nature microbiology

LETTERS

https://doi.org/10.1038/s41564-017-0102-6

Deaminase-mediated multiplex genome editing in *Escherichia coli*

Satomi Banno[1], Keiji Nishida[1*], Takayuki Arazoe[1,3], Hitoshi Mitsunobu[1] and Akihiko Kondo[1,2*]

In eukaryotes, the CRISPR–Cas9 system has now been widely used as a revolutionary genome engineering tool[1,2]. However, in prokaryotes, the use of nuclease-mediated genome editing tools has been limited to negative selection for the already modified cells because of its lethality[3,4]. Here, we report on deaminase-mediated targeted nucleotide editing (Target-AID)[5] adopted in *Escherichia coli*. Cytidine deaminase PmCDA1 fused to the nuclease-deficient CRISPR–Cas9 system achieved specific point mutagenesis at the target sites in *E. coli* by introducing cytosine mutations without compromising cell growth. The cytosine-to-thymine substitutions were induced mainly within an approximately five-base window of target sequences on the protospacer adjacent motif-distal side, which can be shifted depending on the length of the single guide RNA sequence. Use of a uracil DNA glycosylase inhibitor[6] in combination with a degradation tag (LVA tag)[7] resulted in a robustly high mutation efficiency, which allowed simultaneous multiplex editing of six different genes. The major multi-copy transposase genes that consist of at least 41 loci were also simultaneously edited by using four target sequences. As this system does not rely on any additional or host-dependent factors, it may be readily applicable to a wide range of bacteria.

For genome editing in prokaryotes, phage-derived RecET and lambda red recombinases have been employed in recombineering, which facilitates homology-dependent integration/replacement of a donor DNA or oligonucleotide[4,8]. In combination with the strain, which is defective in methyl-directed mismatch repair, highly efficient recombineering can be achieved without integration of a selectable marker[9]. This system is employed in multiplex automated genome engineering to generate a genetic variety of the multiple target loci in *E. coli* within a few days[10]. However, these systems require the genetic background of the target strains such as deficiency of methyl-directed mismatch repair or RecA that involves in the recombinational DNA repair system[11].

Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins have been known to serve as a bacterial adaptive immune system[12] by cleaving target DNA in a single guide RNA (sgRNA)- and protospacer adjacent motif (PAM)-dependent manner. A Cas9 nuclease derived from *Streptococcus pyogenes* has been widely used as a powerful genome engineering tool in eukaryotes that possess repair pathways for DNA double-strand breaks[1,2]. During repair of double-strand breaks by the non-homologous end joining pathway, small insertions and/or deletions (indels) are introduced, leading to a site-specific mutation or gene disruptions in the target DNA. For more precise editing, homology-directed repair can be facilitated by providing donor DNA containing homology arms for the target region, although the efficiency varies depending on the host cells. In most bacteria, chromosomal cleavages by targeted nucleases result in cell death due to a lack of the non-homologous end joining pathway[13,14]. Therefore, CRISPR–Cas9 has been employed to eliminate non-modified cells from a mixed population for enrichment of gene-modified cells by other methods such as the lambda red recombination system[15–18]. Recently, deaminase-mediated targeted base editing has been demonstrated[19–21] to directly edit nucleotides at the target locus without DNA cleavage as well as the template donor DNA. As it employs DNA deamination in place of nuclease-mediated DNA cleavage, it may not induce cell death in bacteria and thus could be applicable in bacterial genome engineering.

To assess whether the deaminase-mediated targeted mutagenesis is applicable in bacteria, we constructed a bacterial Target-AID vector that expresses catalytically inactivated Cas9 (dCas: D10A and H840A mutations) fused to *Petromyzon marinus* cytosine deaminase PmCDA1[19] (CDA) under a temperature-inducible lambda operator system[22] and a 20-nucleotide (nt) target sequence–gRNA scaffold hybrid (sgRNA) under a synthetic constitutive promoter J23119 (Fig. 1b). While nickase Cas9 (nCas: D10A mutation) in combination with deaminase can be used in eukaryotes to perform higher mutational efficiency[19,20], the plasmid expressing nCas(D10A)–CDA showed poor transformation efficiency in *E. coli*, suggesting a severe growth defect or cell death similarly to the full Cas9 nuclease (Supplementary Fig. 1). To quantitatively assess the efficacy of targeted mutagenesis, the *galK* gene was used as a target whose loss of function can be positively selected by a galactose analogue 2-deoxy-D-galactose (2-DOG), which is catalysed by the *galK* gene product galactokinase to form a toxic compound[23]. Target-AID has been reported to induce mutations at cytosine nucleotides (C) located approximately 15–20 bases upstream of the PAM[19] (Fig. 1a). The targeting sequence was selected in *galK* to introduce a stop codon (Fig. 1c) and induced nearly 100% viability against 2-DOG, which suggests highly efficient mutagenesis (Supplementary Fig. 2a). Sequencing analysis revealed that six out of eight colonies from non-2-DOG selective medium contained expected C-to-T substitutions at the −17 and/or −20 positions (Fig. 1c).

Next, an essential gene, *rpoB*, was targeted that encodes the β-subunit of RNA polymerase. While disturbing the function of *rpoB* would cause a growth defect or cell death, specific point mutations in the gene confer rifampicin resistance[24]. Transformed cells with a sgRNA designed to induce one of these mutations gained rifampicin resistance at a frequency of nearly 100% without obvious growth defects (Supplementary Fig. 2b). Sequencing analysis of the clones without rifampicin selection confirmed the C-to-T substitutions at −16 and/or −17 positions from the PAM (position 1545 and 1546 of *rpoB*)

[1]Graduate School of Science, Technology and Innovation, Kobe University, Kobe, Hyogo, Japan. [2]Department of Chemical Science and Engineering, Graduate School of Engineering, Kobe University, Kobe, Hyogo, Japan. [3]Present address: Department of Applied Biological Science, Faculty of Science and Technology, Tokyo University of Science, Noda, Chiba, Japan. *e-mail: keiji_nishida@people.kobe-u.ac.jp; akondo@kobe-u.ac.jp

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

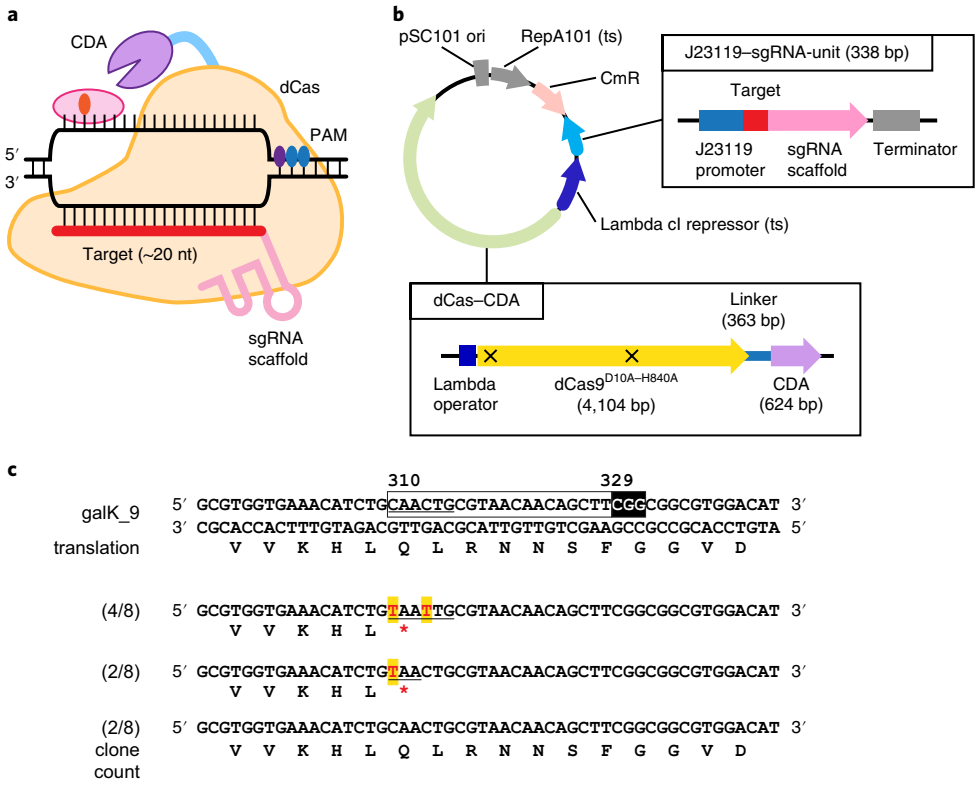


Fig. 1 | Single mutagenesis using Target-AID in bacteria. a, A schematic model for Target-AID (dCas–CDA/sgRNA) base editing. The dCas–CDA/sgRNA complex binds to the double-stranded DNA to form an R-loop in a sgRNA- and PAM-dependent manner. CDA catalyses deamination of cytosines located at the top (non-complementary) strand within 15–20 bases upstream from PAM, which results in C-to-T mutagenesis. **b**, A Target-AID plasmid for a single target in bacteria. The plasmid contains the chloramphenicol-resistant (CmR) gene, temperature-sensitive (ts) lambda cI repressor, the pSC101 origin (ori) and RepA101 (ts). The lambda operator expresses the dCas–CDA fusion at a high temperature (>37 °C) as the cI repressor (ts) gets inactivated. The sgRNA is expressed under a constitutive promoter J23119. The dCas–CDA unit consists of nuclease-deficient Cas9 with D10A and H840A mutations and *Petromyzon marinus* cytosine deaminase. **c**, Sequence alignment of the galK mutants edited by dCas–CDA. The genomic loci and the site targeted by galK_9 sgRNA (box) are shown with a PAM (inverted box). The number on the top indicates the open reading frame number. Mutated sites found from eight randomly picked colonies are highlighted in yellow and mutated bases and amino acids are shown in red. The mutated codons are underlined.

as expected (Fig. 2a). Whole-genome sequencing was performed to assess the possible non-specific mutator effect of Target-AID in *E. coli*. Three independent clones expressing dCas–CDA and the rpoB_1-targeting sgRNA were analysed and found to contain zero to two single-nucleotide variants (SNVs) at an apparently unrelated genomic loci (Fig. 2b). Adjacent sequences of the SNVs did not show any similarity to the *rpoB* target sequence (Fig. 2c).

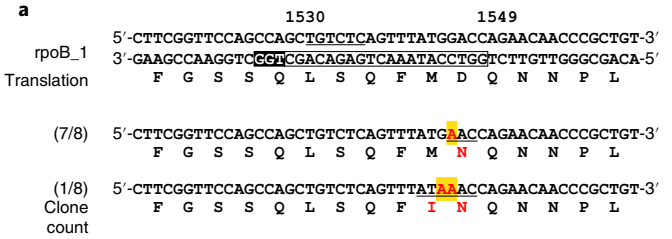
To comprehensively analyse mutational efficiency and positions, a deep-sequencing analysis was performed using 18 targets in *galK* (Fig. 3a and Supplementary Table 1). Seven targets showed highly efficient (61.7–95.1%) mutagenesis while five showed poor (1.4–9.2%) mutagenesis. The most effective mutational positions were 17–20 bases upstream of PAM, which was consistent with a previous study in higher organisms. The mutation frequency also varied depending on the length of the target sequence as the frequency increased at galK_8 and galK_13 and decreased at galK_9 and galK_11 when sgRNAs with longer target sequences were used (Fig. 3a, blue bars).

To improve the mutational efficiency, uracil DNA glycosylase inhibitor (UGI) from bacteriophage PBS2[6] and protein degradation tag (LVA tag)[7] were introduced by fusing them at the carboxy terminus of dCas–CDA, which is named dCas–CDA–UL. UGI increases mutagenesis by cytidine deamination because it inhibits removal of uracil (an immediate product of cytosine deamination) from DNA[19,20]. As dCas–CDA–UGI could be potentially harmful when expressed in excess, the LVA tag was used to decrease the concentration of the protein in cells. dCas–CDA–UL with sgRNAs targeting galK showed robust mutagenesis at all of the targeted sites tested (Fig. 3a, red bars). Using longer sgRNAs, dCas–CDA–UL induced extended mutational spectra towards the 5′ ends (Fig. 3a, galK_9 and galK_11). To further characterize the effect of the length of the sgRNA target sequence, C-rich genomic loci were tested with sgRNAs bearing target sequences varying in length (18, 20, 22 and 24 nt; Supplementary Fig. 3). Mutational spectra for each of the five target sites consistently showed the peak shift and expansion of the active window towards the 5′ end as the target sequences lengthened (Fig. 3b). The intrinsic mutagenic activity of dCas–CDA and dCas–CDA–UL in the absence of sgRNA was assessed by whole-genome sequencing of the cells expressing the proteins after 6-h induction at 37 °C. dCas–CDA and dCas–CDA–UL induced none to two and 21–30 SNV mutations, respectively, with no positional bias in the entire genome (Supplementary Tables 2 and 3), indicating that the dCas–CDA–UL induces genome-wide mutagenesis in this condition.

For multiplex editing, tandem repeats of sgRNA-expression units were assembled onto a separate plasmid from the one expressing dCas–CDA variant proteins (Fig. 4a). The plasmid targeting three sites in *galK* (galK_10, galK_11 and galK_13) was co-introduced into cells with the plasmid expressing dCas, dCas–CDA, dCas–CDA–L or dCas–CDA–UL. The non-specific mutator effect of the variant proteins was assessed by analysis of the occurrence of rifampicin-resistant mutants (Supplementary Fig. 4a). dCas–CDA was

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

a



b

Strain	Clone	Sequence coverage	Parental/ variable mutation	Detected mutations				
				Counts	Regions	Genes	References	Alleles
dCas–CDA /rpoB_1	1	27.9	2	2	1,591,287 4,172,717	*yneO* *rpoB*	C GG	T AA
	1	22.1	3	3	2,170,816 4,056,442 4,172,717	*gatC* *yihN* *rpoB*	G C GG	A A AA
	3	51.1	1	1	4,172,717	*rpoB*	GG	AA

c

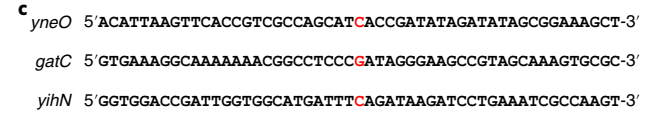


Fig. 2 | Gain-of-function mutagenesis of the *rpoB* gene. a, Sequence alignment of *rpoB* mutants edited by dCas–CDA. The genomic loci and the site targeted by rpoB_1 sgRNA (box) are shown with a PAM (inverted box). The number on the top indicates the open reading frame number. Mutated sites found from eight randomly picked colonies from non-rifampicin selective medium are highlighted in yellow and mutated bases and amino acids are shown in red. Mutated codons are underlined. **b**, Whole-genome sequencing analyses of the rpoB-targeted cells. Three rifampicin-selected clones were subjected to whole-genome sequencing. Sequence coverage was calculated as the sum of the base pairs of the mapped sequence over 4,631 kbp of the *E. coli* BW25113 genome sequence. The parental/variable mutation column indicates the number of variants obtained by subtracting common parental mutations from the variants detected with over a 50% frequency including insertions, deletions, SNVs and multiple-nucleotide variants. The detected mutations columns show the number of mutations (counts), genomic loci (regions/genes), reference genome sequences (references) and mutant alleles (alleles). Variant calling was performed as described in the Methods. **c**, Sequences around the detected mutations listed in **b** except for the *rpoB* locus. Mutated bases are shown in red.

shown to induce an approximately tenfold increase of nonspecific mutation over the background, which is mitigated by the addition of LVA. Addition of UGI–LVA rather induced tenfold mutations over dCas9–CDA. dCas–CDA and dCas–CDA-L were unable to efficiently obtain a triple mutant for the three *galK* target sites at once, while dCas–CDA–UL succeeded in triple mutagenesis in five out of the eight analysed clones (Supplementary Fig. 4b). The multiplex mutational frequency by dCas–CDA–UL seemed to be less efficient compared with those against a single target (the mutations were observed in all clones tested; Supplementary Fig. 4c,d). Six different genes (*galK*, *xylB* (xylulokinase), *manA* (mannose-6-phosphate isomerase), *pta* (phosphate acetyltransferase), *adhE* (aldehyde-alcohol dehydrogenase) and *tpiA* (triosephosphate isomerase)) were targeted to introduce intended missense mutations or stop codons using dCas–CDA–UL with sgRNAs targeting the six genes, which successfully mutated all of the targeted sites in seven out of eight clones (Fig. 4b).

Multi-copy elements occupy a substantial amount of genome sequence. Unlike other methods that involve recombination or genomic cleavage, Target-AID may edit multiple loci simultaneously using the same sgRNA sequence without inducing genomic fragmentation. For a proof of concept, four major transposable element (IS1, 2, 3 and 5 that contain 10, 12, 5 and 14 copies of transposase genes, respectively) in the *E. coli* genome were simultaneously targeted using four sgRNAs. The sgRNAs were designed to contain the common sequences of the transposase genes of each transposable element to introduce stop codons (Supplementary Fig. 5). *E. coli* cells were transformed with two plasmids respectively expressing dCas–CDA–UL and the four sgRNAs. After repeating isolation and verification procedures (Supplementary Fig.6), cells edited at all of the 41 targeted loci were obtained. Two clones of the edited cells were further subjected to whole-genome sequencing to assess the genome-wide off-target effect. Among the 34 potential off-target sites that contain the identical sequences to the targets at 1–8 nt proximal to the PAM, two sites were found to be mutated in both clones (Supplementary Table 4).

Target-AID dominantly induced C-to-T substitutions and enabled simple and multiplex genome editing in *E. coli*. The use of nickase Cas9(D10A) with deaminase was not applicable because of its toxicity in *E. coli*, whereas it has been applied in eukaryotes with its higher mutation efficiency without any growth defect[19,20] probably due to the difference in the DNA repair mechanisms. The nickase Cas9(D10A) introduces a nick on the opposite strand to the deamination site. The strand containing deaminated nucleotides is also nicked by AP endonuclease during the base excision repair pathway. Consequently, these together may cause temporal DNA double-strand breaks, which bacteria have a poorer capacity to repair. Consistently, the other type of nickase(H840A)–CDA that nicks the same strand as the deamination site showed little toxicity and resulted in a modest mutation efficiency in *E. coli* as observed in eukaryotes.

Longer sgRNA could form an extended RNA/DNA hybrid that allows the deaminase access to the unwound single-stranded DNA. Although it has been assumed that the 5′ extension of the sgRNAs is processed into 20-nt lengths in the cell[25], our data consistently showed that the mutational positions and frequency were shifted towards the 5′ side of the target as the length of the sgRNA increases from 18 to 24 nt. It is possible that the processing activity could be host-dependent and/or the transient existence of unprocessed sgRNA is sufficient to carry out the mutagenesis.

UGI certainly promotes mutagenesis by deamination in bacteria, although it may compromise genome integrity when in excess. We were unable to develop a plasmid expressing dCas–CDA–UGI without the LVA tag, as we obtained only plasmids containing mutations, suggesting that the cells cannot retain the plasmid expressing functional dCas–CDA–UGI. By decreasing the half-life of the protein, the LVA tag reduced the toxicity to an allowable level for a cell. However, genome-wide nonspecific mutagenic effects and off-target events were still observed by the use of dCas9–CDA–UL. Cells are capable of repairing deaminated nucleotides correctly, while the excess deamination or compromised repair causes mutation. The observed nonspecific mutations (non-homologous against target sequences) may reflect an excess of dCas–CDA–UL in the cells. The two off-target candidate sites (that is, homologous against target sequences) were found to be mutated by dCas–CDA–UL, which contained at least 13 bases matched to the targets, of which 9 consecutive bases are in the PAM proximal part (seed sequence). Theoretically, the expected number of such off-target sites can be calculated by the equations in Supplementary Fig. 7. Here, off-target sequences that contain 8 matched bases of seed sequences and an additional 5 matched bases in the 12 bases of the distal site may occur approximately 8 times in 5.44 Mb of *E. coli* genome. In the IS experiment using 4 sgRNAs, 2 off-target sites out of the 34 off-target candidates were edited. This estimates that randomly chosen targets may induce an off-target event approximately once every two times in *E. coli*. However, the extended incubation to isolate and characterize the mutations at 41 loci may have increased the chance of off-target events. Moreover, simultaneous

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

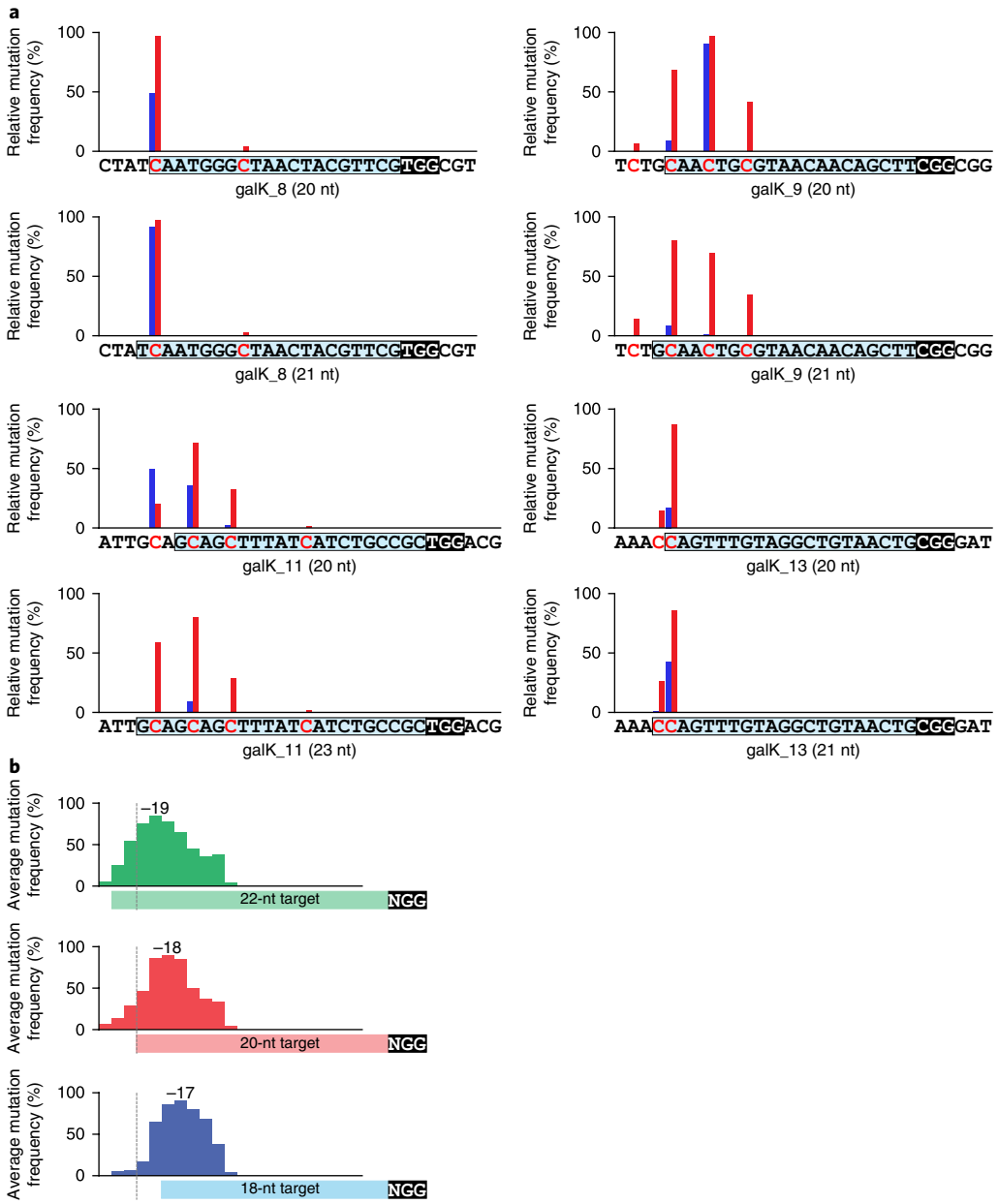


Fig. 3 | Mutated positions and frequency using different lengths of sgRNAs. a, Genomic loci (galK_8, 9, 11 and 13) targeted by dCas-CDA (blue bars) or dCas-CDA-UL (red bars) were analysed by deep sequencing. The average of three independent experiments is plotted. A blue box and an inverted box indicate the target sequences of sgRNA and PAM, respectively. Mutated bases are shown in red. **b**, Average mutational spectra for different lengths of sgRNA. Targeted sites that contain poly-C on the distal site were edited by dCas-CDA-UL and analysed by deep sequencing. Average mutational frequencies at each position for sgRNAs consisting of 18-, 20- or 22-nt target sequence are shown with blue, red and green bars, respectively. The inverted box indicates the PAM and the numbers on the top of each panel indicate the peak positions. The border between 20 and 21 nt upstream of the PAM is shown with a dashed line.

targeting of such a number of genomic loci at once might have loosened the entire genomic structure and allowed less stringent binding of dCas9–gRNA and/or compromised the repair pathways. It is also to be noted that when an off-target event occurs using the Cas9 nuclease in bacteria, it simply kills the cell and can thus be undetermined. Designing target sequences based on the off-target prediction or using a more stringent CRISPR system[26] would help to reduce the number of off-target events. Further studies are also needed to minimize the nonspecific mutator effect of dCas–CDA–UL by fine-tuning the expression of the modifier protein or attenuating the UGI function.

A recent study reported targeted deaminase in bacteria using human cytidine deaminase AID fused to zinc finger protein (ZF–AID) or transcription activator-like effector (TALE–AID)[21], which showed site-specific mutagenesis with up to 13% or 8% efficiency in the *mutS*- and *ung*-disrupted *E. coli*, respectively. Compared to the human AID-fused Cas9-mediated editing method, which showed a diverse mutagenic effect (~500-base-pair (bp) window) in mammalian cells[27,28], Target-AID using PmCDA1 induces specific mutations at a specific window in bacteria and eukaryotes. The combination of the targeting ability of Cas9 and the substrate preference of PmCDA1 to single-stranded DNA secures its specific mutation window.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

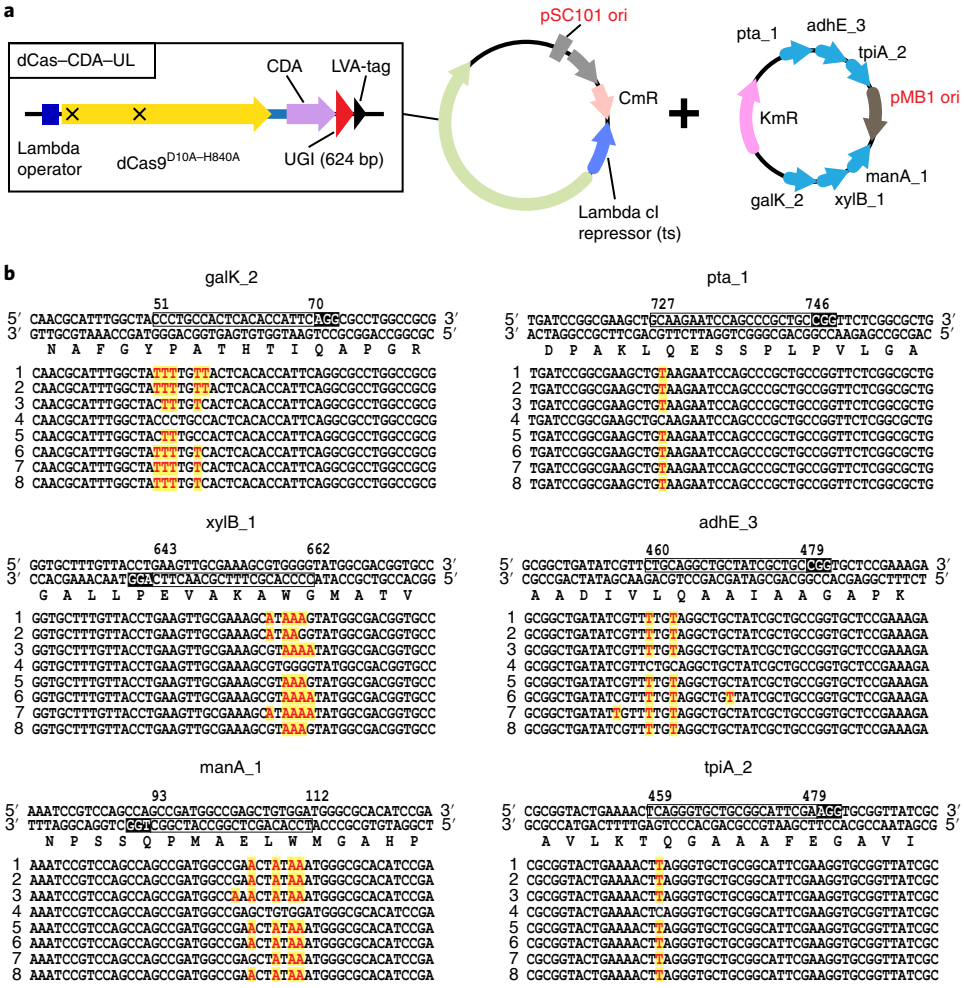


Fig. 4 | Multiplex mutagenesis. a, A schematic drawing of two plasmids for multiplex mutagenesis. A modifier vector expresses dCas–CDA fused with a LVA tag and UGI. The multiple sgRNA-expressing plasmid (pSBP80608) contains two sets of three tandem-repeated sgRNA-expressing units. **b**, Sequence alignments of the targeted loci. The targeted sites from eight randomly picked clones were sequenced and aligned. The clone numbers are shown on the left of the sequences. The translated amino acid sequences are shown underneath each nucleotide sequence. A box and an inverted box indicate the target sequence and PAM, respectively. The mutated bases are shown in red and highlighted in yellow.

As a result of its simple composition and universality, Target-AID may be suited for multiplex and high-throughput mutagenesis and applicable to a wider range of bacterial species including those that were previously not genetically tractable. As in the case of the Base Editor system in mammalian cells[26,29], use of other CRISPR systems and protein engineering will also improve editing capacity with the various selection of PAMs and higher fidelity[30].

Methods

Bacterial strains, plasmids, primers and targeting gRNA design. The following *E. coli* strains were used: DH5α (F⁻ *endA1 supE44 thi-1 recA1 relA1 gyrA96 deoR phoA* Φ80d*lacZ* ΔM15 Δ(*lacZYA-argF*)U169, *hsdR17* (rK⁻, mK⁺), λ⁻) (TaKaRa-Bio); BW25113 (*lacI*⁺ *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR*514 Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$ *rph-1* Δ(*araB–D*)*567* Δ(*rhaD–B*)*568* Δ*lacZ4787*(::*rrnB-3*) *hsdR514 rph-1*); and Top10 (F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX74 nupG recA1 araD139* Δ(*ara-leu*)7697 *galE15 galK16 rpsL*(*Str*^{R}) *endA1* λ–) (Invitrogen). The plasmids and primers used in this study are listed in Supplementary Tables 5 and 6, respectively. A pair of oligonucleotide DNAs for construction of the targeting gRNA vector was designed as follows: 5′-tagc-(target sequence)-3′ and 5′-aaac- (reverse complement of the target sequence)-3′.

Plasmid constructions. pCas9 and pCRISPR plasmids were generously provided by the Marraffini laboratory[15] and obtained via Addgene. Nickase Cas9 (nCas (D10A or H840A)) and nuclease-deficient Cas9 (dCas (D10A H840A))[31] were generated by the PCR method. PmCDA1 was fused to the C terminus of nCas or dCas with a 121-amino-acid peptide linker (Fig. 1b and Supplementary Table 7).

The plasmid pScI_dCas–CDA_J23119–sgRNA carries the sgRNA unit driven by the synthetic constitutive promoter J23119 (BBa_J23119 in the registry for standard biological parts) (http://parts.igem.org/Part:BBa_J23119) (Supplementary Table 7) and is amplified by PCR using p346/p426. The sgRNA-expression unit contains two BsaI restriction sites for insertion of the target sequence. A pair of oligonucleotide DNAs that contains the sgRNA sequence was annealed and ligated into BsaI-digested pScI_dCas–CDA_sgRNA.

pScI contains the lambda operator, and pScI_dCas and pScI_dCas–CDA carry dCas and dCas–CDA genes downstream of the operator. The degradation tag (LVA tag) and the UGI gene are added to the C terminus of the dCas–CDA gene, which results in plasmid pScI_dCas–CDA–L and pScI_dCas–CDA–UL, respectively.

The vector plasmid pTAKN-2 (BioDynamics Laboratory) that possesses the pMB1 origin compatible with pSC101 was used for expression of multiple sgRNAs. A sgRNA–J23119 promoter unit made from synthetic oligonucleotides was digested by EcoRI–HindIII and followed by ligation into pTAKN2. The plasmid harbouring the tandem three target sequences (pSBP804, galK_10–galK_11–gal_13; pSBP806, galK_2–xylB_1–manA_1; and pSBP808, pta_1–adhE_3–tpiA_2) were constructed using Golden Gate assembly of the PCR product using BsaI digestion–ligation[32]. The plasmid harbouring the six different target sequences (pSBP80608) was constructed using Gibson Assembly of the PCR products amplified from pSBP808 (pta_1–adhE_3–tpiA_2 tandem sequence) by primers p597/598 and from pSBP806 (vector and galK_2–xylB_1–manA_1 tandem sequence) by primers p599/p600 (Fig. 4a). For the IS-editing plasmid, four sgRNA-expressing units were tandemly aligned in the order IS1, IS2, IS3 and IS5.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Mutagenesis assay. The DH5α or BW25113 cells were chemically transformed with the appropriate plasmid(s) and were pre-cultured with 1 ml of SOC medium (2% Bacto tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 1 mM $MgSO_4$ and 20 mM glucose). After incubation for 2–3 h at 28 °C, the cell cultures were diluted 1:10 in 1 ml of Luria–Bertani (LB) broth or terrific broth (TB) with the following antibiotics as needed: chloramphenicol (25 μg ml^{-1}) and/or kanamycin (30 μg ml^{-1}), and grown overnight at 28 °C and 1,000 r.p.m. using maximizer (TAITEC). The next day, the cell cultures were diluted 1:10 in 1 ml media again and cultured for 6 h at 37 °C for induction, followed by overnight incubation at 28 °C. The cell cultures were then spotted in serial dilution onto LB or TB agar plates supplemented with appropriate antibiotics and incubated overnight at 28 °C to form single colonies.

For positive-selection of *galK* gene disruption, the cells were grown in M63 minimal medium (2 g l^{-1} $(NH_4)_2SO_4$, 13.6 g l^{-1} KH_2PO_4, 0.5 mg l^{-1} $FeSO_4$–$7H_2O$, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 10 μg ml^{-1} thiamine) containing 0.2% glycerol and 0.2% 2-deoxy-galactose (2-DOG). For the selection of rifampicin-resistant mutation of the *rpoB* gene, cells were grown in LB broth containing 50 μg ml^{-1} rifampicin. For sequencing analysis, the target segments were amplified using appropriate primers by PCR directly from the randomly picked colonies and then analysed by Sanger sequencing using a 3130XL Genetic Analyzer (Applied Biosystems). The *t*-test statistical analysis was conducted using Excel software (Microsoft).

Whole-genome sequencing. BW25113 cells harbouring each expression construct (dCas, dCas–CDA, or dCas–CDA–UL without sgRNA and dCas–CDA with rpoB_1 targeting sgRNA) were pre-incubated overnight at 28 °C, and diluted 1:10 in 1 ml LB media and grown for 6 h at 37 °C for induction, followed by overnight incubation at 28 °C. The cells were spread onto a rifampicin-containing medium plate to isolate single colonies. From three randomly picked colonies cultivated in TB, the genomic DNA was extracted using a Wizard Genomic DNA Purification Kit (Promega) and then fragmented by sonication using a Bioruptor UCD-200 TS Sonication System (Diagenote) to obtain fragments with a size distribution at 500–1,000 bp. The genomic DNA library was prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs) and labelled with Dual Index Primers. The size selection of the library was conducted using the Agencourt AMPure XP (Beckman Coulter) to obtain tagged fragments with a length ranging from 600 to 800 bp. The size distribution was evaluated by the Agilent 2100 Bioanalyzer system (Agilent Technologies). The DNA was quantified using a Qubit HS dsDNA HS Assay Kit and fluorometer (Thermo Fisher Scientific). Sequencing was performed using the MiSeq sequencing system (Illumina) and MiSeq Reagent Kit v3 to obtain 2 × 300 bp read length and approximately 20-fold coverage for the genome size was expected. The data analysis was conducted using the CLC Genomic Workbench 9.0. (CLC bio). The sequence reads were paired and overlapping reads within a read pair were merged and then trimmed on the basis of a quality limit of 0.01 with a maximum of 2 ambiguities. The reads were mapped to the *E. coli* BW25113 reference genome by the following setting (Masking mode = no masking, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3, Length fraction = 0.5, Similarity fraction = 0.8, Global alignment = No, Auto-detect paired distances = Yes and Nonspecific match handling = ignore). Local realignment was conducted with the default settings (Realign unaligned ends = Yes and Multi-pass realignment = 2). The variant calling was performed with the following settings (Ignore positions with coverage = 1,000,000, Ignore broken pairs = Yes, Ignore Nonspecific matches = Reads, Minimum coverage = 5, Minimum count = 2, Minimum frequency = 50%, Base quality filter = No, Read detection filter = No, Relative read direction filter = Yes, Significance = 1%, Read position filter = No, Remove pyro-error variants = No). The output file was rearranged using Excel (Microsoft).

Deep sequencing of the target regions. DH5α cells expressing dCas–CDA or dCas–CDA–UL with the sgRNA targeting *galK*, *gsiA*, *ycbF* or *yfiH* genes were incubated overnight at 28 °C, diluted 1:10 in 1 ml LB media and grown for 6 h at 37 °C for induction. The cell cultures were collected and the genomic DNA was extracted. The target-region-containing fragment (~0.3 kbp) was directly amplified using primer pairs (p685–p696) from the extracted genomic DNA. The amplicon was labelled by the Dual Index Primers. More than 30,000 reads per sample on average were analysed by the MiSeq sequencing system. The sequence reads were paired and trimmed on the basis of a quality limit of 0.01 with a maximum of 2 ambiguities and then overlapping reads within a read pair were merged. The reads were mapped to each reference sequence by the following settings (Masking mode = no masking, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3, Length fraction = 0.5, Similarity fraction = 0.8, Global alignment = No, Auto-detect paired distances = Yes and Nonspecific match handling = Map randomly). An output file was rearranged using Excel.

Isolation and verification procedure for the IS-edited cell. Two plasmids respectively expressing dCas–CDA–UL and four sgRNAs targeting the 41 IS loci were co-introduced into *E. coli* Top10 cells. IS5-1, IS5-2, IS5-11 and IS5-12 sites were firstly analysed as it was found that editing of IS5 targets was less efficient than that of other IS types. Targeted sites were amplified by PCR and sequenced directly from the colonies by using unique primer pairs designed outside the homologous regions. As the cells contained mixed population of genotypes at IS5-1, the cells were re-suspended and spread onto plates to isolate single clones. Eight clones were analysed at IS5-1 to obtain three edited ones, two of which were further analysed by extracting genomic DNA. The isolation and verification procedure was repeated to confirm that all targeted sites were completely edited.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The data that support the findings of this study are available from the corresponding authors upon request. Short-read data in this study have been deposited at the NCBI Sequence Read Archive (SRA) with the accession code PRJNA407291.

Received: 27 January 2017; Accepted: 20 December 2017;
Published online: 05 February 2018

References

1. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
2. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–827 (2013).
3. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
4. Zhang, Y., Buchholz, F., Muyrers, J. P. P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123–128 (1998).
5. Shimatani, Z. et al. Targeted base editing in rice and tomato using a CRISPR–Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
6. Zhigang, W., Smith, D. G. & Mosbaugh, D. W. Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene* **99**, 31–37 (1991).
7. Andersen, J. B. et al. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
8. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
9. Costantino, N. & Court, D. L. Enhanced levels of lambda red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* **100**, 15748–15753 (2003).
10. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
11. Wang, J. et al. An improved recombineering approach by adding RecA to lambda red recombination. *Mol. Biotechnol.* **32**, 43–53 (2006).
12. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
13. Bowater, R. & Doherty, A. J. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* **2**, 93–99 (2006).
14. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Res.* **44**, 4243–4251 (2016).
15. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
16. Jiang, Y. et al. Multigene editing in the *Escherichia coli* genome via the CRISPR–Cas9 system. *Appl. Environ. Microbiol.* **81**, 2506–2514 (2015).
17. Li, Y. et al. Metabolic engineering of *Escherichia coli* using CRISPR–Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).
18. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl. Environ. Microbiol.* **81**, 5103–5114 (2015).
19. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
20. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
21. Yang, L. et al. Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.* **7**, 13330 (2016).
22. Wang, Y. et al. *Bacillus subtilis* genome editing using ssDNA with short homology regions. *Nucleic Acids Res.* **40**, e91 (2012).
23. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).
24. Jin, D. J., Walter, W. A. & Gross, C. A. Characterization of the termination phenotypes of rifampicin-resistant mutants. *J. Mol. Biol.* **202**, 245–253 (1988).
25. Ran, F. A. et al. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

26. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
27. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
28. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
29. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
30. Mitsunobu, H., Teramoto, J., Nishida, K. & Kondo, A. Beyond native Cas9: manipulating genomic information and function. *Trends Biotechnol.* **35**, 983–996 (2017).
31. Jinek, M. et al. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–822 (2012).
32. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4**, e5553 (2009).

Acknowledgements

We would like to thank A. Miyabe and M. Kakimoto for their technical assistance. This work was supported by the Platform Project for Supporting in Drug Discovery and Life Science Research (Platform for Drug Discovery, Informatics, and Structural Life Science) from the Japan Agency for Medical Research and Development (AMED). This work was also partly supported by a Special Coordination Fund for Promoting Science and Technology, Creation of Innovative Centers for Advanced Interdisciplinary Research Areas (Innovative Bioproduction Kobe) from the Ministry of Education, Culture, Sports and Technology (MEXT) of Japan; the Cross-ministerial Strategic Innovation Promotion Program; JSPS KAKENHI (grant number 26119710, 16K14654); and the New Energy and Industrial Technology Development Organization (NEDO).

Author contributions

S.B. and K.N. performed all of the experiments with help from all authors. S.B., K.N. and H.M. wrote the manuscript with input from all authors. T.A. provided technical advice. K.N. and A.K. conceived the project.

Competing interests

A provisional patent has been submitted in part entailing the reported approach.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-017-0102-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.N. or A.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Keiji Nishida

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Biological triplicates or more were performed so that significance in the difference was statistically assessed by using 95% confidence interval for a geometric mean by t-test analysis (Fig S1, S2, S4a). For WGS analysis, 17.3-fold coverage or more was obtained.

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Except for WGS analysis, all experiments are biologically repeated at least three times and statistically analyzed to present 95% confidence interval for a geometric mean by t-test analysis. For WGS analysis, biologically independent clones were individually analyzed and shown as they are. For IS mutant isolation, experiment was repeated twice.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

For sample preparation, average-sized colonies were randomly picked.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

All samples were visually indistinguishable.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	CLC Genomic Workbench 9.0. (CLC bio, Aarhus, Denmark) was used to analyze NGS data. The software is commercially available.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	The study did not involve human research participants.