Published online 20 December 2021

Nucleic Acids Research, 2022, Vol. 50, No. 7 ***e38***
https://doi.org/10.1093/nar/gkab1244

Development of a genome-targeting mutator for the adaptive evolution of microbial cells

Ga-eul Eom, Hyunbin Lee and Seokhee Kim *

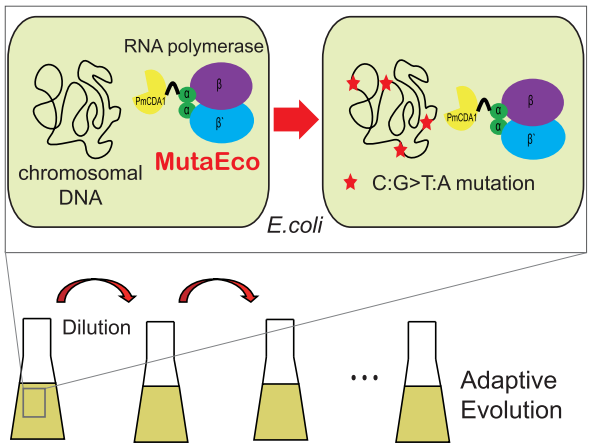
Department of Chemistry, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea

Received August 13, 2021; Revised November 05, 2021; Editorial Decision November 30, 2021; Accepted December 03, 2021

ABSTRACT

Methods that can randomly introduce mutations in the microbial genome have been used for classical genetic screening and, more recently, the evolutionary engineering of microbial cells. However, most methods rely on either cell-damaging agents or disruptive mutations of genes that are involved in accurate DNA replication, of which the latter requires prior knowledge of gene functions, and thus, is not easily transferable to other species. In this study, we developed a new mutator for *in vivo* mutagenesis that can directly modify the genomic DNA. Mutator protein, MutaEco, in which a DNA-modifying enzyme is fused to the α-subunit of *Escherichia coli* RNA polymerase, increases the mutation rate without compromising the cell viability and accelerates the adaptive evolution of *E. coli* for stress tolerance and utilization of unconventional carbon sources. This fusion strategy is expected to accommodate diverse DNA-modifying enzymes and may be easily adapted to various bacterial species.

GRAPHICAL ABSTRACT



INTRODUCTION

Methods for *in vivo* mutagenesis have enabled significant advances in the fields of life sciences, biotechnology, and medicine. One group of tools, including the clustered regularly interspaced palindromic repeat/caspase 9 (CRISPR-Cas9) system (1,2) and its congeners, such as base editors (3–5) and prime editors (6), edit a narrow region of genomic DNA using use their precise DNA-targeting ability. The main goal of these genome editing tools is to introduce a desired DNA sequence into the target DNA region, and they exhibit great potential for biotechnological and therapeutic applications that require the precise disruption or correction of DNA sequences (7). The other group of *in vivo* mutagenesis tools aims to generate mutations at random sites in a gene or genome to promote the directed evolution of biomolecules or the adaptive evolution of cells with desired properties (8). These *in vivo* mutagenesis tools support simultaneous DNA diversification and selection in cells, resulting in the continuous directed evolution of proteins or cells during growth (8).

Adaptive laboratory evolution (ALE) is a powerful method for studying the molecular mechanisms of adaptive evolution of cells and engineering microbial cells for biotechnological applications (9,10). Although traditional ALE relies on spontaneous mutations in the genome during serial passages of microbial cells, several reports demonstrated that hypermutation induced by mutator strains could accelerate the adaptive evolution of cells (11–16). These results are in agreement with the observation that the mutator phenotype is enriched in the bacterial population that undergoes adaptation to a new environment (17,18). High mutation rates, however, may reduce cell viability through the rapid accumulation of deleterious mutations (19), thus hampering the proper selection of adapted cells (20,21). Therefore, the efficient ALE of microbial cells may require an *in vivo* mutagenesis method that confers moderate mutation rates.

Various methods have been developed for genome mutagenesis. Mutagenic compounds and ultraviolet irradiation have been used in classical genetic screens, but their application in ALE is limited because of the additional damage to cellular components (22,23). Deletion of genes

*To whom correspondence should be addressed. Tel: +82 2 880 4123; Fax: +82 2 875 6636; Email: seokheekim@snu.ac.kr

involved in DNA mismatch repair or proofreading processes induces mutations during DNA replication, but suffers from uncontrolled mutagenesis even when mutagenesis is not required (24–26). Although the ectopic expression of dominant-negative variants of these genes grants the temporal control of mutagenesis (11–14,16,27), these variants function as mutators only in the original species and the same strategy can be applied to other microbial species only when the orthologs and their dominant-negative variants have been identified (12,13,16).

Here, we report a new *in vivo* mutagenesis method that does not rely on dominant-negative variants of genes involved in high-fidelity DNA replication. We envisioned that direct DNA modification by enzymes, such as base deaminases, might provide a versatile method that can be applied to various microbial species. Although the expression of cytidine deaminase was previously shown to only marginally increase the mutation rates (27,28), we found that the chimeric mutator protein, in which a cytidine deaminase is fused to the α-subunit of *Escherichia coli* RNA polymerase (RNAP), accelerated the mutation to the level that supports efficient ALE of *E. coli* without reducing the cell viability. We showed that the strain expressing this mutator could develop the tolerance to chemical stress and the utilization of the unconventional carbon sources more effectively than the wild-type cells.

MATERIALS AND METHODS

Materials

All PCR experiments were conducted with KOD Plus neo DNA polymerase (Toyobo, Japan). T4 polynucleotide kinase and T4 DNA ligases were purchased from Enzynomics (South Korea). Plasmids and DNA fragments were purified with LaboPass™ plasmid DNA purification kit mini, LaboPass™ PCR purification kit, and LaboPass™ Gel extraction kit (Cosmogenetech, South Korea). Sequences of the cloned genes were confirmed by Sanger sequencing (Macrogen, South Korea). Antibiotics (ampicillin, chloramphenicol, kanamycin, and tetracycline), arabinose, and IPTG were purchased from LPS solution (South Korea). Carbon sources were obtained from Sigma-Aldrich (D-tartrate, #147-71-7; erythritol, #149-32-6; sucrose, #57-50-1; ethylene glycol, #29810; L-lyxose, #1949-78-6; 2-deoxy-D-glucose, #205-823-0; D-(+)-cellobiose, #528-50-7) and Acros (monomethyl succinate, #3878-55-5; 2-deoxy-D-ribose, #533-67-5; L-sorbose, #87-79-6).

Gene cloning and *E. coli* strain construction

Escherichia coli strains, plasmids, and primers used in this study are listed in Supplementary Tables S1–S3, respectively. Genes for *Petromyzon marinus* cytidine deaminase (PmCDA1) and the XTEN linker were obtained from the plasmid expressing eMutaT7 (pHyo094) (29). Genes for *E.coli* RNAP subunits were obtained from *E. coli* genomic DNA. Genes for deaminase-linker (PmCDA1-XTEN) and *E. coli* RNAP subunits were linked by overlap extention PCR (30). Linker variant plasmids were constructed using the site-directed mutagenesis PCR method (31). For nonnative carbon source evolution experiments, the arabinose induction system was replaced with the tetracycline induction system amplified from the plasmid pSY013 and the IPTG induction system amplified from pET28b, respectively. For sucrose utilization experiments, the plasmid containing the *csc* genes (*csc*A, *csc*B and *csc*K) were purchased from Addgene (#63918).

Escherichia coli W3110_GE △*ung* (the reference strain, cGE048) was used in all evolution experiments. cGE048 was constructed by homologous recombination method (32). The wild-type *ung* gene in W3110_GE was replaced by the kanamycin resistance gene.

Cell culture for *in vivo* mutagenesis

Three biological replicates of strains harboring a mutator plasmid were grown overnight in LB medium with chloramphenicol (35 μg/ml). On the next day, the overnight culture (3.5 μl) was passed into a new fresh LB medium (350 μl) supplemented with arabinose (0.2%) and chloramphenicol (35 μg/ml) in 96-deep well plate (Bioneer, South Korea) at 37°C with shaking. Bacterial cells were diluted every 4 h and the growth cycle was repeated up to 120 rounds. At the end of each cycle, a fraction of cells were taken and stored at −80°C with 20% glycerol.

Rifampicin resistance test

10-fold serial dilutions of cells using LB medium were placed on LB-agar plates either without or with rifampicin (50 μg/ml). After 16 h at 37°C, the number of colonies on the plates was counted to calculate the fraction of cells that develop rifampicin resistance. To compare various mutators, overnight cultures of the reference strain carrying no plasmid, pGE143 (MutaEco), MP1 or MP6 were diluted 100-fold in LB supplemented with chloramphenicol (35 μg/ml) and arabinose (0.2%) except for the reference strain carrying no plasmid, and incubated at 37°C for 16 h. Ten-fold serial dilutions of cells were placed on LB-agar plates either with or without rifampicin (100 μg/ml) and incubated for 16 h. The number of colonies were counted to determine the rifampicin resistance frequency.

Viability test

Overnight cultures of the reference strain carrying no plasmid, pGE143 (MutaEco), MP1 or MP6 were diluted 100-fold in LB supplemented with chloramphenicol (35 μg/ml) except for the reference strain carrying no plasmid, and allowed to grow to log phase (OD_{600} = 0.2–0.5) at 37°C. Cells were diluted 10^6-fold and 200 μl of the diluted cells were placed on LB-agar plates supplemented with or without arabinose (0.2%). After overnight growth at 37°C, the number of colonies on these plates were counted to determine the relative cell viability.

Competition assay

Competition assay was performed by daily dilution after growth with the mixture of the reference strains expressing MutaEco or MP1. Five biological replicates were grown overnight in LB medium with antibiotics. On the next day,

the MutaEco-expressing cells and the MP1-expressing cells were mixed, diluted (1:100) in a fresh medium with antibiotics and arabinose (0.2%), and subjected to the ethanol stress. Cell growth was monitored by OD_{600} using M200 microplate reader (TECAN, Switzerland) and cultures were diluted (1:100) in a fresh medium. The level of each stress was gradually increased. After the 216-h adaptive evolution (5% EtOH), the cells were streaked on agar plates to select single clones for colony PCR. Two pairs of primers specific to the MutaEco- or MP1-expressing plasmids were used to amplify two different sizes of DNA products. PCR products were examined by electrophoresis with 1% (w/v) agarose gel. Ten colonies were checked for each sample ($n = 5$).

Antibiotic resistance test

After 120 rounds of the growth-dilution cycle, the evolved cells were 10-fold serially diluted using LB medium and placed on LB-agar plates containing no antibiotic, rifampicin (50 or 100 μg/ml), carbenicilin (30 μg/ml), cefotaxime (5 μg/ml), streptomycin (30 μg/ml) or tetracycline (10 μg/ml). After 16 h at 37°C, the number of colonies on the plates was counted to determine the resistance frequency.

Whole genome sequencing

The genomic DNA of the evolved cells and the W3110_GE strain were extracted with genomic DNA prep kit (BIONEER, South Korea). Samples were then prepared and sequenced by Macrogen using the manufacturer's reagents and protocol to determine the mutation pattern. DNA was quantified using Quant-IT PicoGreen (Invitrogen, USA). The 2 × 151 paired-end sequencing libraries were prepared using TruSeq DNA Nano sample preparation kit (Illumina, Inc., San Diego, CA, USA), and then sequenced using Novaseq™ (Illumina; operated by Macrogen). After sequencing, FastQC (v0.11.5) was performed in order to check the read quality. Trimmomatic (v0.36) (33) was used to remove adapter sequences and low quality reads in order to reduce biases in analysis. Filtered data were mapped using BWA (v0.7.17) (34) with algorithm to the reference genome in NCBI Reference Sequence database (Macrogen protocol).

The following criteria were set to analyze the mutation pattern of the evolved cells (cycle #120). The analysis was performed in positions with depth not smaller than 3000. The *ung* gene and genes that exist in the mutator plasmid (*araC*, *rpoA*, *rrl*) were excluded from the analysis. Substitutions with a frequency under 0.2% were discarded. Only those presenting the mutation frequency of 0.2% or higher were subjected to the mutation pattern analysis.

RNA sequencing

We extracted RNA from the W3110_GE (wild-type strain) using NucleoSpin RNA kit (MACHEREY-NAGEL, Germany). RNA quality and quantity were assessed using the UV absorbance ratio. RNA sequencing was performed by Macrogen using the manufacturer's reagents and protocol. Sample libraries were independently prepared by Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA, #RS-122-2101). Indexed libraries were then submitted to Illumina NovaSeq (Illumina, Inc., San Diego, CA, USA; operated by Macrogen), and the paired-end (2 × 100 bp) sequencing was performed. The reference genome sequence of *E. coli* str. K-12 substr. W3110 (GCF_000010245.2) and annotation data were downloaded from the NCBI. After alignment, HTSeq v0.10.0 (35) was used to assemble aligned reads into transcripts and to estimate their abundance. It provides the relative abundance estimates as RPKM values (reads per kilobase of transcript per million mapped reads) of transcript and gene expressed in each sample.

Statistical analysis of high-throughput sequencing data

For high-throughput sequencing data (Figure 2), Mann–Whitney test (unpaired Wilcoxon test) was used to assess the significance of the mutation frequency caused by the MutaEco system. Calculation was conducted using Graphad prism5. Statistically significance were determined with P values defined as $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.001$ for this experiment. For other data (Figures 1, 3–6, Supplementary Figures S1, and S3–S6), we assumed the data will follow normal distribution and performed Student's t-test conducted using Graphad prism5.

Adaptive evolution of stress tolerance

Adaptive evolution experiments were performed by daily dilution after growth (the reference strain with or without MutaEco expression). Three biological replicates were grown overnight in LB medium with antibiotics. On the next day, cultures were diluted (1:100) in a fresh medium with antibiotics and arabinose (0.2%), and subjected to the following stresses; heat, isobutanol, or ethanol. Cell growth was monitored by OD_{600} using M200 microplate reader (TECAN, Switzerland) and cultures were diluted (1:100) in a fresh medium. The level of each stress was gradually increased by growing cultures at higher temperature or applying higher concentration of isobutanol or ethanol. The cells grown at $OD_{600} > 0.5$ with the highest amount of stress were transferred to new media with the same or higher levels of stress for the next round. At the end of evolution, a fraction of cells were taken and stored at −80°C. End-point population samples were streaked on agar plates to select single clones for phenotype assay. To test the stress tolerance, overnight cultures of the evolved cells and the reference strain were diluted with fresh media and subjected to stress. Growth was monitored over the 12–24 h.

Evolution of the non-native carbon source utilization

During evolution experiments, strains were grown in M9 minimal media supplemented with glycerol (0–0.2%), a non-native carbon source (0.2%) and anhydrous tetracycline (50 ng/ml). The initial culture contained 0.2% glycerol and was grown for 24 h. At the next round, the culture was diluted (100-fold) to two independent cultures with fresh media containing either the same or half amount of glycerol, respectively. Only the culture that showed significant

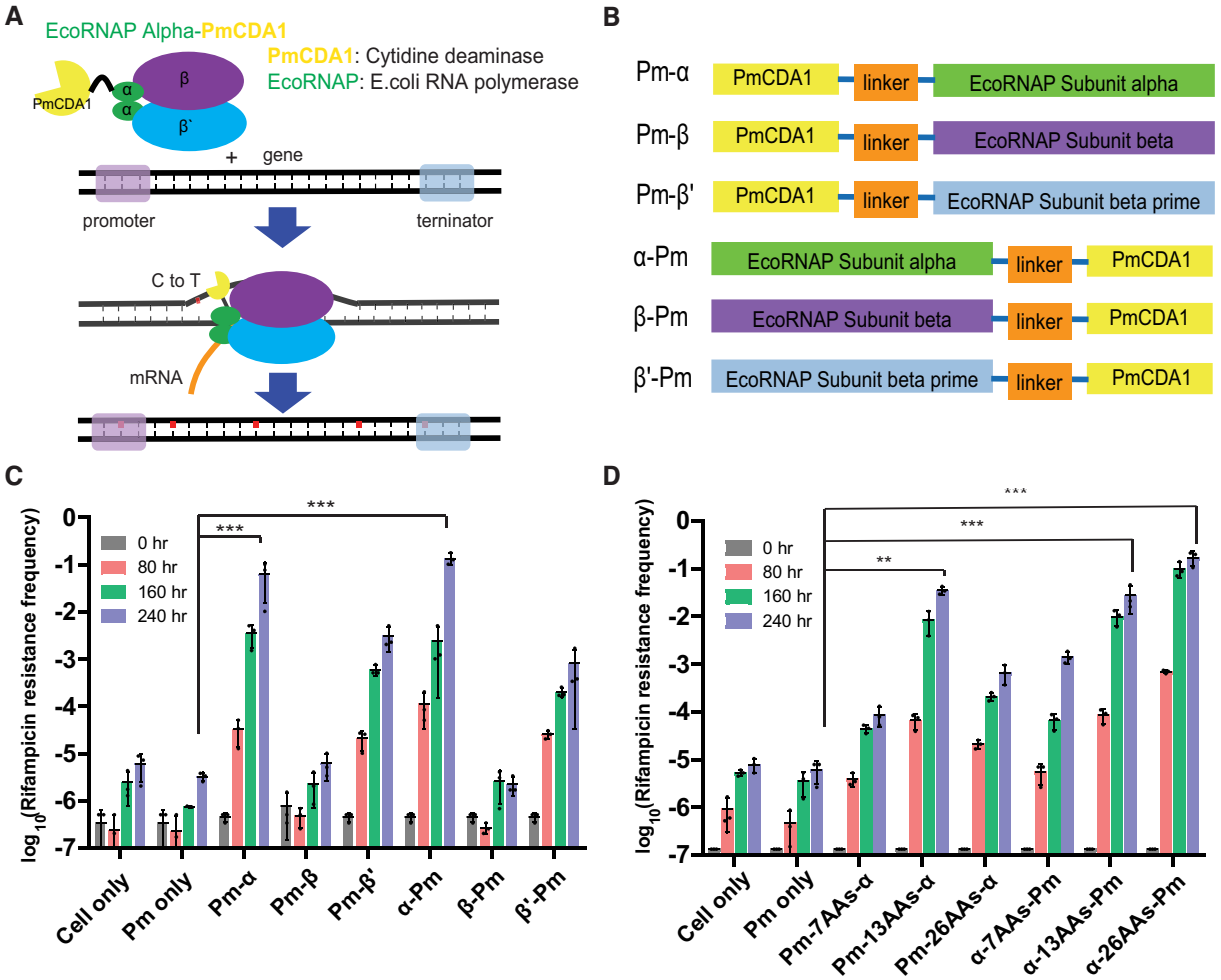


Figure 1. Design and characterization of the genome-targeting *in vivo* mutagenesis system. (**A**) Schematic illustration of the new system. The mutator, a chimeric protein of a cytidine deaminase and a subunit of *Escherichia coli* RNA polymerase (RNAP), induces mutations during the process of transcription. (**B**) Design of chimeric proteins for testing *in vivo* mutagenesis. Pm, *Petromyzon marinus* cytidine deaminase (PmCDA1); linker, an XTEN linker variant; α, *E. coli* RNA polymerase subunit alpha; β, *E. coli* RNA polymerase subunit beta; β′, *E. coli* RNA polymerase subunit beta prime. (C, D) Rifampicin resistance assay for testing the effects of different RNAP subunits (**C**) or linker lengths (**D**). Higher frequency of rifampicin resistance indicates higher mutation rate. Cultures expressing the mutator were taken at several time points and placed on the agar plate in the presence or absence of rifampicin (50 μg/ml). Resistance frequency was calculated as the ratio of the colony numbers grown on plates with and without rifampicin. Data are presented as dot plots with mean ± standard deviation (SD) ($n = 3$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by Student's *t*-test.

growth ($OD_{600} < 0.4$) within 2 days with the lowest amount of glycerol was used for next round of adaptive evolution. Samples were periodically taken and stored at −80°C for later validation. The evolution experiments were concluded once cells grew only in the presence of the non-native carbon source or the evolution time reached 1000 h. End-point population samples were streaked on agar plates to select single clones for validation.

Growth test of the evolved cells

Overnight cultures (2 ml) of the evolved cells (either as a library or as a single clone) and the reference strain were pelleted at 3800 × g and gently resuspended in M9 minimal medium without a carbon source. The sedimentation and resuspension were repeated twice to wash the cells of residual media. Cells were finally resuspended with 2 ml M9 minimal medium without a carbon source (0.2%), and diluted 100-fold to 350 μl of M9 minimal medium with a non-native carbon source in the 96-deep well plate. Cultures were incubated at 37°C, and growth was monitored by OD_{600} using microplate reader. For growth test on the solid media, the washed cells were diluted to $OD_{600} = 0.2$ and 10-fold serial dilutions were spotted on M9-agar plates supplemented with glycerol (0.2%) or a non-native carbon source (0.2%). After overnight growth at 37°C, the number of colonies on the plates were counted.

Whole genome sequencing and analysis of the sucrose-utilizing cells

Two sucrose-utilizing strains (cGE051 and cGE052) and one cheater strain (cGE050) obtained from the evolved cells were subjected to whole genome sequencing. Genomic DNAs were extracted with genomic DNA prep kit (BIONEER). Genomic DNA quality was assessed using

Nanodrop UV absorbance ratios. Samples were then prepared with TruSeq Nano DNA kit and sequenced on Illumina Hiseq (Illumina, USA; operated by Macrogen) in 3 × 151 paired-end runs using the manufacturer's reagents and protocol to determine the mutation. The sequencing and analysis processes are identical to the whole genome sequencing described above.

RESULTS AND DISCUSSION

Cytidine deaminase fused to the α-subunit of *E. coli* RNA polymerase promotes mutagenesis *in vivo*

Pioneered by the Shoulders group (36) and further explored by several other groups (29,37–39), systems using a fusion protein of cytidine deaminase and T7 RNAP have been shown to be simple and efficient methods for gene-specific *in vivo* mutagenesis. Inspired by these studies, we designed a new mutator for genome targeting, the fusion of *Petromyzon marinus* cytidine deaminase (PmCDA1), which introduces the G:C→A:T mutation, and a subunit of *E. coli* RNAP. Previous reports demonstrated that PmCDA1 is the most efficient deaminase in *E. coli* (40) and introduces mutations to the target gene *in vivo* seven times faster than rAPOBEC1 when it is fused to T7 RNAP (29,38). We reasoned that *E. coli* RNAP locates PmCDA1 in proximity to single-stranded DNA, which is transiently generated during transcription and is the main target of PmCDA1 (Figure 1A) (28,41–43).

To determine which subunit of *E. coli* RNAP creates an efficient mutator upon fusion, we tested six chimeric proteins in which PmCDA1 was fused to either the N- or C-terminus of α, β, or β'-subunit of *E. coli* RNAP via a 13-residue flexible linker (Figure 1B). We inserted the mutator plasmids into the *Δung* strain, in which the *ung* gene encoding an uracil–DNA glycosylase (UDG) is deleted and thus the deaminase-mediated G:C→A:T mutagenesis occurs more efficiently (44), and induced the expression of fusion proteins with arabinose. During 60 rounds of growth (4 h per round) and dilution (100-fold with fresh medium) in batch cultures, we did not observe any detectable growth defects of cells expressing various chimeric proteins, and thus, we tested cells taken at different time points for rifampicin resistance. Rifampicin, an inhibitor of β-subunit of *E. coli* RNAP, is often used to monitor random mutagenesis in the genome, because its resistance is easily developed by various point mutations in the target protein (45). Although the early mutation event is likely to be amplified and thus to skew the resistance frequency in this experiment, the results showed that the chimeras fused to α-subunit (Pm-α and α-Pm) generally developed rifampicin resistance faster than those fused to any other subunits (Figure 1C). We also found that cells expressing Pm-α or α-Pm showed much higher resistance frequency ($>10^4$ fold for cells grown for 240 h) than those expressing PmCDA1 alone, indicating that the covalent linkage of PmCDA1 to RNAP significantly enhances its mutagenesis efficiency.

To optimize the linker length, we tested fusion proteins with a 7-, 13- or 26-residue linker at the N- or C-terminus of α-subunit of RNAP. We found that the variant containing C-terminal PmCDA1 and the longer linker (α-26AAs-Pm) showed a slightly higher (∼10-fold) resistance frequency than those with the 13-residue linker (Figure 1D). Therefore, we used α-26AAs-Pm, which we named MutaEco, for the rest of the experiments in this study. We also tested cells that underwent 60 additional growth cycles (total 480 h of growth) for resistance against various antibiotics, and found that cells expressing MutaEco generally showed higher resistance frequency than those expressing PmCDA1, suggesting a higher mutation rate of MutaEco compared to PmCDA1 at multiple loci in the genome (Supplementary Figure S1).

Whole genome sequencing indicates deaminase-dependent mutagenesis

To examine the dependence of the higher mutation rate on the cytidine deaminase, we sequenced genomic DNA from a mixed pool of cells that underwent 480-h *in vivo* mutagenesis with MutaEco using the Illumina method, in which the average depth was ∼5525. To properly discern the mutations from sequencing errors, we also sequenced a single clone of the wild-type strain (W3110_GE) with which we constructed the *Δung* strain (the reference strain) and performed the *in vivo* mutagenesis experiments. W3110_GE is a W3110-based strain, but whole genome sequencing revealed additional single nucleotide substitutions and indels, which were not included in our mutational analysis (Supplementary Table S4). We found 1,002 and 3,608 sites in which the frequency of variant calling (putative mutation) was at least 0.2% (more than ∼10 variant calling) from the wild-type strain and the mutagenized cells, respectively (Supplementary Figure S2A). The analysis of these putative mutations suggests that the function of PmCDA1 is indeed critical for the MutaEco-mediated *in vivo* mutagenesis. First, the number of G:C→A:T mutation sites in the mutagenized cells (2581 sites) was ∼17-fold higher than that in the wild-type strain (154 sites), whereas the number of sites for other types of mutations showed <1.3-fold difference between the two samples (Supplementary Figure S2A). Furthermore, the G:C→A:T mutation sites had a significantly higher average mutation frequency in mutagenized cells (2.13% for 2581 sites) than those in the wild-type strain (0.87% for 154 sites; Figure 2A). Among 1,027 sites with over 0.2% frequency of the other types of mutations in the mutagenized cells, 64% overlapped with those found in the wild-type strain, suggesting that they largely present position-specific sequencing errors (Supplementary Figure S2A). Second, the mutation hotspot analysis revealed that the mutator had a low level of preference for the target sequence, but the preferred sequence pattern, pyrimidine-C-purine (YCR), is very similar to eMutaT7, in which the same cytidine deaminase PmCDA1 was used as the DNA-modifying enzyme (Figure 2B) (29). Collectively, we believe that MutaEco mainly introduces the G:C→A:T transition mutation in the genome using the PmCDA1 deaminase.

These G:C→A:T mutations with 0.2% or higher mutation frequencies were found throughout the *E. coli* genome (Supplementary Figure S2B). Considering the direction of genes, the numbers of the C→T and G→A mutations and their average mutation frequencies were very similar (Supplementary Figure S2C), suggesting that MutaEco behaves differently from MutaT7, in which the C→T

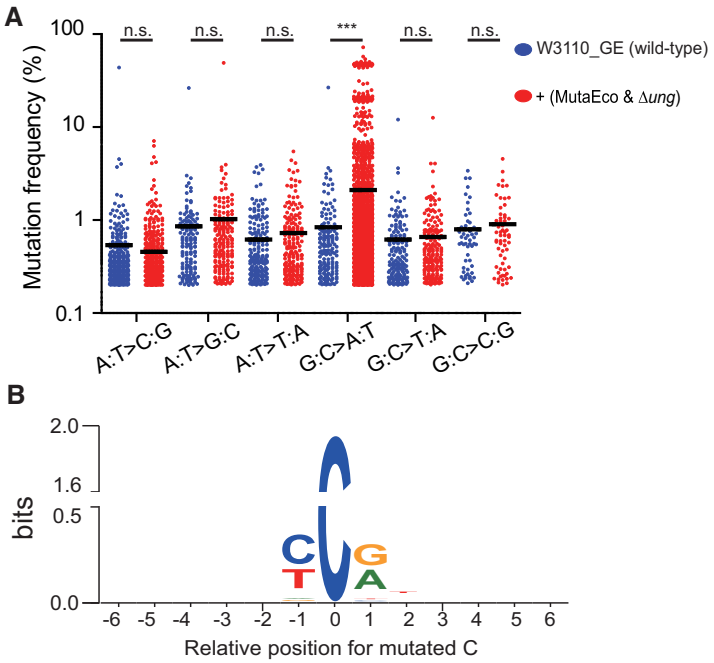


Figure 2. Whole genome sequencing data indicate the deaminase-dependent mutagenesis. (**A**) Frequency of all possible types of substitution shown in the wild-type strain and the cells that underwent 480-h *in vivo* mutagenesis in the presence of MutaEco. Only those with 0.2% or higher mutation frequency are shown. *** $P < 0.001$ by Mann–Whitney's *t*-test. (**B**) Preferred sequence for the mutator. Frequencies of the C-to-T mutation at specific C positions were used to generate the sequence logos (WebLogo, http://weblogo.berkeley.edu).

mutations dominate (29). We also performed RNA sequencing with the total mRNA isolated from the wild-type strain, but found no significant correlation between the expression level and the mutation level of individual genes (Supplementary Figure S2D). This may be because other factors also affect the stable accumulation of mutations or the mutation rate of this mutator is simply too low to generate significant correlation; only about 2400 G:C→A:T mutation sites (approximately one mutation per 2,000 bases in the genome, or 0.5 mutation per gene) had mutation frequencies above 0.2% in this sample.

MutaEco mediates faster adaptive evolution under stress

Next, we tested MutaEco for the ability to develop cellular tolerance to stress. First, we evolved bacterial tolerance to ethanol (EtOH), which is one of the most common biofuels. The genetic basis of EtOH tolerance in *E. coli* has been thoroughly studied (46–52). We tested the growth of two or three parallel cultures in a rich medium with different amounts of EtOH, of which the cells grown at OD_{600} >0.5 with the highest amount of EtOH were diluted to new media with the same or higher amounts of EtOH for the next round of growth (Figure 3A for representative schemes). Three independent cultures of the reference strain with or without the mutator expression displayed divergent paths at a relatively early stage (60 h; 5% EtOH), and the tolerance gap (∼1% EtOH) established at ∼100 h was largely maintained throughout the rest of the 1008-h long (186 generations) evolution (Figure 3A and B). The final tolerance level of the mutator-expressing cells was higher (up to 8%) than that of cells without the mutator (up to 6.85%), which was further confirmed by better growth of the mutator-expressing cells in media containing 5–8% EtOH (Figure 3C). This rate of EtOH tolerance development is comparable to or higher than those reported in the previous study (15). The evolution of EtOH tolerance in a minimal medium showed a similar superiority of the mutator; the evolved cells were tolerant to up to 7% EtOH with the mutator and 6% without the mutator (Supplementary Figure S3).

Second, we performed similar experiments using isobutanol (*i*-BuOH). The major divergence of adaptive paths was observed between 400 and 600 h, and the final tolerance levels after 936-h evolution (∼180 generations) were up to 3% *i*-BuOH with the mutator and 2.4% without the mutator (Figure 3D). The evolved cells with the mutator grew better than those without the mutator in media containing 1–3% *i*-BuOH (Figure 3E). Finally, we evolved heat tolerance in a rich medium; the tolerance level with the mutator reached a temperature of 47°C in 60 h and was almost saturated at 47.7°C in 204 h (Supplementary Figure S4). None of the three independent cultures without the expression of the mutator survived the same path, and the evolved cells grew better than the reference strain at temperatures of 45°C or above (Supplementary Figure S4). Collectively, MutaEco promoted faster adaptive evolution for tolerance against chemical and heat stress.

To compare MutaEco with other mutator systems for the ability to promote adaptive evolution, we selected MP1 and MP6, inducible mutagenesis systems based on the expression of multiple mutator genes, including those disrupting the DNA proofreading or repair system (*dnaQ926, dam, seqA* and *ugi*), mediating translesion mutagenesis upon ultraviolet light or chemical mutagens (*umuD', umuC* and

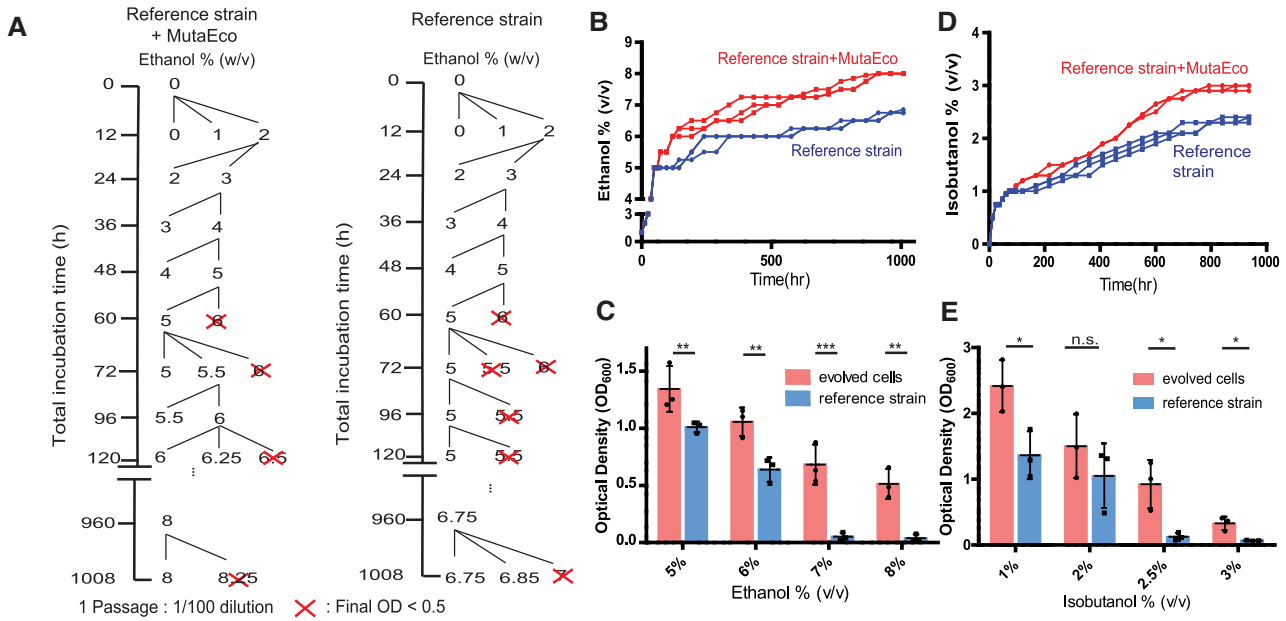


Figure 3. MutaEco promotes faster adaptive evolution of *E. coli* for tolerance against chemical stresses. (**A**) Schemes of the evolutionary pathway of the reference strain for ethanol tolerance with (left) or without (right) the mutator expression. Each number indicates the ethanol concentration (w/v) in media. (B, D) Trajectories of adaptive evolution for ethanol (**B**) or isobutanol (**D**) tolerance ($n = 3$). Individual data points indicate the maximal chemical concentrations at which the cells show substantial tolerance (final $OD_{600} > 0.5$). (**C**, **E**) Final optical density of the evolved cells (red) or the reference strain (blue) that were grown at 37°C for 12 h in the Luria-Bertani (LB) medium containing indicated concentrations of ethanol (C) or isobutanol (E). Data are presented as dot plots with mean ± SD ($n = 3$). * $P < 0.05$ and ** $P < 0.01$ by Student's *t*-test.

recA730), or promoting the C→T transition mutations (*cda1*) (27). MutaEco had a mutation rate 47- and 995-fold slower than MP1 and MP6, respectively, but 70-fold faster than the reference strain (Figure 4A). Cell viabilities of MP6, MP1 and MutaEco showed significant reduction, slight reduction, and no apparent reduction, respectively, indicating that the mutation rates were inversely correlated with cell viability (Figure 4B). We performed the adaptive laboratory evolution for EtOH tolerance and found that the tolerance of the MutaEco-expressing cells was generally better at any time point than those expressing MP1 or MP6, of which MP6-expressing cells stopped growing after 63 h and MP1-expressing cells showed slower adaptation at an early stage (<100 h) but resulted in a tolerance level similar to MutaEco-expressing cells after 1000-h evolution (Figure 4C). Also, we conducted a competition experiment with the MP1- and MutaEco-expressing strains. The 216-h adaptive evolution for EtOH tolerance showed that the MutaEco-expressing strain was significantly enriched in five independent cultures (Supplementary Figure S5). This result suggests that MutaEco promotes more efficient adaptive evolution than MP1 and MP6, presumably because it has an appropriate mutation rate that moderates toxicity but still supports sufficient diversification.

Evolution of nonnative carbon source utilization

We further tested MutaEco for the ability to evolve new strains that can grow solely on a non-native carbon source. To avoid interference from arabinose, which is used to induce the expression of MutaEco but can also be a carbon source, we constructed a new mutator plasmid in which MutaEco expression is induced by anhydrotetracycline. This induction system developed rifampicin resistance at a rate similar to that of the arabinose induction system (Supplementary Figure S6). Previous reports demonstrated that the enhancement of low-level side activities of preexisting enzymes or transcription regulators by overexpression or adaptive evolution could create novel metabolic pathways in *E. coli* (53–55). Based on these previous reports, known native carbon sources of *E. coli* (56), and availability of compounds, we selected ten non-native carbon sources for our experiments: monomethyl succinate, D-lyxose, 2-deoxy-D-ribose, D-tartrate, ethylene glycol, L-sorbose, sucrose, erythritol, 2-deoxy-D-glucose, and cellobiose (Supplementary Figure S7A). Previous reports have successfully identified *E. coli* variants that could utilize the first five compounds as the sole carbon source (53–55). We initially grew the reference strain either with or without MutaEco expression in minimal media containing 0.2% glycerol and 0.2% nonnative carbon source. In the next round of batch cultures, we tested two nutrient conditions in which glycerol was either the same or half the amount of the mother culture, but the concentration of the nonnative carbon source was maintained at 0.2% throughout the evolution experiments (Supplementary Figure S7B). The cells grown at OD_{600} >0.4 with the lowest amount of glycerol in 24 or 48 h were used for the next round of batch cultures. 1000-h-long evolution experiments resulted in three different groups of compounds, which showed different levels of divergence between the cells that expressed and did not express MutaEco. MutaEco-expressing cells grew without glycerol (monomethyl succinate), with no more than 4-fold less glycerol (D-tartrate, ethylene glycol and sucrose),

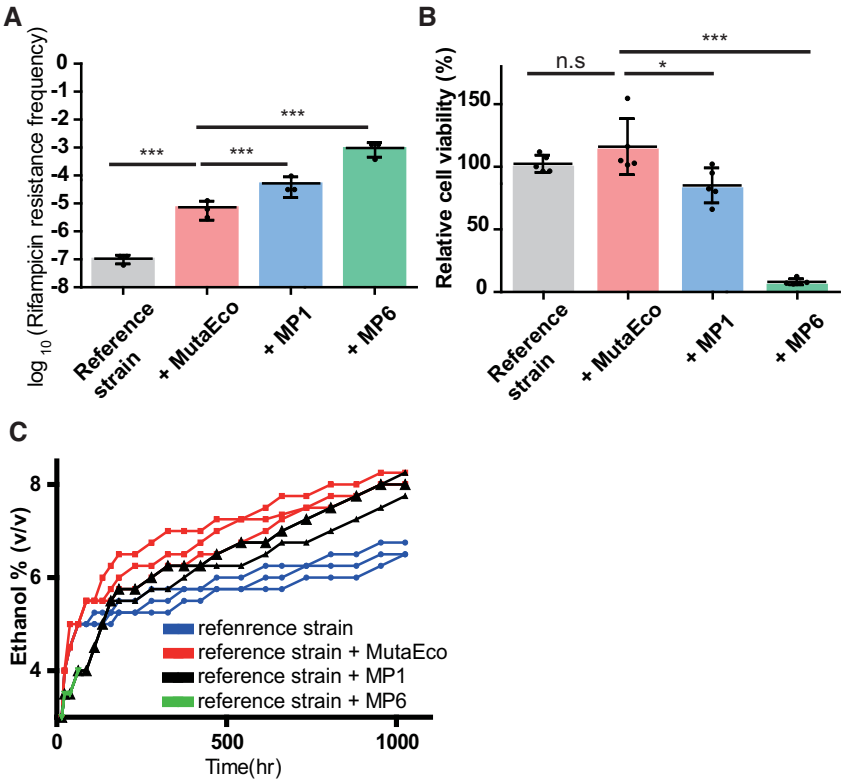


Figure 4. Comparison of MutaEco and other *in vivo* mutagenesis systems. (**A**) Rifampicin resistance frequency of cells grown for 16 h at 37°C with the expression of no mutator, MutaEco, MP1 or MP6. Data are presented as dot plots with mean ± SD ($n = 3$). *** $P < 0.001$ by Student's *t*-test. (**B**) Relative viability of cells expressing no mutator, MutaEco, MP1 or MP6. Data are presented as dot plots with mean ± SD ($n = 3$). * $P < 0.05$ and *** $P < 0.001$ by Student's *t*-test. (**C**) Trajectories of adaptive evolution of cells expressing no mutator (blue), MutaEco (red), MP1 (black), or MP6 (green) for ethanol tolerance ($n = 3$).

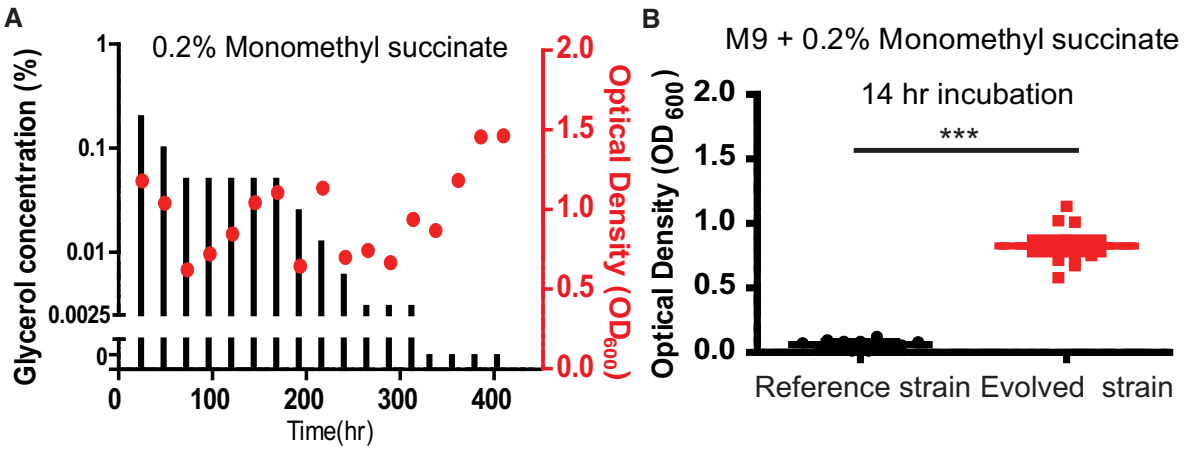
or with the same or half the amount of glycerol (the rest of compounds; Supplementary Figure S7C).

We tested the evolved cells for growth in fresh medium containing only the relevant nonnative carbon source. Although the cells evolved with monomethyl succinate, D-tartrate, ethylene glycol, sucrose or erythritol showed better growth than the reference strain (Supplementary Figure S8A), only those with monomethyl succinate or sucrose maintained proper growth in subsequent rounds of culture (Supplementary Figure S8B). Therefore, we proceeded to further analysis of the cells that evolved with these two compounds.

The evolutionary path with monomethyl succinate showed relatively fast adaptation, as shown by the gradual increase in final cell densities with single glycerol concentrations and successful cell growth without glycerol after the 320-h evolution (Figure 5A). The evolved cells also grew on a solid medium containing 0.2% monomethyl succinate (Supplementary Figure S9A). We tested ten randomly selected clones for growth in a medium containing monomethyl succinate as the sole carbon source, and found that they all grew fine, suggesting that the population of cheaters, which alone does not grow on monomethyl succinate but grow in a mixed population with the monomethyl succinate-utilizing strains, is not significant (Figure 5B). Previously, the utilization of monomethyl succinate as a sole carbon source could be developed by the overexpression of YbfF, an esterase enzyme, or a mutation in the promoter region of *ybfF* (53,55). It was believed that this mutation increased the expression of *ybfF* (55). Thus, we sequenced the *ybfF* region of three clones and found that all of them had a mutation in the promoter region, indicating similar results in our experiment (Figure 5C).

Evolution of sucrose utilization

The evolutionary path of sucrose unitization in our experiment also showed a promising pattern, in which the final cell densities increased over several rounds of growth with single sucrose concentrations (Figure 6A). Although 1000-h evolution resulted in cell growth ($OD_{600} > 0.4$) with 0.003% glycerol and 0.2% sucrose in 48 h, the evolved cells, albeit slowly, grew solely on 0.2% sucrose in 120 h (Supplementary Figure S8B). These evolved cells also grew on a solid medium containing 0.2% sucrose (Supplementary Figure S9B). Previously, the sucrose utilization in *E. coli* was developed by introducing the sucrose-utilizing *csc* genes (*cscB, cscK* and *cscA*) (57,58), and, to our knowledge, our result is, albeit with a low growth rate, the first example of sucrose utilization in *E. coli* without these *csc* genes. We randomly selected ten clones for the growth test on a sucrose-only medium, and found that they had different growth phenotypes: three clones did not grow, indicating they were cheaters (Figure 6B). We chose two clones that grew well



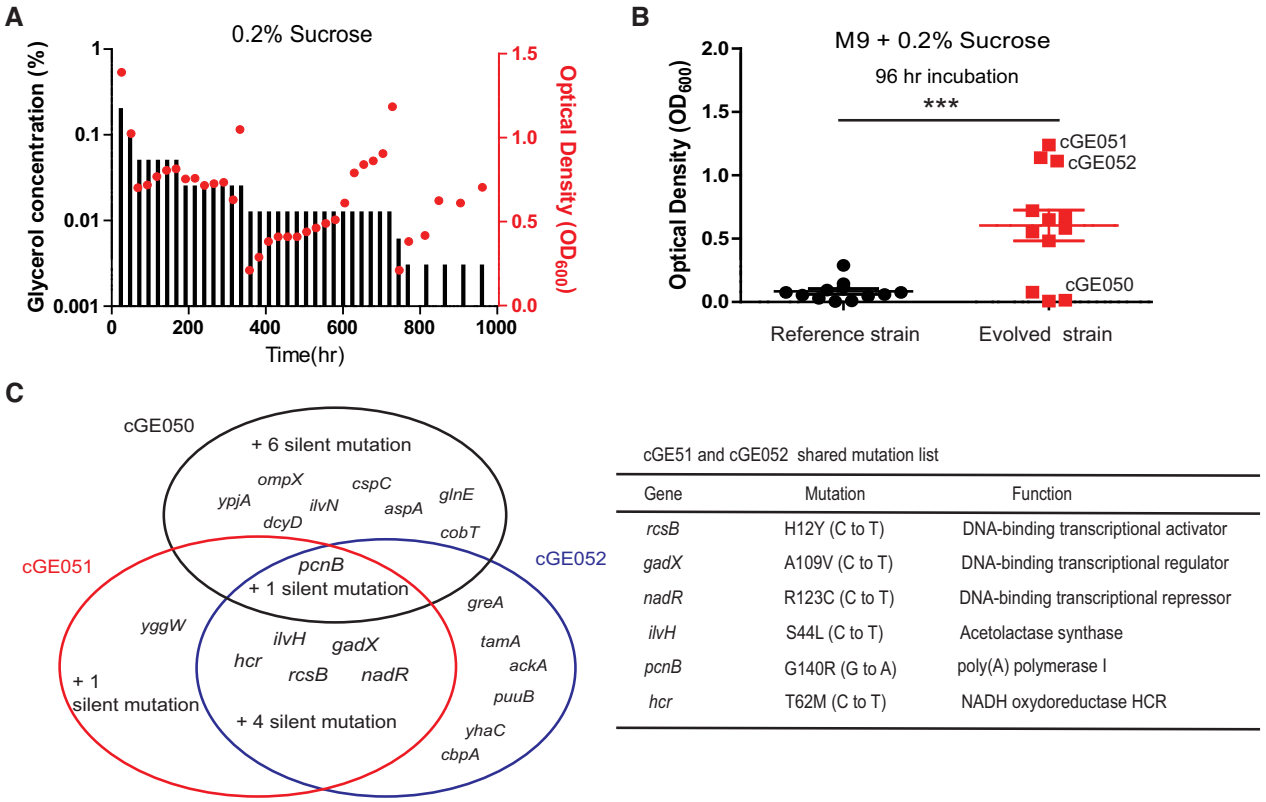
C

Strain	Gene	Mutation
clone 1	*ybfF*	-45 C to T, -76 C to T
clone 2	*ybfF*	-45 C to T, -76 C to T
clone 3	*ybfF*	-45 C to T, -76 C to T

Figure 5. Evolution of monomethyl succinate utilization as the sole carbon source in *E. coli*. (**A**) Detailed evolutionary trajectory for monomethyl succinate utilization. The minimal glycerol concentrations for growth ($OD_{600} > 0.4$) and the final optical densities of cells are presented as the bar graph (black) and the dot plot (red), respectively. (**B**) Overnight cultures of clones obtained from the reference strain or the evolved cells were diluted and incubated at 37°C for 14 h in the M9 minimal medium supplemented with 0.2% monomethyl succinate. Final optical densities are presented as dot plots with mean ± SD ($n = 12$). *** $P < 0.001$ by Student's *t*-test. (**C**) The list of mutations found in the *ybfF* gene of three clones obtained from the evolved cells.

on sucrose (cGE051 and cGE052) and one cheater clone (cGE050), and determined their genome sequences using the Illumina method. As expected, cGE051 and cGE052 shared 11 mutations, including five silent mutations and six missense mutations in *hcr*, *gadX*, *ilvH*, *nadR*, *pcnB*, and *rcsB*, out of a total of 13 and 17 mutations, respectively (Supplementary Table SS5), whereas cGE050 had only two mutations in common with the other two strains out of the total 16 mutations (Figure 6C). It is not clear how these mutations lead to sucrose utilization because these six genes in which the common missense mutations were found do not encode any candidate transporter or hydrolase (Figure 6C). We also found that the combination of the mutations in the evolved cells and the sucrose-utilizing *csc* genes resulted in faster growth or higher final densities of cells on the sucrose-only media (Supplementary Figure S10). Collectively, our mutator protein successfully evolved an *E. coli* strain that grew solely on sucrose and enhanced the efficiency of sucrose utilization when combined with the sucrose-utilizing *csc* genes.

In conclusion, by adapting the previously reported fusion strategy, we have developed a new mutator protein to generate mutations in the bacterial genome. We demonstrated that the chimeric protein, in which a DNA-modifying enzyme is linked to the transcriptional machinery, could introduce mutations at a higher rate than the DNA-modifying enzyme alone, thereby promoting the faster adaptive evolution of cells. MutaEco has several advantages over previously established inducible mutator proteins: First, MutaEco does not require a dominant-negative variant of proteins that function in accurate DNA replication, but is a simple fusion of an exogenous DNA-modifying enzyme with a subunit of the RNA polymerase complex. Because the bacterial RNAP complex is highly conserved (42,59), this strategy may be easily applicable to other bacterial species. Second, MutaEco is a modular mutator, in which the cytidine deaminase directly modifies the DNA and the α-subunit of RNA polymerase helps target the genomic DNA; therefore, the mutation spectrum of MutaEco, which is currently limited to transition mutations, may be expanded by using other DNA-modifying enzymes. Recently, a new mutation system using the fusion of an ssDNA binding protein and a cytidine deaminase was reported for genome evolution in *Saccharomyces cerevisiae* (60). Although we tested our mutator in the *Δung* strain, an uracil–DNA glycosylase inhibitor from a *Bacillus subtilis* bacteriophage, UGI, has been shown to inhibit UDGs from various species including *Bacillus subtilis*, *E. coli*, *Saccharomyces cerevisiae*, rat and human, by binding the conserved DNA-binding groove of UDG (61,62). Therefore,



cGE51 and cGE052 shared mutation list

Gene	Mutation	Function
rcsB	H12Y (C to T)	DNA-binding transcriptional activator
gadX	A109V (C to T)	DNA-binding transcriptional regulator
nadR	R123C (C to T)	DNA-binding transcriptional repressor
ilvH	S44L (C to T)	Acetolactase synthase
pcnB	G140R (G to A)	poly(A) polymerase I
hcr	T62M (C to T)	NADH oxydoreductase HCR

Figure 6. Evolution of sucrose utilization in *E.coli*. (**A**) Evolutionary trajectory for sucrose utilization. The minimal glycerol concentrations for growth (OD_{600} > 0.4) and the final optical densities of cells are presented as the bar graph (black) and the dot plot (red), respectively. (**B**) Overnight cultures of clones obtained from the reference strain or the evolved cells were diluted and incubated at 37°C for 96 h in the M9 minimal medium supplemented with 0.2% sucrose. Final optical densities are presented as dot plots with mean ± SD ($n = 12$). *** $P < 0.001$ by Student's *t*-test. (**C**) The list of mutations found in the three clones obtained from the evolved cells. Two clones (cGE051 and cGE052) that could grow on sucrose and one cheater clone (cGE050) that could not utilize sucrose were selected and subjected to whole genome sequencing.

we believe that UGI expression *in trans* or as a fused protein can be used instead of the UDG deletion in *E. coli* or other bacteria. We believe that MutaEco is not only useful for the evolutionary engineering of microbial cells, but can also be used as a reference to develop novel methods for *in vivo* DNA engineering.

DATA AVAILABILITY

All Illumina sequencing data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession numbers E-MTAB-10368 (Whole genome sequencing) and E-MTAB-10852 (RNA-seq).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Hyojin Park, Daeje Seo, Younghyun Kim, Chanwoo Lee, Hyunjin Cho, Inseok Song, and Hyunsung Nam for helpful discussions. We thank Nam Ki Lee and Soojung Lee for providing DNA plasmid, and Hyunwoo Lee, Yoonbin Lee, Gayoung Lim, Byungmoon Kim, Chulbom Lee and Seungbum Park for providing non-native carbon sources.

FUNDING

National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [2021R1A2C1008730]. Funding for open access charge: Seoul National University.
Conflict of interest statement. None declared.

REFERENCES

1. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
2. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2579–E2586.
3. Komor,A.C., Kim,Y.B., Packer,M.S., Zuris,J.A. and Liu,D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.
4. Nishida,K., Arazoe,T., Yachie,N., Banno,S., Kakimoto,M., Tabata,M., Mochizuki,M., Miyabe,A., Araki,M., Hara,K.Y. *et al.* (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, **353**, aaf8729.
5. Ma,Y., Zhang,J., Yin,W., Zhang,Z., Song,Y. and Chang,X. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods*, **13**, 1029–1035.
6. Anzalone,A.V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A. *et al.*

(2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
7. Anzalone,A.V., Koblan,L.W. and Liu,D.R. (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
8. Morrison,M.S., Podracky,C.J. and Liu,D.R. (2020) The developing toolkit of continuous directed evolution. *Nat. Chem. Biol.*, **16**, 610–619.
9. Dragosits,M. and Mattanovich,D. (2013) Adaptive laboratory evolution–principles and applications for biotechnology. *Microb. Cell Fact.*, **12**, 64.
10. Sandberg,T.E., Salazar,M.J., Weng,L.L., Palsson,B.O. and Feist,A.M. (2019) The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.*, **56**, 1–16.
11. Selifonova,O., Valle,F. and Schellenberger,V. (2001) Rapid evolution of novel traits in microorganisms. *Appl. Environ. Microbiol.*, **67**, 3645–3649.
12. Shimoda,C., Itadani,A., Sugino,A. and Furusawa,M. (2006) Isolation of thermotolerant mutants by using proofreading-deficient DNA polymerase δ as an effective mutator in Saccharomyces cerevisiae. *Genes Genet. Syst.*, **81**, 391–397.
13. Abe,H., Fujita,Y., Takaoka,Y., Kurita,E., Yano,S., Tanaka,N. and Nakayama,K. (2009) Ethanol-tolerant Saccharomyces cerevisiae strains isolated under selective conditions by over-expression of a proofreading-deficient DNA polymerase delta. *J. Biosci. Bioeng.*, **108**, 199–204.
14. Luan,G., Cai,Z., Li,Y. and Ma,Y. (2013) Genome replication engineering assisted continuous evolution (GREACE) to improve microbial tolerance for biofuels production. *Biotechnol. Biofuels*, **6**, 137.
15. Swings,T., Van den Bergh,B., Wuyts,S., Oeyen,E., Voordeckers,K., Verstrepen,K.J., Fauvart,M., Verstraeten,N. and Michiels,J. (2017) Adaptive tuning of mutation rates allows fast response to lethal stress in Escherichia coli. *Elife*, **6**, e22939.
16. Fernandez-Cabezon,L., Cros,A. and Nikel,P.I. (2021) Spatiotemporal manipulation of the mismatch repair system of *Pseudomonas putida* accelerates phenotype emergence. *ACS Synth. Biol.*, **10**, 1214–1226.
17. Sniegowski,P.D., Gerrish,P.J. and Lenski,R.E. (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*, **387**, 703–705.
18. Taddei,F., Radman,M., Maynard-Smith,J., Toupance,B., Gouyon,P.-H. and Godelle,B. (1997) Role of mutator alleles in adaptive evolution. *Nature*, **387**, 700–702.
19. Fijalkowska,I.J. and Schaaper,R.M. (1996) Mutants in the Exo I motif of *Escherichia coli* dnaQ: defective proofreading and inviability due to error catastrophe. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 2856–2861.
20. Loh,E., Salk,J.J. and Loeb,L.A. (2010) Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 1154–1159.
21. Sprouffske,K., Aguilar-Rodríguez,J., Sniegowski,P. and Wagner,A. (2018) High mutation rates limit evolutionary adaptation in Escherichia coli. *PLoS Genet.*, **14**, e1007324.
22. Witkin,E.M. (1976) Ultraviolet mutagenesis and inducible DNA repair in Escherichia coli. *Bacteriol. Rev.*, **40**, 869–907.
23. Alcantara-Diaz,D., Brena-Valle,M. and Serment-Guerrero,J. (2004) Divergent adaptation of *Escherichia coli* to cyclic ultraviolet light exposures. *Mutagenesis*, **19**, 349–354.
24. Notley-McRobb,L., Pinto,R., Seeto,S. and Ferenci,T. (2002) Regulation of mutY and nature of mutator mutations in *Escherichia coli* populations under nutrient limitation. *J. Bacteriol.*, **184**, 739–745.
25. Greener,A., Callahan,M. and Jerpseth,B. (1997) An efficient random mutagenesis technique using an E. coli mutator strain. *Mol. Biotechnol.*, **7**, 189–195.
26. Zhu,L., Cai,Z., Zhang,Y. and Li,Y. (2014) Engineering stress tolerance of *Escherichia coli* by stress-induced mutagenesis (SIM)-based adaptive evolution. *Biotechnol. J.*, **9**, 120–127.
27. Badran,A.H. and Liu,D.R. (2015) Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.*, **6**, 8425.
28. Bhagwat,A.S., Hao,W., Townes,J.P., Lee,H., Tang,H. and Foster,P.L. (2016) Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 2176–2181.
29. Park,H. and Kim,S. (2021) Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. *Nucleic Acids Res.*, **49**, e32.
30. Urban,A., Neukirchen,S. and Jaeger,K.-E. (1997) A rapid and efficient method for site-directed mutagenesis using one-step overlap extension PCR. *Nucleic Acids Res.*, **25**, 2227–2228.
31. Reikofski,J. and Tao,B.Y. (1992) Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.*, **10**, 535–547.
32. Sharan,S.K., Thomason,L.C., Kuznetsov,S.G. and Court,D.L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.*, **4**, 206–223.
33. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
34. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **26**, 589–595.
35. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
36. Moore,C.L., Papa,L.J. 3rd and Shoulders,M.D. (2018) A processive protein chimera introduces mutations across defined DNA regions in vivo. *J. Am. Chem. Soc.*, **140**, 11560–11564.
37. Chen,H., Liu,S., Padula,S., Lesman,D., Griswold,K., Lin,A., Zhao,T., Marshall,J.L. and Chen,F. (2020) Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.*, **38**, 165–168.
38. Alvarez,B., Mencia,M., de Lorenzo,V. and Fernandez,L.A. (2020) In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9. *Nat. Commun.*, **11**, 6436.
39. Cravens,A., Jamil,O.K., Kong,D., Sockolosky,J.T. and Smolke,C.D. (2021) Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering. *Nat. Commun.*, **12**, 1579.
40. Lada,A.G., Krick,C.F., Kozmin,S.G., Mayorov,V.I., Karpova,T.S., Rogozin,I.B. and Pavlov,Y.I. (2011) Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Mosc)*, **76**, 131–146.
41. Murakami,K.S. and Darst,S.A. (2003) Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.*, **13**, 31–39.
42. Werner,F. and Grohmann,D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.*, **9**, 85–98.
43. Salter,J.D., Bennett,R.P. and Smith,H.C. (2016) The APOBEC protein family: united by structure, divergent in function. *Trends Biochem. Sci.*, **41**, 578–594.
44. Petersen-Mahrt,S.K., Harris,R.S. and Neuberger,M.S. (2002) AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature*, **418**, 99–104.
45. Goldstein,B.P. (2014) Resistance to rifampicin: a review. *J. Antibiot. (Tokyo)*, **67**, 625–630.
46. Horinouchi,T., Maeda,T. and Furusawa,C. (2018) Understanding and engineering alcohol-tolerant bacteria using OMICS technology. *World J. Microbiol. Biotechnol.*, **34**, 157.
47. Cao,H., Wei,D., Yang,Y., Shang,Y., Li,G., Zhou,Y., Ma,Q. and Xu,Y. (2017) Systems-level understanding of ethanol-induced stresses and adaptation in *E. coli*. *Sci. Rep.*, **7**, 44150.
48. Swings,T., Weytjens,B., Schalck,T., Bonte,C., Verstraeten,N., Michiels,J. and Marchal,K. (2017) Network-based identification of adaptive pathways in evolved ethanol-tolerant bacterial populations. *Mol. Biol. Evol.*, **34**, 2927–2943.
49. Haft,R.J., Keating,D.H., Schwaegler,T., Schwalbach,M.S., Vinokur,J., Tremaine,M., Peters,J.M., Kotlajich,M.V., Pohlmann,E.L., Ong,I.M. *et al.* (2014) Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E2576–E2585.
50. Woodruff,L.B., Pandhal,J., Ow,S.Y., Karimpour-Fard,A., Weiss,S.J., Wright,P.C. and Gill,R.T. (2013) Genome-scale identification and characterization of ethanol tolerance genes in Escherichia coli. *Metab. Eng.*, **15**, 124–133.
51. Nicolaou,S.A., Gaida,S.M. and Papoutsakis,E.T. (2012) Exploring the combinatorial genomic space in Escherichia coli for ethanol tolerance. *Biotechnol. J.*, **7**, 1337–1345.
52. Goodarzi,H., Bennett,B.D., Amini,S., Reaves,M.L., Hottes,A.K., Rabinowitz,J.D. and Tavazoie,S. (2010) Regulatory and metabolic

rewiring during laboratory evolution of ethanol tolerance in E. coli. *Mol. Syst. Biol.*, **6**, 378.
53. Notebaart,R.A., Szappanos,B., Kintses,B., Pal,F., Gyorkei,A., Bogos,B., Lazar,V., Spohn,R., Csorgo,B., Wagner,A. *et al.* (2014) Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 11762–11767.
54. Szappanos,B., Fritzemeier,J., Csorgo,B., Lazar,V., Lu,X., Fekete,G., Balint,B., Herczeg,R., Nagy,I., Notebaart,R.A. *et al.* (2016) Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat. Commun.*, **7**, 11607.
55. Guzman,G.I., Sandberg,T.E., LaCroix,R.A., Nyerges,A., Papp,H., de Raad,M., King,Z.A., Hefner,Y., Northen,T.R., Notebaart,R.A. *et al.* (2019) Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.*, **15**, e8462.
56. Adler,J., Hazelbauer,G.L. and Dahl,M. (1973) Chemotaxis toward sugars in *Escherichia coli*. *J. Bacteriol.*, **115**, 824–847.
57. Archer,C.T., Kim,J.F., Jeong,H., Park,J.H., Vickers,C.E., Lee,S.Y. and Nielsen,L.K. (2011) The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics*, **12**, 9.
58. Mohamed,E.T., Mundhada,H., Landberg,J., Cann,I., Mackie,R.I., Nielsen,A.T., Herrgard,M.J. and Feist,A.M. (2019) Generation of an *E. coli* platform strain for improved sucrose utilization using adaptive laboratory evolution. *Microb. Cell Fact.*, **18**, 116.
59. Ebright,R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.*, **304**, 687–698.
60. Pan,Y., Xia,S., Dong,C., Pan,H., Cai,J., Huang,L., Xu,Z. and Lian,J. (2021) Random base editing for genome evolution in *Saccharomyces cerevisiae*. *ACS Synth. Biol.*, **10**, 2440–2446.
61. Mol,C.D., Arvai,A.S., Sanderson,R.J., Slupphaug,G., Kavli,B., Krokan,H.E., Mosbaugh,D.W. and Tainer,J.A.J.C. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, **82**, 701–708.
62. Schormann,N., Ricciardi,R. and Chattopadhyay,D. (2014) Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.*, **23**, 1667–1685.