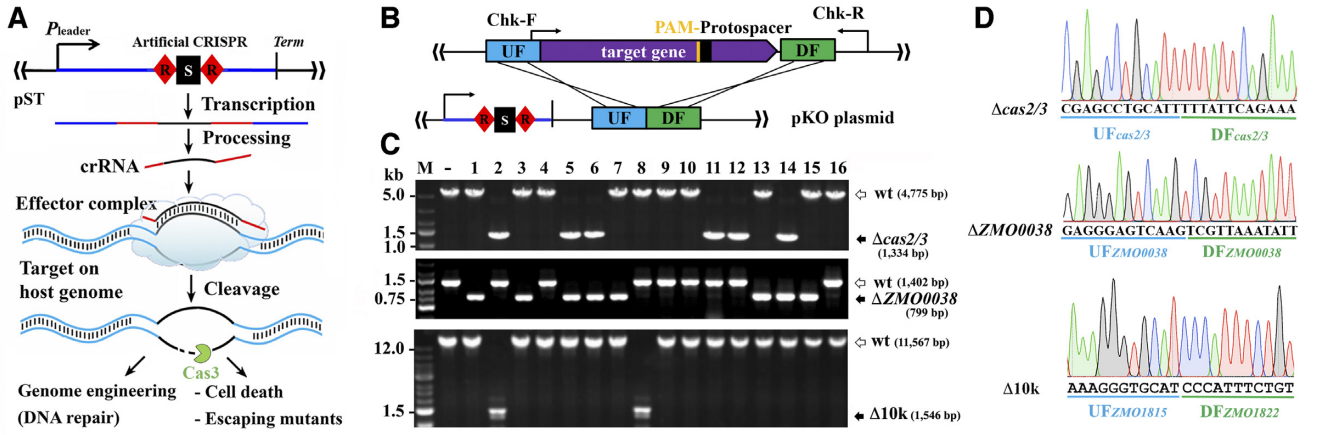



**Figure 3.** Establishment of the Type I-F CRISPR-based genome engineering system for *Z. mobilis* ZM4. (A) A self-targeting plasmid (pST) contained an artificial CRISPR locus. Self-targeting crRNAs were to be produced from the artificial CRISPR and guided the effector complex to bind to the target sequence. Cas3 was then recruited to unwind the double-stranded DNA and nick the non-target strand for degradation. The targeted cells either died due to chromosomal DNA degradation, or survived as escaping mutants. (B) Design of the self-targeting CRISPR and the donor DNA in knockout plasmids (pKO). (C) Colony PCR screening of deletion mutants of *cas2/3*, *ZMO0038* and 10-kb using primer set of Chk-F/Chk-R. Predicted sizes of PCR products in wild-type (wt) and the expected deletion mutants (Δ*cas2/3* or Δ*ZMO0038* or Δ10k) are indicated with unfilled and filled black arrows, respectively. -, PCR amplification using genomic DNA of *Z. mobilis* ZM4 as a DNA template; M, DNA size marker. (D) Representative chromatographs of Sanger sequencing results.

**Table 1.** Transformation efficiencies (TE) and engineering efficiencies (EE) of various genome-engineering plasmids in *Z. mobilis* wild type and a restriction–modification mutant Δ*ZMO0028*. Three replicates were performed for each DNA sample.

| | TE (cfu/μg DNA) | | EE [% (engineered/tested)] | |
|---|---|---|---|---|
| Plasmid | Wild type | Δ*ZMO0028* | Wild type | Δ*ZMO0028* |
| pEZ15Asp | $(3.90 \pm 0.62) \times 10^4$ | $(3.79 \pm 0.23) \times 10^6$ | - | - |
| pKO-*ZMO0028* | $297 \pm 39$ | - | 12.5 (2/16) | - |
| pKO-*cas2/3* | $208 \pm 11$ | $(2.71 \pm 0.33) \times 10^4$ | 37.5 (6/16) | 100 (16/16) |
| pKO-*ZMO0038* | $244 \pm 24$ | $(8.15 \pm 0.93) \times 10^3$ | 50 (8/16) | 100 (16/16) |
| pDel-10k | $256 \pm 44$ | $(1.30 \pm 0.25) \times 10^3$ | 12.5 (2/16) | 50 (8/16) |

-: Not detectable.

viability, were chosen as engineering targets. Self-targeting plasmids (pST plasmids) were firstly constructed to individually carry an artificial CRISPR expression cassette of leader-repeat-spacer-repeat (Figure 3A). To increase the recovery rate of desired genotypes from self-targeting, a donor DNA consisting of two homology arms for facilitating homologous recombination were designed to bear expected mutations (Figure 3B). Using the genome engineering plasmids pKO-*cas2/3* and pKO-*ZMO0038* (Figure 3B, Supplementary Table S1), both target genes were successfully deleted (Figure 3C and D). The transformation rates were >180 folds lower than that with the cloning vector pEZ15Asp (Table 1), indicating that the *cas2/3*- and *ZMO0038*-targeting crRNAs were correctly produced from the plasmid-borne artificial CRISPR loci and had mediated efficient self-targeting by the endogenous Type I-F CRISPR–Cas system. For each assay, the genotypes of 16 randomly chosen transformants were analyzed by colony PCR and Sanger sequencing, among which 6/16 of pKO-*cas2/3* and 8/16 of pKO-*ZMO0038* transformants, respectively, harbored the corresponding deletions (Figure 3C and D).

Next, we evaluated the ability of the Type I-F CRISPR-based toolkit in deleting a genomic fragment of >10 kb, which was hardly achievable with the classical genetic manipulation methods in *Z. mobilis* (47). An array of genes through *ZMO1815-ZMO1822* (10,021 bp) was chosen, which was calculated as ∼5‰ of the *Z. mobilis* ZM4 genome. The corresponding genome engineering plasmid (pDel-10k) was generated by employing the same construction strategy. It was noticed that the transformation rate with pDep-10k was comparable to that obtained with those for single small genes. Subsequently, colony PCR was conducted using Chk-10k-F and Chk-10k-R primers (Supplementary Table S2) to evaluate 16 randomly chosen transformants. Two of the analysed colonies showed the predicted size of 1,546 bp (Δ10k) (Figure 3C), which was further confirmed by Sanger sequencing (Figure 3D).

Despite of the successful genome engineering with the Type I-F CRISPR–Cas system in *Z. mobilis*, the engineering efficiencies were obviously lower than that obtained by repurposing native Type I systems in other microbes (18,19,21). We speculated that the low plasmid transformation efficiency (around $10^4$–$10^5$ CFU/μg DNA) was probably caused by the presence of DNA restriction–modification (R–M) systems in *Z. mobilis* (48). Inactivation of an R–M element (encoded by the *ZMO0028* gene) was reported to lead to an improvement of plasmid transformation effi-

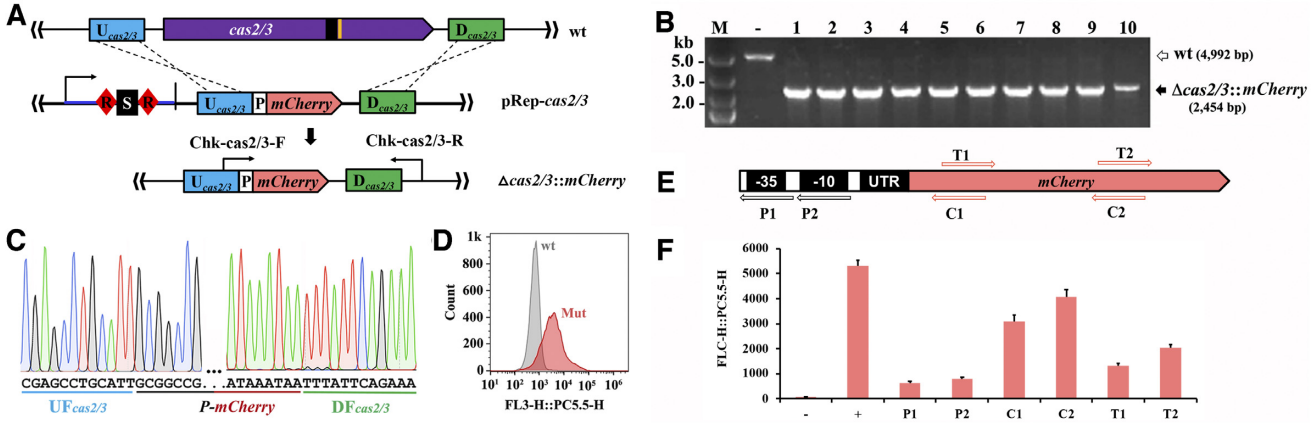



**Figure 4.** Applications of Type I-F CRISPR-based toolkit for gene replacement and gene repression. (A) Schematic showing strategy of *cas2/3* replacement. The donor DNA carried by pRep-*cas2/3* including a P-mCherry block between two homologous arms. The PAM motif is shown in orange. (B) PCR screening of Δ*cas2/3*::*mCherry* recombinants. Predicted sizes of PCR products in wild-type (wt) and the expected Δ*cas2/3*::*mCherry* strains are indicated with unfilled and filled black arrows, respectively. -, PCR amplification of genomic DNA of *Z. mobilis* ZM4; M, DNA size marker. (C) Representative chromatographs of Sanger sequencing results of the *cas2/3* replacement with *mCherry*. (D) Signals of mCherry was detected in two *cas2/3*::*mCherry* mutants (Mut) whereas not in the wild-type strain (wt) by flow cytometry. (E) Schematic showing the positions of protospacers in the promoter (P1 and P2) and the *mCherry* ORF. C1 and C2 targeting the coding strand; T1 and T2 targeting the template strand. (F) Repression of mCherry expression. +, *cas2/3*::*mCherry* cells with a spacer-free plasmid; -, *cas2/3* cells. Three replications were performed for each experiment.

ciency by up to 60-fold (reaching $10^6$ CFU/μg DNA) (48). Therein, a deletion mutant of *ZMO0028* (Δ*ZMO0028*) was generated and reassessed in our studies. Our results showed that the transformation efficiency of Δ*ZMO0028* was ∼90-fold higher than that of the wild-type strain when transformed with the pEZ15Asp vector. Notably, significantly improved transformation efficiencies by 130, 33 and 5 folds, respectively, were yielded with pKO-*cas2/3*, pKO-*ZMO0038* and pDel-10k plasmids being transformed into the Δ*ZMO0028* cells (Table 1). More importantly, under the Δ*ZMO0028* background, substantially enhanced genome engineering rates were obtained for all genome engineering manipulations. The engineering efficiencies of single gene deletion in both cases reached 100% (16/16), while that of the >10 kb fragment deletion was improved from 12.5% (2/16) to 50% (8/16) (Table 1). Due to this highly elevated efficiency of genome engineering up to 100%, this created Δ*ZMO0028* strain, designated *Z. mobilis* DRM1 (short for Delta R–M Strain 1), was thus employed in our subsequent work.

## Application of the Type I-F CRISPR-based toolkit for gene replacement

To further expand its ability of genome editing, gene replacement assay with the Type I-F CRISPR-based toolkit was carried out in *Z. mobilis*. As an example, the *cas2/3* gene was selected to be replaced with the exogenous *mCherry* gene encoding the mCherry red fluorescent protein. The *cas2/3*-replacing plasmid, pRep-*cas2/3*, was generated by inserting a 1,020-bp mCherry-expressing cassette between the homology arms (Figure 4A). Following the genome editing, colony PCR was performed using the primer set of Chk-cas2/3-F and Chk-cas2/3-R (Supplementary Table S2) to amplify a DNA fragment encompassing the engineered region. A smaller PCR product of the predicted 2,454 bp was yielded in all 10 tested colonies (Figure 4B), representing the Δ*cas2/3*::*mCherry* mutants. The successful gene replacement was further confirmed by Sanger sequencing (Figure 4C). Our results also detected the signal of red fluorescent in the mutants (Mut) by flow cytometry, indicating that the heterologously integrated mCherry red fluorescent protein was stably expressed. In contrast, the signal was not detectable in the wild-type strain (wt) (Figure 4D). All these combined results suggested that the Type I-F CRISPR-based toolkit possessed the ability of precisely integrating heterologous gene(s) into the host genome for metabolic engineering practices in *Z. mobilis*.

## Application of the Type I-F Cascade for gene repression

The capability of native Type I-F system for gene repression in *Z. mobilis* was examined with the Δ*cas2/3*::*mCherry* strain. The chromosomally integrated *mCherry*, whose expression is driven by the promoter of the original *ZMO0038* gene ($P_{ZMO0038}$), was used as the reporter gene. Spacers in length of 32 nt targeting elements of the $P_{ZMO0038}$ promoter or the *mCherry* open reading frame (ORF) (Figure 4E) were designed with different gene repression plasmids (pGR) constructed. The repression effect on mCherry expression was then evaluated by measuring the fluorescence of individual pGR transformants through flow cytometry analysis.

pGR plasmids were transformed into the Δ*cas2/3*::*mCherry* cells with similar efficiencies to the spacer-free control plasmid pL2R (Supplementary Table S2). All pGR transformants showed comparable growth to the pL2R transformants, while the mCherry fluorescence intensity was dramatically dropped to various levels. The mCherry fluorescence intensity was mostly decreased when using P1 and P2 crRNAs respectively targeting the −10 and −35 elements of $P_{ZMO0038}$, remaining only around 12% relative fluorescence intensity of that of the pL2R

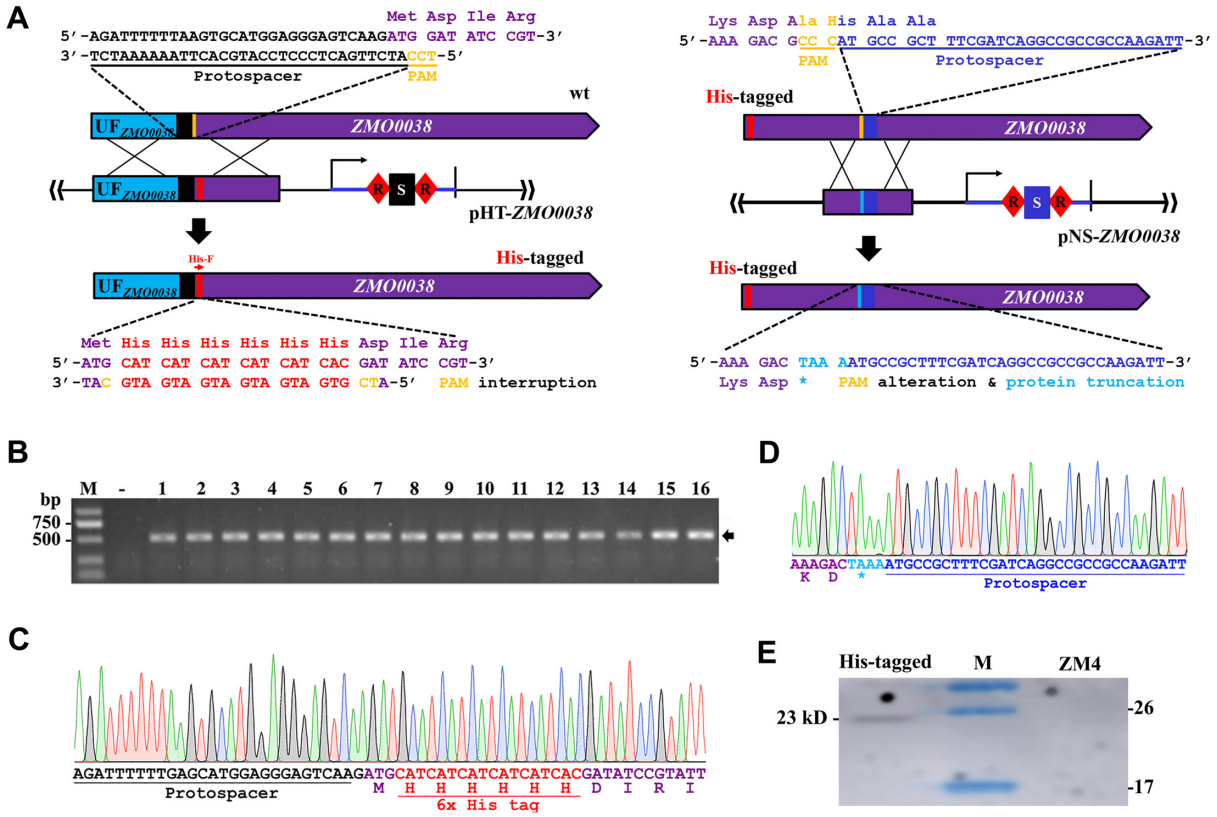



**Figure 5.** Applications of Type I-F CRISPR-based toolkit for *in situ* modification. (A) Schematic showing strategy of *in situ* His-tagging (left panel) and nucleotide substitutions (right panel) of *ZMO0038*. The His-tagging plasmid (pHT-*ZMO0038*) encodes a CRISPR spacer (shown in black) matching the protospacer spanning the start codon of *ZMO0038*. The sequences of His tag are shown and indicated in red. The spacer in the nucleotide substitution plasmid (pNS-*ZMO0038*) and the corresponding protospacer in *ZMO0038* are indicated in blue while the designed mutations in cyan. The PAM motifs are shown in orange. (B) Colony PCR analysis of the pHT-*ZMO0038* transformants. The specific His-F primer is used and the expected target bands are indicated by a black arrow. -, PCR amplification of genomic DNA of *Z. mobilis* ZM4; M, DNA size marker. (C) (D) Representative chromatographs of Sanger sequencing results of the His-tagging (C) and nucleotide substitution mutations (D) in *ZMO0038*. (E) Western blot analysis of His-tagged ZMO0038 using antibody specifically against the His-tag peptide. Peptide of predicted size (ca. 23 kDa) was hybridized (lane His-tagged). ZM4, crude protein sample prepared from the wild-type ZM4 strain as a negative control. M, protein size marker.

transformants (Figure 4F). Comparably, crRNAs targeting the template strand (T1 and T2) or the coding strand (C1 and C2) of the *mCherry* ORF also led to the decrease of fluorescence intensity by 23–75%. Our results also indicated the existence of a strand correlated repression bias that the fluorescence intensity reduced more dramatically by targeting the template strand than that by targeting the coding strand, given that C1 and T1 targeted sequences at approximately the same position and C2 and T2 targeted sequences at exactly the same position on opposite DNA strands (Figure 4E, Supplementary Table S2). This observation is consistent with the results observed for Type I-B in *H. volcanii* (23) and Type I-E in *E. coli* (24).

### Application of the Type I-F CRISPR-based toolkit for *in situ* gene modification

*In situ* gene modifications such as tagging and site-directed mutagenesis are useful methods for studying gene functions. Therefore, the application of the Type I-F CRISPR-based toolkit on performing *in situ* gene modification was explored in *Z. mobilis*.

First, pHT-*ZMO0038*, a genome engineering plasmid for *in situ* His-tagging of *ZMO0038*, was constructed. The crRNAs transcribed from the mini-CRISPR carried by pHT-*ZMO0038* targeted a sequence consisting of the 30-bp upstream flanking sequence and the first two nucleotides of the start codon of *ZMO0038* in the noncoding strand. The corresponding protospacer in pHT-*ZMO0038* was designed to follow a 5′-TCC-3′ PAM, and this plasmid also carried a donor DNA that was designed to include the sequences surrounding the target site, where a 6x His tag was introduced immediately downstream of the start codon to interrupt the 5′-TCC-3′ PAM (Figure 5A). Transformation with pHT-*ZMO0038* into the DRM1 cells yielded hundreds of colonies, suggesting a high recovery rate from recombination. Subsequently, 16 randomly selected colonies were checked by colony PCR with His-F and Chk-ZMO0038-R primers (Supplementary Table S2), and our results showed that a PCR product of the predicted size (505 bp) was consistently amplified in all tested colonies (Figure 5B). Moreover, a DNA fragment covering the edited locus was amplified with the primer set of Chk-ZMO0038-F/Chk-ZMO0038-R (Supplementary Table S2) and the Sanger se-

quencing results confirmed a 100% engineering efficiency (Figure 5C).

Upon the His-tagged *ZMO0038*, we further performed the *in situ* nucleotide substitution. A 5′-GCCC-3′ sequence, occupying the positions 235–238 of the *ZMO0038* coding sequence and fitting to a 5′-CCC-3′ PAM motif, was modified into 5′-TAAA-3′ sequence. After the expected changes, the modified protospacer would be no longer considered as a target, allowing cell surviving, as it turned to follow a non-PAM sequence (5′-AAA-3′). In addition, an internal stop codon could be introduced into the His-tagged *ZMO0038* (after replacing the guanine located at position 235 with a thymine) (Figure 5A). To achieve this goal, a genome engineering plasmid for nucleotide substitutions of *ZMO0038*, pNS-*ZMO0038*, was constructed to include a mini-CRISPR, which bears a spacer matching the 32-nt protospacer immediately downstream of the 5′-GCCC-3′ motif, and a donor containing the nucleotide substitutions. Transformation of this plasmid into the DRM1 cells was carried out, followed by characterizing 16 randomly picked transformants through colony PCR and Sanger sequencing. The results showed that all tested colonies carried precise substitutions in the *ZMO0038* gene allele as designed (Figure 5D), giving again a 100% engineering efficiency. The expression of His-tagged ZMO0038 protein was also confirmed by Western blot. As shown in Figure 5E, a protein band with size of ∼23 kDa was revealed in the crude protein samples of the His-ZMO0038 cells, which was absent from the wild-type ZM4 strain. Together with the sequencing results, it could be concluded that the native Type I-F CRISPR-assisted *in situ* protein tagging represents a useful approach for gene modification in *Z. mobilis*.

### Application of the Type I-F CRISPR-based toolkit for multiplex genome engineering

One challenge of efficient genome engineering is to perform multiple gene editing simultaneously. A multiplex genome engineering plasmid, pMGE-Cs, was thus constructed for multiplex genome engineering in *Z. mobilis* targeting the CRISPR1, CRISPR2 and CRISPR3-4 loci at the same time (Figure 1A). The pMGE-Cs plasmid harboured three donor DNAs with each consisting of sequences immediately upstream and downstream of a target region, and an artificial CRISPR array consisting of three 32-bp spacers and four insulating direct repeats. All three spacers were designed to target on sequences within the vicinity of the corresponding genomic CRISPR regions, respectively (Figure 6A). Transformation of DRM1 cells with pMGE-Cs yielded a total of 263 transformants, getting an average transformation efficiency of 730 ± 38 CFU/μg plasmid DNA.

In order to confirm the genome editing efficiency, 16 transformants were randomly selected for colony PCR analysis using primer sets of Chk-C1-F/R for CRISPR1 (C1), Chk-C2-F/R for CRISPR2 (C2), and Chk-C3-4-F/R for CRISPR3-4 (C3-4), respectively (Supplementary Table S2). As shown in Figure 6B, 10 colonies (*i.e.* Strains 3–5, 7–12 and 16) carried the mutated C1 alleles, four colonies (*i.e.* Strains 3–5 and 16) contained C2 deletion, and 10 colonies (*i.e.* Strains 3–9 and 11–13) had deletion of C3-4 in their genomes, which gave PCR products of precited sizes of 645 bp in ΔC1 strains, 608 bp in ΔC2 strains and 945 bp in ΔC3-4 strains, respectively. These results were further confirmed by Sanger sequencing of the PCR products (Supplementary Figure S2). Collectively, Strains 3, 4 and 5 harbored all three designed mutations, showing an engineering efficiency of 18.75% (3/16). Besides, double deletions occurred in 6 strains, including deletion of C1 and C3-4 in Strains 7–9, 11 and 12, and deletion of C1 and C2 in Strain 16. In addition, Strain 10 carried single deletion of C1, while Strains 6 and 13 contained that of C3-4, showing an engineering efficiency of 37.50% (6/16) (Figure 6C). These results exhibited a considerably high overall engineering efficiency in multiplexing (75%) with 12 out of 16 tested colonies harbored at least one of the expected mutations, and an engineering efficiency of 18.75% (3/16) in simultaneously deleting all three target regions.

### Determination of the minimal donor size for mediating efficient mutant recovery from CRISPR-targeting

Obviously, along with the increasing number of target genes, the size of the engineering plasmid increased due to the accumulation of donor DNA fragments (ca. 1.4 kb each), which would largely hinder plasmid transformation rate and hence engineering efficiency. Therefore, we further determined the minimal donor size for conferring homologous recombination with a considerable rate.

We constructed plasmid derivatives of pKO-*ZMO0038* by replacing the ca. 1.4-kb donor (ca. 700 bp UF + ca. 700 bp DF sequences) to assess the effect of donor size on genome engineering efficiency, including pKO-*ZMO0038*-1 (100 bp UF + 100 bp DF), pKO-*ZMO0038*-2 (200 bp UF + 200 bp DF), pKO-*ZMO0038*-3 (300 bp UF + 300 bp DF), and pKO-*ZMO0038*-4 (400 bp UF + 400 bp DF) (Figure 7A). Transformants obtained with different plasmids of a similar amount (ca. 25 ng) varied greatly. Only one or two transformants could be obtained from each assay with pKO-*ZMO0038*-1 and none of which was correct *ZMO0038* deletants. With pKO-*ZMO0038*-2, totally 14 transformants (2, 6 and 6 transformants from three independent transformations, respectively) were obtained and 8 out of which harbored the *ZMO0038* deletion. Comparably, when transformed with pKO-*ZMO0038*-3 or pKO-*ZMO0038*-4, 48 and 103 transformants could be attained, respectively, yielding relatively much higher transformation rates (Figure 7B). More importantly, all tested transformants for each plasmid contained the precise deletion of *ZMO0038*.

Then we repeated the transformation experiments with a larger amount (250 ng) of pKO-*ZMO0038*-3. We observed further increased number of transformants (in a total of 153) and all 16 randomly chosen transformants were verified to be correct recombinants. In sharp contrast, transformation with 250 ng of pKO-*ZMO0038*-1 yielded only four transformants. And again, none of them contained a mutant *ZMO0038* gene allele (data not shown). Taken together, a 600-bp donor (300 bp UF + 300 bp DF) is sufficient for mediating very efficient mutant recovery from CRISPR-targeting-generated DNA injuries through homologous recombination, while a longer donor (*e.g.* 400

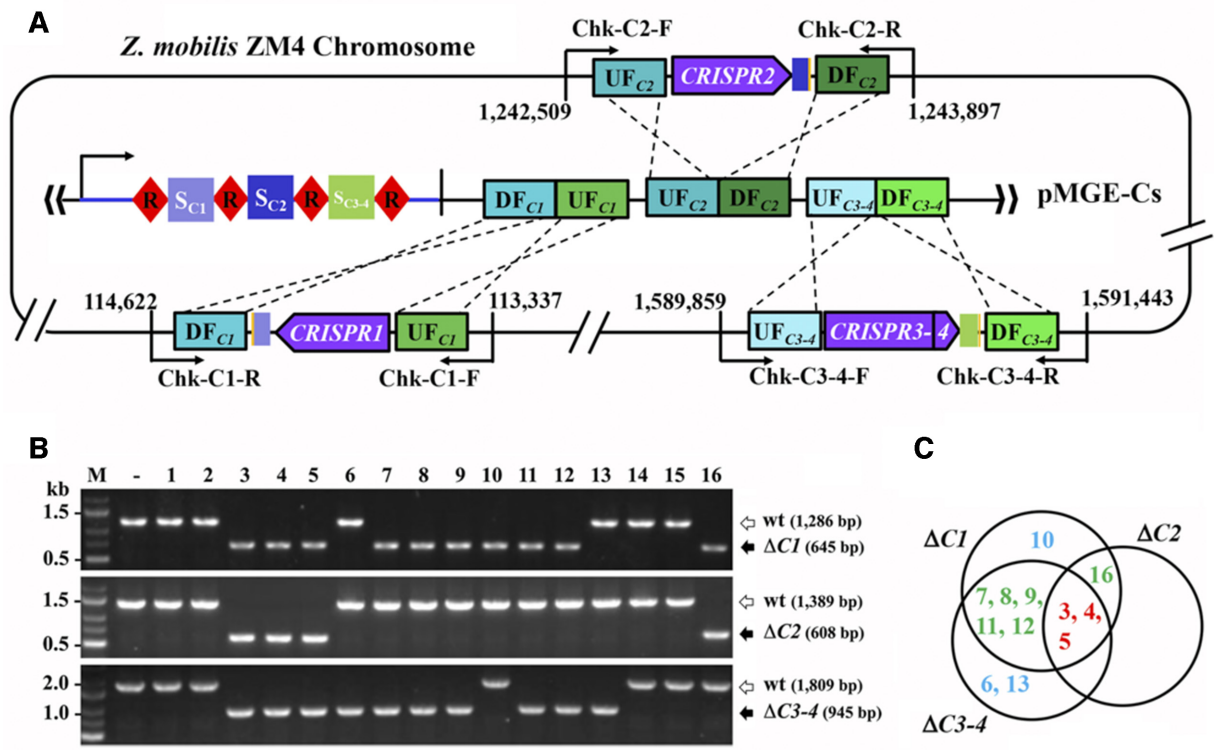


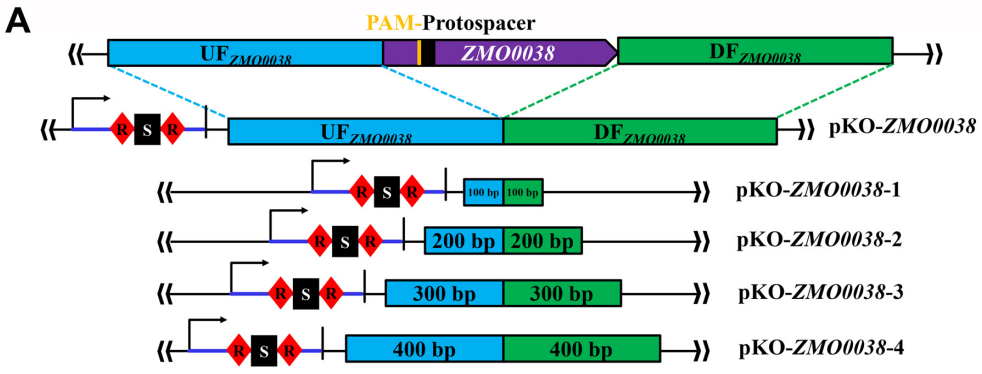

**Figure 6.** Multiplex genome engineering based on the Type I-F CRISPR–Cas system. (A) Schematic showing design of CRISPR-based simultaneous deletion of three CRISPR array-containing regions. The pMGE-Cs plasmid encodes three spacers that respectively match sequences within the vicinity of CRISPR1, CRISPR2, and CRISPR3-4. DNAs up-flanking (UF) or down-flanking (DF) of each target were concatenated on the same plasmid as recombination donors. (B) Colony PCR screening of deletion mutants using the corresponding primer sets as shown in (A), where the start sites of the primers are indicated. Predicted sizes of PCR products in wild-type (wt) and the expected deletion mutants (ΔC1 or ΔC2 or ΔC3-4) are indicated with unfilled and filled black arrows, respectively. -, PCR amplification using genomic DNA of *Z. mobilis* ZM4 as a DNA template; M, DNA size marker. (C) Distribution of gene deletions in the tested transformants. Transformants with all three deletions are shown in red fonts, while those with two or one are indicated in green and cyan fonts, respectively.

bp UF + 400 bp DF) might further improve the yield of transformants, and shorter ones (*e.g.* 200 bp UF + 200 bp DF) reduced both the transformant yield and genome engineering efficiency.

## DISCUSSION

In this study, we have established a native Type I-F CRISPR-based genome engineering toolkit for an important industrial microorganism *Z. mobilis*, which could be potentially expanded to many other prokaryotes. Like the Type I-A of *S. islandicus* (18) and the Type I-E of *L. crispatus* (22), the Type I-F CRISPR–Cas system of *Z. mobilis* is constitutively expressed (Supplementary Figure S1) and hence is ready-to-be-harnessed for genome engineering. By using this toolkit, genome engineering options, including gene or genomic fragment knockout, insertion or replacement, *in situ* tagging and specific point mutagenesis, and multiplexed gene deletion, can be readily achieved in *Z. mobilis*, representing thus far the most efficient and straightforward genome engineering toolkit for this important industrial microorganism. When compared with the currently routinely used genetic manipulation methods in *Z. mobilis* (47), this CRISPR-based toolkit is much simpler, more convenient and time-saving. For instance, due to the strong selection pressure conferred by CRISPR-targeting, no additional selection marker is required, and each edit with desired mutation has been obtained from transformants generated in a single round of transformation. The CRISPR-based toolkit established here will facilitate future effective and efficient genome editing to help understand the effect of single-nucleotide polymorphisms (SNPs) on microbial phenotypic improvement and unravel the underlying mechanism of inhibitor tolerance and stress responses (49,50), which were not able to be easily achieved by other classical genetics approaches. This toolkit might also expedite the development of *Z. mobilis* as a synthetic chassis for sustainable economic biofuel and biochemical productions. More importantly, our work here, and others' (18–22), have provided a framework for exploiting endogenous CRISPR–Cas systems, which widely occur in bacteria and archaea, for genome engineering in their native hosts, and will prompt the development of native CRISPR-based next-generation genome engineering technologies.

Due to lack of an efficient non-homologous end joining (NHEJ) system in most prokaryotes, repair of DNA injuries introduced by CRISPR-directed chromosomal self-targeting in bacteria and archaea mainly relies on the homologous recombination (HR) pathway, where episomal templates are generally required. Therefore, the concentration and stability of DNA templates would determine the homologous recombination rate after DNA injuries. In-




B

| Plasmid | Number of transformants | | | Tested transformants |
|---|---|---|---|---|
| | Assay1 | Assay2 | Assay3 | kb M - 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 |
| pKO-*ZMO0038*-1 | 2 | 1 | 1 | 1.50 0.75 ⇦ 1,334 bp |
| pKO-*ZMO0038*-2 | 2 | 6 | 6 | 1.50 0.75 ⇦ ⬅ 799 bp |
| pKO-*ZMO0038*-3 | 19 | 13 | 16 | 1.50 0.75 ⇦ ⬅ |
| pKO-*ZMO0038*-4 | 36 | 30 | 37 | 1.50 0.75 ⇦ ⬅ |

**Figure 7.** Assessment of the impact of donor size on genome engineering efficiency. (A) Schematic showing structures of the pKO-*ZMO0038* plasmid and its derivatives carrying shortened donors. The size of the donor in each pKO-*ZMO0038* derivative is shown. (B) Number of transformants obtained from three independent assays with the pKO-*ZMO0038* derivatives are present. Randomly selected transformants were colony PCR-analyzed and the predicted sizes of PCR products in wild-type (wt) and the expected *ZMO0038* deletion mutants are indicated with unfilled and filled black arrows, respectively. -, PCR amplification of genomic DNA of *Z. mobilis* ZM4; M, DNA size marker.

deed, chemical modification of the DNA template was observed to dramatically enhance the HR rate by increasing the stability of the DNA template (51). In addition, supplying DNA templates as inserts of plasmids allows their multiplication through plasmid replication, thus conferring more efficient genome editing than as free DNA fragments (19). By taking the similar approach, we achieved deletion of >10 kb genomic fragment, which might provide an efficient method for future construction of minimal microbial genome of *Z. mobilis*. In addition, recombination between identical sequences resulted in internal deletions (52,53), indicating a microhomology-directed homologous recombination mechanism. Using the same mechanism, deletion of >50 kb genomic fragment was achieved in *S. islandicus* (54). Moreover, other mechanisms, *e.g.* the Ligase A-dependent alternative end-joining (A-EJ) first demonstrated for *E. coli* (55), may also potentiate the CRISPR-based toolkit for the removal of large genomic fragments.

Escape from CRISPR-targeting was observed in our studies. The escape modes and the underlying mechanisms would provide possibility for further boosting the overall engineering efficiency. We found that all escaping transformants still harbored an engineering plasmid that conferred the cells antibiotics resistance, while the genome-targeting spacer was deleted in most cells (>70% of the escaping transformants). This was possibly caused by the homologous recombination between direct repeats (Supplementary Table S4) as reported in other studies (56,57). The escape from CRISPR-targeting is particularly evident in the plasmid for multiplex genome engineering (pMGE-Cs) (Supplementary Table S5). It could be speculated that two more direct repeats present in pMGE-Cs compared to other engineering plasmids could improve the recombination possibility and led to various deletion modes. Such an issue might be addressed by producing mature crRNAs through a Cas6-independent pathway, where each crRNA was generated by processing the fusing t-elements by RNase P and tRNase Z (58). Nevertheless, using a single CRISPR array containing tandem repeat-spacer units, all the engineering purposes were obtainable with a considerably high efficiency. In our studies, we carefully examined the sequences of the CRISPR repeats and spacers on the plasmid, and the protospacers and PAMs in the genome. No change in any of these regions was observed (data not shown). Further work can be carried out to identify the exact mutations in the *cas* operon (ca. 8 kb). However, it should not affect our purpose to select the candidates with correct editing since the escaping transformants occurred in a small percentage (<30%).

Apart from the different *cas2/3* knockouts obtained using the Type I-F CRISPR-based genome engineering toolkit in this study, we also successfully introduced nucleotide substitutions at an HD-active site of the Cas2/3 nuclease (Supplementary Figure S3). The nucleotide substitutions were introduced without disrupting the target sequence of the designed crRNA guide, due to the lack of any suitable 5′-NCC-3′ PAM that could render the changes

within the target (Supplementary Figure S3). Instead, the nucleotide substitutions resulted in inactivation of the nuclease activity of Cas2/3 and hence the DNA interference function of the Type I-F CRISPR–Cas system, which thus allowed the screening and obtaining of the intended dCas2/3 mutant. Given the determination of PAM sequences, it would be possible to use alternative PAMs, such as those in Group 2 or 3 (Figure 2), in genome engineering assays where no Group 1 PAM could be used. Also, it would be also good to introduce a silent mutation into the target sequence along with the intended mutations to yield edits with desired phenotypes as previously conducted (59).

Our results also demonstrated that the Type I-F Cascade without Cas2/3 can be directed to a targeted gene sequence, acting as a transcriptional barrier, for efficient *in vivo* gene silencing in *Z. mobilis*. It was reported that while the Type I-B system in *Haloferax volcanii* was harnessed for efficient repression of transcription upon the depletion of Cas3, the expression of a catalytically inactive Cas3 via plasmid gave a further enhanced knockdown effect on an endogenous gene (23), which might suggest the help of a dCas3 on stabilizing the target-bound Cascade. Additionally, upon being fused with the bacterial RNAP activator SoxS, the dCas2/3 constructed could be used to recruit RNAP for transcription activation, which is in parallel with the functions of dCas9/dCas12a in Class 2 systems (60). Collectively, the native CRISPR-dCas2/3 system can be employed as CRISPR interference (CRISPRi) and activation (CRISPRa) toolkits or similarly as CRISPR-assisted multi-dimensional regulation tool for fine-tuning gene expression (61) and metabolic engineering in *Z. mobilis* and related microorganisms in the future.

In this study, we demonstrated that the enhancement of plasmid DNA transformation efficiencies through inactivating the Type IV DNA R–M system in *Z. mobilis* led to boosted genome engineering efficiencies of up to 100% (Table 1). Multiplexed gene deletion, which we failed for several attempts in the wild-type strain, was achieved with a considerably high efficiency under this R–M$^-$ background. R–M systems naturally occur in prokaryotes even more broadly than CRISPR–Cas systems, pointing towards a fact that both adaptive CRISPR–Cas immune systems and innate R–M systems may coexist in many bacteria and archaea (62). Thus, it may explain the perplexity that the diverse CRISPR–Cas systems broadly occur in prokaryotes in nature whereas the native CRISPR-based technologies are much lesser applied. Possibly, highly efficient native CRISPR-based genome engineering in these bacteria and archaea would be only enabled under, at least, an R–M$^-$ background. In addition, other prokaryotic defense modules that also prevent foreign DNA introduction (63) would be taken account as well if occurred. Together, the method developed here may serve as an important reference for the development and deployment of similar toolkits in prokaryotes that harbour active endogenous CRISPR–Cas systems but with low efficiency in the wild-type genetic background.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

*Author Contributions*: W.P., S.Y., L.Y., Y.Z. and L.M. designed the research; Y.Z., J.H., B.W., X.H., R.L. and W.S. performed the experiments; Y.Z., W.P., S.Y. and L.Y. wrote the manuscript. All authors contributed to data analyses, read, revised and approved the final manuscript.

## FUNDING

National Key Technology Research and Development Program of China [2018YFD0500203 to L.Y. and 2018YFA0900300 to S.Y.]; National Science Foundation of China [31870057 to L.Y. and W.P., 21978071 and U1932141 to S.Y.]; Natural Science Foundation of Hubei Province of China [2017CFB538 to W.P.]; Scientific Research Program of Hubei Provincial Department of Education [Q20161007 to W.P.]; Technical Innovation Special Fund of Hubei Province [2019AHB055 and 2018ACA149 to S.Y.]; W.P., S.Y. and L.Y. also acknowledge the support from State Key Laboratory of Biocatalysis and Enzyme Engineering. Funding for open access charge: National Key Technology Research and Development Program of China [2018YFD0500203, 2018YFA0900300].
*Conflict of interest statement.* The authors declare that they have a patent pending to this material.

## REFERENCES

1. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, **8**, 172.
2. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
3. van der Oost,J., Westra,E.R., Jackson,R.N. and Wiedenheft,B. (2014) Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **12**, 479–492.
4. Koonin,E.V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
5. Brouns,S.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J., Snijders,A.P., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
6. van Erp,P.B.G., Patterson,A., Kant,R., Berry,L., Golden,S.M., Forsman,B.L., Carter,J., Jackson,R.N., Bothner,B. and Wiedenheft,B. (2018) Conformational dynamics of DNA binding and Cas3 recruitment by the CRISPR RNA-Guided cascade complex. *ACS Chem. Biol.*, **13**, 481–490.
7. Rollins,M.F., Chowdhury,S., Carter,J., Golden,S.M., Wilkinson,R.A., Bondy-Denomy,J., Lander,G.C. and Wiedenheft,B. (2017) Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E5113–E5121.
8. Mojica,F.J., Diez-Villasenor,C., Garcia-Martinez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
9. Rollins,M.F., Schuman,J.T., Paulus,K., Bukhari,H.S. and Wiedenheft,B. (2015) Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from *Pseudomonas aeruginosa*. *Nucleic Acids Res.*, **43**, 2216–2222.
10. Pickar-Oliver,A. and Gersbach,C.A. (2019) The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.*, **20**, 490–507.
11. Xu,X. and Qi,L.S. (2019) A CRISPR-dCas toolbox for genetic engineering and synthetic biology. *J. Mol. Biol.*, **431**, 34–47.

12. Hidalgo-Cantabrana,C., O'Flaherty,S. and Barrangou,R. (2017) CRISPR-based engineering of next-generation lactic acid bacteria. *Curr. Opin. Microbiol.*, **37**, 79–87.
13. Yang,S., Fei,Q., Zhang,Y., Contreras,L.M., Utturkar,S.M., Brown,S.D., Himmel,M.E. and Zhang,M. (2016) *Zymomonas mobilis* as a model system for production of biofuels and biochemicals. *Microb. Biotechnol.*, **9**, 699–717.
14. Rutkis,R., Kalnenieks,U., Stalidzans,E. and Fell,D.A. (2013) Kinetic modelling of the *Zymomonas mobilis* Entner-Doudoroff pathway: insights into control and functionality. *Microbiology*, **159**, 2674–2689.
15. Wang,X., He,Q., Yang,Y., Wang,J., Haning,K., Hu,Y., Wu,B., He,M., Zhang,Y., Bao,J. *et al.* (2018) Advances and prospects in metabolic engineering of *Zymomonas mobilis*. *Metab. Eng.*, **50**, 57–73.
16. Yang,S., Mohagheghi,A., Franden,M.A., Chou,Y.C., Chen,X., Dowe,N., Himmel,M.E. and Zhang,M. (2016) Metabolic engineering of *Zymomonas mobilis* for 2,3-butanediol production from lignocellulosic biomass sugars. *Biotechnol. Biofuels*, **9**, 189.
17. Cao,Q.H., Shao,H.H., Qiu,H., Li,T., Zhang,Y.Z. and Tan,X.M. (2017) Using the CRISPR/Cas9 system to eliminate native plasmids of *Zymomonas mobilis* ZM4. *Biosci. Biotechnol. Biochem.*, **81**, 453–459.
18. Li,Y., Pan,S., Zhang,Y., Ren,M., Feng,M., Peng,N., Chen,L., Liang,Y.X. and She,Q. (2016) Harnessing Type I and Type III CRISPR–Cas systems for genome editing. *Nucleic Acids Res.*, **44**, e34.
19. Cheng,F., Gong,L., Zhao,D., Yang,H., Zhou,J., Li,M. and Xiang,H. (2017) Harnessing the native type I-B CRISPR–Cas for genome editing in a polyploid archaeon. *J Genet Genomics*, **44**, 541–548.
20. Pyne,M.E., Bruder,M.R., Moo-Young,M., Chung,D.A. and Chou,C.P. (2016) Harnessing heterologous and endogenous CRISPR–Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci. Rep.*, **6**, 25666.
21. Zhang,J., Zong,W., Hong,W., Zhang,Z.T. and Wang,Y. (2018) Exploiting endogenous CRISPR–Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab. Eng.*, **47**, 49–59.
22. Hidalgo-Cantabrana,C., Goh,Y.J., Pan,M., Sanozky-Dawes,R. and Barrangou,R. (2019) Genome editing using the endogenous type I CRISPR–Cas system in *Lactobacillus crispatus*. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 15774–15783.
23. Stachler,A.E. and Marchfelder,A. (2016) Gene repression in *Haloarchaea* using the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas I-B system. *J. Biol. Chem.*, **291**, 15226–15242.
24. Rath,D., Amlinger,L., Hoekzema,M., Devulapally,P.R. and Lundgren,M. (2015) Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.*, **43**, 237–246.
25. Luo,M.L., Mullis,A.S., Leenay,R.T. and Beisel,C.L. (2015) Repurposing endogenous type I CRISPR–Cas systems for programmable gene repression. *Nucleic Acids Res.*, **43**, 674–681.
26. Okamoto,T. and Nakamura,K. (1992) Simple and highly efficient transformation method for *Zymomonas mobilis*: electroporation. *Biosci. Biotechnol. Biochem.*, **56**, 833.
27. Horton,R.M., Cai,Z.L., Ho,S.N. and Pease,L.R. (1990) Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *BioTechniques*, **8**, 528–535.
28. Xia,Y., Li,K., Li,J., Wang,T., Gu,L. and Xun,L. (2019) T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Res.*, **47**, e15.
29. Yang,Y., Shen,W., Huang,J., Li,R., Xiao,Y., Wei,H., Chou,Y.C., Zhang,M., Himmel,M.E., Chen,S. *et al.* (2019) Prediction and characterization of promoters and ribosomal binding sites of *Zymomonas mobilis* in system biology era. *Biotechnol. Biofuels*, **12**, 52.
30. Mohagheghi,A., Linger,J., Smith,H., Yang,S., Dowe,N. and Pienkos,P.T. (2014) Improving xylose utilization by recombinant *Zymomonas mobilis* strain 8b through adaptation using 2-deoxyglucose. *Biotechnol. Biofuels*, **7**, 19.
31. Yang,S., Tschaplinski,T.J., Engle,N.L., Carroll,S.L., Martin,S.L., Davison,B.H., Palumbo,A.V., Rodriguez,M. Jr and Brown,S.D. (2009) Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. *BMC Genomics*, **10**, 34.
32. Hakkila,K., Maksimow,M., Rosengren,A., Karp,M. and Virta,M. (2003) Monitoring promoter activity in a single bacterial cell by using green and red fluorescent proteins. *J. Microbiol. Methods*, **54**, 75–79.
33. Chudakov,D.M., Matz,M.V., Lukyanov,S. and Lukyanov,K.A. (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.*, **90**, 1103–1163.
34. Seo,J.S., Chong,H., Park,H.S., Yoon,K.O., Jung,C., Kim,J.J., Hong,J.H., Kim,H., Kim,J.H., Kil,J.I. *et al.* (2005) The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nat. Biotechnol.*, **23**, 63–68.
35. Yang,S., Pappas,K.M., Hauser,L.J., Land,M.L., Chen,G.L., Hurst,G.B., Pan,C., Kouvelis,V.N., Typas,M.A., Pelletier,D.A. *et al.* (2009) Improved genome annotation for *Zymomonas mobilis*. *Nat. Biotechnol.*, **27**, 893–894.
36. Yang,S., Vera,J.M., Grass,J., Savvakis,G., Moskvin,O.V., Yang,Y., McIlwain,S.J., Lyu,Y., Zinonos,I., Hebert,A.S. *et al.* (2018) Complete genome sequence and the expression pattern of plasmids of the model ethanologen *Zymomonas mobilis* ZM4 and its xylose-utilizing derivatives 8b and 2032. *Biotechnol. Biofuels*, **11**, 125.
37. Richter,C., Dy,R.L., McKenzie,R.E., Watson,B.N., Taylor,C., Chang,J.T., McNeil,M.B., Staals,R.H. and Fineran,P.C. (2014) 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.
38. Haurwitz,R.E., Jinek,M., Wiedenheft,B., Zhou,K. and Doudna,J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
39. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.*, **35**, W52–W57.
40. Couvin,D., Bernheim,A., Toffano-Nioche,C., Touchon,M., Michalik,J., Neron,B., Rocha,E.P.C., Vergnaud,G., Gautheret,D. and Pourcel,C. (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.*, **46**, W246–W251.
41. Biswas,A., Gagnon,J.N., Brouns,S.J., Fineran,P.C. and Brown,C.M. (2013) CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.*, **10**, 817–827.
42. Li,H. (2015) Structural principles of CRISPR RNA processing. *Structure*, **23**, 13–20.
43. Sternberg,S.H., Haurwitz,R.E. and Doudna,J.A. (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA*, **18**, 661–672.
44. Peng,W., Li,H., Hallstrom,S., Peng,N., Liang,Y.X. and She,Q. (2013) Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol.*, **10**, 738–748.
45. Fischer,S., Maier,L.K., Stoll,B., Brendel,J., Fischer,E., Pfeiffer,F., Dyall-Smith,M. and Marchfelder,A. (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J. Biol. Chem.*, **287**, 33351–33363.
46. Leenay,R.T., Maksimchuk,K.R., Slotkowski,R.A., Agrawal,R.N., Gomaa,A.A., Briner,A.E., Barrangou,R. and Beisel,C.L. (2016) Identifying and visualizing functional PAM diversity across CRISPR–Cas systems. *Mol. Cell*, **62**, 137–147.
47. He,M.X., Wu,B., Qin,H., Ruan,Z.Y., Tan,F.R., Wang,J.L., Shui,Z.X., Dai,L.C., Zhu,Q.L., Pan,K. *et al.* (2014) Zymomonas mobilis: a novel platform for future biorefineries. *Biotechnol. Biofuels*, **7**, 101.
48. Kerr,A.L., Jeon,Y.J., Svenson,C.J., Rogers,P.L. and Neilan,B.A. (2011) DNA restriction–modification systems in the ethanologen, *Zymomonas mobilis* ZM4. *Appl. Microbiol. Biotechnol.*, **89**, 761–769.
49. Wu,B., Qin,H., Yang,Y., Duan,G., Yang,S., Xin,F., Zhao,C., Shao,H., Wang,Y., Zhu,Q. *et al.* (2019) Engineered *Zymomonas mobilis* tolerant to acetic acid and low pH via multiplex atmospheric and room temperature plasma mutagenesis. *Biotechnol. Biofuels*, **12**, 10.
50. Mohagheghi,A., Linger,J.G., Yang,S., Smith,H., Dowe,N., Zhang,M. and Pienkos,P.T. (2015) Improving a recombinant *Zymomonas mobilis* strain 8b through continuous adaptation on dilute acid pretreated corn stover hydrolysate. *Biotechnol. Biofuels*, **8**, 55.
51. Renaud,J.B., Boix,C., Charpentier,M., De Cian,A., Cochennec,J., Duvernois-Berthet,E., Perrouault,L., Tesson,L., Edouard,J., Thinard,R. *et al.* (2016) Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR–Cas9 nucleases. *Cell Rep.*, **14**, 2263–2272.
52. Vercoe,R.B., Chang,J.T., Dy,R.L., Taylor,C., Gristwood,T., Clulow,J.S., Richter,C., Przybilski,R., Pitman,A.R. and Fineran,P.C. (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems

can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLos Genet.*, **9**, e1003454.
53. Canez,C., Selle,K., Goh,Y.J. and Barrangou,R. (2019) Outcomes and characterization of chromosomal self-targeting by native CRISPR–Cas systems in *Streptococcus thermophilus*. *FEMS Microbiol. Lett.*, **366**, fnz105.
54. Zhang,C. and Whitaker,R.J. (2018) Microhomology-mediated high-throughput gene inactivation strategy for the hyperthermophilic Crenarchaeon *Sulfolobus islandicus*. *Appl. Environ. Microbiol.*, **84**, e02167-17.
55. Chayot,R., Montagne,B., Mazel,D. and Ricchetti,M. (2010) An end-joining repair mechanism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2141–2146.
56. Gudbergsdottir,S., Deng,L., Chen,Z., Jensen,J.V., Jensen,L.R., She,Q. and Garrett,R.A. (2011) Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.*, **79**, 35–49.
57. Stout,E.A., Sanozky-Dawes,R., Goh,Y.J., Crawley,A.B., Klaenhammer,T.R. and Barrangou,R. (2018) Deletion-based escape of CRISPR–Cas9 targeting in *Lactobacillus gasseri*. *Microbiology*, **164**, 1098–1111.
58. Maier,L.K., Stachler,A.E., Saunders,S.J., Backofen,R. and Marchfelder,A. (2015) An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. *J. Biol. Chem.*, **290**, 4192–4201.
59. Garst,A.D., Bassalo,M.C., Pines,G., Lynch,S.A., Halweg-Edwards,A.L., Liu,R., Liang,L., Wang,Z., Zeitoun,R., Alexander,W.G. *et al.* (2017) Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.*, **35**, 48–55.
60. Yao,R., Liu,D., Jia,X., Zheng,Y., Liu,W. and Xiao,Y. (2018) CRISPR–Cas9/Cas12a biotechnology and application in bacteria. *Synth. Syst. Biotechnol.*, **3**, 135–149.
61. Lu,Z., Yang,S., Yuan,X., Shi,Y., Ouyang,L., Jiang,S., Yi,L. and Zhang,G. (2019) CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in *Bacillus subtilis*. *Nucleic Acids Res.*, **47**, e40.
62. Koonin,E.V., Makarova,K.S. and Wolf,Y.I. (2017) Evolutionary genomics of defense systems in archaea and bacteria. *Annu. Rev. Microbiol.*, **71**, 233–261.
63. Doron,S., Melamed,S., Ofir,G., Leavitt,A., Lopatina,A., Keren,M., Amitai,G. and Sorek,R. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, **359**, eaar4120.