PAM-expanded *Streptococcus thermophilus* Cas9 C-to-T and C-to-G base editors for programmable base editing in mycobacteria

Hongyuan Zhang[1,2], Yifei Zhang[1,2], Wei-Xiao Wang[3], Weizhong Chen[1], Xia Zhang[4], Xingxu Huang[5,6], Wei Chen[3*], Quanjiang Ji[1,6*]

[1]School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China.

[2]University of Chinese Academy of Sciences, Beijing 100049, China.

[3]Clinical Research Center, the Second Hospital of Nanjing, Nanjing University of Chinese Medicine, Nanjing, 210003, China.

[4]Department of Tuberculosis, the Second Hospital of Nanjing, Nanjing University of Chinese Medicine, Nanjing, 210003, China.

[5]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China.

[6]Guangzhou Laboratory, Guangzhou 510120, China.

*To whom corresponding should be addressed:

Quanjiang Ji: quanjiangji@shanghaitech.edu.cn

Wei Chen: njyy039@njucm.edu.cn

Abstract

New therapeutic strategies for rapid and effective treatment of tuberculosis are highly desirable, and their development can be drastically accelerated by facile genetic manipulation methods in *Mycobacterium tuberculosis*. CRISPR base editors allow for rapid, robust, and programmed single base substitutions and gene inactivation, yet no such systems are currently available in *M. tuberculosis*. By screening distinct CRISPR base editors, we identified that only the unusual *Streptococcus thermophilus* Cas9 (St1Cas9) cytidine base editor (CBE), but not the widely used *Streptococcus pyogenes* Cas9 or *Lachnospiraceae bacterium* Cpf1 CBEs, are active in mycobacteria. Despite the notable C-to-T conversions, a high portion of undesired byproducts existed with St1Cas9 CBE. We thus engineered St1Cas9 CBE by uracil DNA glycosylase inhibitor (UGI) or uracil DNA glycosylase (UNG) fusion, yielding two new base editors (CTBE and CGBE), capable of C-to-T or C-to-G conversions with dramatically enhanced editing product purity and multiplexed editing capacity. Because wild-type St1Cas9 recognizes a relatively strict PAM sequence for DNA targeting, we evolved a PAM-expanded St1Cas9 variant by structure-guided protein engineering for the base editors, substantially broadening the targeting scope. Our approaches significantly reduce the efforts and time for precise genetic manipulation and will facilitate functional genomics and drug-target exploration in *M. tuberculosis* and related organisms.

Introduction

As the leading cause of death from a single infectious agent, *Mycobacterium tuberculosis* caused 10 million infections and 1.45 million deaths in 2018 worldwide[1]. The current recommended treatment against drug-susceptible tuberculosis (TB) requires a prolonged 6-month combination therapy containing four first-line drugs. Furthermore, the rapid rise of multi-drug-resistant and totally drug-resistant TB renders this disease even more difficult to cure[2,3]. Thereby, new therapeutic strategies for rapid and effective treatment of TB are highly desirable.

The development of new therapeutic methods against *M. tuberculosis* infections can be drastically accelerated by facile, precise, and markless genetic manipulation methods that allow for rapid identification and characterization of new drug targets in *M. tuberculosis*. Current methods for genetic manipulation in *M. tuberculosis* include allelic exchange using long linear DNA fragments[4], specialized transduction based on mycobacteriophage[5], recombineering mediated by phage-encoded recombination system[6], oligonucleotide-mediated recombineering followed by Bxb1 integrase targeting[7] and non-homologous end joining (NHEJ) induced by highly specific double strand break[8]. These methods either use multiple transformation steps and thus require months to years for editing, or can only generate non-precise mutations and thus are not amenable for gene function exploration at a single-base resolution.

Recently, RNA-guided Cas nucleases have been engineered for rapid and precise genetic manipulation in a variety of organisms, including many eukaryotic cells and

64 diverse bacterial species[9-17]. Through base pairing with a genomic sequence adjacent to
65 a protospacer adjacent motif (PAM), the Cas/gRNA complexes can specifically bind to
66 the target genomic site and generate a double-stranded DNA break (DSB). With the
67 assistance of the cellular homology-directed repair (HDR) pathway or the non-
68 homologous end joining (NHEJ) pathway, precise and non-precise genetic
69 manipulations can be readily achieved, respectively. Moreover, the CRISPR-Cas
70 systems can be further engineered to be transcription inhibition (CRISPRi) systems for
71 direct gene knockdown without relying on cellular DNA repair mechanisms[18,19]. In *M.*
72 *tuberculosis*, CRISPR-assisted NHEJ[8] and CRISPRi[20-23] strategies that allow for non-
73 precise editing and partially gene knockdown, respectively, have been developed for
74 genetic manipulation. However, the highly desirable precise and fully gene inactivation
75 methods are currently unavailable, likely due to the lack of a compatible HDR system.
76 More recently, the development of cytosine base editors (CBEs) and adenine base
77 editors (ABEs) have provided new strategies for precise genetic manipulation[24-26]. The
78 base editors composed of the fusion of a nucleotide deaminase and a dead Cas9 or a
79 Cas9 nickase, can directly achieve precise C-to-T or A-to-G conversions by a catalytic
80 deamination reaction without requiring HDR, and thus are widely applicable for precise
81 genetic manipulation in a variety of microbes with different genetic background[27-35].
82 However, no such systems are currently available in *M. tuberculosis*.
83 In this study, we report the establishment of two cytosine base editors in
84 mycobacterium. By screening distinct CRISPR base editors, we identify that only the

unusual *Streptococcus thermophilus* Cas9 (St1Cas9) cytosine base editor (CBE), but not the widely used *Streptococcus pyogenes* Cas9 or *Lachnospiraceae bacterium* Cpf1 CBEs, are active in mycobacteria. Through systematic engineering of St1Cas9 CBE by DNA glycosylase inhibitor (UGI) or uracil DNA glycosylase (UNG) fusion and PAM expansion of St1Cas9, we created a C-to-T base editor and a C-to-G base editor with enhanced editing product purity, broadened targeting scope, and multiplexed editing capacity. Moreover, we evaluated the off-target effect of the C-to-T base editor by whole-genome sequencing and did not observe detectable off-target editing events. Our approaches require only a single plasmid and one transformation step for efficient and scarless editing, significantly reducing the efforts and time for precise genetic manipulation in mycobacterium.

Results

Identification of active cytosine base editors in *M. smegmatis*.

Streptococcus pyogenes Cas9 (SpCas9) has been widely adapted for genetic manipulation in numerous microbes[16,29,36-39]. However, it is restricted for application in mycobacteria because of the notable cellular toxicity and low DNA targeting efficiency[8,22]. To develop an active CBE in mycobacterium, we screened different CBEs composed of the fusion of rat APOBEC1 cytosine deaminase and different Cas nucleases, including dSpCas9 (catalytically inactive *Streptococcus pyogenes* Cas9), dLbCpf1 (catalytically inactive *Lachnospiraceae bacterium* Cpf1) and dSt1Cas9

106 (catalytically inactive *Streptococcus thermophilus* Cas9). The expression of the Cas
107 nucleases was under the control of an anhydrotetracycline (ATc)-inducible promoter
108 while the expression of the corresponding gRNAs was under the control of a synthetic
109 constitutive promoter J23119.

110 To quantitatively assess the efficiency of targeted base editing, same amount of
111 different CBE plasmids was electroporated into *M. smegmatis*. $CBE_{dSt1Cas9}$ (dSt1Cas9
112 CBE) induced notable C-to-T conversions (from 4% to 15%) with undesired C-to-G
113 conversions (from 18% to 70%) as the major editing products at all the test sites,
114 whereas $CBE_{dSpCas9}$ (dSpCas9 CBE) and $CBE_{dLbCpf1}$ (dLbCas9 CBE) did not show
115 detectable base editing activity at all the target regions (Fig. 1A). Moreover, consistent
116 with previous studies[40,41], no notable cellular toxicity was observed for $CBE_{dSt1Cas9}$ or
117 $CBE_{dLbCpf1}$, whereas high cellular toxicity existed with $CBE_{dSpCas9}$ (Fig. 1B). Thereby,
118 we chose $CBE_{dSt1Cas9}$ as the starting system for further engineering. To cure the plasmid
119 after base editing, we replaced the non-replicating L5 integrating backbone used in the
120 $CBE_{dSt1Cas9}$ system with the replicable pMF1 backbone, and found that the resulting
121 editing plasmid did not show notable cellular toxicity and can be easily cured by
122 culturing the bacterium in the absence of antibiotic (Fig. S1A and B, and Fig. S2A and
123 B).

124

125 **Development of CTBE and CGBE in *M. smegmatis*.**

The G:U mismatch pair in cells is generally repaired by uracil DNA glycosylase (UNG)-mediated base excision repair (BER) process. Inhibition of UNG would protect the edited G:U intermediate from cleavage and thus improve C-to-T conversion efficiency and editing product purity[24,42](Fig. 1C). We fused two uracil glycosylase inhibitor (UGI) to the C terminus of dSt1Cas9 to create the C-to-T base editor ($CTBE_{WT}$). To examine the C-to-T editing efficiency of $CTBE_{WT}$, five $CTBE_{WT}$ plasmids targeting five different loci were separately electroporated into *M. smegmatis*. $CTBE_{WT}$ achieved high levels of C-to-T base editing frequency ranging from 69% to 86% with significantly reduced formation of undesired byproducts (Fig. 1D).

Recent studies revealed that promoting the BER pathway by UNG or other DNA repair proteins fusion to SpCas9 CBEs can convert target C:G base pair into G:C or A:T base pair rather than the expected T:A product[43-46]. Given that a high portion of unexpected C-to-G byproduct existed with $CBE_{dSt1Cas9}$, we anticipated that fusion of UNG to $CBE_{dSt1Cas9}$ would generate new base editors. We thus fused *E. coli* UNG (eUNG) to the C terminus of dSt1Cas9 in $CBE_{dSt1Cas9}$ to create a new base editor $CGBE_{WT}$ and compared the base editing efficiencies of $CGBE_{WT}$ and $CBE_{dSt1Cas9}$. As shown in Fig. 1E, $CGBE_{WT}$ can efficiently achieve C-to-G conversions with enhanced editing product purity at all the five tested sites.

Engineering of a PAM-expanded St1Cas9 variant for the base editors.

St1Cas9 requires a relatively strict PAM sequence (5'-NNRGAA-3', where R is A or G) for DNA targeting[41], significantly restricting the editing scope of CTBE$_{WT}$ and CGBE$_{WT}$. We previously engineered a St1Cas9 variant KLKL by introducing D939K/E1057L/N1081K/K1086L mutations to relieve the PAM specificity[41]. In wild-type St1Cas9, Q1084 and K1086 form bidentate hydrogen bond with the third A and fourth G when 5'-NNAGGA-3' PAM is used, and Q1084 is further stabilized by E1057 via a hydrogen bond (Fig. 2A). Mutations of E1057L and K1086L would disrupt base-specific interactions, while mutations of D939K and N1081K would introduce non-base specific interactions[41]. We recently identified that mutation of L1057Q in the KLKL variant would increase the in vitro DNA cleavage activity towards 5'-NNTTAA-3' PAM- and 5'-NNCTAA-3' PAM-containing DNAs (Fig. 2B). We therefore engineered a St1Cas9 KQKL variant containing D939K/E1057Q/N1081K/K1086L mutations to further relieve the PAM specificity. Comprehensive in vitro DNA cleavage assay revealed that mutation of D939K/E1057Q/N1081K/K1086L substantially expanded PAM recognition scope of the KQKL variant with the 5'-NNNNAA-3' PAM specificity compared with the 5'-NNRGAA-3' specificity of the wild-type St1Cas9 (Fig. 2C).

Next, we replaced wild-type St1Cas9 of CTBE$_{WT}$ with the KQKL variant to expand the editing scope of the C-to-T base editor (Fig. 3A). The resulting CTBE$_{evolve}$ system was first subjected to ATc concentration optimization because ATc is the inducer for the expression of the engineered St1Cas9 variant. We screened six different

167 concentrations of ATc ranging from 5 ng/mL to 100 ng/mL and no significant
168 differences in the editing efficiencies were noticed (Fig. S3). We selected 20 ng/mL as
169 the inducer concentration for base editing in the subsequent experiments.

170 To systemically characterize the PAM preference of $CTBE_{evolve}$ in vivo, we
171 assembled 48 spacers targeting 48 different endogenous genomic sites with 3 spacers
172 for each 5'-NNNNAA-3' PAM. We collected all the colonies after each transformation
173 and subjected the PCR products of the target sites for sequencing to evaluate the editing
174 efficiencies. Consistent with the in vitro DNA cleavage assay, $CTBE_{evolve}$ has an
175 expanded PAM preference with the 5'-NNNNAA-3' PAM specificity (Fig. 3B). We
176 also noticed that when targeting the sites with less-active PAMs, such as 5'-NNTAAA-
177 3', 5'-NNTCAA-3', 5'-NNTGAA-3', and 5'-NNTTAA-3', the spacer sequence content
178 can significantly affect the editing efficiency of $CTBE_{evolve}$ (Fig. 3B). Moreover, we
179 analyzed the editing window of $CTBE_{evolve}$, revealing that $CTBE_{evolve}$ preferred to edit
180 Cs within the window from positions 4 to 12 (Fig. 3C). When the Cs were located
181 outside the window from positions 4 to 12, but inside the window from 2 to 15, the Cs
182 could sometimes still be edited with high efficiencies (Fig. 3C).

183 Similarly, we engineered $CGBE_{WT}$ with $St1Cas9_{evolve}$ to construct the $CGBE_{evolve}$
184 system (Fig. 3A). We comprehensively characterized $CGBE_{evolve}$ by testing the editing
185 efficiencies of 29 endogenous genomic sites. Interestingly, only the Cs within the
186 window from positions 5 to 8 could be edited with the $CGBE_{evolve}$ system (Fig. 3D)
187 although the same Cas9 protein and deaminase were used for $CGBE_{evolve}$ and

188 CTBE$_{evolve}$. Moreover, only Cs with the TC motif could be edited even if they were
189 located within the editing window from positions 5 to 8 (Fig. 3D).
190 Additionally, we replaced the eUNG in CGBE$_{evolve}$ with an orthologous UNG from
191 *M. smegmatis* (mUNG) to facilitate C-to-G conversions in *M. smegmatis* (Fig. S4). Four
192 versions of C-to-G base editors with the fusion of eUNG or mUNG to the C or N
193 terminus of the St1Cas9 protein were constructed for base editing in *M. smegmatis*. We
194 found that fusion of mUNG or eUNG to the C terminus of St1Cas9 gave similar editing
195 efficiencies at all the tested sites, whereas fusion of mUNG or eUNG at the N terminus
196 did not yield efficient C-to-G editing (Fig. S4). Given the similar editing efficiency
197 using mUNG or eUNG, we kept eUNG in CGBE$_{evolve}$ for further characterization.
198
199 **Gene inactivation by CTBE$_{evolve}$ in *M. smegmatis*.**
200 CRISPR C-to-T base editors can convert CAA, CGA, CAG, and TGG codons to
201 premature stop codons, and thereby are promising tools for gene inactivation[47,48]. We
202 examined the gene inactivation capacity of CTBE$_{evolve}$ by designing three different
203 spacers targeting the essential L-leucine biosynthesis genes *leuB* and *leuC*. Efficient
204 editing was observed for all the designed spacers (Fig. 3E and F and Fig. S5) and the
205 isolated pure mutants were subjected to the phenotypical assay. Generating a premature
206 stop codon in *leuB* or *leuC* renders the bacterium incapable of growth in the absence of
207 L-leucine, confirming the inactivation of L-leucine biosynthesis (Fig. 3F). Moreover,
208 similar strategy was successfully applied for inactivating *ctpE* by generating a

209 premature stop codon with CTBE$_{evolve}$. Consistent with previous studies[49], inactivation
210 of *ctpE* by introducing the premature stop codon increased bacterium aggregation in
211 the presence of EGTA (Fig. S6). Together, these results confirmed that CTBE$_{evolve}$ is a
212 powerful and reliable tool for gene inactivation in *M. smegmatis*.
213
214 **Multiplexed editing in *M. smegmatis*.**
215 Serial genome editing for multiple genes in slow-growing pathogens is extremely time
216 consuming, and multiplexed genome editing can drastically expedite the genome
217 editing progress. We assembled two sgRNA-expression cassettes into a single
218 CTBE$_{evolve}$ plasmid to test the multiplexed editing capacity of CTBE$_{evolve}$ in the *M.*
219 *smegmatis* mc^{2}155 strain (Fig. S7A). As shown in Fig. S7B, two genes (*sigF* and
220 Ms6753) were simultaneously mutated in seven out of eight randomly picked colonies.
221 Moreover, three different genomic sites (*cysS*, *sigF* and Ms6753) were targeted
222 simultaneously by assembling three sgRNAs into the CTBE$_{evolve}$ plasmid, and all of the
223 three targeted sites were successfully mutated in seven out of eight randomly picked
224 colonies (Fig. 4A, B). Similarly, we assembled two or three sgRNA-expression
225 cassettes into a single CGBE$_{evolve}$ plasmid and tested the multiplexed editing capacity
226 of C-to-G conversion in the *M. smegmatis* mc^{2}155 strain (Fig. S8A and Fig. 4C). As
227 shown in Fig. S8B and Fig. 4D, highly efficient multiplexed editing of C-to-G
228 conversion was observed in both cases. Together, these results demonstrated that both

CTBE$_{evolve}$ and CGBE$_{evolve}$ are amenable for multiplexed editing with high efficiencies in *M. smegmatis*.

Genome-wide off-target evaluation for CTBE$_{evolve}$.

To evaluate the genome-wide off-target editing of the CTBE$_{evolve}$ system, two edited *M. smegmatis* colonies were randomly selected and subjected to whole-genome sequencing. No potential off-target editing sites that contain the identical sequences to the PAM proximal 1–8 nt of the protospacer were detected (Table S1). These results demonstrated the high editing fidelity of the CTBE$_{evolve}$ system and are consistent with previous discoveries that St1Cas9 is a high-fidelity enzyme and its editing is highly sensitive to the mutations in the spacer sequences[41].

Genome editing in *M. tuberculosis*.

Given the success of base editing with CTBE$_{evolve}$ in *M. smegmatis*, we sought to examine the base editing capacity of this system in *M. tuberculosis*. Four different spacers targeting four different endogenous sites were separately cloned into the CTBE$_{evolve}$ system, and the resulting editing plasmids were separately electroporated into the *M. tuberculosis* H37Rv competent cells. C-to-T conversion efficiencies were measured by collecting all the transformants and sequencing the target sites. Notable C-to-T conversions were observed for all the four tested sites with the editing frequencies ranging from 12% to 95% (Fig. 5A). Moreover, we applied this system for

gene inactivation by introducing a premature stop codon into *katG* (Fig. 5B). Inactivation of katG was further confirmed by a phenotypic assay as the *katG* mutant was more resistant to Isoniazid (INH) treatment (16 μg/mL) than that of the wild-type strain[50] (Fig. 5B). Because the relative position of the introduced premature stop codon in an open reading frame (ORF) can significantly affect the gene inactivation efficiency, we systematically calculated the possible targetable codons of $CTBE_{evolve}$ in mycobacteria with CRISPR-CBEI[48]. As shown in Fig. 5C, more than 75%, 60%, and 40% ORFs of the analyzed mycobacterium species (*M. tuberculosis*, *M. smegmatis*, and *M. marinum*) can be possibly targeted by $CTBE_{evolve}$ to introduce at least one premature stop codon within top 75%, 50%, and 25% of the ORF body, respectively, demonstrating that numerous genes can possibly be inactivated by $CTBE_{evolve}$ in mycobacteria.

Discussion

Genetic manipulation is of vital importance in facilitating the study of *M. tuberculosis* biology and drug-target exploration. However, the highly desirable scarless, precise, and markless editing in *M. tuberculosis* relies on HDR and requires months to years for editing. CRISPR-assisted HDR methods have been developed for rapid and precise genome editing in a number of bacterial species[13,15,17,34]. However, it is not applicable in *M. tuberculosis* likely due to the lack of a CRISPR-compatible HDR system. CRISPRi systems using catalytic inactive St1Cas9 or SpCas9 or *Francisella novicida*

Cas12a (FnCas12a) has been developed for gene silencing in mycobacteria. However, these systems can only achieve partial gene knockdown and will cause polar effect that the operonic genes downstream of the Cas protein binding sites will also be silenced[21,22].

To address these challenges in mycobacteria gene editing, we developed highly efficient PAM-expanded St1Cas9 C-to-T and C-to-G base editors for programmed base editing in mycobacteria. These systems can achieve precise single base substitutions by a single transformation step, thereby substantially reducing the time and efforts for genetic manipulation. Moreover, the expression of the operonic genes downstream of the Cas protein binding sites that can be silenced by the CRISPRi system[21,22], will not be affected by the base editing systems. In addition, the base editing systems are amenable for highly efficient multiplexed editing, which are extremely difficult to achieve using serial editing that is prohibitively time consuming in slow-growing pathogens. Because only a 20 nt spacer sequence is required for targeting, in addition to single gene editing, the systems can be further engineered to be a high-throughput gene knockout screening method, which would allow for the systematical discovery of new drug targets and facilitate new therapeutic method development in mycobacteria.

Experimental

Bacterial strains and plasmids

M. tuberculosis strain H37Rv, *M. smegmatis* strain mc^{2}155, and their derivatives were used in this study. All the bacterial strains used in this study are listed in Table S2. The plasmids used in this study are listed in Table S3. The targeting sequences are listed in Table S4. The *E. coli* strains were cultured in Luria-Bertani broth (LB). *M. tuberculosis* and *M. smegmatis* were grown at 37 °C in Middlebrook 7H9 broth or 7H10 plates supplemented with 0.2% glycerol, 0.05% Tween80, 1× albumin dextrose catalase (ADC) (*M. smegmatis*) or oleic acid ADC (*M. tuberculosis*) and the appropriate antibiotics. When noted, antibiotics or chemicals were used at the following concentrations: kanamycin (20 μg/mL), leucine (50 μg/mL), Isoniazid (INH) (16 μg/mL) and anhydrotetracycline (ATc) (20 ng/mL).

Plasmid construction and cloning

Cas9 and Cpf1 sequences were obtained from the published studies[10,22,40]. dCas9 and dCpf1 protein sequences used in this study are listed in Table S5. The CTBE$_{evolve}$ plasmid was constructed using the following steps: The dSt1Cas9 gene, kanamycin-resistance marker, ColE1 origin, and TetR-regulated promoter were amplified from pLJR962 plasmid[22]; The pMF1 replicon and Tet repressor were amplified from the pYC1640 plasmid[8]; The sgRNA scaffold and rAPOBEC1 were amplified from pBECKP-St1[41]; The fragments were assembled by Gibson assembly. Mutations in

St1Cas9 were introduced by site specific mutagenesis. The successful construction of the $CTBE_{evolve}$ plasmid was verified by PCR, sequencing, and enzyme digestion. For base editing, targeting sequences were cloned into the $CTBE_{evolve}$ plasmid by Golden Gate Assembly and verified by sequencing. For multiplexed editing, the second and third sgRNA expression cassettes were PCR-amplified and inserted into the backbone of the first sgRNA-containing $CTBE_{evolve}$ plasmid by Gibson assembly. The successful construction of the plasmid was verified by PCR, sequencing, and enzyme digestion.

Competent cell preparation and electroporation.

M. tuberculosis strain H37Rv (ATCC27294) was inoculated in Lowenstein-Jensen slant from frozen stock and incubated at 37 ℃ for 2 weeks, and then transferred and grown in 100 mL Middlebrook 7H9 broth supplemented with 0.05% Tween 80, 0.2% glycerol, and OADC (Oleic Albumin Dextrose Catalase) for another 2 weeks. The culture was cooled on ice for 5 min and collected by centrifugation. The pellets were washed twice with 30 mL 10% precooled glycerol and resuspended in 5 mL 10% cold glycerol. For electroporation, 1 μg recombinant plasmid was mixed with 100 μL competent cells in 0.2 cm cuvette, and the transformation was conducted with the Gene Pulser Xcell Electroporation Systems (Bio-Rad) with the following conditions: 2.5 kV, 1,000 Ω, and 25 μF. After shock, 1 mL of 7H9 broth supplemented with OADC was immediately added into the cuvette. The culture was incubated for 2 days at 37°C and plated on Middlebrook 7H10 agar supplemented with OADC, 20 μg/mL kanamycin

and 100 ng/mL anhydrous tetracycline. The plates were sealed by parafilm and incubated for 20 to 30 days at 37°C.

M. smegmatis strain mc^{2}155 (ATCC700084) was inoculated in7H10 with 10% ADC enrichment from frozen stock and incubated at 37 ℃ for 4 days. A single colony of a *M. smegmatis* strain was inoculated into 2 mL Middlebrook 7H9 broth supplemented with 0.05% Tween 80, 0.2% glycerol and ADC (Albumin Dextrose Catalase) at 37 ℃ for 24 h. The cells were 1:100 diluted into 100 mL Middlebrook 7H9 broth supplemented with 0.05% Tween 80, 0.2% glycerol and ADC for another 12 to 15 hours. When the optical density at 600 nm of the culture reached 0.8 to 1, the culture was cooled on ice for 20 min and collected by centrifugation in 50-mL conical tubes at 4,000 rpm for 10 min. The pellets were washed twice with 30 mL 10% precooled glycerol and resuspended in 10 mL 10% cold glycerol. For electroporation, 100 ng plasmid was mixed with 100 μL competent cells in 0.2 cm cuvette, and the transformation was conducted with the Gene Pulser Xcell Electroporation Systems (Bio-Rad) with the following conditions: 2.5 kV, 1,000 Ω, and 25 μF. After shock, 1 mL of 7H9 broth supplemented with ADC was immediately added into the cuvette. The culture was incubated for 3 hours at 37°C and 10% portions of the culture were plated on Middlebrook 7H10 agar supplemented with ADC, 20 μg/mL kanamycin and 20 ng/mL anhydrous tetracycline. The plates were sealed by parafilm and incubated for 7 to 10 days at 37°C.

Editing efficiency evaluation

After transformation, all the colonies of *M. smegmatis* or *M. tuberculosis* were collected from the plates and the genomic DNAs were extracted using the Rapid Bacterial Genomic DNA Isolation Kit (Sangon Biotech, Shanghai, China). The target region was amplified with the Easy Taq DNA Polymerase (TransGen, Beijing, China) using specific primers for the target region. The PCR products were sent out for Sanger sequencing and the editing efficiency was calculated by using EditR[51].

leuB or _leuC_ knockout using CTBE$_{evolve}$

Spacers were designed and inserted into the CTBE$_{evolve}$ plasmid. The successfully constructed plasmids were electroporated into *M. smegmatis* mc^{2}155 competent cells. After shock, 1 mL of Middlebrook 7H9 broth with ADC and leucine was immediately added into the cuvette. The culture was incubated for 3 hours at 37°C and 10% portions of the culture were plated onto a Middlebrook 7H10 agar plate supplemented with ADC, 20 μg/mL kanamycin, 20 ng/mL anhydrous tetracycline and 50 μg/mL leucine. The plates were sealed by parafilm and incubated at 37°C. Seven days after electroporation, the overall editing efficiencies were evaluated using the method described in the **Editing efficiency evaluation** Section of the experimental procedures and a single colony was cultured in 5 mL of 7H9 broth supplemented with ADC and leucine at 37°C for two days in the absence of kanamycin. Then the cells were plated onto Middlebrook

7H10 agar plates supplemented with ADC and leucine in the absence of kanamycin. The *leuB* or *leuC* mutant strains were separated and confirmed by sequencing.

Plasmid curing

To cure the editing plasmid in *M. smegmatis* after base editing, one colony was cultured in Middlebrook 7H9 broth with ADC in the absence of kanamycin. After growing to the stationary phase (4 days), the cells were plated onto a Middlebrook 7H10 agar plated supplemented with ADC. A single colony was picked and diluted into 5 mL Middlebrook 7H9. A fraction of the diluted cells were plated onto a Middlebrook 7H10 agar supplemented with ADC without kanamycin and another fraction was plated onto a Middlebrook 7H10 agar supplemented with ADC containing kanamycin. The cells whose plasmid was successfully cured could only grow on the plate without kanamycin.

Leucine auxotrophy assay

The *leuB* or *leuC* gene was knocked out by introducing a premature stop codon using the $CTBE_{evolve}$ plasmid. The strains were grown in Middlebrook 7H9 broth supplemented with ADC and leucine to the stationary phase. A fraction of the cells was plated onto a Middlebrook 7H10 agar plate supplemented with ADC without leucine and another fraction was plated onto a Middlebrook 7H10 agar supplemented with ADC and leucine. The plates were sealed by parafilm and incubated for 4 days at 37°C.

INH resistance assay

The *katG* gene was knocked out by introducing a premature stop codon at Gln3 (CAA to TAA) using the $CTBE_{evolve}$ plasmid. The strains were grown in Middlebrook 7H9 broth with OADC to the stationary phase. A fraction of the cells was plated onto the Middlebrook 7H10 agar plate supplemented with OADC without INH and another fraction was plated onto a Middlebrook 7H10 agar supplemented with OADC and INH (16 μg/mL). The plates were sealed by parafilm and incubated for 20 days at 37°C.

Aggregation assay

The *ctpE* gene was knocked out by introducing a premature stop codon at Gln16 (CAG to TAG) using the $CTBE_{evolve}$ plasmid. The strains were grown in Middlebrook 7H9 broth to the mid-log phase. The cells were 1:100 diluted and cultured in Middlebrook 7H9 broth containing 1.0 mM EGTA or 0 mM EGTA. The cells were grown at 37°C for 48 h with shaking at 200 rpm and then the cells were left undisturbed for 1.0 h at room temperature for cell aggregation.

Whole-genome sequencing

Genomes of the wild-type and two edited strains were sent out for whole-genome sequencing by the Illumina HiSeq/Nova 2x150bp platform at Genewiz. The paired-end fragment libraries were sequenced according to the Illumina HiSeq/Nova 2x150bp platform's protocol. The assembly scaffolds were aligned with reference genome

(Accession number: NC_008596) for detection of single nucleotide polymorphism (SNP).

Preparation of sgRNA and DNA.

The transcription templates of sgRNA were chemically synthesized by GENEWIZ (Suzhou, China). The templates were further amplified by PCR. sgRNA was prepared by in vitro transcription using the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) following manufacturers' instructions and further purified using phenol/chloroform extraction followed by ethanol precipitation.

In vitro DNA cleavage assay for St1Cas9 proteins

The preparation of the proteins and the cleavage assay were performed by following a previous study[41]. In brief, the sequence containing the AAAGAA PAM was first cloned into vector pUC19 (referred as pUC-AGAA). The plasmids with distinct PAM sites were constructed by site specific mutagenesis, using pUC-AGAA as the template (Table S6). The plasmids were linearized by KpnI digestion and the products were used as the cleavage substrates. In vitro cleavage reactions were performed in a 10 μL reaction buffer (10 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT) containing 5 nM linearized plasmids, 250 nM purified St1Cas9 or its variants and 500 nM sgRNA. The reactions were incubated at 37 °C for 40 min and quenched by 25 mM

EDTA and 10 μg Proteinase K for incubation at 58 °C for 10 min. Reaction products were analyzed by a 1% agarose gel.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

We thank Dr. Yi-Cheng Sun for providing the pYC1640 plasmid. This work was supported by grants 21922705 (to Q.J.), 91753127 (to Q.J.), and 2207783 (to Q.J.) from the National Natural Science Foundation of China, 19QA1406000 (to Q.J.) from the Shanghai Committee of Science and Technology, China, EKPG21-18 (to Q.J.) m the Emergency Key Program of Guangzhou Laboratory, and M2020019 (to W.C.) from General program of Jiangsu Health Committee Foundation.

References

(1) World Health Organization Global tuberculosis report 2018. *WHO* **2018**.

(2) Gandhi, N. R.; Nunn, P.; Dheda, K.; Schaaf, H. S.; Zignol, M.; van Soolingen, D.; Jensen, P.; Bayona, J. Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* **2010,** *375* (9728), 1830.

(3) Udwadia, Z. F.; Amale, R. A.; Ajbani, K. K.; Rodrigues, C. Totally drug-resistant tuberculosis in India. *Clin Infect Dis* **2012,** *54* (4), 579.

(4) Balasubramanian, V.; Pavelka, M. S., Jr.; Bardarov, S. S.; Martin, J.; Weisbrod, T. R.; McAdam, R. A.; Bloom, B. R.; Jacobs, W. R., Jr. Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *J Bacteriol* **1996,** *178* (1), 273.

(5) Bardarov, S.; Bardarov, S.; Pavelka, M. S.; Sambandamurthy, V.; Larsen, M.; Tufariello, J.; Chan, J.; Hatfull, G.; Jacobs, W. R. Specialized transduction: an efficient method for generating marked and unmarked targeted gene disruptions in *Mycobacterium tuberculosis*, *M. bovis* BCG and *M. smegmatis*. *Microbiology* **2002,** *148* (Pt 10), 3007.

(6) van Kessel, J. C.; Hatfull, G. F. Recombineering in *Mycobacterium tuberculosis*. *Nat Methods* **2007,** *4* (2), 147.
(7) Murphy, K. C.; Nelson, S. J.; Nambi, S.; Papavinasasundaram, K.; Baer, C. E.; Sassetti, C. M. ORBIT: a new paradigm for genetic engineering of mycobacterial chromosomes. *mBio* **2018,** *9* (6).
(8) Yan, M. Y.; Li, S. S.; Ding, X. Y.; Guo, X. P.; Jin, Q.; Sun, Y. C. A CRISPR-assisted nonhomologous end-joining strategy for efficient genome editing in *Mycobacterium tuberculosis*. *mBio* **2020,** *11* (1).
(9) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012,** *337* (6096), 816.
(10) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **2013,** *339* (6121), 819.
(11) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-guided human genome engineering via Cas9. *Science* **2013,** *339* (6121), 823.
(12) Ge, X.; Xi, H.; Yang, F.; Zhi, X.; Fu, Y.; Chen, D.; Xu, R. H.; Lin, G.; Qu, J.; Zhao, J.et al. CRISPR/Cas9-AAV mediated knock-in at NRL locus in human embryonic stem cells. *Mol Ther Nucleic Acids* **2016,** *5* (11), e393.
(13) Cobb, R. E.; Wang, Y.; Zhao, H. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth Biol* **2015,** *4* (6), 723.
(14) Yang, F.; Liu, C.; Chen, D.; Tu, M.; Xie, H.; Sun, H.; Ge, X.; Tang, L.; Li, J.; Zheng, J.et al. CRISPR/Cas9-loxP-mediated gene editing as a novel site-specific genetic manipulation tool. *Mol Ther Nucleic Acids* **2017,** *7*, 378.
(15) Tong, Y.; Charusanti, P.; Zhang, L.; Weber, T.; Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol* **2015,** *4* (9), 1020.
(16) Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **2013,** *31* (3), 233.
(17) Chen, W.; Zhang, Y.; Yeo, W. S.; Bae, T.; Ji, Q. Rapid and efficient genome editing in *Staphylococcus aureus* by using an engineered CRISPR/Cas9 system. *J Am Chem Soc* **2017,** *139* (10), 3790.
(18) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013,** *152* (5), 1173.
(19) Gilbert, L. A.; Larson, M. H.; Morsut, L.; Liu, Z.; Brar, G. A.; Torres, S. E.; Stern-Ginossar, N.; Brandman, O.; Whitehead, E. H.; Doudna, J. A.et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **2013,** *154* (2), 442.

(20) Choudhary, E.; Thakur, P.; Pareek, M.; Agarwal, N. Gene silencing by CRISPR interference in mycobacteria. *Nat Commun* **2015,** *6*, 6267.
(21) Singh, A. K.; Carette, X.; Potluri, L. P.; Sharp, J. D.; Xu, R.; Prisic, S.; Husson, R. N. Investigating essential gene function in *Mycobacterium tuberculosis* using an efficient CRISPR interference system. *Nucleic Acids Res* **2016,** *44* (18), e143.
(22) Rock, J. M.; Hopkins, F. F.; Chavez, A.; Diallo, M.; Chase, M. R.; Gerrick, E. R.; Pritchard, J. R.; Church, G. M.; Rubin, E. J.; Sassetti, C. M.et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* **2017,** *2*, 16274.
(23) Fleck, N.; Grundner, C. A Cas12a-based CRISPR interference system for multigene regulation in mycobacteria. *J Biol Chem* **2021**, DOI:10.1016/j.jbc.2021.100990 10.1016/j.jbc.2021.100990.
(24) 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* **2016,** *533* (7603), 420.
(25) Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; Liu, D. R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **2017,** *551* (7681), 464.
(26) Jin, S.; Zong, Y.; Gao, Q.; Zhu, Z.; Wang, Y.; Qin, P.; Liang, C.; Wang, D.; Qiu, J. L.; Zhang, F.et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **2019,** *364* (6437), 292.
(27) Tong, Y.; Whitford, C. M.; Robertsen, H. L.; Blin, K.; Jørgensen, T. S.; Klitgaard, A. K.; Gren, T.; Jiang, X.; Weber, T.; Lee, S. Y. Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc Natl Acad Sci U S A* **2019,** *116* (41), 20366.
(28) Gu, T.; Zhao, S.; Pi, Y.; Chen, W.; Chen, C.; Liu, Q.; Li, M.; Han, D.; Ji, Q. Highly efficient base editing in *Staphylococcus aureus* using an engineered CRISPR RNA-guided cytidine deaminase. *Chem Sci* **2018,** *9* (12), 3248.
(29) Wang, Y.; Wang, Z.; Chen, Y.; Hua, X.; Yu, Y.; Ji, Q. A highly efficient CRISPR-Cas9-based genome engineering platform in *Acinetobacter baumannii* to understand the H_2O_2-sensing mechanism of OxyR. *Cell Chem Biol* **2019,** *26* (12), 1732.
(30) Zheng, K.; Wang, Y.; Li, N.; Jiang, F. F.; Wu, C. X.; Liu, F.; Chen, H. C.; Liu, Z. F. Highly efficient base editing in bacteria using a Cas9-cytidine deaminase fusion. *Commun Biol* **2018,** *1*, 32.
(31) Li, Q.; Seys, F. M.; Minton, N. P.; Yang, J.; Jiang, Y.; Jiang, W.; Yang, S. CRISPR-Cas9(D10A) nickase-assisted base editing in the solvent producer *Clostridium beijerinckii*. *Biotechnol Bioeng* **2019,** *116* (6), 1475.
(32) Chen, W.; Zhang, Y.; Zhang, Y.; Pi, Y.; Gu, T.; Song, L.; Wang, Y.; Ji, Q. CRISPR/Cas9-based genome editing in *Pseudomonas aeruginosa* and cytidine

deaminase-mediated base editing in pseudomonas species. *iScience* **2018,** *6*, 222.

(33) Wang, Y.; Wang, S.; Chen, W.; Song, L.; Zhang, Y.; Shen, Z.; Yu, F.; Li, M.; Ji, Q. CRISPR-Cas9 and CRISPR-assisted cytidine deaminase enable precise and efficient genome editing in *Klebsiella pneumoniae*. *Appl Environ Microbiol* **2018,** *84* (23).

(34) Banno, S.; Nishida, K.; Arazoe, T.; Mitsunobu, H.; Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat Microbiol* **2018,** *3* (4), 423.

(35) Wang, Y.; Liu, Y.; Liu, J.; Guo, Y.; Fan, L.; Ni, X.; Zheng, X.; Wang, M.; Zheng, P.; Sun, J.et al. MACBETH: Multiplex automated *Corynebacterium glutamicum* base editing method. *Metab Eng* **2018,** *47*, 200.

(36) Altenbuchner, J. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 System. *Appl Environ Microbiol* **2016,** *82* (17), 5421.

(37) Huang, H.; Chai, C.; Li, N.; Rowe, P.; Minton, N. P.; Yang, S.; Jiang, W.; Gu, Y. CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synth Biol* **2016,** *5* (12), 1355.

(38) Leenay, R. T.; Vento, J. M.; Shah, M.; Martino, M. E.; Leulier, F.; Beisel, C. L. Genome editing with CRISPR-Cas9 in *Lactobacillus plantarum* revealed that editing outcomes can vary across strains and between methods. *Biotechnol J* **2019,** *14* (3), e1700583.

(39) Jiang, Y.; Chen, B.; Duan, C.; Sun, B.; Yang, J.; Yang, S. Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* **2015,** *81* (7), 2506.

(40) Li, X.; Wang, Y.; Liu, Y.; Yang, B.; Wang, X.; Wei, J.; Lu, Z.; Zhang, Y.; Wu, J.; Huang, X.et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **2018,** *36* (4), 324.

(41) Zhang, Y.; Zhang, H.; Xu, X.; Wang, Y.; Chen, W.; Wang, Y.; Wu, Z.; Tang, N.; Wang, Y.; Zhao, S.et al. Catalytic-state structure and engineering of Streptococcus thermophilus Cas9. *Nat Catal* **2020,** *3*, 813

(42) Komor, A. C.; Zhao, K. T.; Packer, M. S.; Gaudelli, N. M.; Waterbury, A. L.; Koblan, L. W.; Kim, Y. B.; Badran, A. H.; Liu, D. R. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **2017,** *3* (8), eaao4774.

(43) Zhao, D.; Li, J.; Li, S.; Xin, X.; Hu, M.; Price, M. A.; Rosser, S. J.; Bi, C.; Zhang, X. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* **2021,** *39* (1), 35.

(44) Kurt, I. C.; Zhou, R.; Iyer, S.; Garcia, S. P.; Miller, B. R.; Langner, L. M.; Grünewald, J.; Joung, J. K. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* **2021,** *39* (1), 41.

(45) Chen, L.; Park, J. E.; Paa, P.; Rajakumar, P. D.; Prekop, H. T.; Chew, Y. T.; Manivannan, S. N.; Chew, W. L. Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat Commun* **2021,** *12* (1), 1384.

(46) Koblan, L. W.; Arbab, M.; Shen, M. W.; Hussmann, J. A.; Anzalone, A. V.; Doman, J. L.; Newby, G. A.; Yang, D.; Mok, B.; Replogle, J. M.et al. Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat Biotechnol* **2021**, DOI:10.1038/s41587-021-00938-z.

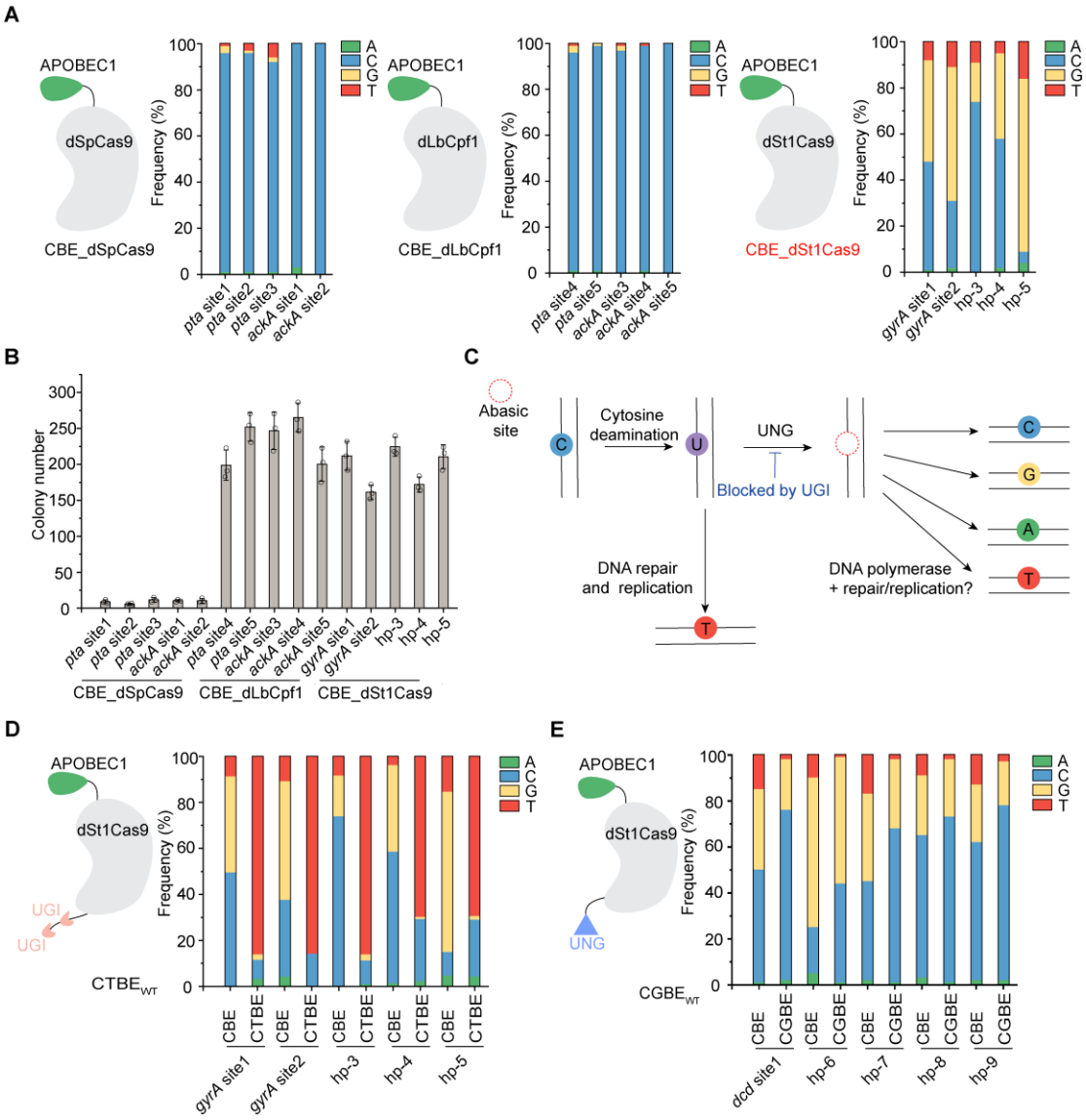
(47) Billon, P.; Bryant, E. E.; Joseph, S. A.; Nambiar, T. S.; Hayward, S. B.; Rothstein, R.; Ciccia, A. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol Cell* **2017,** *67* (6), 1068.

(48) Yu, H.; Wu, Z.; Chen, X.; Ji, Q.; Tao, S. CRISPR-CBEI: a designing and analyzing tool kit for cytosine base editor-mediated gene inactivation. *mSystems* **2020,** *5* (5).

(49) Gupta, H. K.; Shrivastava, S.; Sharma, R. A novel calcium uptake transporter of uncharacterized P-type ATPase family supplies calcium for cell surface integrity in *Mycobacterium smegmatis*. *mBio* **2017,** *8* (5).

(50) Unissa, A. N.; Subbian, S.; Hanna, L. E.; Selvakumar, N. Overview on mechanisms of isoniazid action and resistance in *Mycobacterium tuberculosis*. *Infect Genet Evol* **2016,** *45*, 474.

(51) Kluesner, M. G.; Nedveck, D. A.; Lahr, W. S.; Garbe, J. R.; Abrahante, J. E.; Webber, B. R.; Moriarity, B. S. EditR: a method to quantify base editing from sanger sequencing. *The CRISPR Journal* **2018,** *1* (3), 239.



631

632 **Figure 1. St1Cas9-mediated base editing in *M. smegmatis*.** (A) Identification of
633 active CRISPR base editors in *M. smegmatis*. Three distinct base editors, *S.*
634 *thermophilus* Cas9 cytidine base editor (CBE_dSt1Cas9), *S. pyogenes* Cas9 cytidine
635 base editor (CBE_dSpCas9) and *L. bacterium* Cpf1 cytidine base editor
636 (CBE_dLbCpf1) were screened. The APOBEC1 cytidine deaminase was fused to the
637 N-terminus of the Cas nucleases. (B) Transformation efficiencies of the distinct base
638 editors in *M. smegmatis*. One hundred nanogram of each plasmid was used for
639 electroporation. (C) Possible cellular DNA repair mechanisms of cytidine deamination.
640 The initial editing product, U:G mismatch pair, can be directly converted to the T:A pair
641 by DNA repair and replication, or it can be excised by endogenous uracil glycosylase
642 (UNG), leading to the formation of diverse editing products. Uracil glycosylase
643 inhibitor (UGI) can block endogenous UNG activity. (D, E) St1Cas9 CBE was
644 engineered with UGI or UNG fusion, yielding two new base editor CTBE (D) and
645 CGBE (E), capable of C-to-T or C-to-G conversions in *M. smegmatis* with drastically
646 enhanced editing purity.

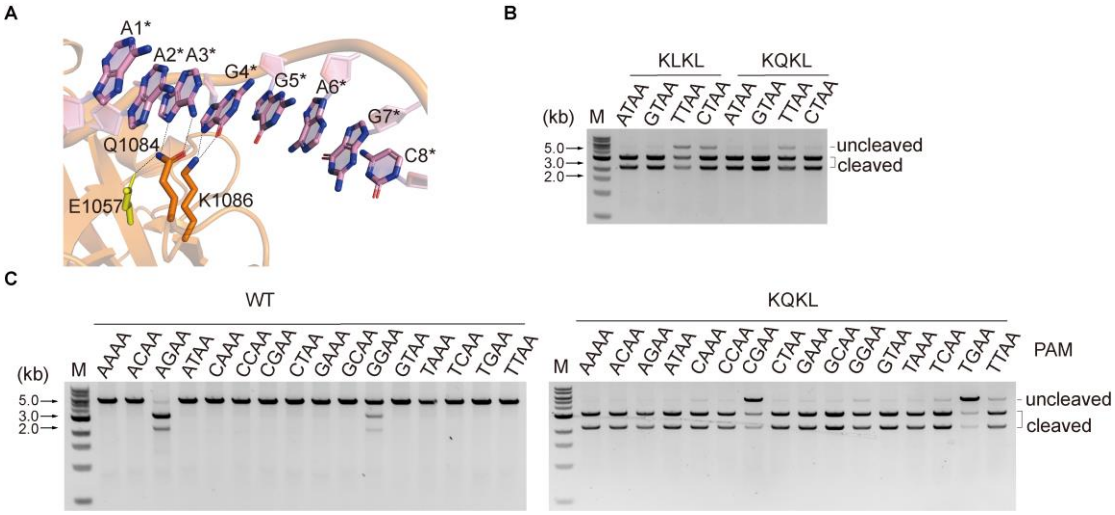


Figure 2. PAM expansion of St1Cas9 via structure-guided engineering. (A) PAM recognition mechanism of wild-type St1Cas9 (PDB: 6M0X). (B) In vitro DNA cleavage assay of two engineered mutants of St1Cas9. (C) In vitro DNA cleavage assay of the wild-type St1Cas9 and the KQKL variant.

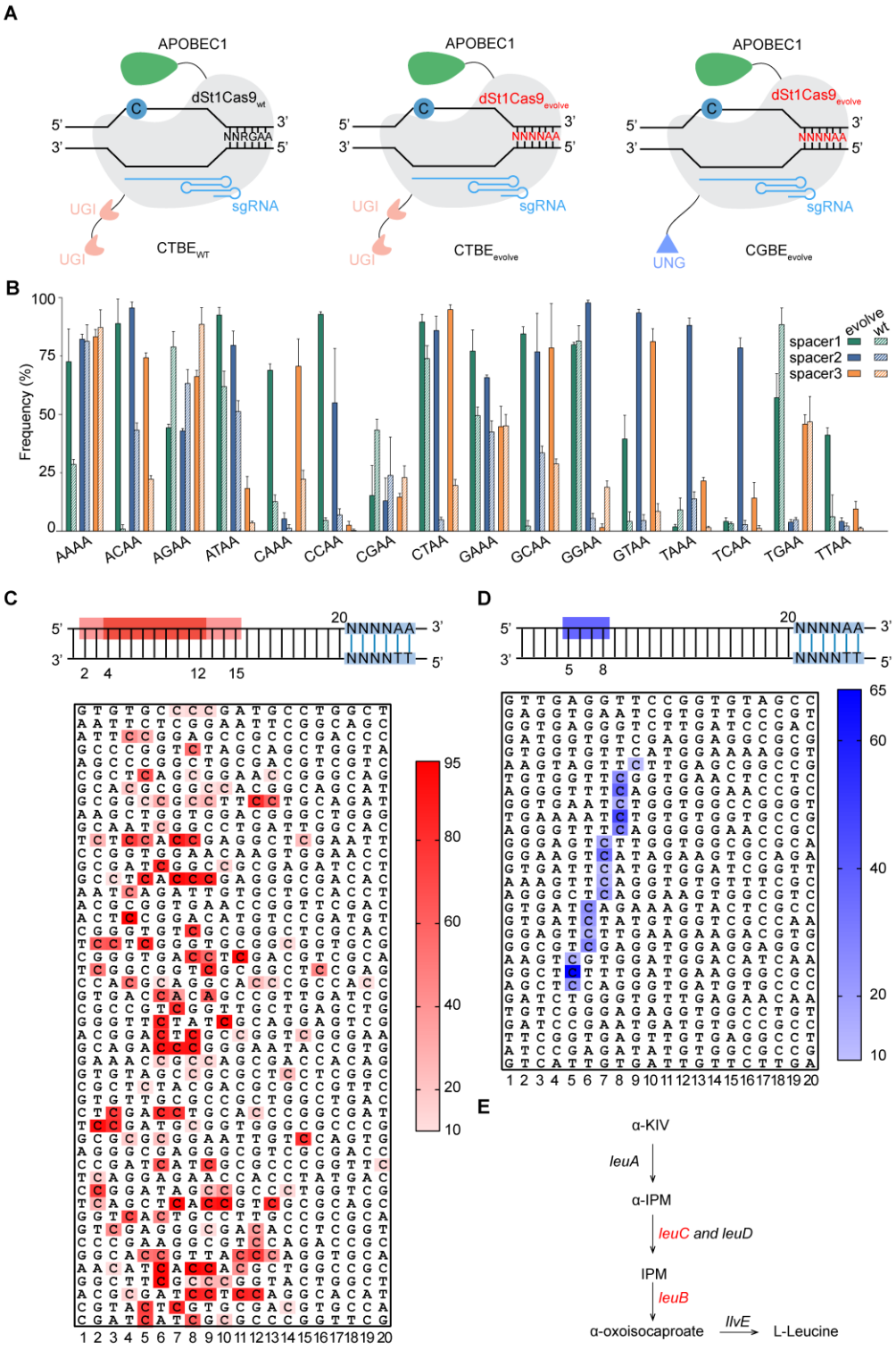

A
APOBEC1
dSt1Cas9wt
NNRGAA
sgRNA
UGI
CTBEWT
dSt1Cas9evolve
NNNNAA
CTBEevolve
UNG
CGBEevolve
B
Frequency (%)
spacer1
spacer2
spacer3
evolve
wt
AAAA
ACAA
AGAA
ATAA
CAAA
CCAA
CGAA
CTAA
GAAA
GCAA
GGAA
GTAA
TAAA
TCAA
TGAA
TTAA
C
D
NNNNAA
NNNNTT
E
α-KIV
leuA
α-IPM
leuC and leuD
IPM
leuB
α-oxoisocaproate
IlvE
L-Leucine
F
WT 5′- ttcgcgggcagcctgtggtcgcgcac
3′- aagcgccgtcggacaccagcgcgtg
Trp193
CTBEevolve
ΔleuB 5′- ttcgcgggcagcctgtaatcgcgcac
overall editing efficiency: 73%
- Leucine
+ Leucine
ΔleuB
ΔleuC
WT 5′- accaggtcgagacattgcggcgcaa
3′- tgggtccagctctgtaacgccgcgtt
Gln96
ΔleuC 5′- atttaggttgagacattgcggcgcaa
overall editing efficiency: 16%

Figure 3. Comprehensive characterizations of CTBE and CGBE in *M. smegmatis*. (A) Compositions of different St1Cas9 CBEs. (B) Editing activity comparison of $CTBE_{wt}$ and $CTBE_{evovle}$ in *M. smegmatis*. Bar plots reflect the maximum editing frequency within the editing widow. (C, D) Editing windows of $CTBE_{evolve}$ (C) and $CGBE_{evolve}$ (D). (E) The L-leucine biosynthesis pathway in *M. smegmatis*. The targeted genes (*leuC* and *leuB*) were highlighted in red. α-KIV, α-ketoisovalerate; α-IPM, α-isopropylmalate; IPM, 3-isopropylmalate. (F) Inactivation of *leuB* and *leuC* by $CTBE_{evolve}$ via generating premature stop codons. Inactivation of *leuB* or *leuC* causes auxotroph in the absence of L-Leucine. A box and an inverted box indicate the target sequence and PAM, respectively.

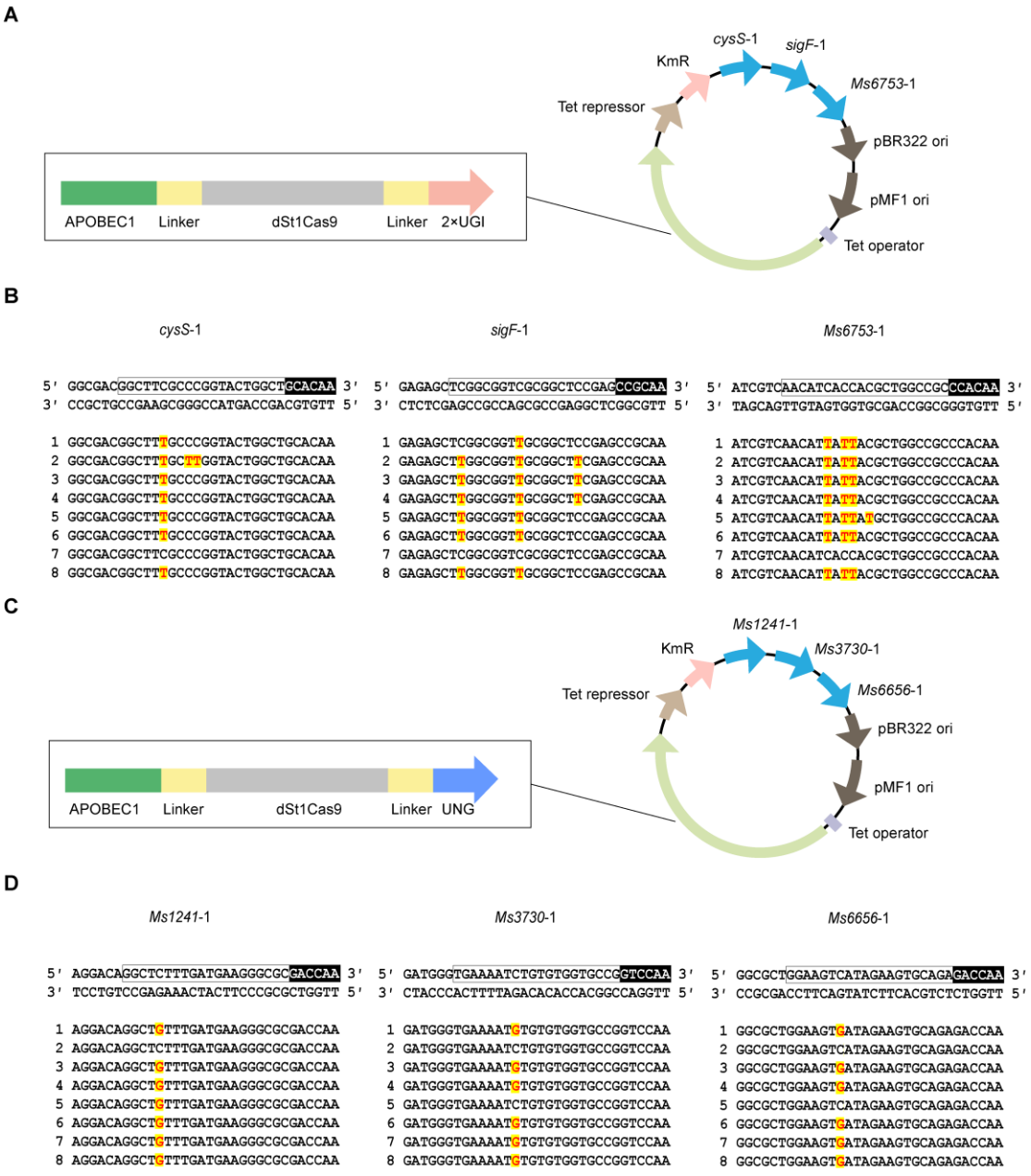


Figure 4. Multiplexed editing in *M. smegmatis*. (A) Map of the single plasmid system for $CTBE_{evolve}$ mediated multiplexed mutagenesis. (B) Editing results of the $CTBE_{evolve}$ mediated multiplexed editing assay. A box and an inverted box indicate the target sequence and PAM, respectively. The edited bases are shown in red and highlighted in yellow. (C) Map of the single plasmid system for $CGBE_{evolve}$ mediated multiplexed mutagenesis. (D) Editing results of the CGBE mediated multiplexed editing assay. A box and an inverted box indicate the target sequence and PAM, respectively. The edited bases are shown in red and highlighted in yellow.

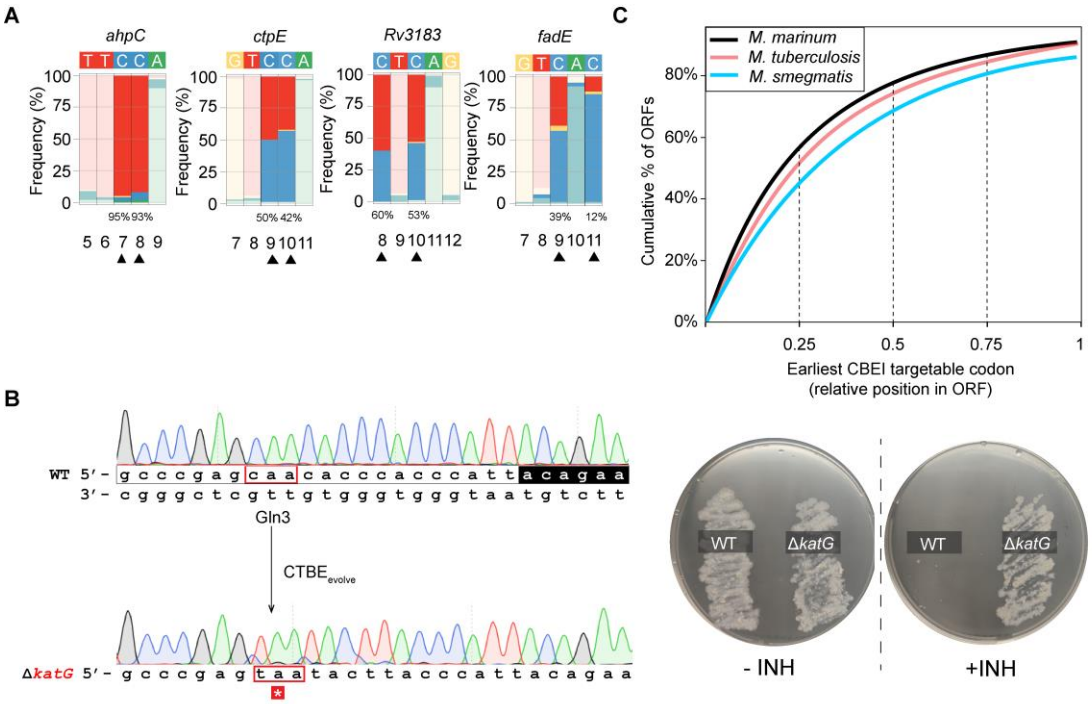


Figure 5. Base editing in *M. tuberculosis* with CTBE$_{evolve}$. (A) Bar plots showing the editing efficiencies of CTBE$_{evolve}$ in *M. tuberculosis* on different sites. Numbering on the bottom indicates the position of the bases in a protospacer with 1 being the most PAM-distal base. Arrowheads indicate cytosines with C-to-T conversions. (B) Inactivation of *katG* by CTBE$_{evolve}$ via generating a premature stop codon. Inactivation of *katG* makes *M. tuberculosis* resistant to INH. A box and an inverted box indicate the target sequence and PAM, respectively. (C) Relative position of the earliest induction of stop codons targetable in mycobacteria ORFs (cumulative percentage) by CTBE$_{evolve}$.